I J C T A, 9(11) 2016, pp. 5529-5540
© International Science Press

K MEANS CLUSTERING FOR GENE-GENE INTERACTION IN EPISODIC MEMORY

Sudhakar Tripathi[*] , Anand Kumar Sharma[**] , R. B. Mishra[***] and Babita Pandey[****]

Abstract: In this paper K means Clustering Algorithm is used for clustering of candidate genes related to human episodic memory. The clustering of genes is based on gene-gene interaction score. The clusters are supposed to be formed so that distribution of cluster as well as overall interaction Score of clusters should be better. The K-means clustering technique applied to cluster the genes such as in tool STRING 9.1 provides cluster outcome. We compare the results of K means Clustering provided by STRING 9.1 with our K means Clustering Algorithm .The results obtained using K-means shows that clusters formed have better distribution of genes.

Key Words: gene-gene interaction, episodic memory, clustering, K-means, interaction score.

1. INTRODUCTION

Episodic memory is a polygenic behavioural trait with substantial heritability estimates. Despite its complexity, recent empirical evidence supports the notion that behavioural genetic studies of episodic memory might successfully identify trait-associated molecules and pathways. The development of high-throughput genotyping methods via GWAS, of elaborated statistical analyses and of phenotypic assessment methods at the neural systems level will facilitate the reliable identification of novel memory-related genes. Importantly, a necessary crosstalk between behavioural genetic studies and investigation of causality by molecular genetic studies will ultimately pave the way towards the identification of biologically important genes and molecular pathways related to human episodic memory [1].

In this paper we present candidate genes related to episodic memory, gene-gene interactions with score and important and significant pathways. We have proposed a K-Means Clustering Algorithm to optimize gene clusters specific to signalling pathways based on global interaction score. STRING 9.1 tool [2] and related databases have been used for interaction network, score and primitive cluster identification. Section 2 We describe the data acquisition and problem description consisting of Candidate genes related to episodic memory and gene-gene interaction network. Section 3 contains STRING 9.1 Clustering results. Section 4 describes the K means Clustering methods and the corresponding results. Section 5 contains the conclusion of the work.

[*] Department of Computer Science and Engineering, NIT Patna
[**] Department of Computer Science and Engineering, IIT BHU Varanasi
[***] Department of Computer Science and Engineering, IIT BHU Varanasi
[****] Department of Computer Science and Engineering Lovely Professional University, Phagwara, India

2. DATA ACQUISITION AND PROBLEM DESCRIPTION

2.1 Candidate genes related to Episodic Memory

The candidate gene approach and deal with sets of genes in biologically meaningful candidate pathways. Human homologues of genes with well-established molecular and biological functions in synaptic plasticity and animal memory led to the identification of gene cluster associated with episodic memory [3].This gene cluster represents important memory- related molecules such as adenylyl cyclases, kinases and glutamate receptors. An aggregate, individual gene score based on the gene cluster was also associated with activations in memory-related brain regions, such as the hippocampus and parahippocampal gyrus. Experimental work in animals has shown that memory formation depends on a cascade of molecular events [4,5].

Human memory performance is related to variability in genes encoding proteins of this signaling cascade, including the NMDA and metabotrobic glutamate receptors, adenylyl cyclase, CAMKII, PKA, and PKC. The individual profile of genetic variability in these signaling molecules correlated significantly with episodic memory performance. The table 4.1 indicates genes and variability in the human homologues of memory related signaling genes contributes to inter individual differences in human memory performance and memory-related brain activations. The human genes encoding adenylyl cyclase 8 (ADCY8), the catalytic subunit of cAMP dependent protein kinase (PRKACG), the subunit of calcium calmodulin dependent protein kinase II (CAMK2G), 2a and 2b subunits of the ionotropic NMDA glutamate receptor (GRIN2A, GRIN2B), metabotropic glutamate receptor 3 (GRM3), and protein kinase C (PRKCA) are important for human memory function [6-12]. The variability among these genes was specifically associated with memory performance and with activation in memory-related brain regions. Thus, the genes described herein appear to form a cluster with strong impact on human memory performance [1,3].

