Introduction

According to 2011 survey Alzheimer's disease (AD) afflicts around 24 million people worldwide. Alzheimer’s disease (AD) is a multifactorial disease and involves several different mechanisms. It is characterised by memory loss and cognitive dysfunction. Examination of sufferer’s brains reveals abundance of two neuro-pathological features – senile plaques (β-amyloid (Aβ)) and neurofibrillary tangles (abnormally phosphorylated tau (P-tau)), which are thought to be central to disease pathogenesis.

HSV-1 is a neurotropic virus that infects most human population. However, HSV-1 infection is common only in the peripheral nervous system; at least 80% of the population have been infected with HSV-1 by age 60. HSV-1 can undergo a latent infection in which no viral particles are produced and HSV-1 gene expression is limited. Stimuli such as stress and immune-suppression can reactivate the virus, leading to a productive infection. It is also found that in later life cycle, virus spreads and becomes latent in the brain and during events such as immune-suppression, head trauma and peripheral infection; it reactivates and causes damage, including Aβ and P-tau formation. After each reactivation event, HSV-1 would spread (via intraneuronal pathways) and would thus become latent in additional cells. Repeated reactivation would lead to further spread and cumulative damage, which would develop into AD. Several researches demonstrated that the Herpes simplex virus-1 (HSV-1) is present in the brain of most elderly people with and without AD pathology and is also present within the brain regions affected by AD.

Moreover, HSV-1 is found to be a strong risk factor for the disease and its severity increases with specific genetic factor – the type 4 allele of the apolipoprotein E gene. However, APOE-4 is risk factor for less than 50% population; rest of the population is having other genetic factors which are still not known. HSV-1 is responsible for a no. of diseases including Ocular infection (Keritits), Herpes Labialis (cold sores), Herpes simplex encephalitis (HSE) and some cases of Genital herpes.

Further, Aβ and P-tau accumulation are not specific to HSV-1. Indeed, other pathogens can produce AD-like changes – for example HIV causes Aβ and P-tau formation and measles virus causes neurofibrillary tangles formation. Moreover, Varicella zoster virus (VZV) belongs to HSV family and is also neurotropic virus and reside latently in the peripheral nervous system (PNS) but no connection has been found between VZV and AD. So, currently only HSV-1 has been detected in normal and AD brains and so it is the only candidate viral agent.

Interaction of viral and host proteins suggests that some form of synergy is present between the pathogen and host genetic factors. This synergy is playing a role in the pathology of sporadic form of Alzheimer's disease. However, HSV-1 infection is not the only cause of the disease, but it participates in the pathogenic process of Alzheimer.

Chapter-1 Protein-Protein Interaction Network

1.1 Introduction

Proteins are the main catalysts, structural elements, signalling messengers and molecular machines of biological tissues.[1] Protein–protein interactions (PPIs) are extremely important in orchestrating the events in a cell.[2] A fundamental challenge in human health is the identification of disease-causing genes. Recently, several studies have tackled this challenge via a network-based approach and observed that genes causing the same or similar diseases tend to lie close to one another or possibly are neighbours in network of protein-protein interactions.[3] Comprehensive knowledge of protein-protein interactions provides a framework for understanding the biology of complex diseases as an integrated system.[4] Understanding the genetic background of diseases is crucial to medical research with implications in diagnosis, treatment and drug development. Molecular approaches to this challenge are time consuming and costly, computational approaches offer an efficient alternative.[3] The availability of complete and annotated genome sequences of several organisms has led to a paradigm shift from the study of individual proteins in an organism to large-scale proteome-wide studies of proteins, which interact in a complex network of metabolic, signalling and regulatory pathways in a cell. Several efforts have been made to identify these interactions, in an attempt to understand biological systems better.[5, 6]

Recent advances in high throughput genome-wide screening techniques have increased not only the amount of generated data, but also its quality. However, as often is the case in life science, the devil is in the details.[7] Meanwhile, hoping to solve this problem, researchers have been broadening their view and have been looking elsewhere to solve these problem in simplified way. One of these is the field of network science. This new field has emerged from graph theory and has proved to be a powerful method for the mathematical representation, visualization and analysis of complex data that involves many interacting components. In this area powerful concepts have been developed, such as network centrality and network motifs that have enabled us to understand a system through its network topology.[8, 9] Subsequently many fields have benefited from these advances. For example in epidemiology the mapping of human interactions into social networks gave insight into how sexually transmitted disease spread in a population.[10,11] In developmental biology the representation of interactions among different genes as gene regulatory networks has been widely accepted.[12, 13] However, the field of virology has not yet received the full attention it deserves from network research, despite the availability of abundant data and ready to use scientific methodology.[14]

1.2 Methods to construct PPI

With the advent of high-throughput experiments to identify PPIs, more knowledge on protein function has been obtained, together with the development of several methods to predict and study the interactions between proteins. A wide variety of methods have been used to identify protein–protein associations; these associations may range from direct physical interactions inferred from experimental methods to functional linkages predicted on the basis of computational analyses. These methods range from identifying a single pair of interacting proteins at one end, to the identification and analysis of a large network of thousands of proteins, the latter as large as that of an entire proteome of a given cell.[15]

There are two ways to construct protein-protein interaction network and they are as follow:

- ✓ **Experimental methods**
- ✓ **Computational methods**

1.2.1 Experimental methods

There are a number of experimental techniques such as:

- ✓ Yeast-two hybrid.[16]
- ✓ Affinity purification/ mass spectrometry.[5]
- ✓ Protein microarrays.[17] and many more.

These form the basis of several large-scale datasets on PPIs. In the yeast-two hybrid assay, two fusion proteins are created: the ‘bait’ (a protein of interest with a DNA binding domain attached to its N-terminus) and the ‘prey’ (its potential interaction partner, fused to an activation domain). If the ‘bait’ and the ‘prey’ interact, their binding forms a functional transcriptional activator, which in turn activates reporter genes or selectable markers.[16] This assay has been adapted for high-throughput analyses of PPIs.[18]

Gavin and collaborators have described the purification of complexes of 1739 proteins from *S. cerevisiae* (including the complete set of 1143 human orthologous) using tandem affinity purification coupled to mass spectrometry, illustrating the complexity of connectivity between protein complexes.[5]

Protein microarrays aid in the detection of in-vitro binary interactions of various type such as: protein–protein, protein–lipid or antigen–antibody interactions. Proteins covalently attached to a solid support are screened with fluorescently labelled probes (proteins or lipids), to identify interactions. A high density yeast protein microarray comprising 5800 yeast proteins was developed and used to identify novel calmodulin and phospholipid binding proteins.[17] Although many of these assays can identify PPIs with high confidence, they still have their share of false positives and can suffer from a limited reproducibility.[15]

1.2.2 Computational methods

There are two ways:

- ✓ **When protein-protein interaction data is not available**
- ✓ **When protein-protein interaction data is available.**

When protein-protein interaction data is not available then interaction between two or more genes/proteins can be derived from genomic context. This can be done by any of following methods:

- ✓ **Domain Fusion/ Rosetta Stone**

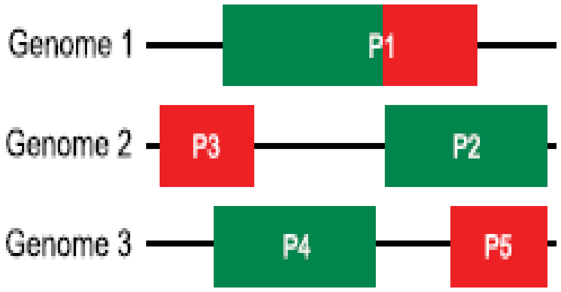


Fig: 1.1 Showing Domain Fusion/Rosetta Stone method. Each protein is shown with boxes representing domains. Protein P3 and P2 in genome 2 and 3 are predicted to interact with each other because their homologues are fused in genome 1.[19]

✓ Gene Neighborhood

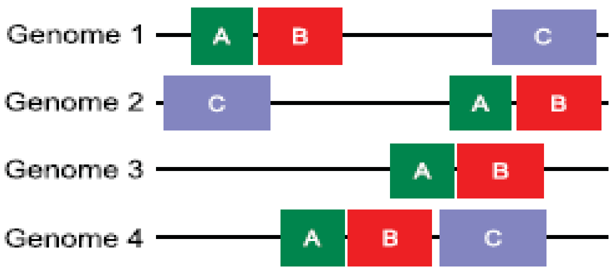
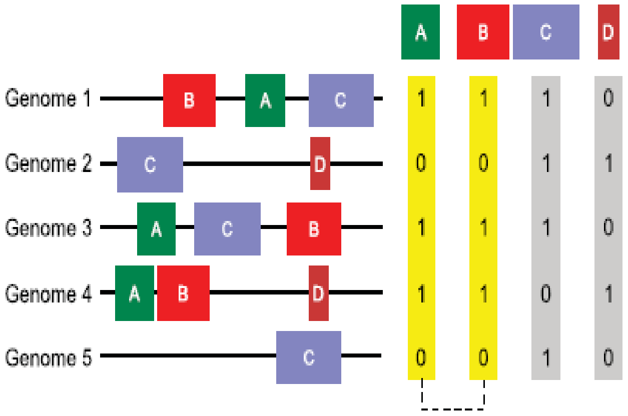


Fig: 1.2 Showing Gene Neighborhood method. There are four hypothetical genomes, containing one or more of the genes A, B and C. Since the genes A and B are co-localised in multiple genomes (1–4), they are likely to be functionally linked with one another.[20]

✓ Phylogenetic Analysis



	A	B	C	D
Genome 1	1	1	1	0
Genome 2	0	0	1	1
Genome 3	1	1	1	0
Genome 4	1	1	0	1
Genome 5	0	0	1	0

Fig: 1.3 Showing Domain Fusion/Rosetta Stone method. There are5 hypothetical genomes, each containing combination of proteins A, B, C and D. The presence or absence of each protein is indicated by 1 or 0, respectively in phylogenetic profiles. Proteins A and B are functionally linked (dotted line), whereas proteins C and D, which have different phylogenetic profiles (shown in grey) are not likely to be functionally linked.[21]

✓ Correlated Mutation

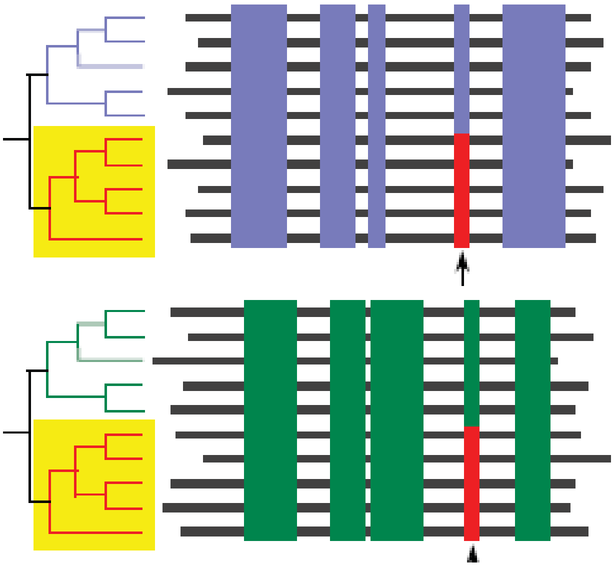


Fig: 1.4 Showing Domain Fusion/Rosetta Stone method. The alignments of two protein families are shown; conserved residues in either alignment are shown in the same colour (blue and green). Correlated mutations in either alignment (coloured red) are indicated by arrow marks. Common sub-trees of the phylogenetic trees are highlighted in yellow. The presence of correlated mutations in each family suggests that the corresponding sites may be involved in mediating interactions between the proteins from each family.[22,23]

When protein-protein interaction data is available then it can be simply be taken from databases (house of PPI data) or literature and by using network generation tools network can be created, visualized and analyzed.

✓ **List of PPI databases.**

a. DIP (**D**atabase of **I**nteracting **P**roteins) [24]

b. **B**iomolecular **I**nteraction **N**etwork **D**atabase (BIND) [25]

c. IntAct [26]

d. **M**olecular **Int**eractions Database (MINT) [27]

e. **H**uman **P**rotein **R**eference **D**atabase (HPRD) [28] and many more.

✓ **List of tools for network generation, visualization and analysis.**

a. Cytoscape (http://www.cytoscape.org/) [29]

b. Pajek (http://pajek.imfm.si/) [30]

c. PINA or **P**rotein **I**nteraction **N**etwork **A**nalysis **P**latform (http://cbg.garvan.unsw.edu.au/pina/) and many more [31]

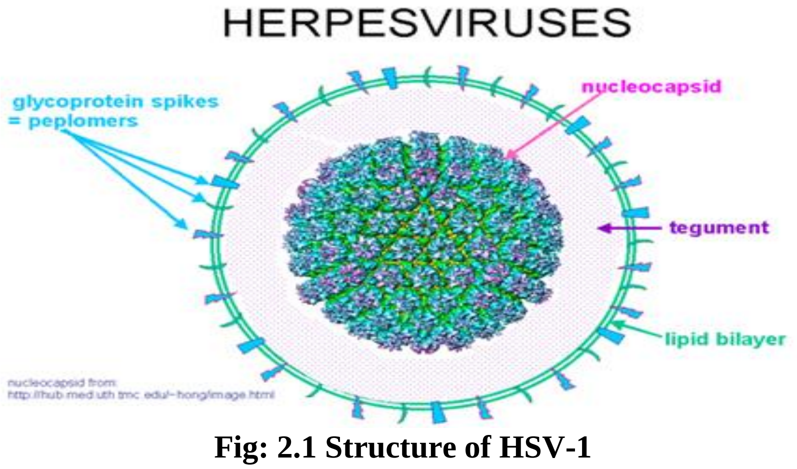
Chapter-2 Herpes Simplex Virus-1

2.1 Introduction

Herpes simplex virus also known as Human herpes virus is members of the herpes virus family, Herpesviridae, that infect humans. Herpes-viruses constitute a family of large DNA viruses widely spread in vertebrates and causing a variety of different diseases. They possess ds-DNA genomes ranging from 120 to 240 Kbp encoding between70 to 74 open reading frames.[32] Herpes simplex is divided into two types: HSV type 1 and HSV type 2.[33] HSV1 primarily causes mouth, throat, face, eye, and central nervous system infections, while HSV2 primarily causes genital infections.[33]

2.2 Viral Structure

The structure of herpes viruses consists of a relatively large double-stranded, linear DNA genome encased within an icosahedral protein cage called the capsid, which is wrapped in a lipid bilayer called the envelope. The envelope is joined to the capsid by means of a tegument. This complete particle is known as the virion.[34] HSV-1 and HSV-2 each contain at least 74 genes (or open-reading frames, ORFs) within their genomes, though speculation over gene crowding allows as many as 84 unique protein coding genes by 94 putative ORFs. These genes encode a variety of proteins involved in forming the capsid, tegument and envelope of the virus, as well as controlling the replication and infectivity of the virus.[35, 36]



nucleocapsid from:
http://hub.med.uth.tmc.edu/~hong/image.html

Fig: 2.1 Structure of HSV-1

The genomes of HSV1 and HSV2 are complex and contain two unique regions called the long unique region (U_L) and the short unique region (U_S). Of the 74 known ORFs, U_L contains 56 viral genes, whereas U_S contains only 12. Transcription of HSV genes is catalysed by RNA polymerase II of the infected host. Immediate early genes, which encode proteins that regulate the expression of *early* and *late* viral genes, are the first to be expressed following infection. Early gene expression, allow the synthesis of enzymes involved in DNA replication and the production of certain envelope glycoproteins. Expression of late genes occurs last; this group of genes predominantly encode proteins that form the virion particle.[35]Five proteins from (U_L) form the viral capsid; UL6, UL18, UL35, UL38 and the major capsid protein UL19.[36,37]

2.3 **Cellular Entry**

Entry of HSV into the host cell involves interactions of several glycoproteins which are present on the surface of the enveloped virus, with receptors on the surface of the host cell. The envelope covering the virus particle, when bound to specific receptors on the cell surface, will fuse with the host cell membrane and create an opening or *pore*, through which the virus enters the host cell.[38]

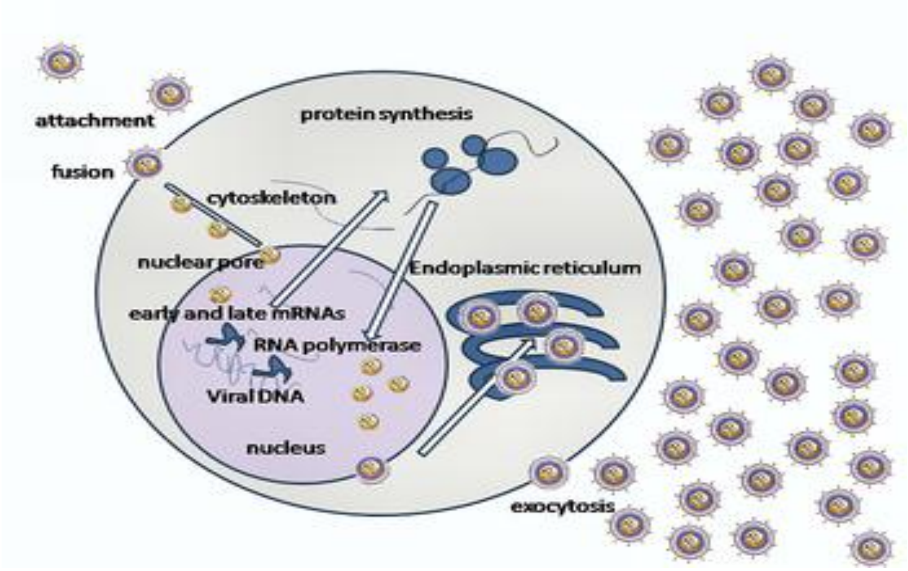


Fig: 2.2 Mode of cellular entry of HSV-I

The sequential stages of HSV entry are analogous to those of other viruses.

- At first, complementary receptors on the virus and the cell surface bring the viral and cell membranes into close proximity.
- In an intermediate state, the two membranes begin to merge, forming a *hemifusion state.*
- Finally, a stable *entry pore* is formed through which the viral envelope contents are introduced to the host cell.

