**Project Plan:**

**Gene Ontology – Alzheimer’s Disease Biological Process Domains**

### Introduction

Alzheimer’s disease (AD) is a devastating illness, affecting over 35 million people worldwide and expected to increase to 115 million by 2050. Strategies to delay the disease onset or prevent the disease altogether would reduce the burden on worldwide healthcare systems. Our overall goal is to develop a method that efficiently uses magnetic resonance imaging (MRI)-based neuroanatomical measures to better understand how genetic risk may impact the brain. Rather than use a data-driven hypothesis free approach with all single nucleotide polymorphisms available to me, the burden of multiple comparisons will be reduced by focusing only on specific genetic data. Through this process, I hope to identify robust biomarkers that can be used to shed light on the aetiology of AD and AD risk.

I selected genes associated with the aetiology and risk of AD through previous genome wide association studies (GWAS). Gene Ontology (GO) was then used to derive common biological process domains from the selected genes based on the current data in the GO.

A total of 21 relevant genes were selected from a meta-analysis done by Lambert, JC., et al 2013. The list of genes is seen in the table below.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Gene Name</th>
</tr>
</thead>
<tbody>
<tr>
<td>ApoE</td>
<td>Apolipoprotein E</td>
</tr>
<tr>
<td>BIN1</td>
<td>Myc box-dependent-interacting protein 1</td>
</tr>
<tr>
<td>CLU</td>
<td>Clusterin (Apolipoprotein J)</td>
</tr>
<tr>
<td>ABCA7</td>
<td>ATP-binding cassette sub-family A member 7</td>
</tr>
<tr>
<td>CR1</td>
<td>Complement receptor type 1</td>
</tr>
<tr>
<td>PICALM</td>
<td>Phosphatidylinositol-binding clathrin assembly protein</td>
</tr>
<tr>
<td>MS4A6A</td>
<td>Membrane-spanning 4-domains subfamily A member 6A</td>
</tr>
<tr>
<td>CD33</td>
<td>Myeloid cell surface antigen CD33</td>
</tr>
<tr>
<td>CD2AP</td>
<td>CD2-associated protein</td>
</tr>
<tr>
<td>EPHA1</td>
<td>Ephrin type-A receptor 1</td>
</tr>
<tr>
<td>HLA DRB5-DRB1</td>
<td>Human leukocyte antigen</td>
</tr>
<tr>
<td>SORL1</td>
<td>Sortilin-related receptor</td>
</tr>
<tr>
<td>PTK2B</td>
<td>Protein-tyrosine kinase 2-beta</td>
</tr>
<tr>
<td>SLC24A4</td>
<td>Sodium/potassium/calcium exchanger 4</td>
</tr>
<tr>
<td>ZCWPW1 (NYAP1)</td>
<td>Encode zinc finger CW type with PWWP domain 1 region</td>
</tr>
<tr>
<td>CELF1 (MADD)</td>
<td>Encode CUGBP, Elav-like family member 1 region</td>
</tr>
<tr>
<td>NME8</td>
<td>Thioredoxin domain-containing protein 3</td>
</tr>
<tr>
<td>FERMT2</td>
<td>Fermitin family homolog 2</td>
</tr>
<tr>
<td>INPP5D</td>
<td>Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 1</td>
</tr>
<tr>
<td>MEF2C</td>
<td>Myocyte-specific enhancer factor 2C</td>
</tr>
<tr>
<td>CASS4</td>
<td>Cas scaffolding protein family member 4</td>
</tr>
</tbody>
</table>
From the current data in GO the common biological process (BP) domains for the 21 genes are:

- vesicle-mediated / transport / Endocytosis
- steroid / cholesterol metabolic process
- cell membrane processes / cell migration
- Neuron development / synaptic function
- regulation of calcium-mediated signaling
- immune system process