Here we found that genes GRIN1, GRIN2A, GRIN2B, GRIN2C, GRIN2D, GRIN3A, GRIN3B, and GRINA are NMDA receptor genes. GRM1, GRM3, GRM4, GRM5 are glutamate receptor (GluR) genes. GRIA1 and GRIA4 are AMPA receptor genes. CAMK2A, CAMK2B, CAMK2D, CAMK2G and CALM2 are Ca Signaling genes. PRKCA, PRKCB1, PRKCG, PRKACA, PRKACB, PRKACG, PRKCD, PRKCE, PRKCH, PRKAG1, PRKAR1A and PRKAR2B are genes related to protein kinase activities. All these genes mostly play an important role in Long term Potentiation (LTP)[1,13,14].

2.2 Gene-gene Interaction Network

Gene interactions are crucial components of all cellular, Molecular, biological processes and signalling pathways related to a gene group. Recently, high-throughput methods have been developed to obtain a global description of the interactome (the whole network of gene/protein interactions for a given organism). This estimate is based on the integration of data sets obtained by various methods (mass spectrometry, two-hybrid methods, genetic studies)[15-17].

The interactome (genes/proteins) can be represented as a graph where nodes correspond with genes/proteins and edges with pairwise interactions. More recently, high-throughput methods have been developed for large-scale detection of pairwise interactions (two hybrid systems , the split ubiquitin method) or multi-protein complexes (TAP-TAG, HMS-PCI) [18-21].

In recent years clustering methods have been developed and applied in order to extract relevant modules from such graphs. These algorithms require the specification of parameters that may drastically affect the results. The network of interactions between proteins is generally represented as an interaction graph, where nodes represent proteins and edges represent pairwise interactions[2,15,16].

Evidence view shows the interactions based on neighborhood, gene fusion, co-occurrence, homology, co-expression, experiments, databases and text mining individually and aggregated and normalized score view results in confidence view[2].

2.3 Gene-Gene interaction Network in confidence view:

The network and scores are accessed from STRING 9.1 tool and database. STRING 9.1 tool is used to find interaction network, scores, significant biological, cellular, molecular processes and mapped input genes to these processes and pathways. In the gene-gene interaction network (graph) the nodes show the input genes mapping and the edges connecting them show interaction. The thicker the edge higher the interaction and vice-versa. The thickness of edge in ratio of normalized overall score of interactions. The interactions are shown in three views i.e. confidence, action and evidence view having overall, action specific and evidence specific scoring edges.

The confidence view shows over all interaction score corresponding to neighbourhood, gene fusion, co-occurrence, homology, co-expression, experiments, databases and text mining. The individual parameter interaction scores are between [0 1]. For global (overall) interaction score all individual parameter specific scores are added and normalized between [0 1]. Global interaction Scores for gene interaction is not unidirectional rather it is bidirectional (score Adjacency matrix is not symmetric) i.e. score for edge gene(i) --> gene(j) is not equal to gene(j) --> gene(i). But in the graph it is shown by adding both.