In the case of a herpes virus, initial interactions occur when a viral envelope glycoprotein called glycoprotein C (gC) binds to a cell surface particle called heparansulfate.[39] A second glycoprotein, glycoprotein D (gD), binds specifically to at least one of three known entry receptors. These include herpesvirus entry mediator (HVEM), nectin-1 and 3-O sulphated heparansulfate. The receptor provides a strong, fixed attachment to the host cell. These interactions bring the membrane surfaces into mutual proximity and allow for other glycoproteins embedded in the viral envelope to interact with other cell surface molecules. Once bound to the HVEM, gD changes its conformation and interacts with viral glycoproteins H (gH) and L (gL), which form a complex. The interaction of these membrane proteins results in the hemi-fusion state. Afterward, gB interaction with the gH/gL complex creates an entry pore for the viral capsid. Glycoprotein B interacts with glycosaminoglycans on the surface of the host cell.[38,39,40]

After the viral capsid enters the cellular cytoplasm, it is transported to the cell nucleus. Once attached to the nucleus at a *nuclear entry pore*, the capsid ejects its DNA contents via the capsid *portal*. The capsid portal is formed by twelve copies of portal protein, UL6, arranged as a ring; the proteins contain a leucine zipper sequence of amino acids which allow them to adhere to each other.[41] Each icosahedral capsid contains a single portal, located in one vertex.[42,43] The DNA exits the capsid in a single linear segment.[44]

2.4 Immune Evasion

HSV evades the immune system through interference with MHC class I presentation of antigen on the cell surface.[45] It achieves this through blockade of the TAP transporter induced by the secretion of ICP-4 by HSV. TAP maintains the integrity of the MHC class I molecule before it is transported via the Golgi apparatus for recognition by CD8+ CTLs on the cell surface. ICP-47 disrupts this integrity, preventing the capture of cytosolic proteins for CTL recognition and thus evades CTL destruction.[46]

2.5 Replication

Following infection of a cell, herpes virus proteins, called *immediate-early*, *early*, and *late*, are produced. Research indicates the possibility of an additional lytic stage, *delayed-late*.[13] These stages of lytic infection, particularly *late lytic*, are distinct from the latency stage. In the case of HSV-1, no protein products are detected during latency, whereas they are detected during the lytic cycle.[47]

The early proteins transcribed are used in the regulation of genetic replication of the virus. On entering the cell, α-TIF protein joins the viral particle and aids in immediate-early transcription. The virion host shutoff protein (VHS or UL41) is very important to viral replication.[48] This enzyme shuts off protein synthesis in the host, degrades host mRNA, helps in viral replication, and regulates gene expression of viral proteins. The viral genome immediately travels to the nucleus but the VHS protein remains in the cytoplasm.[49,50]

The late proteins are used in to form the capsid and the receptors on the surface of the virus. Packaging of the viral particles — including the genome, core and the capsid - occurs in the nucleus of the cell. Here, concatemers of the viral genome are separated by cleavage and are placed into pre-formed capsids. HSV-1 undergoes a process of primary and secondary envelopment. The primary envelope is acquired by budding into the inner nuclear membrane of the cell. This then fuses with the outer nuclear membrane releasing a naked capsid into the cytoplasm. The virus acquires its final envelope by budding into cytoplasmic vesicles.[51]

Major obstacle in HSV treatment is its reoccurrence. HSVs may persist in a quiescent but persistent form known as *latent infection*, notably in neural ganglia[32].HSV-1 tends to reside in the trigeminal ganglia, while HSV-2 tends to reside in the sacral ganglia. During such latent infection of a cell, HSVs express Latency Associated Transcript (LAT)RNA. LAT is known to regulate the host cell genome and interferes with natural cell death mechanisms. By maintaining the host cells, LAT expression preserves a reservoir of the virus, which allows subsequent, usually symptomatic, periodic recurrences or "outbreaks" characteristic of non-latency. Whether or not recurrences are noticeable (symptomatic) or not, viral shedding occurs to produce further infections. A protein found in neurons may bind to herpes virus DNA and regulate latency. Herpes virus DNA contains a gene for a protein called ICP4, which is an important transactivator of genes associated with lytic infection in HSV-1.[52] Elements surrounding the gene for ICP4 bind a protein known as the human neuronal protein or Neuronal Restrictive Silencing Factor (NRSF) or human Repressor Element Silencing Transcription Factor (REST). When bound to the viral DNA elements, histone deacetylation occurs atop the ICP4 gene sequence to prevent initiation of transcription from this gene, thereby preventing transcription of other viral genes involved in the lytic cycle.[52, 53] Another HSV protein reverses the inhibition of ICP4 protein synthesis. ICP0 dissociates NRSF from the *ICP4* gene and thus prevents silencing of the viral DNA.[54]

Chapter-3 Alzheimer

3.1 Introduction

Alzheimer's disease (AD) is a subset of neurodegenerative diseases and is common form of dementia. Currently there is no cure for the disease; however extensive research is going in this field. Dementia becomes worse as it progresses, and eventually leads to death.[55] Most often, it is diagnosed in people over 65 years of age, although the less-prevalent early-onset Alzheimer's can occur much earlier.[56]

3.2 Characteristics and Classification

Symptoms includes stress, difficulty in remembering recent events, confusion, irritability and aggression, mood swings, trouble with language, and long-term memory loss.[57,58]

The disease course is divided into four stages, with progressive patterns of cognitive and functional impairments.

- Pre-dementia
- Early
- Moderate
- Advanced

Pre-dementia: The first symptoms are often attributed to aging or stress.[57,59] The most noticeable deficit is memory loss, which shows up as difficulty in remembering recently learned facts and inability to acquire new information.[60,61]

Early: In people with AD the increasing impairment of learning and memory eventually leads to a definitive diagnosis. AD does not affect all memory capacities equally. Older memories of the person's life are affected to a lesser degree than new facts or memories.[62,63]

Moderate: Progressive deterioration eventually hinders independence; with subjects being unable to perform most common activities of daily living.[64] Speech difficulties become evident due to an inability to recall vocabulary, which leads to frequent incorrect word substitutions (paraphasias). During this phase, memory problems worsen, and the person may fail to recognise close relatives.[64]

Advanced: During this last stage of AD, the person is completely dependent upon caregivers [64]. Language is reduced to simple phrases or even single words, eventually leading to complete loss of speech.[64,65]

3.3Theories related to Alzheimer

The cause for Alzheimer is still essentially unknown (except for 1% to 5% of cases where genetic differences have been identified).[66]

Several hypotheses exists to explain the cause of the disease.

- The oldest is the ***cholinergic hypothesis,*** which proposes that AD is caused by reduced synthesis of the neurotransmitter acetylcholine.[67]
- The ***amyloid hypothesis*** postulated, that amyloid beta (Aβ) deposits are the fundamental cause of the disease.[68,69]
- Next is ***tau hypothesis,*** which supports the idea that tau protein abnormalities initiate the disease cascade. In this model, hyper-phosphorylated tau begins to pair with other threads of tau. Eventually, they form neurofibrillary tangles inside nerve cell bodies.[70] When this occurs, the microtubules disintegrate, collapsing the neuron's transport system. This may result first in malfunctions in biochemical communication between neurons and later in the death of the cells.[71]

3.4 Risk factors

There are several risk factors associated with Alzheimer, combination of 2 or more risk factors is sufficient to initiate the pathology of HSV-1. These risk factors are as follow:

- ✓ Age; For persons between ages 65–69, 70–74, 75–79, 80–84, and 85 and older the incidence of AD has been estimated at 0.6%, 1.0%, 2.0%, 3.3%, and 8.4%. [72]
- ✓ Family history; a person reported to have family history for Alzheimer is more susceptible for disease.[73]
- ✓ Sex; female are more susceptible to suffer from Alzheimer.[74]

- ✓ Head trauma.[73]
- ✓ Vascular diseases.[73]
- ✓ Genetic factors; APOE e4 allele increases the chances of Alzheimer.[75]
- ✓ Environmental risk factors; HSV-1 is found to be strong risk factor for initiating Alzheimer pathology.[76]
- ✓ Stress inside cell.[73]

3.5 Genetics

The vast majority of cases of Alzheimer's disease are sporadic, meaning that they are not genetically inherited, although some genes may act as risk factors. On the other hand, around 0.1% of the cases are familial forms of autosomal dominant (not sex-linked) inheritance, which usually have an onset before age 65. This form of the disease is known as Early onset familial Alzheimer's disease [77].

Most of autosomal dominant familial AD can be attributed to mutations in one of three genes: Amyloid Precursor Protein (APP) and Presenilins 1 (PSEN1) and 2 (PSEN2).Most mutations in the APP and Presenilin genes increase the production of a small protein called Aβ42, which is a toxic compound for neurons and is main component of senile plaques [78].

Most cases of Alzheimer's disease do not exhibit autosomal-dominant inheritance and are termed sporadic AD. Nevertheless genetic differences may act as risk factors. The best known and widely studies genetic risk factor is ε4 allele of the Apolipoprotein E (APOE) [79].

3.6 Mechanism

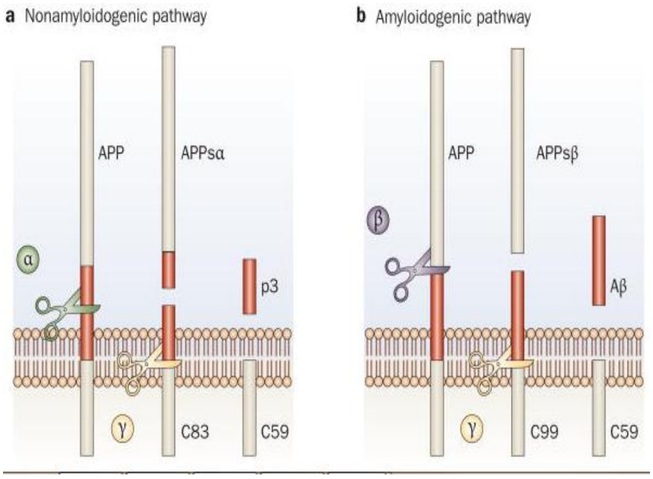


Fig-3.1 Showing production of toxic A-beta by APP.

There are 3 classes of secretases (enzyme that cleaves APP); Alpha, beta and gamma. In non-amyloidogenic pathway APP is firstly cleaved by Gamma secretase and then by alpha; releasing APPsa (non-toxic). In amyloidogenic pathway alpha secretase is not available to cut APP then in this case, beta secretase cuts the APP at site different from where alpha cuts, then gamma cuts at same site where it cuts APP during this process a small fragment is released this is known as amyloid beta and is toxic for cell [80].

Chapter-4 Methodology

Methodology is divided into following three phases:

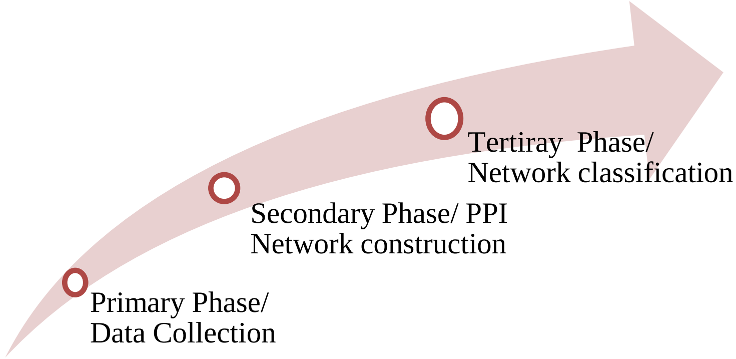
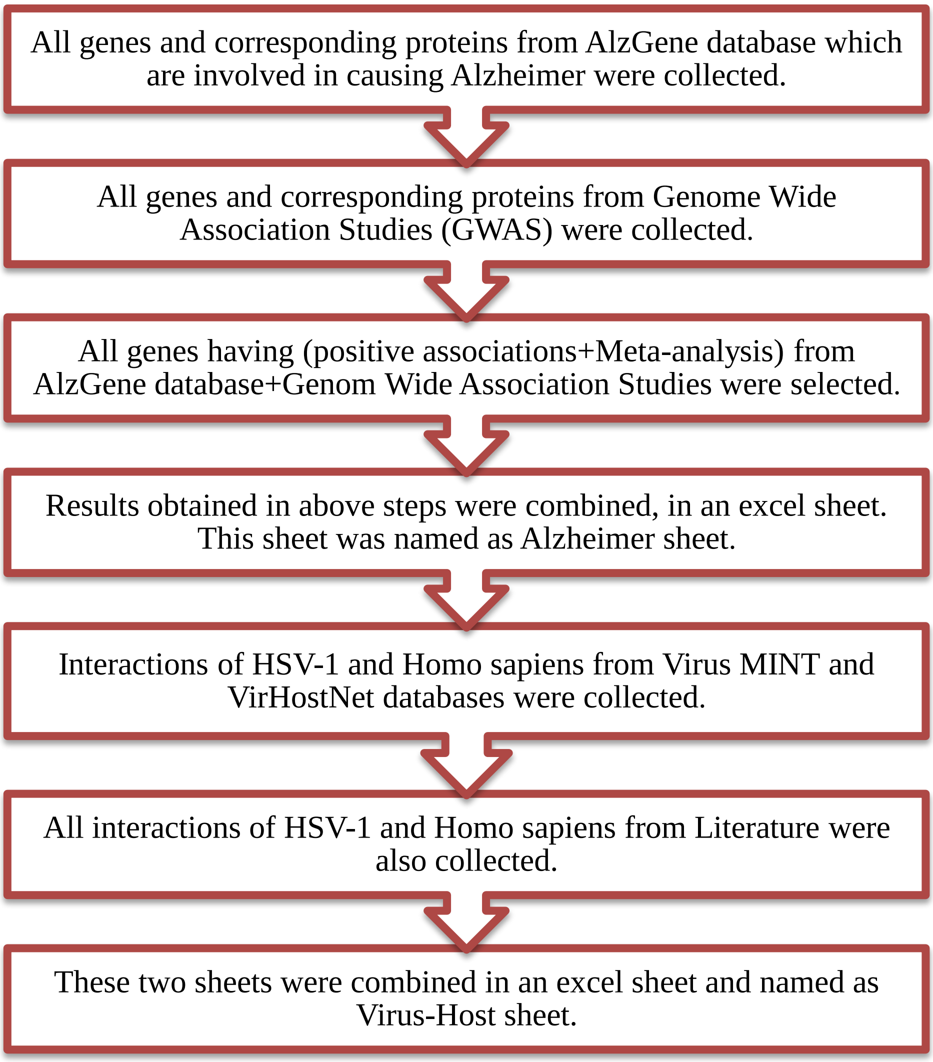


Fig-4.1 General Overview of Methodology



All genes and corresponding proteins from AlzGene database which are involved in causing Alzheimer were collected.

All genes and corresponding proteins from Genome Wide Association Studies (GWAS) were collected.

All genes having (positive associations+Meta-analysis) from AlzGene database+Genom Wide Association Studies were selected.

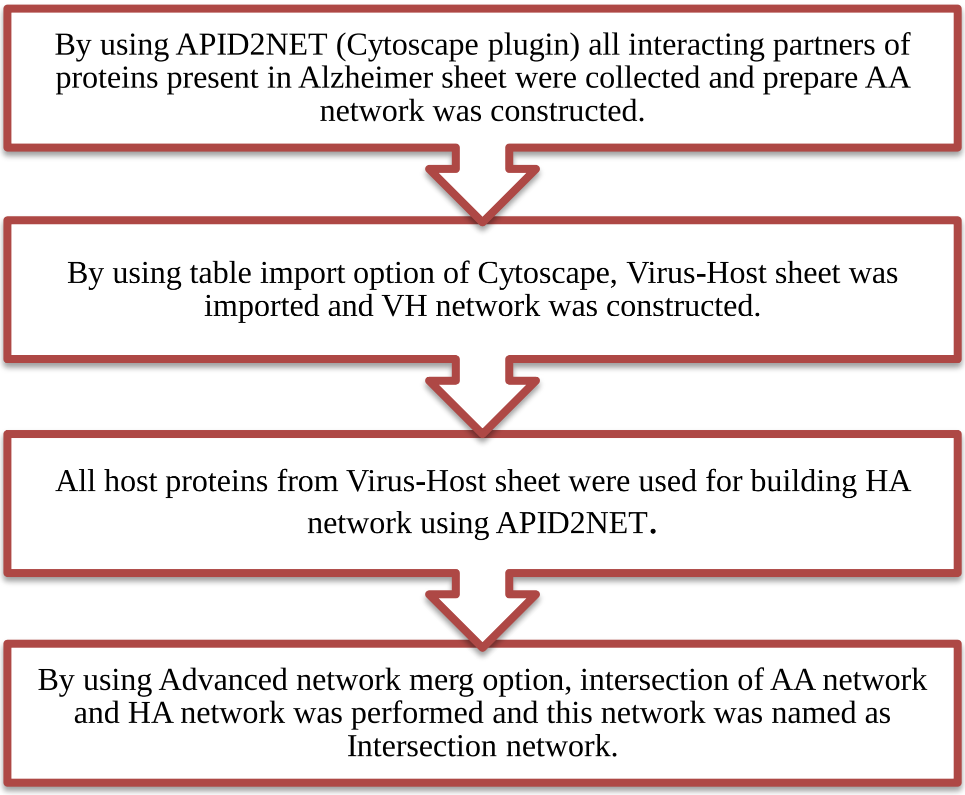
Results obtained in above steps were combined, in an excel sheet. This sheet was named as Alzheimer sheet.

Interactions of HSV-1 and Homo sapiens from Virus MINT and VirHostNet databases were collected.

All interactions of HSV-1 and Homo sapiens from Literature were also collected.

These two sheets were combined in an excel sheet and named as Virus-Host sheet.

Fig-4.2 Primary Phase or Data Collection



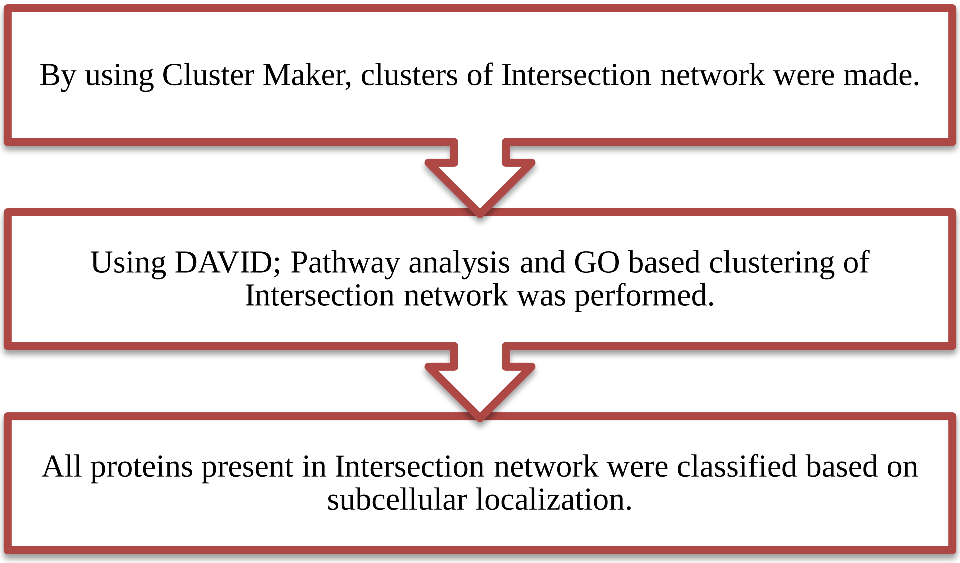
By using APID2NET (Cytoscape plugin) all interacting partners of proteins present in Alzheimer sheet were collected and prepare AA network was constructed.

By using table import option of Cytoscape, Virus-Host sheet was imported and VH network was constructed.

All host proteins from Virus-Host sheet were used for building HA network using APID2NET.

By using Advanced network merg option, intersection of AA network and HA network was performed and this network was named as Intersection network.

Fig-4.3 Secondary Phase or Network Construction



By using Cluster Maker, clusters of Intersection network were made.

Using DAVID; Pathway analysis and GO based clustering of Intersection network was performed.

All proteins present in Intersection network were classified based on subcellular localization.