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein ID</th>
<th>vesicle-mediated Transport/Endocytosis</th>
<th>steroid / cholesterol metabolic process</th>
<th>immune system process</th>
<th>cell membrane processes / cell migration</th>
<th>neuron development / Synaptic function</th>
<th>regulation of calcium-mediated signaling</th>
<th>Aβ metabolism</th>
<th>Tangle formation</th>
<th>Cell death</th>
<th>Apoptosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRB5</td>
<td>Q30154</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>SORT1</td>
<td>Q92673</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td>X</td>
<td>X</td>
<td>X</td>
</tr>
<tr>
<td>PTK2B</td>
<td>Q14289</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SLC24A4</td>
<td>Q8NFF2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NYAP1</td>
<td>Q6ZVC0</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>NME8</td>
<td>Q8N427</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>FERM2</td>
<td>Q96AC1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>INPP5D</td>
<td>Q92835</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MEF2C</td>
<td>Q06413</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CASS4</td>
<td>Q9NQ75</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ApoE</td>
<td>P02649</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>BIN1</td>
<td>O00499</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CLU</td>
<td>P10909</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>ABCA7</td>
<td>Q8IZY2</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CR1</td>
<td>P17927</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>PICALM</td>
<td>Q13492</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MS4A6A</td>
<td>Q9H2W1</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD33</td>
<td>P20138</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>CD2AP</td>
<td>Q9Y5K6</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>EPHA1</td>
<td>P21709</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td>X</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Goal**

Most of the annotations done to the selected genes contained evidence code of IEA (automated biocuration process). We now plan to manually curate the genes of interest to further explore the aetiology of AD and develop the content in the GO database within these biological process domains.

By extracting data from primary articles relating to genes of interest and focusing on the biological processes these genes are involved we will be able to annotate these genes to not only the Biological Process (BP) ontology but also to the cellular component and molecular function ontologies.

Ideally in this project I would like to focus on the 6 (shaded yellow) biological process domains listed above. Based on the common BP domains that are shared between the selected genes, focusing on the 6 domains will help us further explore the aetiology of AD and investigate how the domains may interconnect with each other through the manual biocuration process. In addition, two main pathological markers seen in AD are amyloid plaque and neurofibrillary tangles. Impairment in these 6 biological process domains may be connected to the physical properties seen in AD in terms of amyloid plaque and neurofibrillary tangles which lead to cell death (figure1 below).
BP Domains

Steroid / cholesterol metabolic process

Cell membrane processes
  Cell migration

Vesicle-mediated Transport/Endocytosis

Regulation of calcium-mediated signaling

Immune system process

Neuron development
  Synaptic function

Aβ metabolism

Tangle formation

Cell death / Apoptosis

Figure 1: Using GO to explore how the 6 domains (in yellow) results in amyloid plaque and neurofibrillary tangle formation (in green) which leads to cell death (in purple).

Limitations

The main limitation of the manual biocuration process is if there is limited amount of published functional data on some of the selected genes. (Currently I am looking for primary articles of functional data for the selected genes to be used for annotation process when I am at EBI)

I believe that it may be overly difficult to complete annotations for all of the 6 proposed domains during the time frame in which I will be at EBI. Therefore I propose to focus on the manual annotation of one or two BP domains during my visit at EBI and continue this process back home once I have a better understanding of the process taken in developing the GO database and the different tools used.

Below are the selected genes associated with the different domains based on BP GO terms currently in the database. In addition, some background information from Olgiati, P., et al 2011 is given on the key BP domains that are shared between the selected genes.
**BP Domains Associated with AD**

**Cholesterol metabolism:**

The table below shows the selected genes that I want to annotate in the sterol / cholesterol metabolic process domain.

<table>
<thead>
<tr>
<th>Gene</th>
<th>protein ID</th>
<th>sterol / cholesterol metabolic process</th>
</tr>
</thead>
<tbody>
<tr>
<td>SORL1</td>
<td>Q92673</td>
<td>X</td>
</tr>
<tr>
<td>APOE</td>
<td>P02649</td>
<td>X</td>
</tr>
<tr>
<td>CLU</td>
<td>P10909</td>
<td>X</td>
</tr>
<tr>
<td>ABCA7</td>
<td>Q8IZY2</td>
<td>X</td>
</tr>
</tbody>
</table>

**Background information**

Epidemiological finding suggested that having high levels of cholesterol at midlife becomes a risk factor for developing AD. Amyloidogenic proteins such as amyloid plaque are found in lipid-based structures. There are three groups of lipid in the brain glycerophospholipids, sphingolipids, and cholesterol. Sphingolipids and cholesterol self-aggregate into membrane domains called lipid rafts, which are involved in trafficking, signaling, and sites for host interaction for toxins and pathogens. There are 2 main mechanisms that transfer amyloidogenic protein from cytosolic/extracellular to neuronal plasma membrane:

1. **Reduction of dimensionality of the protein conformation from 3D to 2D**: Beta amyloid lacks defined 3D structure. Therefore, distance between two proteins is shorter in 2D than 3D which increases protein concentration and increases protein-protein interaction. Therefore, this helps to explain amyloidogenic protein self-aggregating after a conformation change.