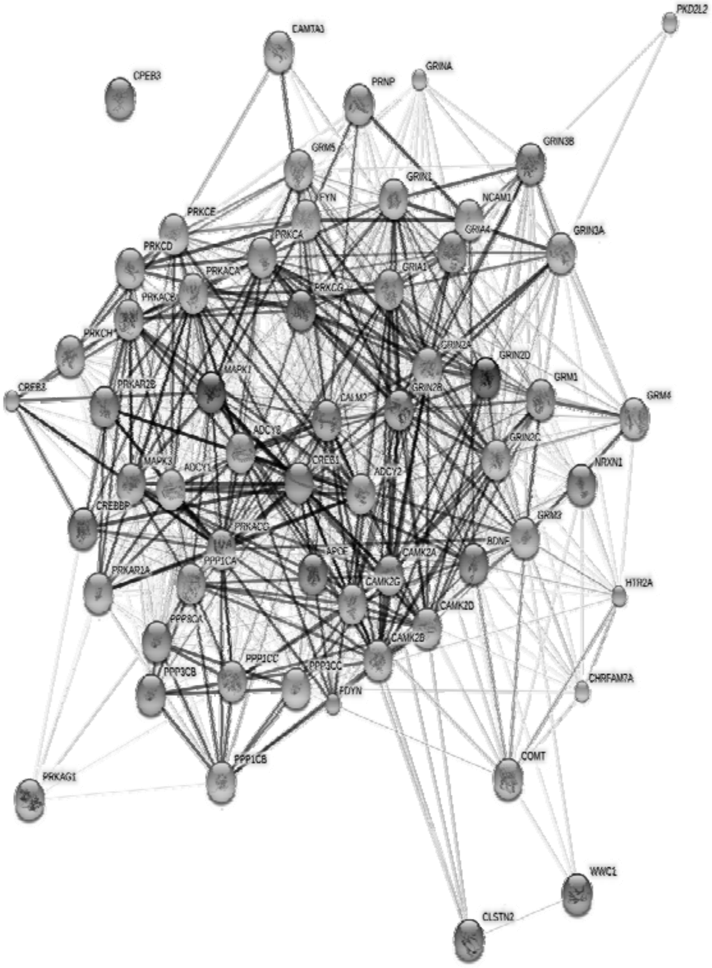

Figure 1 : Gene-Gene interaction Confidence View

Likewise interactions of genes in action and evidence view can also be shown. Action view interactions are based on activation, inhibition, Binding, Phenotype, catalysis, post translation mechanism, reaction and expression.

Evidence view shows the interactions base on neighbourhood, gene fusion, co-occurrence, homology, co-expression, experiments, databases and text mining individually and aggregated and normalized score view results in confidence view.

3. MATERIALS AND METHODS

Gene clustering, the process of grouping related genes in the same cluster, is at the foundation of different genomic studies that aim at analysing the function of genes. Microarray technologies have made it possible to measure gene expression levels for thousands of genes simultaneously. For knowledge to be extracted from the datasets generated by these technologies, the datasets have to be presented to a scientist in a meaningful way. Gene clustering methods serve this purpose.

3.1 K-Means Clustering using STRING 9.1

In string network gene clustering is done using k means method based on global interaction score and it provides cluster with close signaling group. The groups needed to be optimized so that their overall interaction score is highest to result in optimal signaling cascade.

3.2 Cluster output by STRING 9.1

STRING 9.1 Tool and database[2] provides non optimal cluster of genes using k-means algorithm based on global interaction score . It is needed to optimize the clusters for Best global interaction scores of cluster groups. It gave 8 clusters having 4, 5, 6, 6, 6, 7, 12 and 13 gene instances. The genes belonging to clusters are shown in box1.

K-Means output showing clusters and genes belonging to them

Cluster1: GRM5, PRKCE, PRKCH, CAMTA1

Cluster2: NCAM1, PRNP, GRIN2B, FYN, PRKCD

Cluster3: PRKACB, CREB3, ADCY2, PRKAR2B, PRKAR1A, ADCY8

Cluster4: PPP1CA, PPP1CB, PPP1CC, PPP3CA, PPP3CB, PRKAG1

Cluster5: GRM1, GRM4, GRM3, COMT, WWC1, GRINA

Cluster6: CPEB3, PKD2L2,CHRFAM7A,NRXN1, PPP3CC,HTR2A, CLSTN2

Cluster7: GRIA1,GRIA4,GRIN1,GRIN2A,GRIN2C, GRIN2D, GRIN3A, GRIN3B,CAMK2D, CAMK2A, PRKCG,PRKCA

Cluster8: ADCY1,APOE,BDNF, CALM2,CAMK2G, CAMK2B,CREBBP,CREB1, MAPK1, MAPK3, PRKACA, PDYN,PRKACG

Here, K-Means Score = 111.345

4. EPISODIC MEMORY CANDIDATE GENE INDEXING

All 59 candidate genes are indexed from 1 to 59, and their indexes are used in the chromosomes. The indexes and corresponding gene are shown in table 4.6.

Table 1
Gene Indexing

Gene Index	*Gene Name*	*Gene Index*	*Gene Name*	*Gene Index*	*Gene Name*	*Gene Index*	*Gene Name*
1	ADCY1	16	CREB1	31	GRM3	46	PPP3CC
2	ADCY2	17	CREB3	32	GRM4	47	PRKACA
3	ADCY8	18	CREBBP	33	GRM5	48	PRKACB
4	APOE	19	FYN	34	HTR2A	49	PRKACG
5	BDNF	20	GRIA1	35	MAPK1	50	PRKAG1
6	CALM2	21	GRIA4	36	MAPK3	51	PRKAR1A
7	CAMK2A	22	GRIN1	37	NCAM1	52	PRKAR2B
8	CAMK2B	23	GRIN2A	38	NRXN1	53	PRKCA
9	CAMK2D	24	GRIN2B	39	PDYN	54	PRKCD
10	CAMK2G	25	GRIN2C	40	PKD2L2	55	PRKCE
11	CAMTA1	26	GRIN2D	41	PPP1CA	56	PRKCG
12	CHRFAM7A	27	GRIN3A	42	PPP1CB	57	PRKCH
13	CLSTN2	28	GRIN3B	43	PPP1CC	58	PRNP
14	COMT	29	GRINA	44	PPP3CA	59	WWC1
15	CPEB3	30	GRM1	45	PPP3CB		

