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.

Gene	Gene Name
ApoE	Apolipoprotein E
BIN1	Myc box-dependent-interacting protein 1
CLU	Clusterin (Apolipoprotein J)
ABCA7	ATP-binding cassette sub-family A member 7
CR1	Complement receptor type 1
PICALM	Phosphatidylinositol-binding clathrin assembly protein
MS4A6A	Membrane-spanning 4-domains subfamily A member 6A
CD33	Myeloid cell surface antigen CD33
CD2AP	CD2-associated protein
EPHA1	Ephrin type-A receptor 1
HLA DRB5-DRB1	Human leukocyte antigen
SORL1	Sortilin-related receptor
PTK2B	Protein-tyrosine kinase 2-beta
SLC24A4	Sodium/potassium/calcium exchanger 4
ZCWPW1	Encode zinc finger CW type with PWWP domain 1 region
(NYAP1)	Neuronal tyrosine-phosphorylated phosphoinositide-3-kinase adapter 1
CELF1	Encode CUGBP, Elav-like family member 1 region
(MADD)	MAP kinase-activating death domain protein
NME8	Thioredoxin domain-containing protein 3
FERMT2	Fermitin family homolog 2
INPP5D	Phosphatidylinositol 3,4,5-trisphosphate 5-phosphatase 1
MEF2C	Myocyte-specific enhancer factor 2C
CASS4	Cas scaffolding protein family member 4

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

Gene	Protein ID	vesicle-mediated Transport/ Endocytosis	steroid / cholesterol metabolic process	immune system process	cell membrane processes/ cell migration	Neuron development/ Synaptic function	regulation of calcium-mediated signaling	Aβ metabolism	Tangle formation	Cell death Apoptosis
DRB5	Q30154			X						
SORL1	Q92673	X	X		X			X		X
PTK2B	Q14289			X	X	X	X			X
SLC24A4	Q8NFF2	X					X			
NYAP1	Q6ZVC0					X				
MADD	Q8WXG6									X
NME8	Q8N427									
FERMT2	Q96AC1								X	
INPP5D	Q92835			X						X
MEF2C	Q06413					X				X
CASS4	Q9NQ75							X	X	
ApoE	P02649	X	X		X	X	X	X		X
BIN1	O00499	X			X					
CLU	P10909	X	X	X				X		X
ABCA7	Q8IZY2	X	X	X				X		
CR1	P17927			X						
PICALM	Q13492	X		X		X				
MS4A6A	Q9H2W1			X						
CD33	P20138			X	X					
CD2AP	Q9Y5K6				X					
EPHA1	P21709			X	X					

