

## **InterMine Workshop Use-Case: Integrative analysis of a set of candidate genes for Alzheimer's disease:**

### **Introduction**

We will examine a set of genes from a hypothetical Alzheimer's disease genome wide association study and analyse them both in humanMine and in model organisms. The analysis will involve both some exploratory discovery and some candidate gene filtering. Note that this will not be an exhaustive analysis, but should give you an idea of the different ways you can use InterMine - if you were doing this for "real" you could include further data and analysis in your investigation.

We will also use the following information about Alzheimer's disease for our investigation:

1. **Amyloid plaques:** Plaques found in Alzheimer's patients brains consist predominantly of abnormal deposits of a protein fragment called beta-amyloid. This protein is formed from the breakdown of a larger protein called amyloid precursor protein (APP).
2. **Loss of neuronal connections and cell death:** The synaptic connections between certain groups of neurons stop functioning and begin to degenerate. The neurons eventually die.
3. **Cell adhesion molecules** mediate interactions between cells and their surroundings and are considered central to the pathogenesis and progression of Alzheimer's disease. App is involved in cell adhesion.

### **Section1: Upload the candidate set and convert to a set of genes:**

1. **Upload the candidate list:** The set of neurodegeneration proteins has been uploaded into HumanMine for you. Find this in the lists tab:
  - a. The list is called **InterMine\_workshop\_ListA**
  - b. You should have a list of **199 proteins**.
2. **Convert the list to a set of genes**
  - a. Navigate to the **List Analysis page** for InterMine\_workshop\_ListA. The set of identifiers you have are PROTEIN ids. You need to convert these to a set of genes - most data in humanMine and the other MOD-InterMines is attached to genes.

- b. In the top right hand corner of the list analysis page you will see “**Convert to a different type**”. Select Gene. This will give us a results table showing all the equivalent genes for our list of proteins.
- c. Select “**Save as list**” and save the set of Genes (the Protein > Genes set).
- d. You should now have a set of **197 genes (LIST B)**.

## **Section 2: Remove those already strongly associated with Alzheimers disease**

We are mostly interested in looking at potentially new associations of genes with Alzheimer’s disease, so we will remove from our list those that already have a strong Alzheimer’s annotation.

### **1. Find genes from our set (LIST B) annotated with an Alzheimer’s pathway and that have been associated with Alzheimer’s through GWAS (genome wide association studies)**

- A. Run the template **Gene → Pathway** with **LIST B**.  
*Navigate to the templates tab. If you can’t see the Gene → Pathway template then type “pathway” into the filter box. Click on the template name. To run with your list, select the checkbox and LIST B from the drop-down list. Select “Show results”.*
- B. Filter the “Pathways name” column for “Alzheimer” .  
*Click on the column summary icon. In the “filter values” box, type “Alzheimer”. Select the checkbox next to “Alzheimer’s disease” and “filter”. The table will be re-drawn with the 19 genes annotated to Alzheimer’s disease.*
- C. Save the 19 genes annotated to the Alzheimer’s disease pathway as a list.  
*Use the “Save as list” function to create a list of the 19 genes. **LIST C (19 genes)***
- D. Run the template **Gene → GWAS hit** on **LIST B**, with the default P-value, and filter the “Results Phenotype” column for “Alzheimer”. You can use the column summary to do this.  
*If you can’t see the template, try putting “GWAS” into the filter box. Remember to select LIST B from the drop-down. When you type “Alzheimer” into the filter box in the column summary you should see 5 different Alzheimer phenotypes. Select them all and filter the results. Your table should now have 20 rows, but note that only three genes are associated with these phenotypes. Use the “Save as list” function to save the 3 genes. **LIST D (4 genes)**.*
- E. Remove genes in **LIST C** and **LIST D** from the original gene list (**LISTB**).

To do this, combine **LIST C** and **LIST D** and then remove this combined list from our original gene list, **LIST B**, as follows:

- i. Navigate to the lists “view” tab. Select **LIST C** and **LIST D**. Select the “union” function, provide a new temporary list name (temp1, 22 genes) and save the combined list.
- ii. Select **LIST B** and your new temporary list. Select “Asymmetric Difference” and “list **LIST B** minus temp1”. Label your new list **LIST E**.