4.1 Gene Interaction Score

Gene interaction score is shown in table2,3,4. The gene interaction score matrix is an adjacency matrix which is not symmetric. The scores are normalized between 0 and 1. First row and first column in table contains gene index. Rest of the cells contains gene-gene interaction score in the genes having index in corresponding row and column of the gene interaction score matrix. Gene interaction scores are base on homology, neighbourhood, coregulation, coexpression, gene fusion and experimentations. All the scores are added and normalized between 0 and 1 by the tool STRING 9.1.

Table 2: Interaction Score Matrix: (Part I)

Gene Index	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
1	0	0.963	0	0	0.179	0	0	0	0	0.195	0	0	0	0	0	0.907	0	0	0	0
2	0	0	0	0	0.514	0	0.218	0	0	0	0	0	0	0.229	0	0.613	0	0	0.318	0
3	0.907	0.907	0	0	0.286	0	0	0	0.462	0.768	0	0	0	0	0	0.922	0	0	0	0
4	0	0.263	0	0	0.466	0.425	0	0	0	0.229	0	0.256	0	0.429	0	0.489	0.221	0.179	0.317	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0.38	0.954	0.97	0	0.651	0	0.607	0	0.936	0.968	0.884	0	0	0.198	0	0.929	0	0	0.516	0.565
7	0	0	0	0	0.265	0	0	0	0	0	0	0	0.378	0	0	0.966	0	0	0	0
8	0.269	0.38	0.732	0	0.43	0.964	0.941	0	0.984	0.902	0	0	0.378	0	0	0.989	0	0.374	0	0.833
9	0	0	0	0	0	0	0.975	0	0	0	0	0	0.378	0	0	0.902	0	0	0	0
10	0	0.507	0	0	0.542	0	0	0	0.902	0	0	0	0.378	0	0	0.916	0	0	0.166	0

Table 2 Contd…

11	0	0	0	0	0	0	0	0	0	0.155	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0.227	0	0	0	0	0	0	0	0	0.318	0	0.193	0	0	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0.609	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0.947	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.276	0	0	0	0
18	0	0.366	0	0	0.424	0.505	0	0	0	0.379	0	0	0	0	0	0.999	0.803	0	0	0
19	0	0	0	0	0.43	0	0	0	0	0	0	0	0	0	0	0.288	0	0	0	0
20	0	0.358	0.168	0	0.54	0	0.877	0	0.8	0.96	0.154	0	0	0	0	0.543	0	0	0.228	0
21	0	0.27	0	0	0.379	0	0.683	0	0	0.743	0	0	0	0	0	0.333	0	0	0	0.89
22	0	0	0	0	0.364	0	0.397	0	0	0	0	0	0	0	0	0.307	0	0	0	0
23	0	0.191	0	0	0.428	0	0.968	0	0.934	0.861	0	0	0	0.202	0	0.34	0	0	0.999	0
24	0.241	0.379	0.677	0	0.659	0	0.991	0	0.97	0.193	0	0.227	0	0.31	0	0.507	0	0	0.998	0.845
25	0	0	0	0	0.361	0	0.908	0	0.899	0.256	0	0	0	0	0	0.305	0	0	0	0
26	0	0	0.297	0	0.406	0.2	0.899	0	0.899	0	0	0	0	0	0	0.326	0	0	0.168	0.833
27	0	0	0	0	0.291	0	0.202	0	0	0	0	0	0	0	0	0.26	0	0	0.202	0
28	0	0	0	0	0.285	0.185	0	0	0	0	0	0	0	0	0	0.26	0	0	0	0.744
29	0	0	0	0	0.259	0	0	0	0	0	0	0	0	0	0	0.229	0	0	0	0
30	0	0.563	0	0	0.371	0	0.864	0	0.8	0.861	0	0	0	0	0	0.24	0	0	0.16	0.613
31	0	0	0	0	0.424	0	0	0	0	0	0	0	0	0	0	0.411	0	0	0	0
32	0	0	0	0	0.243	0	0	0	0	0	0	0	0	0	0	0.302	0	0	0	0
33	0	0.534	0	0	0.425	0	0	0	0	0.193	0	0	0	0	0	0.317	0	0	0	0
34	0	0	0	0	0.411	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	0	0.497	0.241	0.348	0.468	0.609	0.813	0.826	0.82	0.256	0	0	0	0	0	0.992	0.829	0.427	0.922	0.16
36	0	0.374	0.257	0	0.34	0.38	0.814	0	0.821	0.822	0	0	0	0	0	0.986	0.831	0	0.207	0
37	0	0.18	0	0	0.803	0	0	0	0	0	0	0	0	0	0	0.241	0	0	0.994	0
38	0	0	0	0	0.193	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0.9	0.951	0.914	0	0.534	0.428	0.291	0	0	0.26	0	0	0	0.243	0	0.878	0	0.19	0	0.271
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0.81	0	0.805	0	0	0	0	0	0	0.672	0.152	0	0.157	0
42	0	0	0	0	0	0	0.8	0	0.803	0.821	0	0	0	0	0	0	0	0	0	0
43	0	0	0	0	0	0	0.808	0	0.808	0	0	0	0	0	0	0	0	0	0.161	0
44	0	0	0	0	0	0	0.198	0	0	0	0	0	0	0	0	0.193	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.154	0	0	0	0
46	0	0	0	0.195	0.205	0.536	0	0	0	0.205	0	0.324	0	0.259	0	0	0	0	0	0
47	0	0.917	0	0	0.195	0	0	0	0	0	0	0	0	0	0	0.93	0.8	0	0	0
48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.8	0	0	0	0
49	0	0	0	0	0.296	0	0	0	0	0	0	0	0	0	0	0.921	0	0	0	0
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51	0	0	0	0	0.193	0	0.187	0	0	0	0	0	0	0	0	0.317	0	0	0	0
52	0.905	0.922	0.9	0	0	0.151	0.212	0	0	0.195	0	0	0	0	0	0	0	0	0	0.195