Fig: 4.4 Tertiary Phase or Network Classification

4.1 Primary phase or Data collection

4.1.1 Alzheimer data collection from AlzGene database

a. The AlzGene database provides a comprehensive, unbiased and regularly updated field synopsis of genetic association studies performed in Alzheimer disease. In addition, hundreds of **meta-analyses** are also available for all eligible official polymorphisms with sufficient data. It can be freely accessed by visiting their website: http://www.alzgene.org/

   The database contain summary of 695 genes which are found to be associated with Alzheimer, besides, it also includes pseudo-genes, some non-coding RNA and some mitochondrion non-coding genes.

b. From database detail of every gene is collected chromosome vise. Detail includes Gene symbol,Entrez id, Chromosome location, Linkage region and UniProt id.

c. Database stated that there were 695 genes but after checking database 2 times we found only 685 genes and after excluding pseudo-genes, non-coding RNA and mitochondrion non-coding genes we are left with 636 genes.

d. For those 636 gene non-family based number of positive association, number of negative association, number of trends and family based number of positive association, number of negative association, number of trends are collected, these associations were taken for Caucasian population because they are genetically very similar to Indians, rest all populations were ignored.

e. The database also provides meta-analysis for some genes, all these meta-analysis and their polymorphism along with the allele frequency were collected. The meta-analysis studies showed that these polymorphisms play an important role in Alzheimer Disease.

f. Using this information an excel sheet was prepared and was named this sheet as **AlzGene sheet** (Appendix Table 1).

g. For each gene, number of positive association and number of negative association were added to give net result for both non-family based studies and family based studies.

h. If the net result was found to be positive then that particular gene was selected for further studies and if net result was negative then that gene was discarded. The following formula could better explain this:

If, (**N**=positive) **or** (**F**=positive) **or** (**M**=Y)

Then, Select that particular gene

N= Net result of non-family based studies, this come out to be 105

F= Net result of family based studies, this come out to be 26

M= Meta-analysis, this come out to be 154

i. Now all the Genome Wide Association Studies (GWAS) were collected from the same database and each association was checked from corresponding literature.

j. In the database GWAS was available for 60 genes, details of every gene was collected. Details includethe Gene symbol, Entrez id, Chromosome location, Linkage region, Uniprot id.

k. From this data an excel sheet was prepared and named the sheet as **GWAS sheet** (Appendix Table 2).

l. Final sheet was prepared by combining Alz Gene sheet and GWAS sheet and after that all duplicate genes were removed.

m. After removing all duplicate values we were left with 239 genes. From these genes final sheet was prepared.

n. This sheet was named as **Alzheimer sheet** (Appendix Table 3).

4.1.2 Collection of Virus-Host interactions

i. From Virus MINT

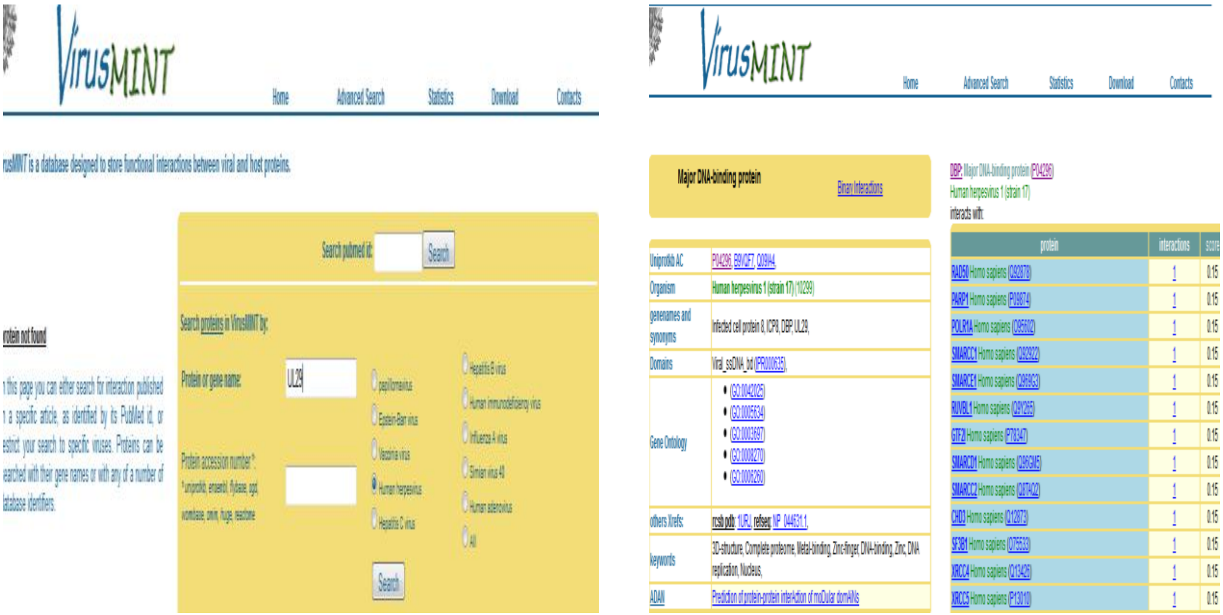
a. Virus MINT (**M**olecular **INT**eraction) is a database of virus and host interactions, it can be accessed freely through the following website: http://mint.bio.uniroma2.it/virusmint/

b. Virus MINT aims at collecting and annotating (in a structured format) all interactions between human and viral proteins.

c. It was searched for retrieval of Herpes simplex virus-1 and Homo sapiens specific interactions.

d. In order to collect interactions from this database following procedure was followed:

Open home page > Enter Gene name > Select Human Herpesvirus> Click on Search



rusMINT is a database designed to store functional interactions between viral and host proteins.

rotein not found

n this page you can either search for interaction published n a specific article, as identified by its PubMed id, or estrict your search to specific viruses. Proteins can be earched with their gene names or with any of a number of latabase identifiers.

Search pubmed id:

Search proteins in VirusMINT by:

Protein or gene name: UL29

Protein accession number*:
*uniprotkb, ensembl, flybase, sgd, wormbase, omim, huge, reactome

○ papillomavirus
○ Epstein-Barr virus
○ Vaccinia virus
● Human herpesvirus
○ Hepatitis C virus
○ Hepatitis B virus
○ Human immunodeficiency virus
○ Influenza A virus
○ Simian virus 40
○ Human adenovirus
○ All

Major DNA-binding protein

Binary Interactions

Uniprotkb AC	P04296, B9VQF7, Q09IA4,
Organism	Human herpesvirus 1 (strain 17) (10299)
genenames and synonyms	Infected cell protein 8, ICP8, DBP, UL29,
Domains	Viral_ssDNA_bd (IPR000635),
Gene Ontology	(GO:0042025) (GO:0005634) (GO:0003697) (GO:0008270) (GO:0006260)
others Xrefs:	rcsb pdb: 1URJ, refseq: NP_044631.1,
keywords	3D-structure, Complete proteome, Metal-binding, Zinc-finger, DNA-binding, Zinc, DNA replication, Nucleus,
ADAN	Prediction of protein-protein interAction of moDular domAiNs

DBP: Major DNA-binding protein (P04296)
Human herpesvirus 1 (strain 17)
interacts with:

protein	interactions	score
RAD50 Homo sapiens (Q92878)	1	0.15
PARP1 Homo sapiens (P09874)	1	0.15
POLR1A Homo sapiens (O95602)	1	0.15
SMARCC1 Homo sapiens (Q92922)	1	0.15
SMARCE1 Homo sapiens (Q969G3)	1	0.15
RUVBL1 Homo sapiens (Q9Y265)	1	0.15
GTF2I Homo sapiens (P78347)	1	0.15
SMARCD1 Homo sapiens (Q96GM5)	1	0.15
SMARCC2 Homo sapiens (Q8TAQ2)	1	0.15
CHD3 Homo sapiens (Q12873)	1	0.15
SF3B1 Homo sapiens (O75533)	1	0.15
XRCC4 Homo sapiens (Q13426)	1	0.15
XRCC5 Homo sapiens (P13010)	1	0.15

Fig: 4.5 Shows search pages for UL29

e. Each gene of HSV-1 was searched individually and its interactions with Human proteins was collected.

f. 64 such interactions were found.

g. Along with this, other relevant information such as Gene symbol, Uniprot id, Taxonomy id, Source, Reference, PubMed id, Interaction type and number of evidences were also collected.

h. From this information an excel sheet was prepared. This sheet is named as **VMINT sheet** (Appendix table 4).

ii. From VirHostNet

a. **VirHostNet** (**Vir**us-**HostNet**work) is a knowledgebase system dedicated to the curation, the integration, the management and the analysis of virus-host molecular (mainly protein-protein) interaction networks.

b. It is freely available and can be accessed through the following website: http://pbildb1.univ-lyon1.fr/virhostnet/index.php/

c. This is a comprehensive database of viruses and human interaction and is updated yearly.

d. It can be searched for a particular type of virus and human interaction or list of viral proteins and their targets.

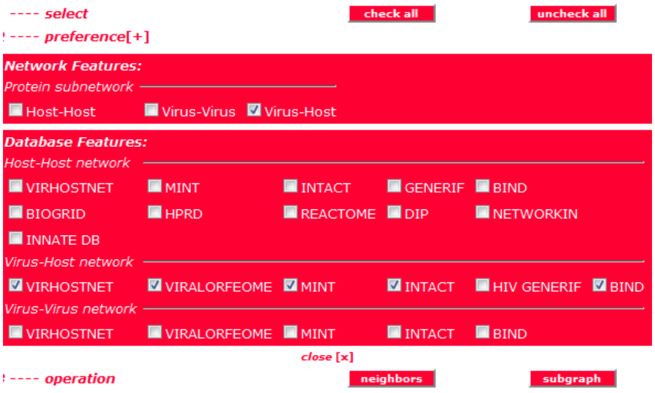
e. We searched the entire database for virus (HSV-1) and Host interactions.

f. In order to collect interactions from this database, the following procedure was followed:

VirHostNet > Browse > Taxonomy > Select HSV-1(10298) > Click on Proteins

g. Her we found 86 viral proteins. For all the 86 viral proteins data was collected with the following **Preferences.** The procedure has been discussed below:

- In network features; select Virus-Host.
- In Database features; select all databases (except HIV GENERIF).



Network Features:
Protein subnetwork
□ Host-Host □ Virus-Virus ☑ Virus-Host

Database Features:
Host-Host network
□ VIRHOSTNET □ MINT □ INTACT □ GENERIF □ BIND
□ BIOGRID □ HPRD □ REACTOME □ DIP □ NETWORKIN
□ INNATE DB
Virus-Host network
☑ VIRHOSTNET ☑ VIRALORFEOME ☑ MINT ☑ INTACT □ HIV GENERIF ☑ BIND
Virus-Virus network
□ VIRHOSTNET □ VIRALORFEOME □ MINT □ INTACT □ BIND

Fig: 4.6 Showing VirHostNet Prefrences

h. Now click on **Neighbors,** this gave interacting partners (Human proteins) of selected viral proteins.

i. Here all105 interactions were found, but most part of the data was redundant.

j. After removing duplicate values we were left with85 interactions, many of these were common to those in Virus MINT.

k. Along with this, other relevant information such as Uniprot id, Gene name Taxonomy id, Interactor, Source, Reference, PubMed id, Interaction type and number of evidences were also collected.

l. From this information an excel sheet was prepared. This sheet was named as **VirHostNet sheet** (Appendix Table 5).

iii. From Literature

a. All PPI interactions of HSV-1 and human were collected by exploring literature.

b. Around 133 interactions were found in literature; many of these were also present in Virus MINT and VirHostNet.

c. Along with this, other relevant information such as Uniprot id, Gene symbol, Taxonomy id, Interactor, Source, Reference, PubMed id, Interaction type and No. of evidences were also collected.

d. From this information an excel sheet was prepared. This sheet was named as **Literature sheet** (Appendix Table 6).

e. Final sheet was prepared by combining data from the three methods mentioned above and all duplicate interactions were removed. Final sheet contained 226 unique Virus-Host interactions. This sheet was named as **Virus-Host sheet** (Appendix Table 7).

4.2 Secondary phase or Protein-Protein Interaction Network construction

4.2.1 Network construction for Alzheimer

a. Uniprot id for each protein present in **Final Alzheimersheet** was taken and were submitted in APID2NET, an implemented plug-in of Cytoscape

b. APID2NET retrieves all possible information on protein-protein interaction from five interaction databases namely, **D**atabase of **I**nteracting **P**roteins (DIP), **B**iomolecular**I**nteraction **N**etwork **D**atabase (BIND), IntAct, **M**olecular **Int**eractions Database (MINT) and **H**uman **P**rotein **R**eference **D**atabase (HPRD).

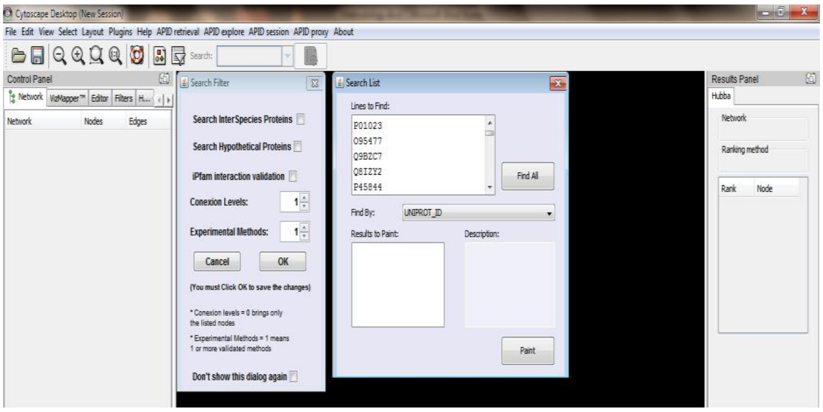
c. In order to import data from APID2NET following procedure was followed:

Cytoscape > Plugins > Click on APID2NET > APID retrieval > Search List

It will open two APID window panel; first one is **FILTER panel** where we can set filters and second one is **SEARCH LIST panel** where we can enter list to find PPI.

d. In FILTER panel deselect Interspecies proteins, hypothetical proteins and iPfam interactions. Set **connection level=1** and **experimental level=2** andclick on "**OK**".

e. In SEARCH LIST panel enter the Uniprot id of all proteins taken from **Final Alzheimersheet** and click on FINDS ALL.


Cytoscape Desktop (New Session)
File Edit View Select Layout Plugins Help APID retrieval APID explore APID session APID proxy About
Search:
Control Panel
Network VizMapper™ Editor Filters H...
Network Nodes Edges
Search Filter
Cancel
OK
Search List
Find All
Paint
Results Panel
Hubba
Network
Ranking method
Rank Node

Search InterSpecies Proteins □

Search Hypothetical Proteins □

iPfam interaction validation □

Conexion Levels: 1

Experimental Methods: 1

(You must Click OK to save the changes)

* Conexion levels = 0 brings only the listed nodes

* Experimental Methods = 1 means 1 or more validated methods

Don't show this dialog again □

Lines to Find:

```
P01023
O95477
Q9BZC7
Q8IZY2
P45844
```

Find By: UNIPROT_ID

Results to Paint:

Description:

Fig: 4.7 Showing APID2NET panels

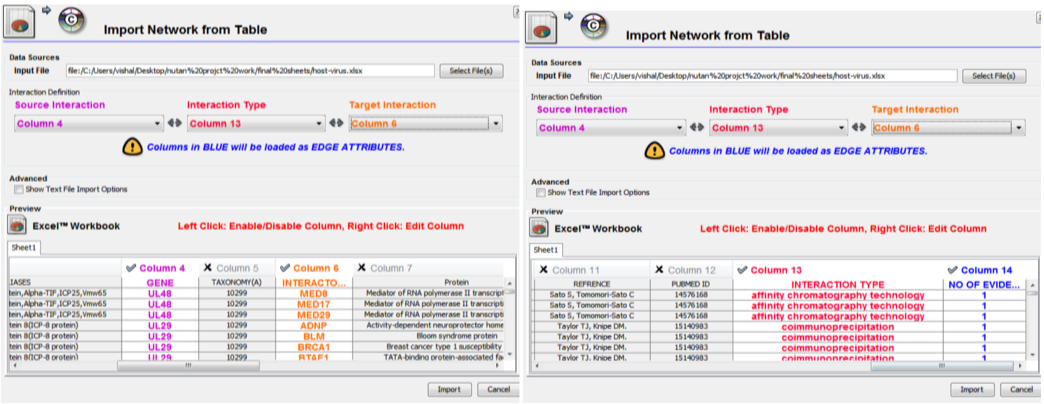
f. It searched all proteins in above mentioned databases and then will generate a list in which description of all proteins were mentioned.

g. APID2NET also gives some extra associations; so the results are taken from RESULTS TO PANT panel and extra proteins found in list were deleted. This had to be done manually.

h. Now click on **PAINT**, it will download all relevant information and protein-protein interaction network for list of proteins that we entered and will generate Protein-Protein Interaction network.
i. This network contained 1199 nodes and 4126 edges.
j. There were many duplicate edges and self-loop, theywere removed, this was done by Network Modification Plugin of Cytoscape.
k. We found 294 self-loop and duplicate edges. The final network had 1199 nodes and 3832 edges.
l. In addition, all single nodes and separated mini networks were also deleted .These mini network not within theenriched network area and hence were insignificant.
m. Our final network contained with 1077 nodes and 4091 edges.
n. This network was exported in .sif format for further analysis and was named as **AA network.**

4.2.2 Network construction for Host-Virus

a. A different kind of approach was followed for virus and host network construction.
b. By using table Import Option in Cytoscape, **Virus-Host sheet** was imported in Cytoscape.
c. In order to import data in Cytoscape following procedure was followed:
   File > Import > Network from Table > Select File
d. It opened**Virus-Hostsheet**, then we selected the column from Excel sheet to store information in nodes and edges.
e. Here, Column 4 = Source Interaction (purple colour).
   Column 6 = Target Interaction (orange colour).
   Column 13 = Interaction type (red colour).
   Column 14 = Edge Attribute (blue colour).



Import Network from Table

Data Sources
Input File: file:/C:/Users/vishal/Desktop/nutan%20projct%20work/final%20sheets/host-virus.xlsx

Interaction Definition
Source Interaction: Column 4
Interaction Type: Column 13
Target Interaction: Column 6

Columns in BLUE will be loaded as EDGE ATTRIBUTES.