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

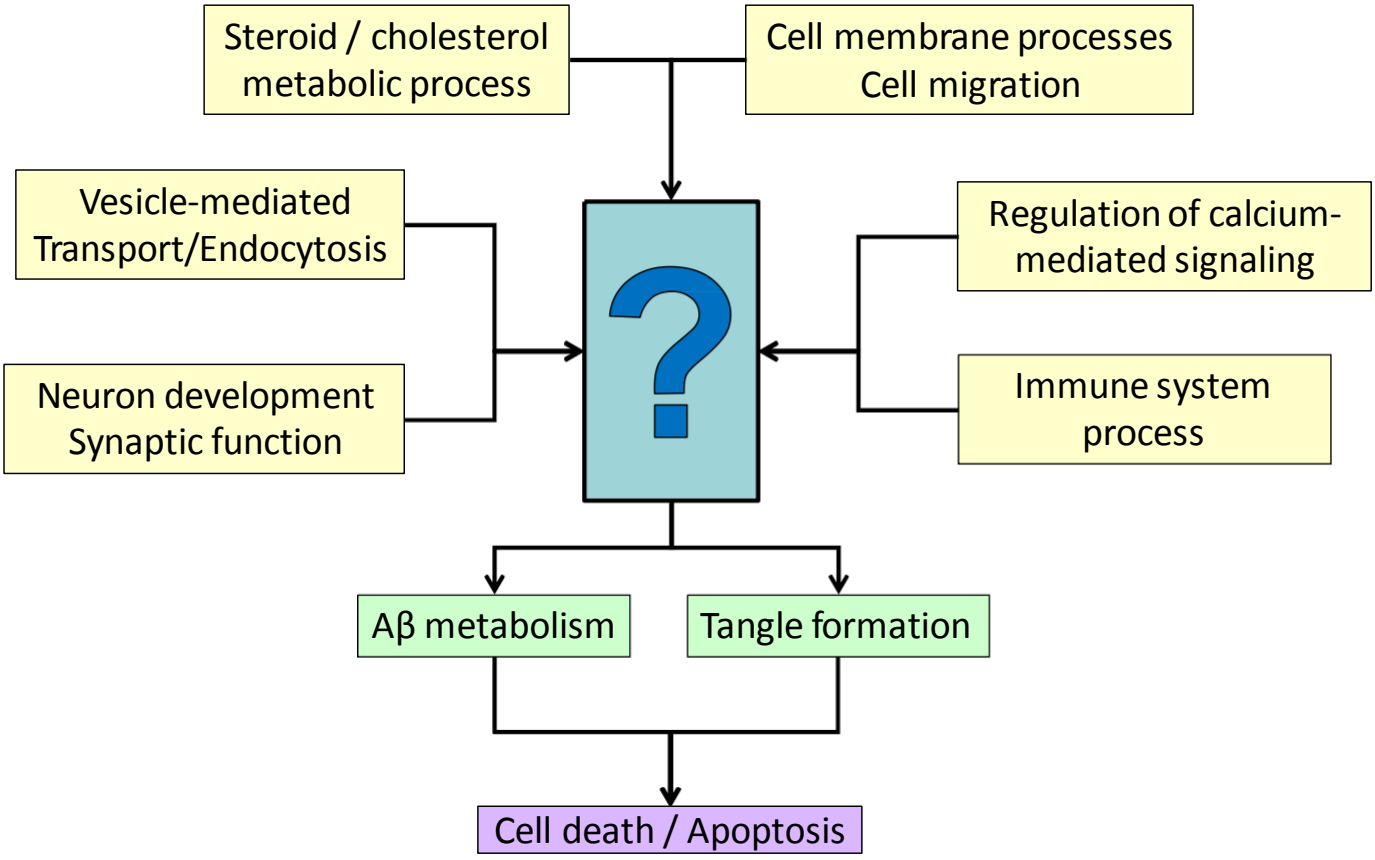

Figure 1: Using GO to explore how the 6 domains (in yellow) results in amyloid plaque and neurofibrillary tangle formation (in green) which leads to cell death (in purple).

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 steroid / cholesterol metabolic process domain

Gene	protein ID	steroid / cholesterol metabolic process
SORL1	Q92673	X
APOE	P02649	X
CLU	P10909	X
ABCA7	Q8IZY2	X

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 amyloiogenic 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

Gene	protein ID	Vesicle-mediated/ Transport / Endocytosis
SORL1	Q92673	X
SLC24A4	Q8NFF2	X
APOE	P02649	X
BIN1	O00499	X
CLU	P10909	X
ABCA7	Q8IZY2	X
PICALM	Q13492	X

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

Gene	protein ID	Immune system process
DRB5	Q30154	X
PTK2B	Q14289	X
INPP5D	Q92835	X
CLU	P10909	X
ABCA7	Q8IZY2	X
CR1	P17927	X
PICALM	Q13492	X
MS4A6A	Q9H2W1	X
CD33	P20138	X
EPHA1	P21709	X

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.

Gene	Protein ID	Cell membrane processes / cell migration
SORL1	Q92673	X
PTK2B	Q14289	X
APOE	P02649	X
BIN1	O00499	X
CD33	P20138	X
CD2AP	Q9Y5K6	X
EPHA1	P21709	X

Gene	Protein ID	Neuron development / synaptic function
PTK2B	Q14289	X
NYAP1	Q6ZVC0	X
MEF2C	Q06413	X
APOE	P02649	X
PICALM	Q13492	X

Gene	Protein ID	Regulation of calcium-mediated signaling
PTK2B	Q14289	X
SLC24A4	Q8NFF2	X
APOE	P02649	X

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:

Gene	protein ID	Aβ metabolism
SORL1	Q92673	X
CASS4	Q9NQ75	X
APOE	P02649	X
CLU	P10909	X
ABCA7	Q8IZY2	X

Gene	protein ID	Tangle formation
FERMT2	Q96AC1	X
CASS4	Q9NQ75	X

Gene	protein ID	cell death/apoptosis
SORL1	Q92673	X
PTK2B	Q14289	X
(MADD)	Q8WXG6	X
INPP5D	Q92835	X
MEF2C	Q06413	X
APOE	P02649	X
CLU	P10909	X

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

Barber, R. C. (2012). The Genetics of Alzheimer's Disease. *Scientifica*, *2012*.

Lambert, J. C., Ibrahim-Verbaas, C. A., Harold, D., Naj, A. C., Sims, R., Bellenguez, C., ... & Evans, D. (2013). Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease. *Nature genetics*.

Olgiati, P., Politis, A. M., Papadimitriou, G. N., De Ronchi, D., & Serretti, A. (2011). Genetics of late-onset Alzheimer's disease: update from the alzgene database and analysis of shared pathways. *International journal of Alzheimer's disease*, *2011*.