Table 2 Contd…

53	0.307	0.34	0.675	0	0.246	0	0	0	0	0.243	0	0	0	0	0	0.286	0	0	0.91	0.66
54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.286	0	0	0.994	0
55	0	0.903	0	0	0	0	0	0	0	0	0.514	0	0	0	0	0.412	0	0	0.66	0
56	0	0.26	0.195	0	0.307	0.51	0	0	0	0	0	0	0	0	0	0.201	0	0	0	0.187
57	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.276	0	0	0.699	0
58	0	0	0	0	0.286	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
59	0	0	0	0	0.185	0	0	0	0	0	0	0	0.319	0.201	0	0	0	0	0	0

Table 3
Interaction Score Matrix: (Part II)

Gene Index	21	22	23	24	25	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40
1	0	0	0.195	0	0	0	0	0	0	0	0.227	0	0.16	0	0	0	0	0	0	0
2	0	0	0	0	0	0	0	0	0	0	0.162	0.514	0	0.195	0	0	0	0	0	0
3	0	0.369	0.766	0	0.215	0	0	0	0	0	0.794	0	0	0	0	0	0	0	0	0
4	0	0	0.193	0.263	0	0	0	0	0	0	0	0	0	0.193	0	0.235	0.243	0.191	0	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0.43	0.514	0.437	0.677	0.24	0	0.187	0	0	0.428	0.283	0.165	0.412	0.314	0	0	0.318	0.193	0	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.427	0	0
8	0.185	0.53	0.985	0.985	0.914	0.921	0	0	0	0.808	0.738	0	0.16	0	0	0.817	0	0.254	0	0
9	0	0.268	0	0	0	0	0	0	0	0	0.463	0	0	0	0	0	0	0.251	0	0
10	0	0.431	0.733	0	0	0	0	0	0	0	0.778	0	0	0	0	0	0	0.251	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.307	0	0
12	0	0	0	0	0	0	0	0	0	0	0.329	0	0	0.462	0	0	0.208	0.283	0	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0.613	0	0	0.563	0	0	0	0.227	0	0
15	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0.2	0	0	0	0	0	0	0	0	0	0	0	0.219	0	0	0	0
19	0	0.31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
20	0	0.826	0.848	0	0.833	0	0.745	0	0.16	0	0.379	0.424	0.58	0	0	0	0	0.255	0	0
21	0	0.739	0.776	0	0.428	0	0.741	0	0	0.514	0.378	0.347	0.425	0	0	0	0	0	0	0
22	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
23	0	0.986	0	0	0	0	0.965	0	0	0	0.778	0.424	0	0.201	0	0	0	0.213	0	0
24	0.772	0.997	0.938	0	0.907	0	0.964	0	0.235	0.699	0.712	0.395	0.602	0.227	0	0	0.195	0.297	0	0
25	0	0.465	0.907	0	0	0	0.466	0	0.16	0	0.419	0.366	0.422	0	0	0	0	0	0	0
26	0.452	0.783	0.908	0.908	0.906	0	0.466	0	0.162	0.428	0.366	0.378	0.386	0	0	0	0	0	0	0
27	0	0.986	0	0	0	0	0	0	0	0	0.33	0.334	0	0.196	0	0	0	0	0	0
28	0.761	0.906	0.907	0.764	0.456	0.451	0.737	0	0.306	0.34	0.307	0.34	0.283	0	0	0	0	0	0	0.152
29	0	0	0.203	0	0	0	0.255	0	0	0	0.271	0	0	0	0	0	0	0	0	0
30	0	0.553	0.603	0	0.456	0	0.271	0	0	0	0.904	0.907	0	0	0	0	0	0	0	0