Advanced
□ Show Text File Import Options

Preview
Excel™ Workbook — Left Click: Enable/Disable Column, Right Click: Edit Column

Sheet1

	✓ Column 4	✗ Column 5	✓ Column 6	✗ Column 7
IASES	GENE	TAXONOMY(A)	INTERACTO...	Protein
tein,Alpha-TIF,ICP25,Vmw65	UL48	10299	MED8	Mediator of RNA polymerase II transcript
tein,Alpha-TIF,ICP25,Vmw65	UL48	10299	MED17	Mediator of RNA polymerase II transcripti
tein,Alpha-TIF,ICP25,Vmw65	UL48	10299	MED29	Mediator of RNA polymerase II transcripti
tein 8(ICP-8 protein)	UL29	10299	ADNP	Activity-dependent neuroprotector home
tein 8(ICP-8 protein)	UL29	10299	BLM	Bloom syndrome protein
tein 8(ICP-8 protein)	UL29	10299	BRCA1	Breast cancer type 1 susceptibility
tein 8(ICP-8 protein)	UL29	10299	BTAF1	TATA-binding protein-associated fa

✗ Column 11	✗ Column 12	✓ Column 13	✓ Column 14
REFRENCE	PUBMED ID	INTERACTION TYPE	NO OF EVIDE...
Sato S, Tomomori-Sato C	14576168	affinity chromatography technology	1
Sato S, Tomomori-Sato C	14576168	affinity chromatography technology	1
Sato S, Tomomori-Sato C	14576168	affinity chromatography technology	1
Taylor TJ, Knipe DM.	15140983	coimmunoprecipitation	1
Taylor TJ, Knipe DM.	15140983	coimmunoprecipitation	1
Taylor TJ, Knipe DM.	15140983	coimmunoprecipitation	1
Taylor TJ, Knipe DM.	15140983	coimmunoprecipitation	1

Fig: 4.8 Showing columns which are selected

f. Then click on **Import**, Cytoscape would load network attribute and would generate network.

g. Within the network virus genes were forming hubs and were surrounded by human proteins.

h. This network contained 226 nodes and 221 edges.

i. Export this network as .sif format for further analysis and was named as **VH network.**

4.2.3 Network construction for HOST

a. Same procedure (as mentioned in **Network construction for Alzheimer)**was followed and network of all host proteins present in **Virus-Host sheet**wasprepared.

b. Here also, **connection level=1** and **experimental level=2** were taken.

c. This network contained all Human/Host proteins present in **Virus-Host sheet** and their immediate neighbours.

d. The network contained 1756 nodes and 8044 edges.

e. There were many duplicate edges and self-loop, so they need to be removed, this was done by Network Modification Plugin of Cytoscape.

o. 412 self-loop and duplicate edges were found. After removing duplicate edges our network had 1756 nodes and 7632 edges.

p. In addition, all single nodes and separated mini networks are also deleted because they were not falling within the enriched network and hence were insignificant.

q. Finally we were left with 1658 nodes and 7486 edges.

f. This network was exported in .sif format for further analysis and was named as **HA network.**

4.2.4 Perform Intersection

a. Cytoscape provides us a platform for performing variety of operations, such an operation is Advanced network merge, an implemented plug-in of Cytoscape.

b. This allow user to perform Union, Intersection and Difference of two or more networks.

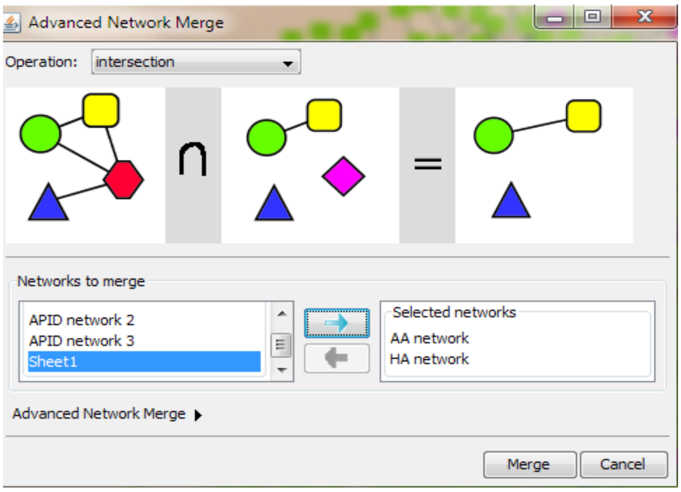
c. Intersection finds common nodes and edges present in two or more networks.

d. For this reason intersection of **HA network and AA network** was found using this.

e. In order to find Intersection following procedure was followed:

   Plugin > Advanced Network Merge > Operations > Intersection

f. Then select the network i.e. **AA network and HA network** for performing intersection and click on **Merge.**



Operation: intersection

Networks to merge

APID network 2
APID network 3
Sheet1

Selected networks

AA network
HA network

Fig: 4.9 Shows advanced network merge plugin

g. The network contained 515 nodes and 2358 edges.

h. All separated nodes were removed; the final network had only 508 nodes and 2358 edges.

i. This network was exported in the .sif format for further analysis.The name of the network was **Intersection network.**

4.3 Tertiary phase or Network Classification

4.3.1 Clustering of Intersection network

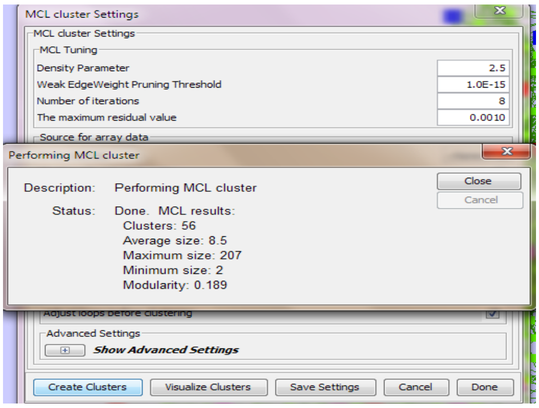
a. Clustering is a method to cluster same type of attributes in same group. Cytoscape provides many clustering plugin to create clusters.

b. By using Cluster maker(a Cytoscape plugin),an Interaction network was clustered.

c. The variety of algorithms were present to perform clustering, we selected MCL (Markov Chain Clustering) algorithm. For doing so, following procedure was followed:

Open Cytoscape > Load Network > Go to Plugin > Open Cluster maker > Select parameters > Click on Create clusters > Click on visualize Clusters.

d. Here 56 clusters are formed with maximum size=207 nodes and minimum size=2 nodes.



MCL cluster Settings

MCL cluster Settings

MCL Tuning

Density Parameter	2.5
Weak EdgeWeight Pruning Threshold	1.0E-15
Number of iterations	8
The maximum residual value	0.0010

Source for array data

Performing MCL cluster

```
Description:   Performing MCL cluster

     Status:   Done.  MCL results:
                 Clusters: 56
                 Average size: 8.5
                 Maximum size: 207
                 Minimum size: 2
                 Modularity: 0.189
```

Adjust loops before clustering ☑

Advanced Settings

Fig: 4.10 Shows result panel of MCL algorithm

e. From the result, all clusters, which were showing interaction between Blue-Red or Purple-Red nodes and Cyanine colour nodes were selected and saved.

4.3.2 Pathway Analysis

a. This was done by using DAVID (**D**atabase for **A**nnotation, **V**isualization and**I**ntegrated **D**iscovery) database; which is an online resource and is widely used for Functional gene classification and for Functional Annotation.

b. It provides a comprehensive set of functional annotation tools for investigators to understand biological meaning behind large list of genes.

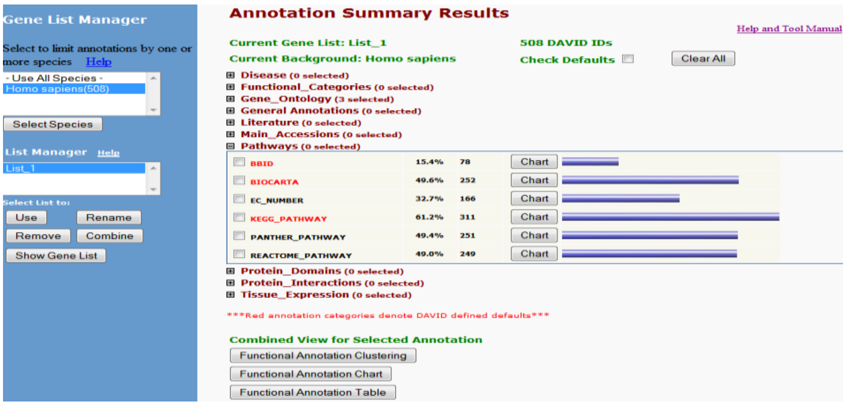
c. It can be assessed freely by visiting following website:

http://david.abcc.ncifcrf.gov/home.jsp

d. For this purpose, Entrez gene id for each protein present in Intersection network was downloaded From UniProt (Repository of proteins).

e. For pathway analysis following process was followed:

Open home page > Click on Start Analysis > Submit gene list > Select identifier as ENTREZ_GENE_ID > Select list type as Gene list > click on submit > Click on Functional Annotation tool.



Gene List Manager

Select to limit annotations by one or more species Help

- Use All Species -
Homo sapiens(508)

Select Species

List Manager Help

List_1

Select List to:

Use Rename Remove Combine Show Gene List

Annotation Summary Results

Help and Tool Manual

Current Gene List: List_1 508 DAVID IDs

Current Background: Homo sapiens Check Defaults □ Clear All

⊞ Disease (0 selected)
⊞ Functional_Categories (0 selected)
⊞ Gene_Ontology (3 selected)
⊞ General Annotations (0 selected)
⊞ Literature (0 selected)
⊞ Main_Accessions (0 selected)
⊟ Pathways (0 selected)

□	BBID	15.4%	78	Chart
□	BIOCARTA	49.6%	252	Chart
□	EC_NUMBER	32.7%	166	Chart
□	KEGG_PATHWAY	61.2%	311	Chart
□	PANTHER_PATHWAY	49.4%	251	Chart
□	REACTOME_PATHWAY	49.0%	249	Chart

⊞ Protein_Domains (0 selected)
⊞ Protein_Interactions (0 selected)
⊞ Tissue_Expression (0 selected)

Red annotation categories denote DAVID defined defaults

Combined View for Selected Annotation

Functional Annotation Clustering
Functional Annotation Chart
Functional Annotation Table

Fig: 4.11 Showing Annotation summary page for KEGG pathway

f. It will open Annotation summary page, then:

Select KEGG pathway > Click on Functional Annotation Clustering.

g. It clustered the genes in different pathways, based on Benjamini score and p-value.

h. Here 7 clusters were found; from this top 3 were analyzed.

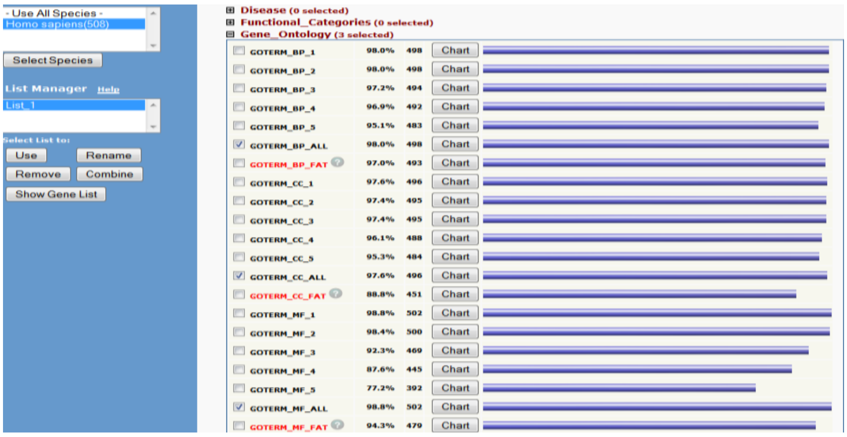
4.3.3 Clustering based on GO

a. For Go classification same procedure was followed as mentioned above.

b. The difference step that was performed, was in Annotation summary page GO classification was selected instead of Pathways.

c. Procedure for this is as follow:

Select Gene_ Ontology > Then select GOTERM_BP_ALL, GOTERM_CC_ALL, GOTERM_MF_ALL > Click on Functional Annotation Clustering.



⊞ Disease (0 selected)
⊞ Functional_Categories (0 selected)
⊟ Gene_Ontology (3 selected)

□	GOTERM_BP_1	98.0%	498	Chart
□	GOTERM_BP_2	98.0%	498	Chart
□	GOTERM_BP_3	97.2%	494	Chart
□	GOTERM_BP_4	96.9%	492	Chart
□	GOTERM_BP_5	95.1%	483	Chart
☑	GOTERM_BP_ALL	98.0%	498	Chart
□	GOTERM_BP_FAT	97.0%	493	Chart
□	GOTERM_CC_1	97.6%	496	Chart
□	GOTERM_CC_2	97.4%	495	Chart
□	GOTERM_CC_3	97.4%	495	Chart
□	GOTERM_CC_4	96.1%	488	Chart
□	GOTERM_CC_5	95.3%	484	Chart
☑	GOTERM_CC_ALL	97.6%	496	Chart
□	GOTERM_CC_FAT	88.8%	451	Chart
□	GOTERM_MF_1	98.8%	502	Chart
□	GOTERM_MF_2	98.4%	500	Chart
□	GOTERM_MF_3	92.3%	469	Chart
□	GOTERM_MF_4	87.6%	445	Chart
□	GOTERM_MF_5	77.2%	392	Chart
☑	GOTERM_MF_ALL	98.8%	502	Chart
□	GOTERM_MF_FAT	94.3%	479	Chart

Fig: 4.12 Showing GO annotation summary page

d. It opened page that was having 236 clusters. Genes in our list were clustered based on Benjamini score and p-value.

e. Top 10 clusters were taken.

4.3.4 Sub-cellular localization classification

a. UniProt is repository of proteins and is highly curated database, it can be accessed freely by visiting following website:

http://www.uniprot.org/

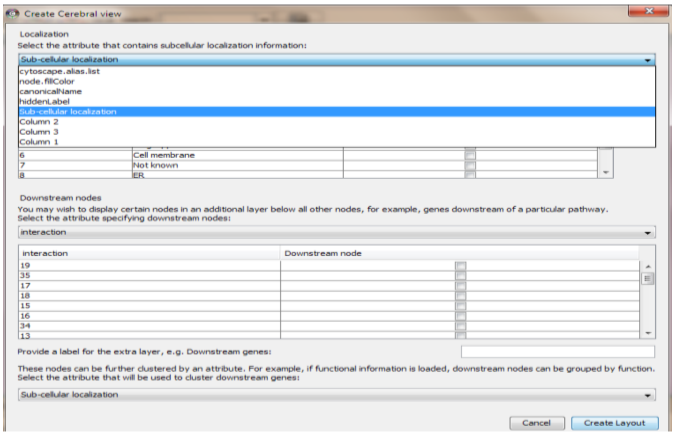
b. For subcellular localization classification, cellular location for each protein present in intersection network was downloaded from UniProt database.

c. By this data sheet we prepared a file in .txt format in order to load it in Cytoscape.

d. Then by using, Cerebral, Cytoscape plugin this sheet was loaded in Cytoscape.

e. For subcellular classification following procedure was followed:

Open network > Load Subcellular localization file > Go to Plugin > Click on Create Cerebral View > In Cerebral window Select Subcellular localization > Click on Create View.



Select the attribute that contains subcellular localization information:

6	Cell membrane		□
7	Not known		□
8	ER		□

You may wish to display certain nodes in an additional layer below all other nodes, for example, genes downstream of a particular pathway.
Select the attribute specifying downstream nodes:

interaction	Downstream node
19	□
35	□
17	□
18	□
15	□
16	□
34	□
13	□

Provide a label for the extra layer, e.g. Downstream genes:

These nodes can be further clustered by an attribute. For example, if functional information is loaded, downstream nodes can be grouped by function.
Select the attribute that will be used to cluster downstream genes:

Fig: 4.13 Showing Cerebral Panel

f. After around 50 seconds view will be generated.

g. Click on Save view and save file in jpeg format.

Chapter-5 Results

Alzheimer is a very well-studied disease. Huge amount of data inferred from various experiments is available to study Alzheimer. AlzGene is repository of genes involved in causing Alzheimer; here associations are shown on the basis of positive and negative results.

5.1 More positive association refers to strong link between Alzheimer and Gene/Protein

In AlzGene database, each gene was given any of the following three values:

- Positive (P), there was association between disease and gene.
- Negative (N), there was no association between disease and gene.
- Trend (T), result was inbetween positive and negative.

Again for each gene there was ***Q*** no. of studies showing association (N+P+T) of particular gene with the diseases. So, if we take net result of positive and negative association and it came out to be Positive, then we will be left with only those genes/proteins for which ***R*** no. of studies were showing positive result ***R***<<***Q*** .It suggested that, the particular gene was associated with Alzheimer. Greater value of ***R***, suggested the strong link between disease and gene. Here we found such 105 genes showing net positive results for non-family based studies and 26 genes showing net positive results for family based studies. However; by this method there are chances that we can miss some of the important genes because there is less number of studies available for some genes.

All results showing Trend were ignored because; either they were not associated and if then by chance and not by evidence.

Again if for a particular gene polymorphism sufficient data is available then Meta-Analysis is performed and based on its significance in relation to disease the polymorphism is regarded as Significant. So, all those genes for which Meta-Analysis is performed were also taken; there are such 115 genes. However, there is some overlap between net positive results and Meta-Analysis.

Genome Wide Association Studies (GWAS) are more potential in locating disease with particular gene or SNP. Taking this point into consideration, all GWAS studies were taken from AlzGene database. They came out to be 60 and overlaps with meta-analysis result and net positive result.

Description	Count
No. of entries in AlzGene database	695
No. of entries found	685
No. of entries (remove pseudo-genes, tRNA and non-coding mitochondrion gene)	637
No. of net non-family based positive results	105
No. of net family based positive results	26
No. of meta-analysis performed	154
No. of GWAS found	60
Total No. of dataset	239

Table: 5.1 Showing count for each step

5.2 HSV-1 and Host (Human) data collection needs much attention

There is several experimental data supporting that HSV-1 is a strong risk factor associated with Alzheimer and with other genetic factors is sufficient to initiate Alzheimer pathology.

Virus MINT: Here we found 64 interactions. All interactions present in database are physical interactions and are confirmed by several experimental techniques. Since none of the databases alone is sufficient and these databases are not updated regularly, so there arose need to search for other sources of data.

VirHostNet: Here we found 86 interactions having some overlap with Virus MINT. This database also contains physical interactions from publically available literature and by their in-house laboratory and is confirmed by several experimental techniques.

Literature: All available **literature** was searched from PubMed. In literature mining we found 133 interactions overlapping with Virus MINT and VirHostNet.

Finally by combining all data we found 226 unique interactions.

Description	Count
Entries from Virus MINT	64
Entries from VirHostNet	86
Entries from Literature	133
Total no. of dataset	226

Table: 5.2 Showing count for each step

5.3 Protein-protein interaction network

In network organization each gene or protein is represented as a node. The number of interactions that a node has with other nodes is defined as a degree of that node. The protein-protein interaction network obtained from APID includes information on co-interacting proteins, defined as proteins that have physical interaction. APID provides known experimentally validated protein-protein interactions. The edge value corresponds to no. of experimental evidences available.

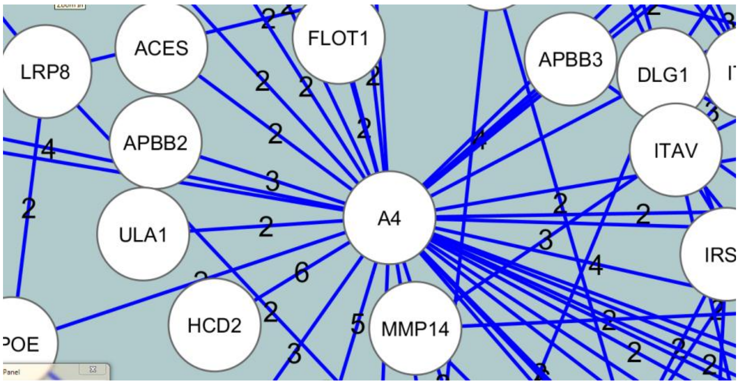


Fig: 5.1 Showing number of evidences available for particular interaction.

For this reason APID2NET, Cytoscape plugin was used for construction of PPI. An additional advantage of using APID2NET is that we can choose Connection level which corresponds to number of neighbours of a particular node; if connection level= 1 then output would be having our input nodes and all of its first neighbours. Along with this, we can also choose Experimental validation which corresponds to number of experiments to conform that particular interaction. In order to exclude false positive in this step here we selected Experiment validation=2, because if Experiment validation=1 and let's say it is Y2H then it would increase the false positive results.