2. **Amyloidogenic proteins have affinity for neuronal membranes which is lipid specific**: High affinity for sphingolipids results in the concentration of amyloid protein in lipid rafts area at the extracellular leaflet of the plasma membrane. Therefore, it has been hypothesized that sphingolipids and cholesterol play a role in confirmation, oligomerisation, and aggregation of amyloidiogenic proteins.

**Vesicle mediated and Transport/Endocytosis**

The table below shows the selected genes that I want to annotate in the vesicle-mediated/ transport/ Endocytosis domain.

<table>
<thead>
<tr>
<th>Gene</th>
<th>protein ID</th>
<th>Vesicle-mediated/ Transport / Endocytosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>SORL1</td>
<td>Q92673</td>
<td>X</td>
</tr>
<tr>
<td>SLC24A4</td>
<td>Q8NFF2</td>
<td>X</td>
</tr>
<tr>
<td>APOE</td>
<td>P02649</td>
<td>X</td>
</tr>
<tr>
<td>BIN1</td>
<td>O00499</td>
<td>X</td>
</tr>
<tr>
<td>CLU</td>
<td>P10909</td>
<td>X</td>
</tr>
<tr>
<td>ABCA7</td>
<td>Q8IZY2</td>
<td>X</td>
</tr>
<tr>
<td>PICALM</td>
<td>Q13492</td>
<td>X</td>
</tr>
</tbody>
</table>
**Background information**

Changes in intracellular transport of APP (Amyloid Precursor Protein) is based on how APP is cleaved by either:

- α-secretase which is found at the cell surface, releases a nontoxic soluble APP when APP is cleaved (sAPPα)
- β-secretase which is found in the Golgi apparatus and endosomes, releases neurotoxic form of beta amyloid
- γ-secretase which is found in the endoplasmic reticulum, lysosomes and cell surface, releases neurotoxic form of beta amyloid

When APP is moved into the endosome it is cleaved by β-secretase and transported to the cell surface or lysosome which is further cleaved by γ-secretase which forms beta amyloid. Therefore APP accumulating at the cell surface which will be cleaved by α-secretase and as a result form sAPPα which is a nontoxic soluble peptide.

**Immune System**

The table below shows the selected genes that I want to annotate to the immune system domain

<table>
<thead>
<tr>
<th>Gene</th>
<th>protein ID</th>
<th>Immune system process</th>
</tr>
</thead>
<tbody>
<tr>
<td>DRB5</td>
<td>Q30154</td>
<td>X</td>
</tr>
<tr>
<td>PTK2B</td>
<td>Q14289</td>
<td>X</td>
</tr>
<tr>
<td>INPP5D</td>
<td>Q92835</td>
<td>X</td>
</tr>
<tr>
<td>CLU</td>
<td>P10909</td>
<td>X</td>
</tr>
<tr>
<td>ABCA7</td>
<td>Q8IZY2</td>
<td>X</td>
</tr>
<tr>
<td>CR1</td>
<td>P17927</td>
<td>X</td>
</tr>
<tr>
<td>PICALM</td>
<td>Q13492</td>
<td>X</td>
</tr>
<tr>
<td>MS4A6A</td>
<td>Q9H2W1</td>
<td>X</td>
</tr>
<tr>
<td>CD33</td>
<td>P20138</td>
<td>X</td>
</tr>
<tr>
<td>EPHA1</td>
<td>P21709</td>
<td>X</td>
</tr>
</tbody>
</table>