- F. This is the list we will now work with. It contains genes from our “screen” which are not obviously associated with Alzheimer’s disease.  
**LIST E (175 genes).**

### **Section 3: Exploratory Data Analysis in HumanMine:**

The first thing we want to do is explore our candidate set of genes. A good place to start is the list analysis page for an overall summary of our list. Click on **LIST E** to view its list analysis page

1. **List Analysis:** We will take a closer look at three aspects of the page:
  - A. **esyN Network Diagram:** This shows interactions between the genes in our list. Note that many of the genes interact with each other, suggesting many may operate in the same pathway.
  - B. **Gene Ontology Enrichment widget:** Scroll through the set of ontology terms enriched for this list. Note a lot of the initial terms are high-level terms and we may need to scroll a bit to find more specific annotations. We want to find information related to our set of Alzheimer’s criteria above. Scrolling down (you need to scroll quite a bit, or use CTRL+F to search) you will find “Neuron death” is enriched with a significant P-value. The “matches” column shows that 16 of the genes in our list are enriched with this term. Click on this number to create a list of this set of genes. (**LIST F, 16 genes**).  
*Click on the number 16 in the matches column. This will bring up a box with the 16 genes with that GO annotation. Click on “view results”. This will take you to a results table from where you can use the “Save as list” function to create a set of the genes.*
  - C. **Interactions:** Return to our main gene list (**LIST E**) and now take a look at the interactions widget. Unlike the ESN network diagram this widget shows interactions for genes in our list and other genes. The 4th entry in the widget shows that some of our genes interact with **amyloid beta precursor protein (APP)**. Knowing this to be a key player in Alzheimer’s disease, we are going to also investigate this set of genes further. As above, make a list of this set of genes (**LIST G, 22 genes**). *In this case, select the*

*checkbox for amyloid beta precursor protein and use the “view” link above to access the table.*

- D. Also, browse **other information** available on this page. We will not use it now, but if you were carrying out a full investigation you could take a look at all the data on this page. Don't forget there is also a set of pre-run templates towards the bottom of the page.

## 2. Cell adhesion association.

We will now see if any of the genes identified in our screen are associated with cell adhesion. To do this we will look at the Gene Ontology annotations through a template search:

- A. Run the template **Gene -> GO terms** with **LIST E**.
- B. Filter the “Parents Name” column for “Cell Adhesion”. NOTE that we are filtering the **parent** ontology term field (second to last column) rather than the “Ontology term name” column. This way we will return genes annotated to “Cell adhesion” AND any child terms of “Cell adhesion”.  
*Since there are a large number of gene ontology terms, it is necessary to use the filter function rather than the column summary to filter the results. Select the filter icon on the “Parents name” column. Click on “Add filter to ontology term name”. Type “Cell Adhesion” into the filter box - available terms should appear automatically. Finally, click “Add constraint”.*
- C. Use the “Save as list” function to create a list of the genes with Cell adhesion annotations (**LIST H, 25 genes**)

## 3. Further analysis of the sub-lists:

We have created three sub-lists of interesting genes from our initial list:

**List F:** Neuron death genes

**List G:** Interact with APP genes

**List H:** Cell adhesion genes

First, are there any genes in common between our lists?

- A. List Overlap Analysis: Examine the overlap between **LIST H** and **LIST F** (Cell adhesion list and neuron death list) and then **LIST H** and **LIST G** (Cell adhesion list and interact with APP list):  
*You can use the intersect operation for this on the lists “view” page. Intersect **LIST H** and **LIST F** to give **LIST HF** and **LIST H** and **LIST G** to give **LIST HG**. Union **LIST HF** and **LIST HG** to give list **LIST HFG**. Note that 3 and 5 genes from each set respectively*

*overlap with our cell adhesion set and this gives us 7 potential candidate genes that we will keep for possible further analysis. Can you find anything else interesting about these genes - in particular see the mouse section below.*

#### **Section 4. Model organism analysis - Fly.**

From our exploratory analysis above we have created three sets of interesting genes (**LIST F (neuron death genes)**, **LIST G (interact with APP genes)** and **LIST H (cell adhesion genes)**) that we are interested in analysing further, which include our 5 potential candidate genes in **LIST HFG** above. For the purpose of this exercise we are going to combine these sets and take a look at the equivalent genes in Fly and Mouse. (If you were carrying out a “real” analysis you may keep these sets distinct and analyse them separately).