Disadvantage of APID2NET is that, along with input nodes it also gives some additional nodes which are not present in our protein list, possible reason behind this is that they are interacting partner of input proteins; it increases the size and complexity of network; so these extra nodes need to be deleted manually.

i. For Alzheimer

This network shows proteins associated with Alzheimer and their first interacting protein partners.

- ✓ Purple node represents Human proteins involved in Alzheimer.
- ✓ Green nodes represent Human proteins interacting with Purple nodes.
- ✓ Network shows a spatial kind of relationship between Alzheimer and its interacting Proteins.

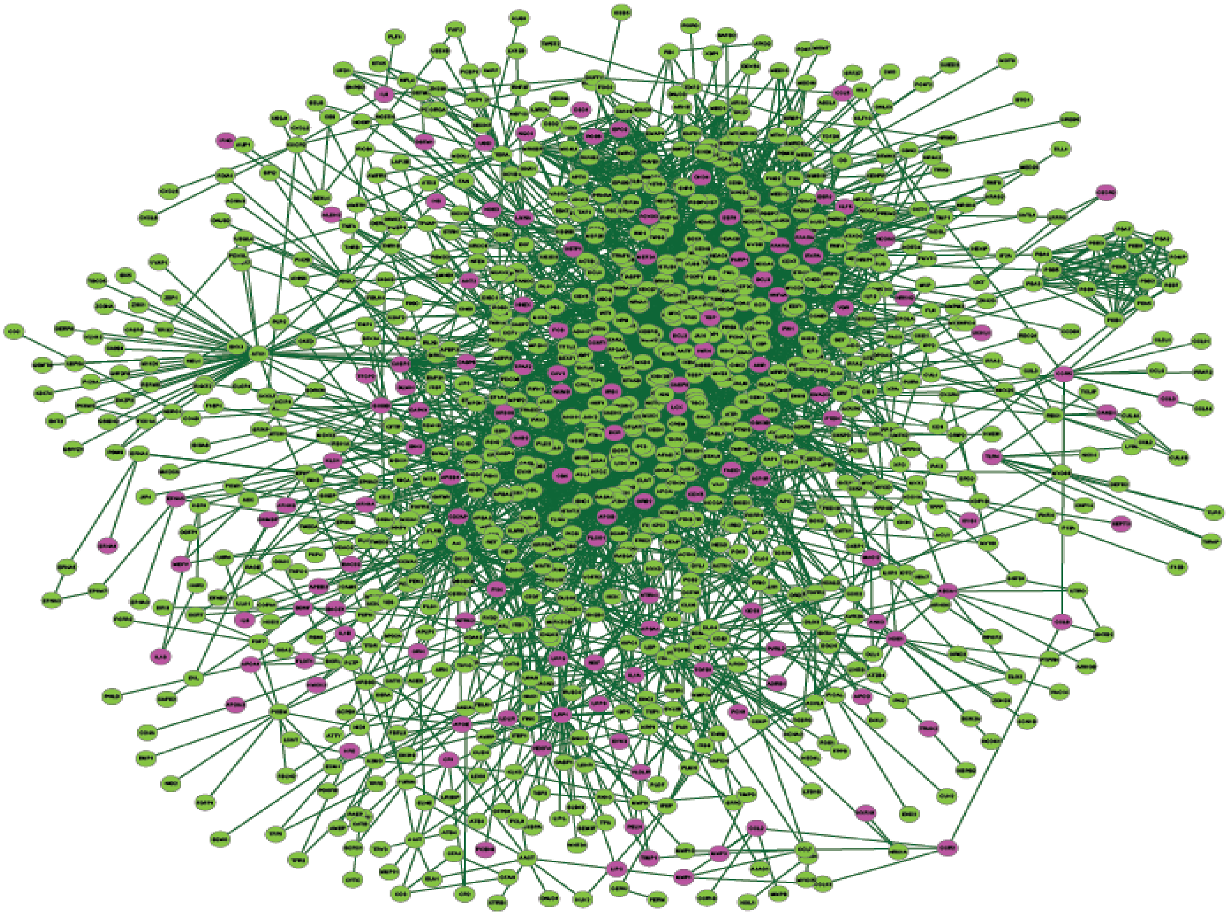


Fig: 5.2 Showing AA network

Total number of nodes	1077
Number of Green nodes	838
Number of Purple nodes	239
Number of edges	4091

Table: 5.3 Count for number of nodes and edges in AA network

ii. For Host-Virus

Virus and Host interaction is crucial for our work. For this purpose, an integrated network of Virus proteins and its target Human proteins was constructed.

- ✓ Yellow nodes represent HSV-1 proteins. There is so much confusion in protein name so in order to omit this confusion every HSV-1 protein is represented by its Gene name.
- ✓ Red nodes represent Human proteins which are direct target of Virus.
- ✓ This network provides linear relationship between HSV-1 and Human targets.

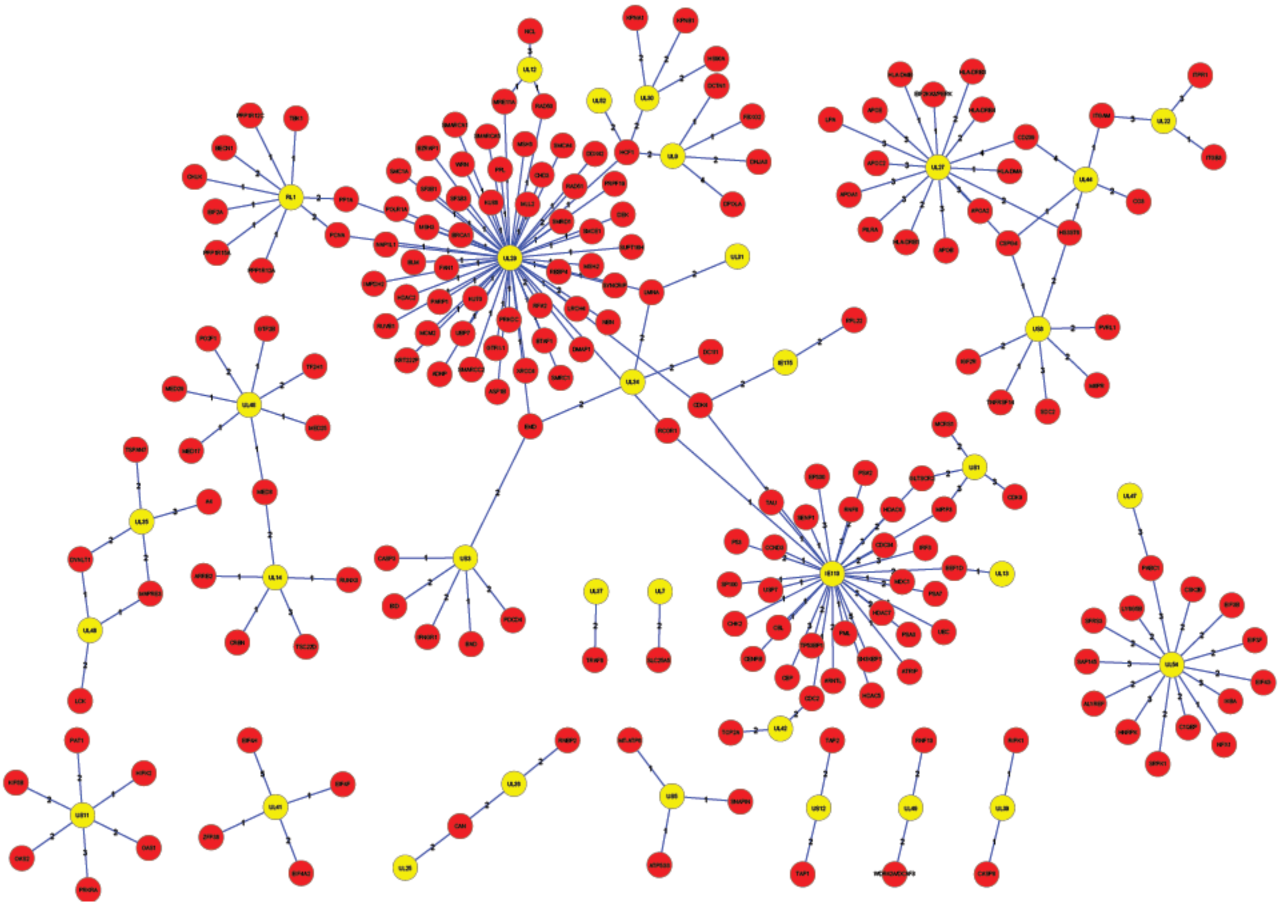


Fig: 5.3 Showing VH network

Total number of nodes	226
Number of Yellow nodes	34
Number of Red nodes	182
Number of edges	221

Table: 5.4 Count for number of nodes and edges in VH network

iii. For Host

Linear relationship tells about direct interaction, but we are interested in dynamic or indirect interactions also, for this purpose interaction network of HSV-1 targeted Human proteins (shown in Red colour in above step) was constructed.

- ✓ Red nodes represent Human proteins which are direct target of Virus.
- ✓ Green nodes represent Human proteins interacting with Red nodes.

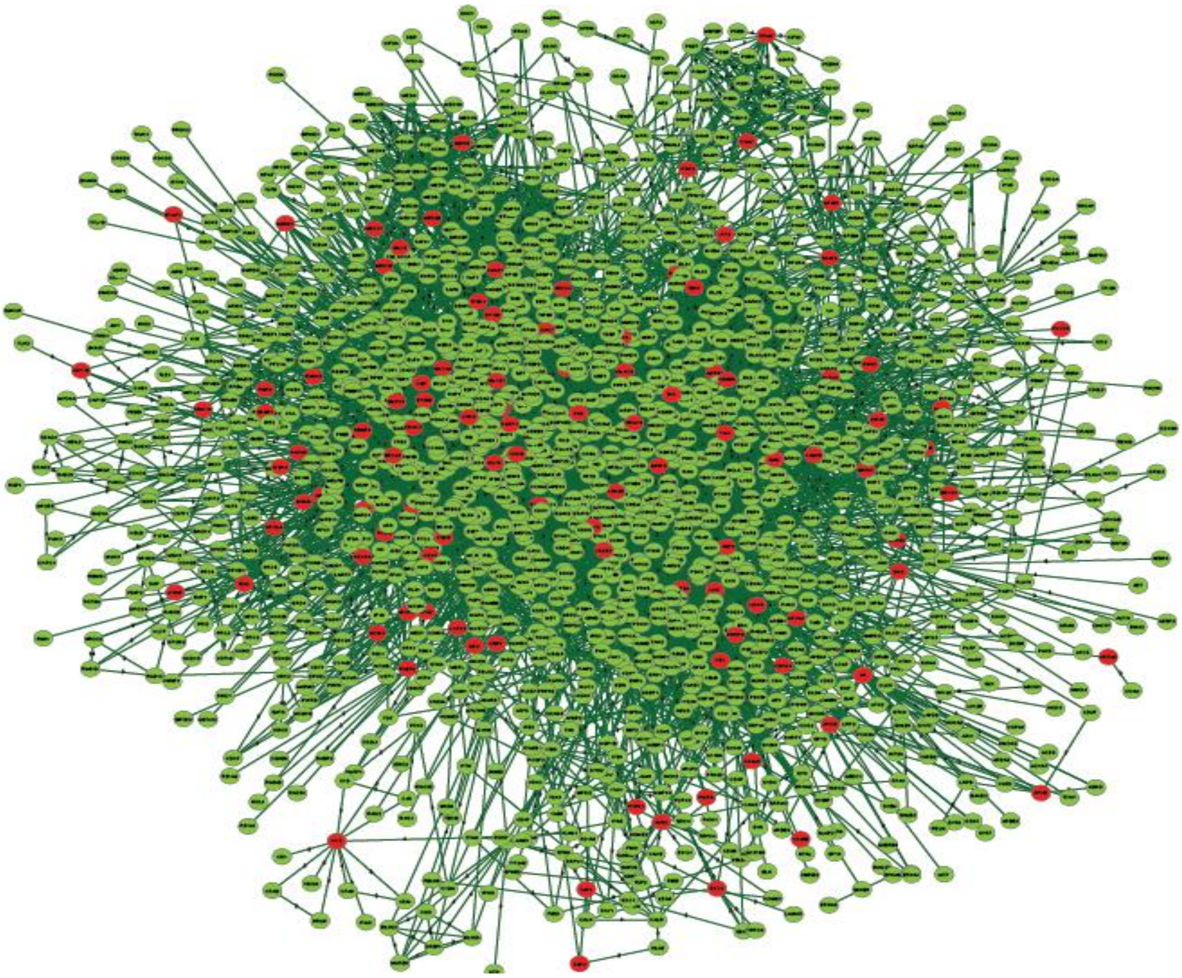


Fig: 5.4 Showing HA network

Total number of nodes	1658
Number of Green nodes	1432
Number of Red nodes	226
Number of edges	7486

Table: 5.5 Count for number of nodes and edges in HA network

5.4 Common interacting partners can solve the biological puzzle

The role of HSV-1 in pathology of Alzheimer is 30 year old, but still very little knowledge is available in this field. One of the major reason behind this is that, it is a very slow and gradual process, generally symptoms appear at the age of 60 or later but inflammation starts about 20 year ago i.e. at the age of 40 (an approximation).

At this point 2 basic questions arise:

1. Why there is such a long time for symptoms to appear?
2. What is happening during these 20 years?

The possible reason for first question may be that; there are several risk factors associated with Alzheimer and symptoms will appear only when 2 or 3 risk factors are playing role in association. In short it is like team work, if any of worker is missing, then more time will be required to build Amyloid.

Till date, there is no perfect answer to second question; but it could be concluded from the results of Intersection. Intersection can be defined as subset of two sets. Intersection fond all common proteins and their interacting partners present in AA network and HA network. This network provided clear idea that there were so many common proteins present in these two networks and contains HSV-1 direct targets, Alzheimer proteins, GWAS proteins and those proteins which were present in both Final Alzheimer table (Appendix table-3) and Host-Virus table (Appendix table-7).

- ✓ Red nodes represent Human proteins which are direct target of Virus.
- ✓ Green nodes represent Human proteins interacting with Red nodes.
- ✓ Purple nodes represent Human proteins involved in Alzheimer.
- ✓ Blue nodes represent GWAS proteins involved in Alzheimer.
- ✓ Cyanine nodes represent common nodes which are direct target of Virus and are also directly linked to Alzheimer.
- ✓ To our knowledge this is for the first time that, such Alzheimer-HSV-1 protein-protein interaction network is constructed.

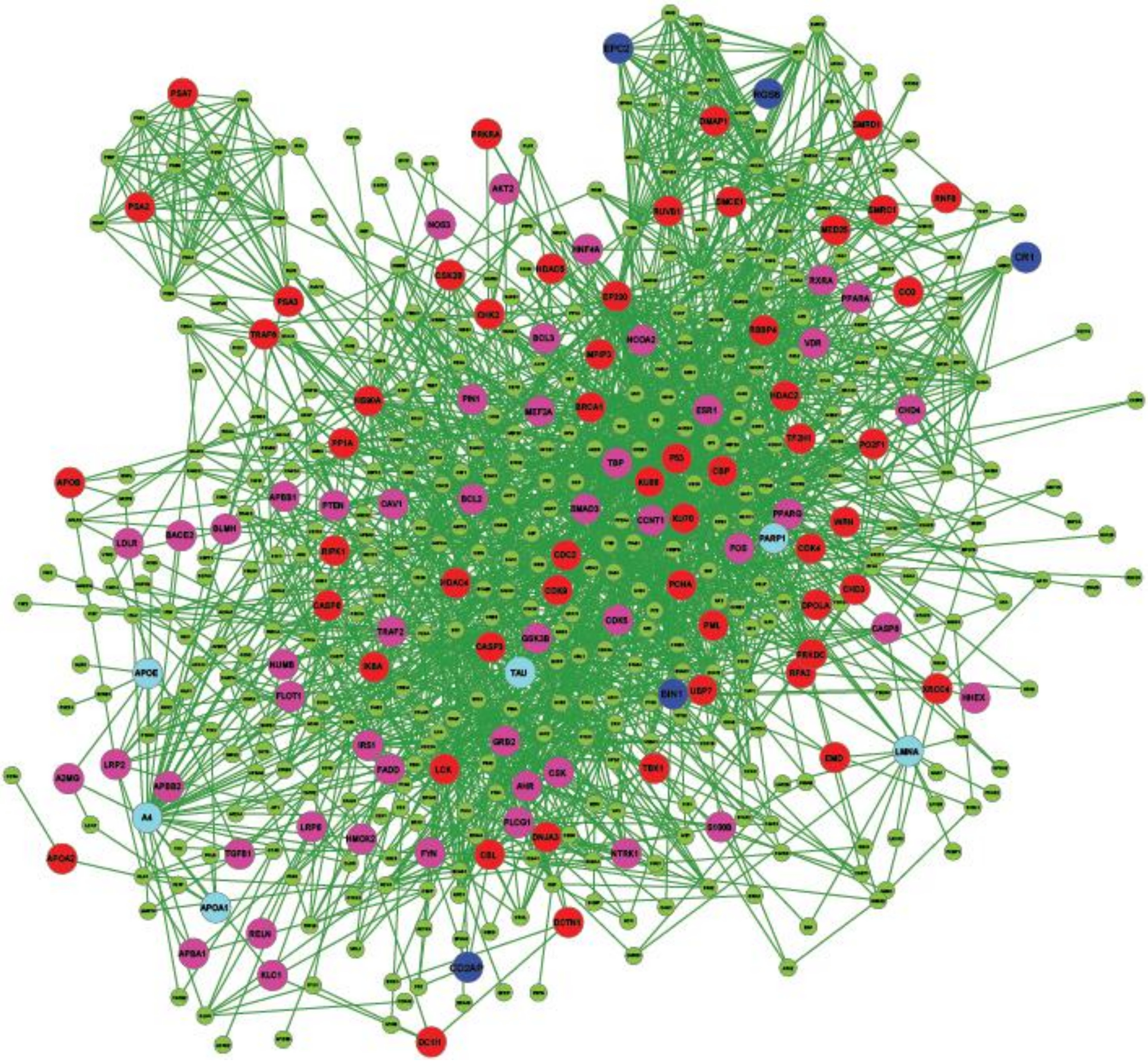


Fig: 5.5 Intersection network

Total number of nodes	508
Number of edges	2358
Number of Purple nodes	49
Number of Red nodes	55
Number of Blue nodes	6
Number of Cyanine nodes	6

Table: 5.6 Count for number of nodes and edges in Intersection network

In intersection network we found 6 proteins that were directly targeted by HSV-1 and were also involved in AD.

S.No.	Protein	Full Name	No. of disease associated (OMIM)	Alzheimer Disease Association	Interactor Protein of HSV-1	PMID	No. of evidences available
1	APOE	Apolipoprotein E	4	Yes	UL23/gB	2842273	3
2	A4	Amyloid beta A4 protein/ APP	2	Yes	UL35/ VP26	21483850	3
3	APOA1	Apolipoprotein A-I	3	No	UL23/gB	2842273	3
4	TAU	Microtubule-associated protein TAU	4	Probably	IE110/ICP0	Mingyu Liu, Edward E. Schmidt	1
5	PARP1	Poly [ADP-ribose] polymerase 1	0	No	UL29/ Major DNA binding protein	15140983	1
6	LMNA	Prelamin-A/C	12	No	UL29/ Major DNA binding protein	15140983	1

Table: 5.7 Showing Relation between new candidate proteins, Alzheimer and HSV-1.