**Background information**

It has been hypothesized that amyloidogenesis leads to neuroinflammation. Beta amyloid deposition in the central nervous system activates microglia which starts a proinflammatory cascade releasing neurotoxic substances (cytokines; chemokines; reactive oxygen and nitrogen species; proteolytic enzymes) which amplify neuronal damage. Activated microglia lead to phosphorylation of tau which leads to formation of neurofibrillary tangles. In addition, inflammatory cells mediate clearance of beta amyloid by phagocytosis. As a result AD patients may increase proinflammatory cytokines and activate microglia as a compensatory for a defective clearance of beta amyloid therefore the inflammatory cascade will cause brain damage.
**Other Known BP Domains:**
There have been other known common BP domains the selected gene have been involved in such as cell membrane process and cell migration, neuron development and synaptic function and regulation of calcium mediated signaling.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein ID</th>
<th>Cell membrane processes / cell migration</th>
</tr>
</thead>
<tbody>
<tr>
<td>SORL1</td>
<td>Q92673</td>
<td>X</td>
</tr>
<tr>
<td>PTK2B</td>
<td>Q14289</td>
<td>X</td>
</tr>
<tr>
<td>APOE</td>
<td>P02649</td>
<td>X</td>
</tr>
<tr>
<td>BIN1</td>
<td>O00499</td>
<td>X</td>
</tr>
<tr>
<td>CD33</td>
<td>P20138</td>
<td>X</td>
</tr>
<tr>
<td>CD2AP</td>
<td>Q9Y5K6</td>
<td>X</td>
</tr>
<tr>
<td>EPHA1</td>
<td>P21709</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein ID</th>
<th>Neuron development / synaptic function</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTK2B</td>
<td>Q14289</td>
<td>X</td>
</tr>
<tr>
<td>NYAP1</td>
<td>Q6ZVC0</td>
<td>X</td>
</tr>
<tr>
<td>MEF2C</td>
<td>Q06413</td>
<td>X</td>
</tr>
<tr>
<td>APOE</td>
<td>P02649</td>
<td>X</td>
</tr>
<tr>
<td>PICALM</td>
<td>Q13492</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein ID</th>
<th>Regulation of calcium-mediated signaling</th>
</tr>
</thead>
<tbody>
<tr>
<td>PTK2B</td>
<td>Q14289</td>
<td>X</td>
</tr>
<tr>
<td>SLC24A4</td>
<td>Q8NFF2</td>
<td>X</td>
</tr>
<tr>
<td>APOE</td>
<td>P02649</td>
<td>X</td>
</tr>
</tbody>
</table>

**BP domains that lead to final outcome:**
Based on the BP domains above, the final outcome or consequences of the 6 domains when impaired lead to Aβ metabolism and tangle formation which results in cell death/apoptosis. The tables below show which genes are involved in the following final outcomes:

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein ID</th>
<th>Aβ metabolism</th>
</tr>
</thead>
<tbody>
<tr>
<td>SORL1</td>
<td>Q92673</td>
<td>X</td>
</tr>
<tr>
<td>CASS4</td>
<td>Q9NQ75</td>
<td>X</td>
</tr>
<tr>
<td>APOE</td>
<td>P02649</td>
<td>X</td>
</tr>
<tr>
<td>CLU</td>
<td>P10909</td>
<td>X</td>
</tr>
<tr>
<td>ABCA7</td>
<td>Q8IZY2</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein ID</th>
<th>Tangle formation</th>
</tr>
</thead>
<tbody>
<tr>
<td>FERMT2</td>
<td>Q96AC1</td>
<td>X</td>
</tr>
<tr>
<td>CASS4</td>
<td>Q9NQ75</td>
<td>X</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Gene</th>
<th>Protein ID</th>
<th>cell death/apoptosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>SORL1</td>
<td>Q92673</td>
<td>X</td>
</tr>
<tr>
<td>PTK2B</td>
<td>Q14289</td>
<td>X</td>
</tr>
<tr>
<td>(MADD)</td>
<td>Q8WXG6</td>
<td>X</td>
</tr>
<tr>
<td>INPP5D</td>
<td>Q92835</td>
<td>X</td>
</tr>
<tr>
<td>MEF2C</td>
<td>Q06413</td>
<td>X</td>
</tr>
<tr>
<td>APOE</td>
<td>P02649</td>
<td>X</td>
</tr>
<tr>
<td>CLU</td>
<td>P10909</td>
<td>X</td>
</tr>
</tbody>
</table>
Acknowledgments

- University of Toronto and Centre for Addiction and Mental Health
  Dr. Jo Knight
  Dr. Mallar Chakravarty
- Gene Ontology Team at EBI
  Dr. Jane Lomax

References