##### **1. Create a combined candidate gene set**

###### **A. Create a combined candidate set of **LIST F**, **LIST G** and **LIST H**.**

*Navigate to the lists “view” page. Select the three lists and combine these (union) to create **LIST I (52 genes)**.*

##### **2. Analysis of the orthologous fly genes (FlyMine).** First we are going to analyse our set of genes in the fly.

###### **A. Navigate to the orthologous fly set in FlyMine:**

*From the list analysis page for **List I**, find the link to FlyMine under “View homologues in other mines”. This will open flyMine with the fly orthologues for our set of genes. You should have 74 genes. The list will be called **link\_x**. but you can rename it under your myMine account. **LIST J (62 genes)***

##### **3. Examine the list analysis page for LIST J.**

- A. Publications:** Navigate to the publications widget: as you scroll down notice a publication “Functional screening in Drosophila identifies Alzheimer's disease susceptibility genes and implicates Tau-mediated mechanisms”. Four of the genes from our candidate list are mentioned in this paper. These would therefore be potential genes to follow up.
- B. Protein Domains:** Which protein domains are enriched? Does this have any significance for our alzheimer's study?

**4. Examine the fly genes for involvement in cell adhesion and neuron death using the Gene Ontology:**

A. Run the template “**Gene → GO terms and parents of those terms**” with the fly list (**LIST J**).

B. Filter the results for the terms “Neuron death” and “Cell adhesion”.

Use the **Parent’s Name** column for your filter.

*Note this will require two separate filter steps on the original list. You will have to use the filter function rather than the column summary. Remember you will need to “undo” your first filter before adding the second. Alternatively you could try altering this query using the query builder - you can add both constraints to the parents ontology term name field and modify the constraint logic (A and (B or C)).*

C. Create new lists of the filtered genes (**LIST K (2 genes, Neuron death) and LIST L (16 genes, Cell adhesion)**). *(This will be one step if you used the query builder as above).*

D. Combine the two filtered lists to create **LIST M (9 genes)**.

**5. List analysis of LIST M - fly candidate set.** Examine the list analysis page for this list. We are looking for any potential candidates that may be worth following up experimentally in the fly. Note that the list includes cell adhesion genes and members of the wnt signaling pathway, including the Armadillo protein:

A. **Interactions:** Note that the interaction network for this set of genes is much smaller than for the human set but that interactions between 4 of the genes in our list is shown in the Esyn network. These genes may be interesting to follow up you could make a list of these and examine further.

B. **Other data:** Look at the other data on the page. Is there anything else we could use to narrow down our set of genes to those we wish to study further. For example, are any of the genes expressed in the brain? If you have time, you could run some template searches too.

**Section 5: Analysis in the mouse.**

List HG contains five genes involved in cell adhesion that also interact with APP. Examine this set in mouseMine. Have these interactions been studied in the mouse? In particular look at the interaction and publication data in mouseMine.

Use your knowledge of InterMine to explore the other candidate sets in the mouse (starting either from humanMine **LIST I** or FlyMine **LIST J**). Use the list analysis page, templates and list operations to filter the set to a potentially interesting set for further study. You could also take a look at other model organisms if you have time.

---

List	Count	Contents
List A	199 proteins	original set of proteins
List B	197 genes	original set of genes
List C	19 genes	genes from original set in Alzheimer's pathway
List D	4 genes	genes from original set associated with Alzheimer's via GWAS
List E	173 genes	original set minus any genes already associated with Alzheimer's
List F	16 genes	genes annotated with "neuron death" GO term
List G	22 genes	genes that interact with gene known to be involved in Alzheimer's
List H	25 genes	genes annotated with "Cell adhesion" GO term
List I	52 genes	Union of Lists F, G and H
List J / link_1	62 genes (FlyMine)	List I from HumanMine viewed in FlyMine
List K	2 genes (FlyMine)	Genes annotated with Neuron Death GO term
List L	16 genes (FlyMine)	Genes annotated with Cell adhesion GO term
List M	9 genes (FlyMine)	Union of List K and L