5.5 Clustering of Intersection network

The network that we obtained after intersection was very complex that all direct interactions between two or more proteins were not visible. In order to simplify these interactions clustering was performed. Clustering is just like normalization of data, here dense networks are normalized based on degree of a node. Degree represents direct number of neighbors that node is connected to.

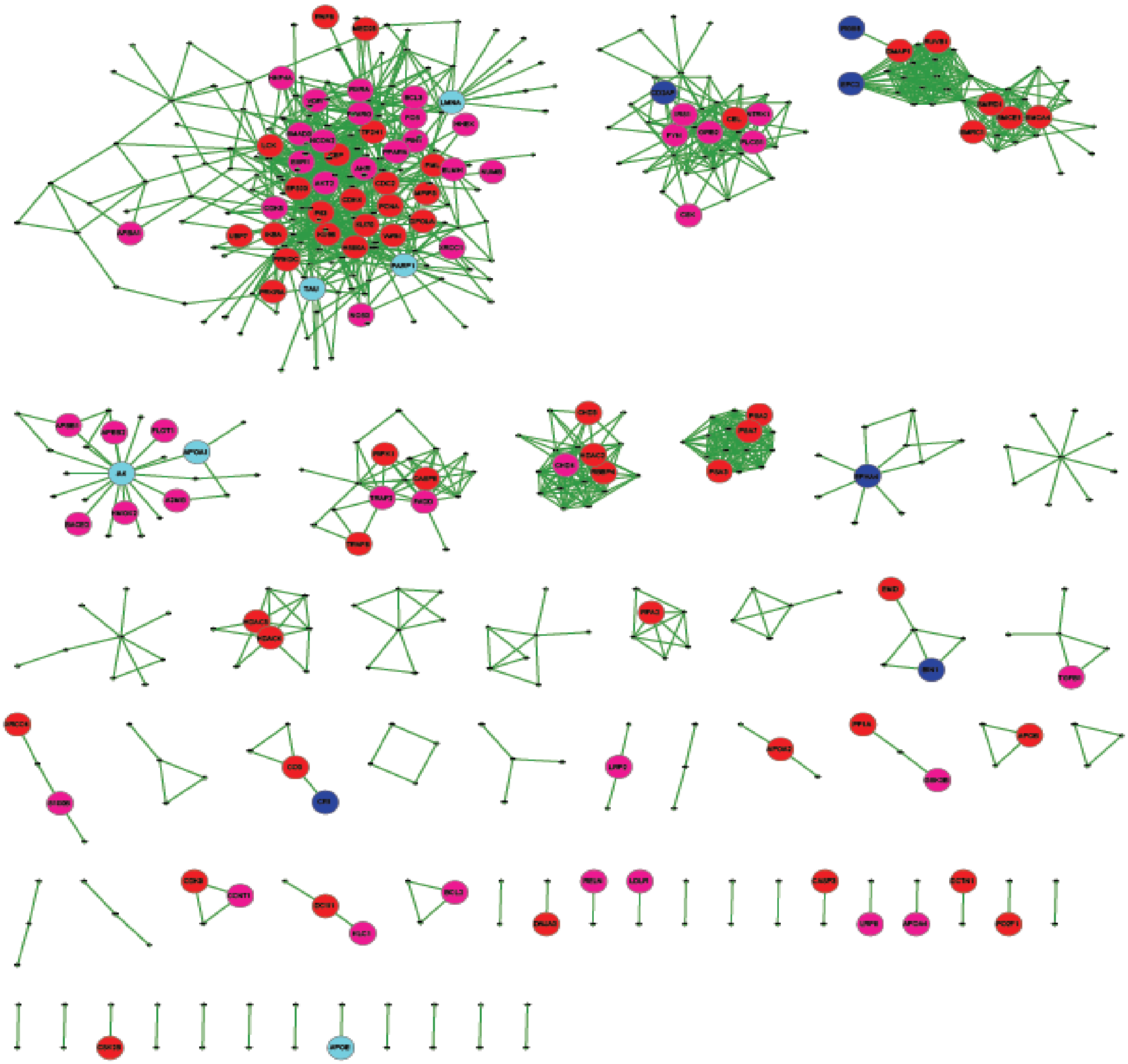
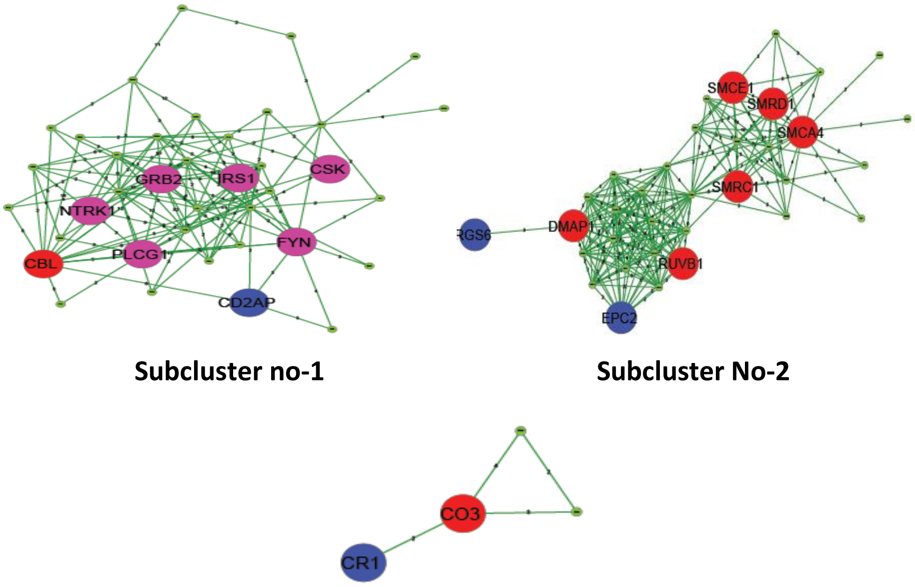


Fig: 5.6 Showing clusters formed by MCL algorithm

From this result, all those clusters which were showing direct connection Red-Blue nodes were taken. They are as follow:

We had already discussed about all those proteins involved in AD and direct target of HSV-1 (Cyanine nodes). Here we are giving a brief description of Blue nodes (GWAS proteins) that are interacting partner of HSV-1.



Subcluster no-1

Subcluster No-2

Subcluster no-3

Fig: 5.7 Subcluster no.-1, 2, 3.

In the last step we had already identified those proteins which were direct target of HSV-1 and were also involved in AD. Now here we are taking all those proteins which are targeted by HSV-1(Red) and are interacting with GWAS proteins (Blue).

All these gene are very well known to be associated with Alzheimer, here we are just showing that how they are interacting with HSV-1 targeted proteins, thus may play role in inititating Alzheimer pathology. Hence are new risk factors.

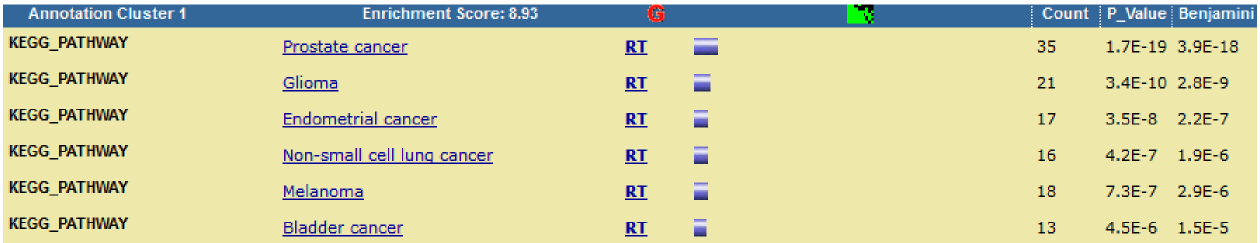
5.6 Pathway and Go based clustering can reveal hidden meaning of PPI network

Although protein-protein interaction network represents data in much simplified way, but just at looking at it we cannot identify enriched terms associated with network or clusters of proteins having same functions. So it becomes important to identify these clusters.

i. KEGG pathway results:

Here, input genes were grouped into 7 clusters, and top 3 were taken:

- Cluster-1; Enrichment score: 8.93



Annotation Cluster 1	Enrichment Score: 8.93	G			Count	P_Value	Benjamini
KEGG_PATHWAY	Prostate cancer	RT			35	1.7E-19	3.9E-18
KEGG_PATHWAY	Glioma	RT			21	3.4E-10	2.8E-9
KEGG_PATHWAY	Endometrial cancer	RT			17	3.5E-8	2.2E-7
KEGG_PATHWAY	Non-small cell lung cancer	RT			16	4.2E-7	1.9E-6
KEGG_PATHWAY	Melanoma	RT			18	7.3E-7	2.9E-6
KEGG_PATHWAY	Bladder cancer	RT			13	4.5E-6	1.5E-5

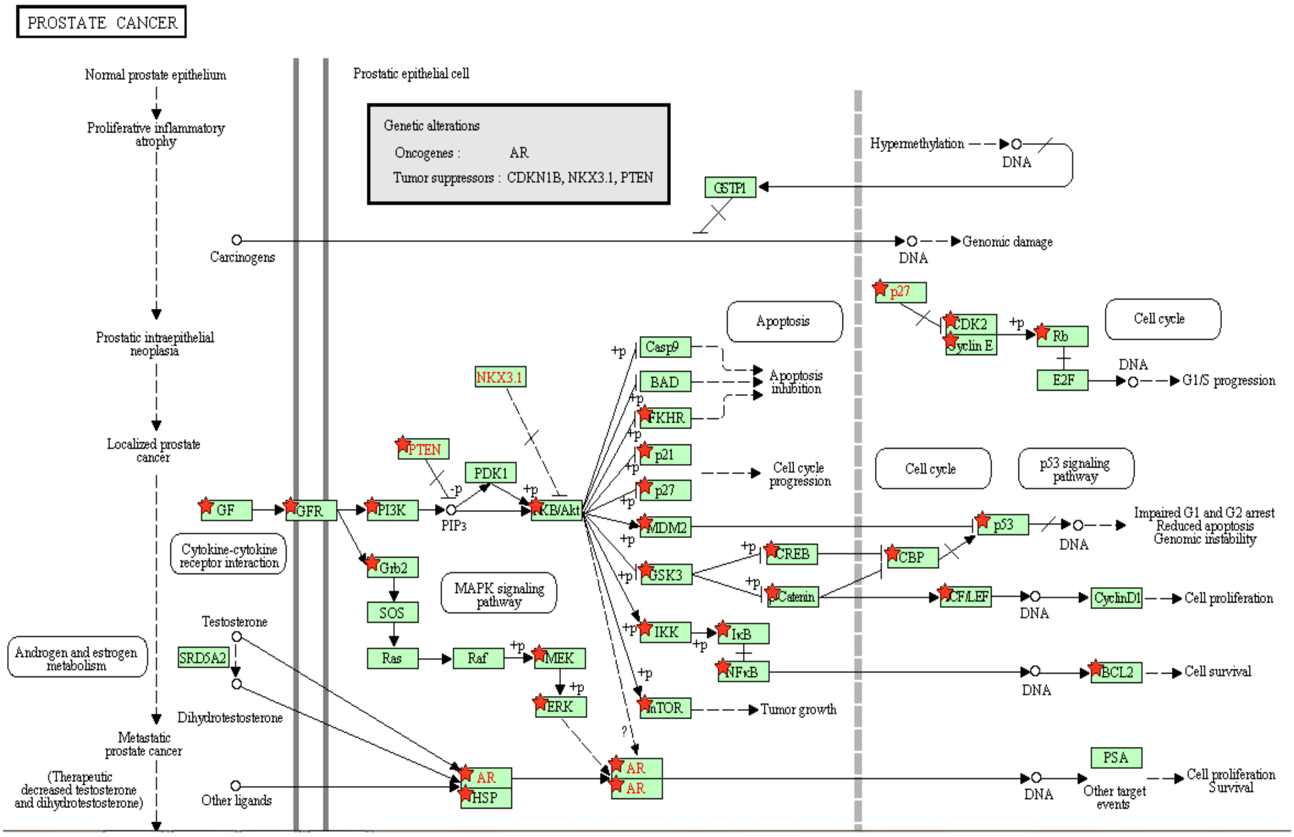
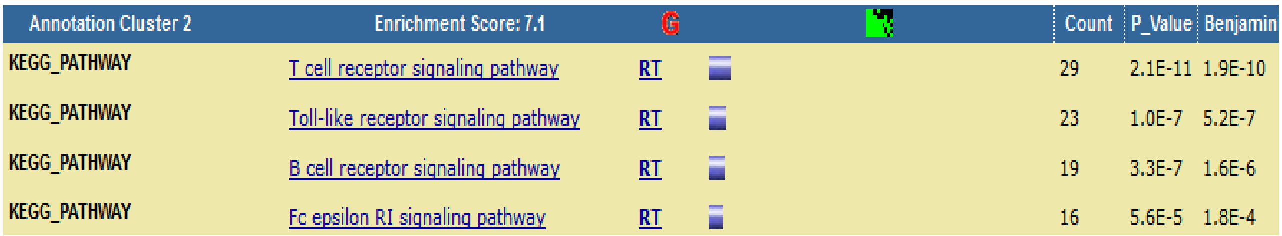


Fig: 5.8 Showing Prostate Cancer Pathway

As we can see cluster-1 is representing different type of cancers. Here count represents number of genes present in our list and falling in corresponding pathway i.e. 35 gene from our list are present in Prostate cancer pathway. Reason behind this is that HSV-1 infection also utilizes same proteins which are falling into cancer pathway. After entering in cells, HSV-1 controls the cell cycle and cancer also controls cell cycle but the difference is that HSV-1 negatively regulates cell cycle and causes apoptosis (productive phase) and cancer positively regulates cell cycle and causes cells to divide and proliferate. In above pathway all those proteins which are marked with Red star are same proteins which are also present in our input list.

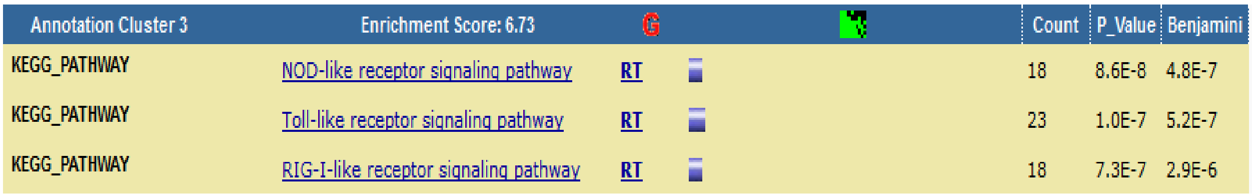
- Cluster-2; Enrichment score: 7.1



Annotation Cluster 2	Enrichment Score: 7.1	G	🟩	Count	P_Value	Benjamin
KEGG_PATHWAY	T cell receptor signaling pathway	RT		29	2.1E-11	1.9E-10
KEGG_PATHWAY	Toll-like receptor signaling pathway	RT		23	1.0E-7	5.2E-7
KEGG_PATHWAY	B cell receptor signaling pathway	RT		19	3.3E-7	1.6E-6
KEGG_PATHWAY	Fc epsilon RI signaling pathway	RT		16	5.6E-5	1.8E-4

In second cluster, 23 genes from our list are falling in T-cell receptor signaling pathway, showing immune response of host upon HSV-1 infection. We had already discussed in review that HSV-1 interact with TAP and inhibit MHC-1 dependent T-cell response for its survival. This information might be useful in unmasking the exact mechanism behind viral survival and might also reveals inflammatory signals during viral infection which contribute to AD pathology.

- Cluster-3; Enrichment score: 6.73



Annotation Cluster 3	Enrichment Score: 6.73	G	🟩	Count	P_Value	Benjamini
KEGG_PATHWAY	NOD-like receptor signaling pathway	RT		18	8.6E-8	4.8E-7
KEGG_PATHWAY	Toll-like receptor signaling pathway	RT		23	1.0E-7	5.2E-7
KEGG_PATHWAY	RIG-I-like receptor signaling pathway	RT		18	7.3E-7	2.9E-6

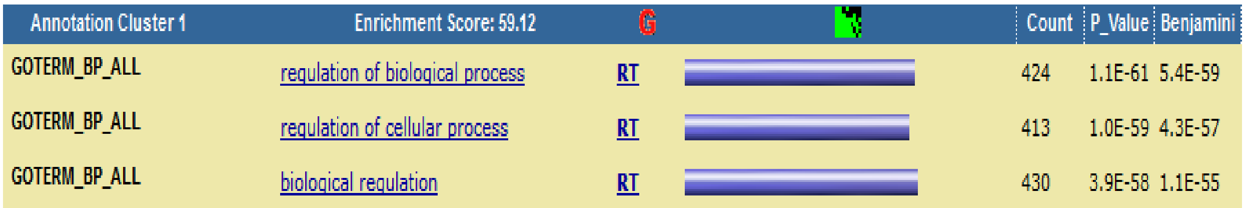
In this cluster, 23 genes from our list are falling in Toll- like receptor (TLRs) pathway. TLRs provokes rapid activation of innate immunity by inducing production of pro-inflammatory cytokines and up-regulation of co-stimulatory molecules. It clears the view that by interacting with these gene product HSV-1 inhibit cytokine production.

To our surprise Alzheimer pathway is not present in pathway results, reason behind this is that Alzheimer is secondary pathological feature of virus and not the primary.

ii. Gene Ontology Results

Based on Biological Process (BP), Molecular Function (MF) and Cellular Component (CC) all genes present in our list were classified into 239 clusters. Out of these top 10 were taken. Each cluster is showing enriched biological term and count represents number of genes from our list falling within the term.

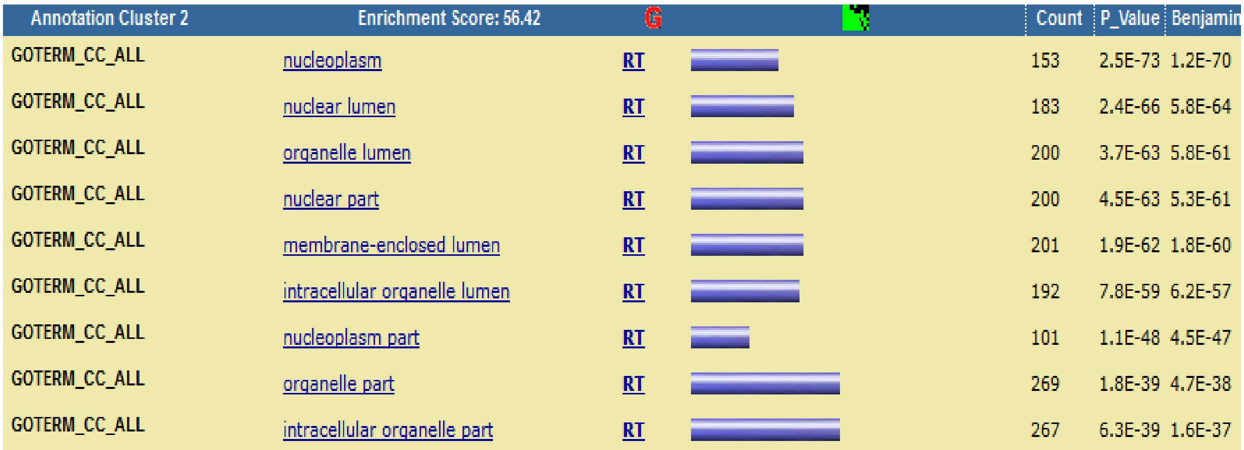
- Cluster-1; Enrichment score= 59.12



Annotation Cluster 1	Enrichment Score: 59.12	G		Count	P_Value	Benjamini
GOTERM_BP_ALL	regulation of biological process	RT		424	1.1E-61	5.4E-59
GOTERM_BP_ALL	regulation of cellular process	RT		413	1.0E-59	4.3E-57
GOTERM_BP_ALL	biological regulation	RT		430	3.9E-58	1.1E-55

It shows that 430 genes of our input genes are involved in “Regulation of biological process” such as cell cycle control etc.

