Whitney Domico

Pr. Hansen

BIO247

5 December 2022

Research Project Reflection

*Changes From Original Proposal*

When I originally proposed my research project, I had not decided on whether I would focus on the single nucleotide polymorphism (SNP) loci associated with schizophrenia or the *N*-methyl-D-aspartate (NMDA) pathway. I had identified 16 peer-reviewed articles that contained information for one or both of these topics, and I intended to run them through a machine reader and see if any patterns arose to help me decide what to focus on for the project. Unfortunately, the reader did not find anything significant in my articles, and so this approach did not work. Instead, I decided to focus on the loci because those had more numeric-looking data, and so I imported the data from 10 of the original 16 articles into R (the remaining 6 articles were ones that solely focused on the NMDA pathway and did not have loci data).

After successfully cleaning the SNP loci data, I followed my initial proposal and discretized the loci into their respective chromosomes and then respective genes. I decided against my original intention to do an inverse-weighted fixed effects model because I was surprised at the outcome of my analysis being so autoimmune-based that I instead decided to follow that path. So, at the end of my research, I had identified common SNP locis, genes associated with them, and diseases associated with these genes.

*What I Learned*

First and foremost, I became much more comfortable with R. When I began this research, I realized that I did not have the knowledge to even begin doing what I had planned, and so I spent a lot of time doing Euler problems until I was more comfortable. I also learned the value of coding diagrams (which I had admittedly initially thought were not very useful), and now I feel confident in my R skills. I got really good at loops. Most of my code is nested loops, and it is something that I am now very comfortable with.

Secondly, I learned that cleaning data is no easy task. Cleaning my data was the most time-consuming of everything that I did. I got really good at string-splitting, and this also really cemented my knowledge of the differences between lists and vectors.

Similarly, I also learned that coding is very tedious and time-consuming. I learned that there are some tasks that seem simple that actually take hours and some tasks that seem like they will take hours but actually take a singular trip to Stack Exchange (I also got really good at articulating what I needed in terms that Stack Exchange could understand).

Not regarding coding, I also learned that schizophrenia (and mental disorders as a group) are grossly under researched (at least in terms of genetics and biochemistry), and I also became much more confident that this is the area that I want to go into for the rest of my life.

*What Problems I Encountered*

The first major problem that I encountered was that very few of the articles that I used had their data in an easy-to-find location. I spent a long time digging through supplementary information, appendices, and, in some cases, emailing PIs to get all of the data for my project.

Similarly, very few articles had organized their SNP data in the same way. Some had separate columns for chromosomes and location (ex. “chr6”, “156789”), while others put them in the same column (ex. “6:156789”). Some articles used megabase-pairs, some used kilobase-pairs, some used base-pairs, etc. Some articles called the X chromosome “X”, others called it “x”, and some called it “0X.”

Another problem that I encountered was that I was unable to copy and paste much of the SNP data from the articles into Excel (due, usually, to formatting). Instead, I had to download a PDF-to-text program.

I had many problems with the chromosomes as a list (not formatting correctly in graphs, not sorting properly, etc) until I changed all single-digit numbers (ex. 1, 2, 3) to double-digit numbers (ex. 01, 02, 03). For some reason, this solved all of these issues.

I got very comfortable using “temp” as a temporary vector during loops; however, I often forgot to reset them to temp <- c() either between terms or between loops and this was the cause of many (many, many) bugs.

It took me a very long time to figure out how to make a column that defined all of the genes involved in each pathway. It turned out to be 10 separate loops, and it was an extremely tedious process that took many coding diagrams and lots of trial and error.

*Final Outcome of My Project*

Overall, my project supports the claim that schizophrenia is caused by more than just an overabundance of dopamine by showing that 88% of the known SNPs of its sufferers are within the histocompatibility complex, which is consistent with autoimmune disorders (and, similarly, about 43% of the related disorders that I found were autoimmune in nature). Therefore, this provides evidence that schizophrenia may be more related to autoimmune dysfunction than mental dysfunction, and perhaps this could suggest that treatment options should be shifted to reflect this.

Here is a link to my repository 🡪 <https://github.com/domicowe/BIO247> -> BIO247Project -> BIO247ProjectSubmissions and download zip file “DomicoProjectZip.zip”

*Note: There are instructions called “PROJECT\_INSTRUCTIONS” for detailed help with usage.*

*Future Expansion*

If I had more time, I would try the same process with other mental disorders (depression, anxiety, obsessive-compulsive disorder, etc) to make sure that it is just schizophrenia that is linked this closely to autoimmune disorders (otherwise, this would just suggest that mental disorders and autoimmune disorders are closely related which would be interesting but also basically negate all of the work I have done). Then I would look into autoimmune disorders and their common SNPs and compare these against schizophrenia to find more overlaps or possible divergences. Overall, the goal would be to discover just how similar schizophrenia is to autoimmune diseases to then determine if there is a certain subset of autoimmune diseases that are most similar. These most similar ones would be the key to finding a more specialized medication for those suffering from schizophrenia.

Originally (before my final research proposal), I had expressed interest in comparing schizophrenia to autism because I had read that there is some speculation that they may be related. If I had more time, I may consider doing an identical analysis of the SNPs associated with autism and comparing them with schizophrenia (and, potentially, autoimmune disorders if it happened that autism was also associated with autoimmune dysfunction).