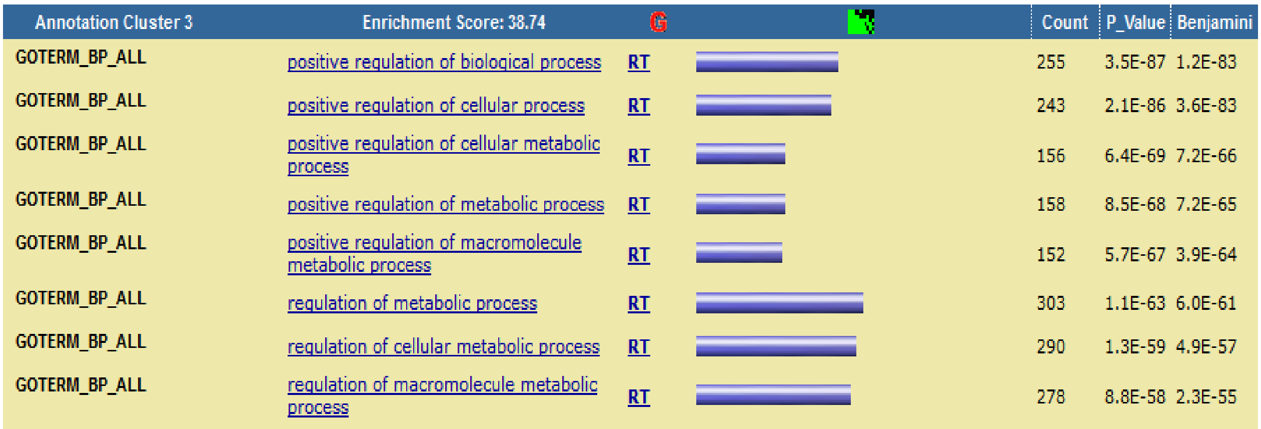
- Cluster-2; Enrichment score= 56.42



Annotation Cluster 2	Enrichment Score: 56.42	G		Count	P_Value	Benjamin
GOTERM_CC_ALL	nucleoplasm	RT		153	2.5E-73	1.2E-70
GOTERM_CC_ALL	nuclear lumen	RT		183	2.4E-66	5.8E-64
GOTERM_CC_ALL	organelle lumen	RT		200	3.7E-63	5.8E-61
GOTERM_CC_ALL	nuclear part	RT		200	4.5E-63	5.3E-61
GOTERM_CC_ALL	membrane-enclosed lumen	RT		201	1.9E-62	1.8E-60
GOTERM_CC_ALL	intracellular organelle lumen	RT		192	7.8E-59	6.2E-57
GOTERM_CC_ALL	nucleoplasm part	RT		101	1.1E-48	4.5E-47
GOTERM_CC_ALL	organelle part	RT		269	1.8E-39	4.7E-38
GOTERM_CC_ALL	intracellular organelle part	RT		267	6.3E-39	1.6E-37

The CC clustering is showing that most of our intersection genes are present within “Organelle part” and “Nuclear part”, it provides clear idea that majority of AD and HSV-1 targeted proteins are clustered in nucleus and organelle parts.

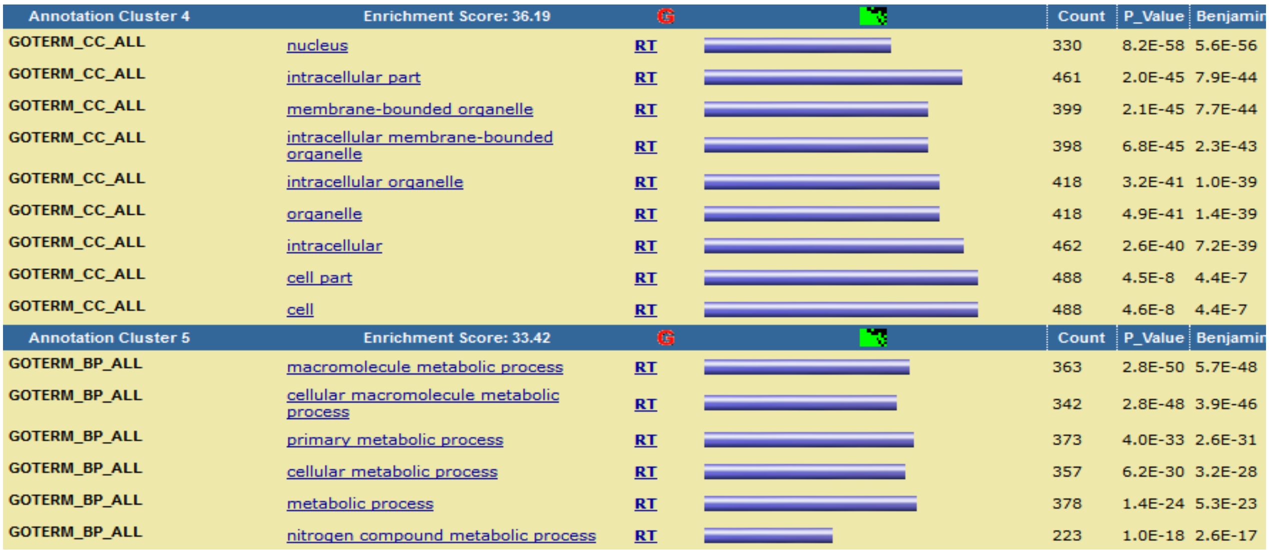
- Cluster-3; Enrichment score= 38.74



Annotation Cluster 3	Enrichment Score: 38.74	G	🟩	Count	P_Value	Benjamini
GOTERM_BP_ALL	positive regulation of biological process	RT		255	3.5E-87	1.2E-83
GOTERM_BP_ALL	positive regulation of cellular process	RT		243	2.1E-86	3.6E-83
GOTERM_BP_ALL	positive regulation of cellular metabolic process	RT		156	6.4E-69	7.2E-66
GOTERM_BP_ALL	positive regulation of metabolic process	RT		158	8.5E-68	7.2E-65
GOTERM_BP_ALL	positive regulation of macromolecule metabolic process	RT		152	5.7E-67	3.9E-64
GOTERM_BP_ALL	regulation of metabolic process	RT		303	1.1E-63	6.0E-61
GOTERM_BP_ALL	regulation of cellular metabolic process	RT		290	1.3E-59	4.9E-57
GOTERM_BP_ALL	regulation of macromolecule metabolic process	RT		278	8.8E-58	2.3E-55

Cluster number 3 shows that 303 genes are falling in “Regulation of metabolic process” and out of these 255 genes are involved in “Positive regulation of biological process”. It shows that genes involved in AD and HSV-1 infection are involved in metabolic process e.g. APP processing and A-beta production.

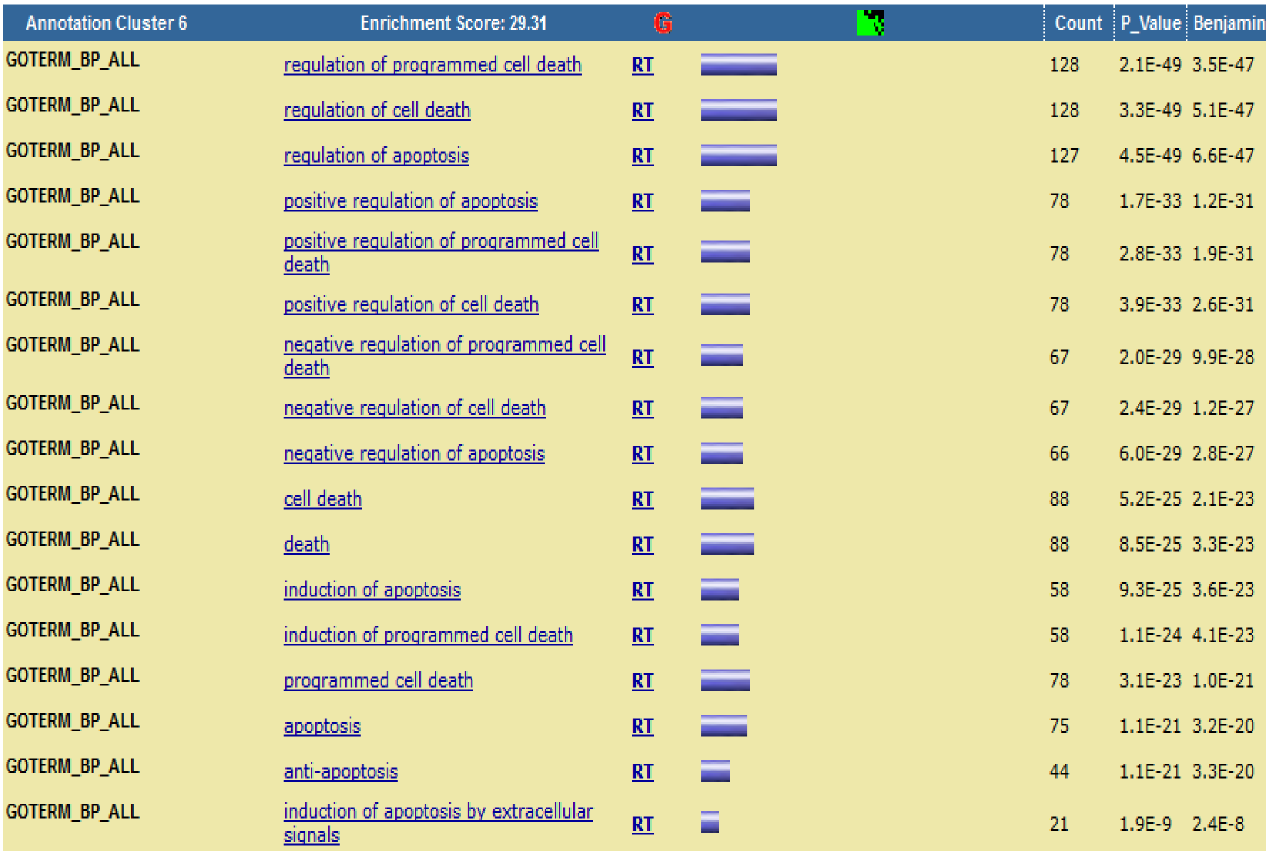
- Cluster-4; Enrichment score= 36.19 and Cluster-5; Enrichment score= 33.42



Annotation Cluster 4	Enrichment Score: 36.19	G	🟩	Count	P_Value	Benjamin
GOTERM_CC_ALL	nucleus	RT		330	8.2E-58	5.6E-56
GOTERM_CC_ALL	intracellular part	RT		461	2.0E-45	7.9E-44
GOTERM_CC_ALL	membrane-bounded organelle	RT		399	2.1E-45	7.7E-44
GOTERM_CC_ALL	intracellular membrane-bounded organelle	RT		398	6.8E-45	2.3E-43
GOTERM_CC_ALL	intracellular organelle	RT		418	3.2E-41	1.0E-39
GOTERM_CC_ALL	organelle	RT		418	4.9E-41	1.4E-39
GOTERM_CC_ALL	intracellular	RT		462	2.6E-40	7.2E-39
GOTERM_CC_ALL	cell part	RT		488	4.5E-8	4.4E-7
GOTERM_CC_ALL	cell	RT		488	4.6E-8	4.4E-7
Annotation Cluster 5	Enrichment Score: 33.42	G	🟩	Count	P_Value	Benjamin
GOTERM_BP_ALL	macromolecule metabolic process	RT		363	2.8E-50	5.7E-48
GOTERM_BP_ALL	cellular macromolecule metabolic process	RT		342	2.8E-48	3.9E-46
GOTERM_BP_ALL	primary metabolic process	RT		373	4.0E-33	2.6E-31
GOTERM_BP_ALL	cellular metabolic process	RT		357	6.2E-30	3.2E-28
GOTERM_BP_ALL	metabolic process	RT		378	1.4E-24	5.3E-23
GOTERM_BP_ALL	nitrogen compound metabolic process	RT		223	1.0E-18	2.6E-17

In cluster 4 and 5, are showing repetitive terms.

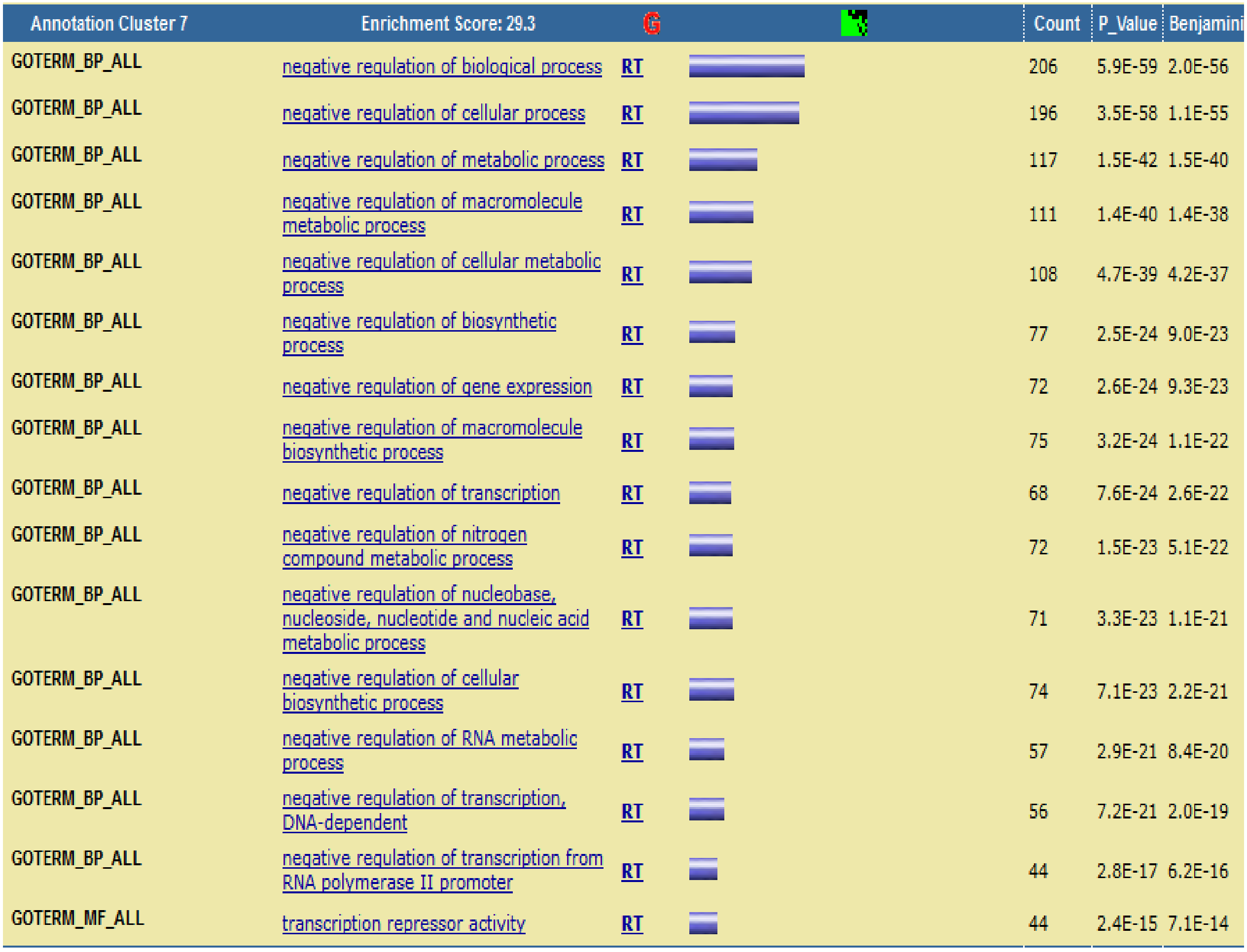
- Cluster-6; Enrichment score= 29.31



Annotation Cluster 6	Enrichment Score: 29.31	G		Count	P_Value	Benjamin
GOTERM_BP_ALL	regulation of programmed cell death	RT		128	2.1E-49	3.5E-47
GOTERM_BP_ALL	regulation of cell death	RT		128	3.3E-49	5.1E-47
GOTERM_BP_ALL	regulation of apoptosis	RT		127	4.5E-49	6.6E-47
GOTERM_BP_ALL	positive regulation of apoptosis	RT		78	1.7E-33	1.2E-31
GOTERM_BP_ALL	positive regulation of programmed cell death	RT		78	2.8E-33	1.9E-31
GOTERM_BP_ALL	positive regulation of cell death	RT		78	3.9E-33	2.6E-31
GOTERM_BP_ALL	negative regulation of programmed cell death	RT		67	2.0E-29	9.9E-28
GOTERM_BP_ALL	negative regulation of cell death	RT		67	2.4E-29	1.2E-27
GOTERM_BP_ALL	negative regulation of apoptosis	RT		66	6.0E-29	2.8E-27
GOTERM_BP_ALL	cell death	RT		88	5.2E-25	2.1E-23
GOTERM_BP_ALL	death	RT		88	8.5E-25	3.3E-23
GOTERM_BP_ALL	induction of apoptosis	RT		58	9.3E-25	3.6E-23
GOTERM_BP_ALL	induction of programmed cell death	RT		58	1.1E-24	4.1E-23
GOTERM_BP_ALL	programmed cell death	RT		78	3.1E-23	1.0E-21
GOTERM_BP_ALL	apoptosis	RT		75	1.1E-21	3.2E-20
GOTERM_BP_ALL	anti-apoptosis	RT		44	1.1E-21	3.3E-20
GOTERM_BP_ALL	induction of apoptosis by extracellular signals	RT		21	1.9E-9	2.4E-8

This cluster explains the reason of getting different kind of Cancers in Pathway analysis; it combines all genes whose biological process is to cause apoptosis (characteristic of replicating HSV-1) and to survive from apoptosis (characteristic of latent virus). This is one of very important finding of our study, 128 genes present within this cluster are responsible for “Regulation of programmed cell death”. Out of these 78 positively regulated it and 67 are involved in negative regulation of apoptosis. Here 88 are specific for “cell death”; 44 gene are specifically clustered in “Anti-apoptotic process”. The fact that replicating HSV-1 always causes apoptosis and by taking this fact into consideration researchers are also trying to use HSV-1 as vaccine to target cancerous cells.[81] Present report helps to identify all those genes, which could be further used in developing oncolytic herpes for targeting cancers.

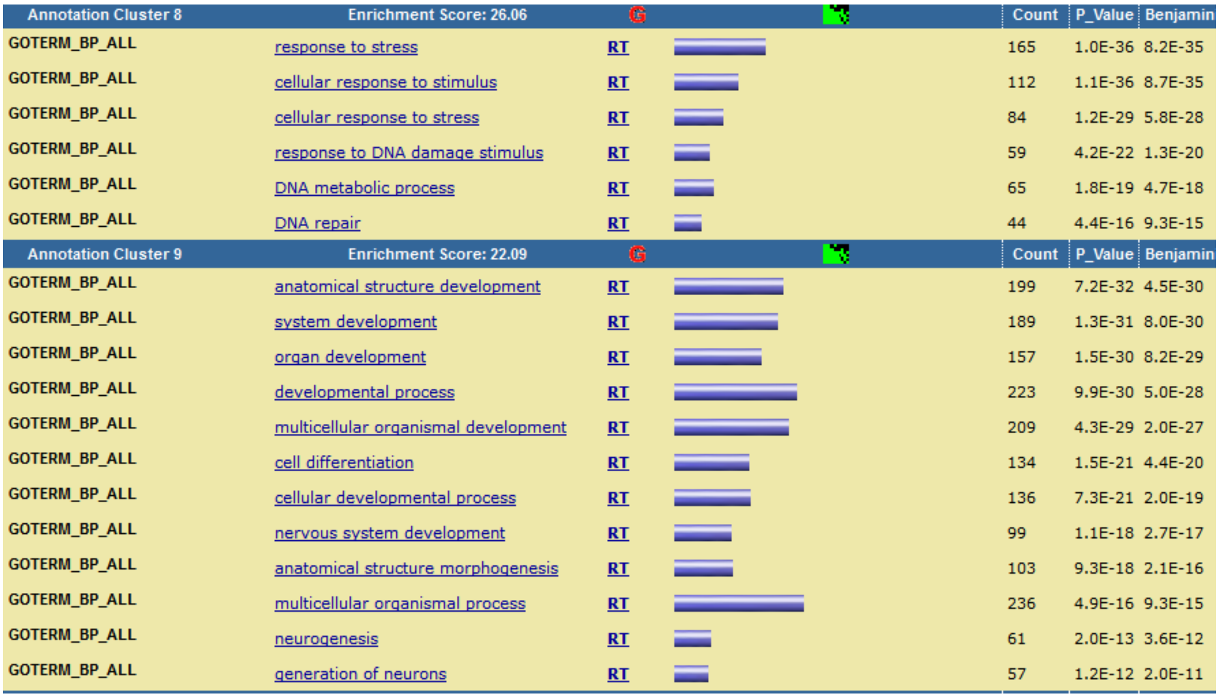
Cluster-7; Enrichment score: 29.3



Annotation Cluster 7	Enrichment Score: 29.3	G	🟩	Count	P_Value	Benjamini
GOTERM_BP_ALL	negative regulation of biological process	RT		206	5.9E-59	2.0E-56
GOTERM_BP_ALL	negative regulation of cellular process	RT		196	3.5E-58	1.1E-55
GOTERM_BP_ALL	negative regulation of metabolic process	RT		117	1.5E-42	1.5E-40
GOTERM_BP_ALL	negative regulation of macromolecule metabolic process	RT		111	1.4E-40	1.4E-38
GOTERM_BP_ALL	negative regulation of cellular metabolic process	RT		108	4.7E-39	4.2E-37
GOTERM_BP_ALL	negative regulation of biosynthetic process	RT		77	2.5E-24	9.0E-23
GOTERM_BP_ALL	negative regulation of gene expression	RT		72	2.6E-24	9.3E-23
GOTERM_BP_ALL	negative regulation of macromolecule biosynthetic process	RT		75	3.2E-24	1.1E-22
GOTERM_BP_ALL	negative regulation of transcription	RT		68	7.6E-24	2.6E-22
GOTERM_BP_ALL	negative regulation of nitrogen compound metabolic process	RT		72	1.5E-23	5.1E-22
GOTERM_BP_ALL	negative regulation of nucleobase, nucleoside, nucleotide and nucleic acid metabolic process	RT		71	3.3E-23	1.1E-21
GOTERM_BP_ALL	negative regulation of cellular biosynthetic process	RT		74	7.1E-23	2.2E-21
GOTERM_BP_ALL	negative regulation of RNA metabolic process	RT		57	2.9E-21	8.4E-20
GOTERM_BP_ALL	negative regulation of transcription, DNA-dependent	RT		56	7.2E-21	2.0E-19
GOTERM_BP_ALL	negative regulation of transcription from RNA polymerase II promoter	RT		44	2.8E-17	6.2E-16
GOTERM_MF_ALL	transcription repressor activity	RT		44	2.4E-15	7.1E-14

This cluster represents all genes that negatively regulate the biological process. This information can also be used for targeting cancerous cells.

Cluster-8; Enrichment score: 26.06 and Cluster-9; Enrichment score: 22.09

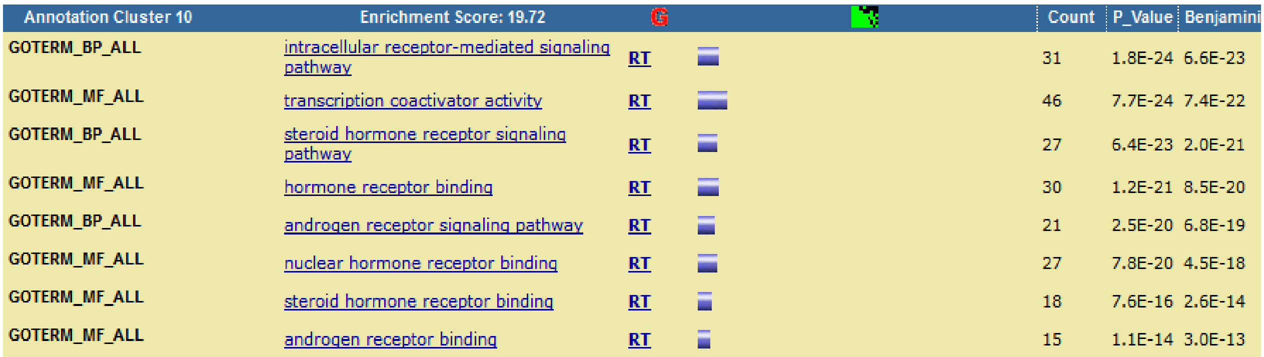


Annotation Cluster 8	Enrichment Score: 26.06	G		Count	P_Value	Benjamin
GOTERM_BP_ALL	response to stress	RT		165	1.0E-36	8.2E-35
GOTERM_BP_ALL	cellular response to stimulus	RT		112	1.1E-36	8.7E-35
GOTERM_BP_ALL	cellular response to stress	RT		84	1.2E-29	5.8E-28
GOTERM_BP_ALL	response to DNA damage stimulus	RT		59	4.2E-22	1.3E-20
GOTERM_BP_ALL	DNA metabolic process	RT		65	1.8E-19	4.7E-18
GOTERM_BP_ALL	DNA repair	RT		44	4.4E-16	9.3E-15
Annotation Cluster 9	Enrichment Score: 22.09	G		Count	P_Value	Benjamin
GOTERM_BP_ALL	anatomical structure development	RT		199	7.2E-32	4.5E-30
GOTERM_BP_ALL	system development	RT		189	1.3E-31	8.0E-30
GOTERM_BP_ALL	organ development	RT		157	1.5E-30	8.2E-29
GOTERM_BP_ALL	developmental process	RT		223	9.9E-30	5.0E-28
GOTERM_BP_ALL	multicellular organismal development	RT		209	4.3E-29	2.0E-27
GOTERM_BP_ALL	cell differentiation	RT		134	1.5E-21	4.4E-20
GOTERM_BP_ALL	cellular developmental process	RT		136	7.3E-21	2.0E-19
GOTERM_BP_ALL	nervous system development	RT		99	1.1E-18	2.7E-17
GOTERM_BP_ALL	anatomical structure morphogenesis	RT		103	9.3E-18	2.1E-16
GOTERM_BP_ALL	multicellular organismal process	RT		236	4.9E-16	9.3E-15
GOTERM_BP_ALL	neurogenesis	RT		61	2.0E-13	3.6E-12
GOTERM_BP_ALL	generation of neurons	RT		57	1.2E-12	2.0E-11

In Cluster 8, 165 genes are responsible for “response to stress”. These genes and their information can explain the mechanism behind apoptosis caused by AD pathology and HSV-1 infection.

In Cluster 9, 99 genes are responsible for “nervous system development”, this clusters all genes which are targeted by HSV-1 and are also involved in AD pathology.

Cluster-10; Enrichment score: 19.72



Annotation Cluster 10	Enrichment Score: 19.72	G		Count	P_Value	Benjamini
GOTERM_BP_ALL	intracellular receptor-mediated signaling pathway	RT		31	1.8E-24	6.6E-23
GOTERM_MF_ALL	transcription coactivator activity	RT		46	7.7E-24	7.4E-22
GOTERM_BP_ALL	steroid hormone receptor signaling pathway	RT		27	6.4E-23	2.0E-21
GOTERM_MF_ALL	hormone receptor binding	RT		30	1.2E-21	8.5E-20
GOTERM_BP_ALL	androgen receptor signaling pathway	RT		21	2.5E-20	6.8E-19
GOTERM_MF_ALL	nuclear hormone receptor binding	RT		27	7.8E-20	4.5E-18
GOTERM_MF_ALL	steroid hormone receptor binding	RT		18	7.6E-16	2.6E-14
GOTERM_MF_ALL	androgen receptor binding	RT		15	1.1E-14	3.0E-13

In Cluster 10, all those genes are clustered which acts as transcription co-factor and helps in HSV-1 replication. Since HSV-1 infection is recurrent, so information from this cluster can be used for developing of better anti-viral drugs specific for HSV-1.

5.7 Sub-cellular localization classification

Sub-cellular localization determines the location of particular protein within the biological system. This information can be effectively used for-

- ✓ Identification of localization of Alzheimer proteins.
- ✓ Identification of localization of proteins that are target of HSV-1.

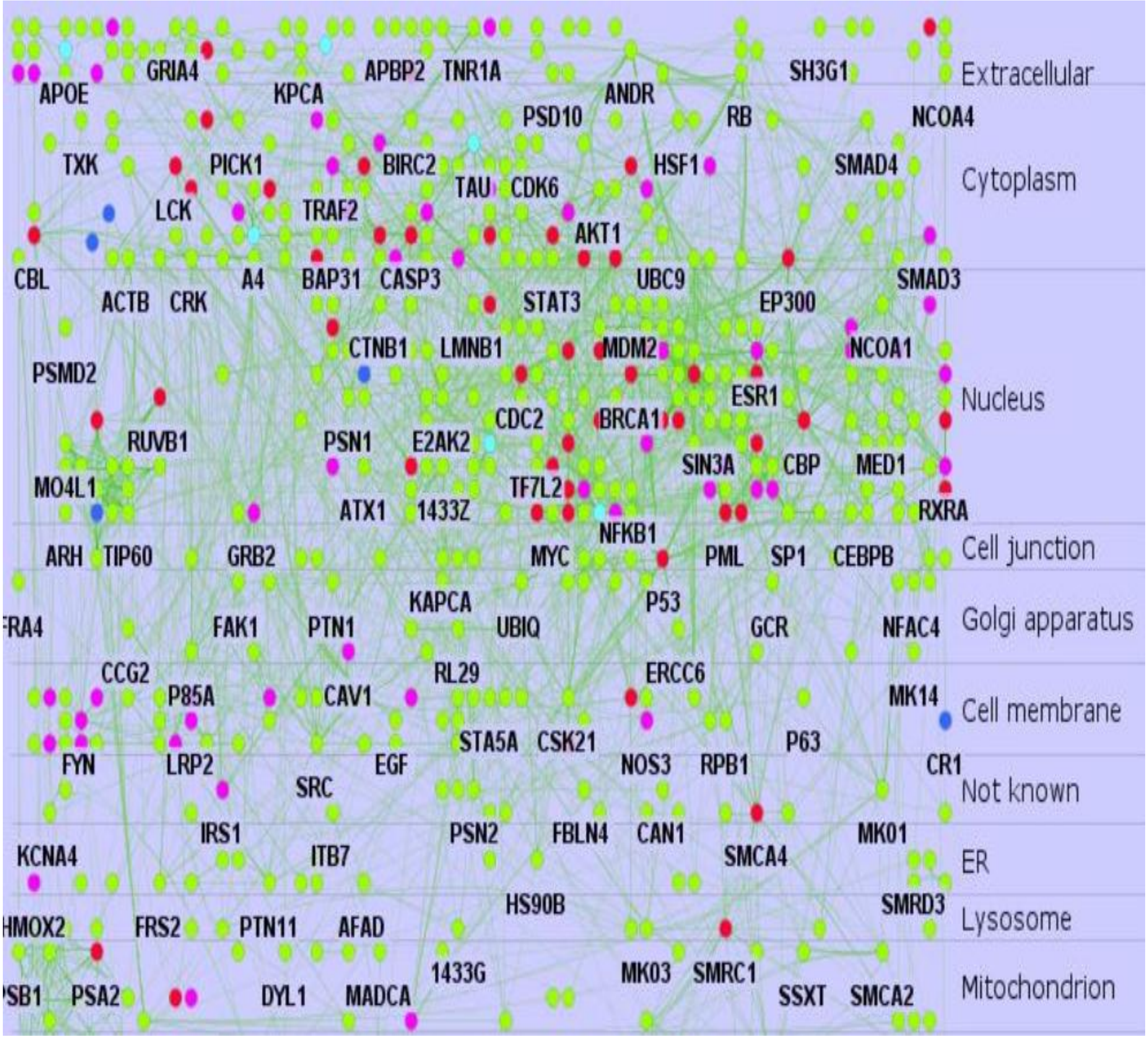


Fig: 5.9 Showing classification based on sub-cellular localization

Although GO clustering, clustered all genes in Nuclear part and Organelle part, now we are taking over all distribution of proteins with in cell.

From the sub-cellular classification it is clear that;

- Majority of AD proteins (Purple) are clustering themselves in Cell Membrane, Cytoplasm and Nucleus.
- While, HSV-1 targeted proteins (Red) are widely distribute, however most of them are falling within Cytoplasm, Cell Junction and Nucleus.
    - ✓ It can be seen that HSV-1 also target cell junction proteins and hence we can say that there occur retrograde and anterograde movement of virus during reactivation. And virus targets 8 proteins for this kind of movement. Out of these some are specifically expressed in brain.

By this classification we came to conclusion that proteins involved in AD pathology and HSV-1 infection are present within same cell compartments and also show that they have some degree of synergy between them. This information can be utilized for the development of better antiviral drugs. However, further study in this field is required.

Discussion

In any disease model it is very important to find protein-protein interaction network. Huge data of protein-protein interaction is available but it becomes difficult to deal with multiple genes/proteins at a time, a simple approach for this is to use protein-protein interaction network approach, where nodes represent proteins and edge represents interaction between two proteins.

Positive results from various GWAS and other studies shows significant level of disease (Alzheimer) and gene/protein association. It is very well known and experimentally proved that HSV-1 is a strong risk factor for Alzheimer and can initiate Alzheimer pathology. Taking this into consideration a network of proteins involved in Alzheimer and its interacting proteins is constructed, it is significant because it shows relationship between Alzheimer and other Human proteins. In order to find virus interaction with human, HSV-1 and Human PPI is constructed, it reveals all Human proteins which are targeted by virus. This network contain only virus target, now we want to know about interacting partners of these targets. For this reason, Host network is constructed, it contains all proteins targeted by virus and their immediate neighbors. Intersection of Alzheimer network and Host network gives all common proteins that are found to be associated with Alzheimer and that are utilized by virus or targeted by virus for its survival. In this network we found 6 proteins that are directly associated with Alzheimer and are also direct target of HSV-1. They are as follow: APOE, A4, APOA1, LMNA, TAU and PARP1. In addition 4 GWAS proteins are also interacting with HSV-1 targeted proteins, so they are indirectly involved in AD. They are RGS6, EPC2, CR1, CD2AP.

In OMIM database, only APOE and A4 is confirmed to be associated with Alzheimer while TAU is risk factors. In our report we confirm that HSV-1 infection initiates Alzheimer pathology by utilizing various proteins. In our network we also found that PARP1, APOA1 LMNA, RGS6, EPC2, CR1 and CD2AP are new possible risk factors which have potential to initiate Alzheimer pathology upon HSV-1 infection.

In pathway analysis of proteins (present in intersection network) many of proteins are clustered in various cancer pathways this is because cancer regulates cell cycle via many proteins i.e. CBP, CRBP etc. and HSV-1 also do same by interacting with same proteins. The difference is that cancer positively regulates cell cycle while HSV-1 negatively/positively regulates it depending on productive/latent phase respectively. In next two clusters (as specified in results) B-cell and T-cell receptor signaling pathway and Toll-like receptor signaling pathway is present which corresponds to inflammatory response triggered by HSV-1 infection. In Gene Ontology analysis most of the proteins (430) are clustered in Regulation of biological process, regulation of programmed cell death (128), negative regulation of biological process (206), response to stress (106), neurogenesis (66) and generation of neurons (57). This represents common mechanism of Alzheimer and HSV-1 infection.

The above mentioned targets for HSV-1 that have been shown by our network are further studied for sub-cellular localization. Here, APOE and APOA1 are lying in Extracellular location; APOE e4 is linked to HSV-1 and increases the chances of Alzheimer by 16 fold. Here we found APOA1 as a new target, so it may also have some significant role in Alzheimer pathology. PARP1, LMNA and EPC2 are lying in nucleus; virus utilizes these proteins to enter in nucleus which mean they are targeted by virus just after it enters in cell suggesting that they may be associated with early onset of Alzheimer. TAU, A4, RGS6 and CD2AP are localizing in cytoplasm; suggesting that when present in cytoplasm virus will interact with these and will form neurofibrillary tangles and amyloid beta (plaque) and may initiate Alzheimer pathology. CR1 is present in cell membrane, it suggest that when virus enters then it might interact with this indirectly and may be involved in early inflammation of AD.

Conclusion

Network approach represents complex interaction data in much simplified way and allow user to understand the meaning of network in biological context. In the present study protein-protein interaction data generated by experimental methods was utilized to construct network of host (Human) and virus (HSV-1) protein-protein interaction. This network was constructed by computational tool e.g. Cytoscape. Here, network of Alzheimer, Host-Virus and Host (Network of viral targets and its neighbors) was constructed. Then common proteins present in Alzheimer and host network were computed (Intersection network). Here we found 8 candidate proteins (APOA1, LAMNA, PARP1, TAU, RGS6, EPC2, CR1 and CD2AP) involved in Alzheimer and which are targets of HSV-1 either direct or indirect. Out of these; APOE and A4 are known risk factors of Alzheimer and TAU may be involved. Here we conclude that APOA1, LMNA, PARP1, RGS6, CR1 and CD2AP may be the new possible risk factor which may initiate AD pathology upon HSV-1 infection. Further, pathway, GO and sub-cellular localization classification helps to identify the location where majority of Alzheimer proteins and HSV-1 targets are clustered. This could help to identify new targets for targeting Alzheimer.

Till date no dedicated medicine for Alzheimer is available, reason being, exact cause behind disease is not known. This study is shedding light on new risk factors of Alzheimer and on HSV-1 interaction with Alzheimer risk factors (Human proteins). Further, these new candidate risk factors need to be confirmed in wet lab in Alzheimer patients.