Table 3 Contd…

31	0	0.24	0	0	0	0	0	0	0	0	0	0.903	0	0.403	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0.228	0	0	0	0	0	0
33	0	0.429	0.55	0	0	0	0.181	0	0	0	0	0	0	0	0	0	0	0.182	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.302	0	0
35	0	0.372	0.161	0.452	0	0	0	0	0	0.18	0	0	0.826	0	0	0.934	0.246	0	0.235	0
36	0	0.368	0.161	0.369	0	0	0	0	0	0	0	0	0.228	0	0	0	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0.159	0	0	0	0.164	0	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0.173	0.38	0	0	0	0	0	0.34	0	0.154	0.318	0	0	0.191	0	0	0	0
40	0	0	0	0	0	0	0.244	0	0	0	0	0	0	0	0	0	0	0	0	0
41	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
42	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
43	0	0.227	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46	0	0.244	0.177	0.294	0.158	0	0	0	0	0	0.26	0	0	0.308	0	0.326	0.179	0.302	0	0
47	0	0.227	0.22	0	0	0	0	0	0.208	0	0	0	0	0	0	0	0	0	0	0
48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
51	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
52	0	0.387	0.234	0.357	0	0	0	0	0	0	0	0	0.18	0	0	0	0	0	0	0
53	0	0.839	0.967	0	0.824	0	0	0	0	0	0.677	0	0.646	0	0	0	0.193	0	0	0
54	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
55	0	0.416	0.269	0	0	0	0	0	0	0	0	0	0.628	0	0	0	0	0	0	0
56	0.769	0.872	0.829	0.939	0.817	0	0	0	0	0.202	0.195	0	0.703	0	0	0	0	0	0	0
57	0	0	0.198	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
59	0	0	0	0	0	0	0	0	0	0	0.26	0	0	0	0	0	0	0	0	0

Table 4
Interaction Score Matrix: (Part III)

Gene Index	41	42	43	44	45	46	47	48	49	50	51	52	53	54	55	56	57	58	59
1	0	0	0	0	0	0	0.915	0.904	0.92	0	0.903	0	0	0	0.899	0	0	0	0
2	0	0	0	0.159	0.152	0	0	0.903	0.914	0	0.919	0	0	0	0	0	0	0	0
3	0	0	0	0	0	0	0.915	0.926	0.978	0	0.904	0	0	0	0.902	0	0	0	0
4	0	0	0	0	0	0	0	0	0	0	0	0	0.173	0	0.188	0.195	0.219	0.758	0
5	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	0.509	0.508	0.51	0.638	0.577	0	0.171	0	0.288	0	0.155	0	0.573	0.392	0.399	0	0.321	0.226	0
7	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	0.814	0.813	0.813	0.294	0.219	0	0	0	0.189	0	0	0	0.257	0	0	0	0	0	0
9	0	0	0	0	0	0	0	0	0.172	0	0	0	0	0	0	0	0	0	0

Table 4 Contd…

10	0.82	0	0.821	0.324	0.241	0	0	0.168	0.222	0	0.167	0	0	0	0	0	0	0	0
11	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.177	0
13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
15	0	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0	0	0
16	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
17	0	0	0	0	0	0	0	0.8	0	0	0	0	0	0	0	0	0	0	0
18	0	0	0	0	0	0	0.828	0.806	0.8	0	0	0	0.17	0.623	0	0	0	0	0
19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.874	0
20	0.7	0	0	0	0	0	0.919	0.8	0.8	0	0	0	0	0	0	0	0	0	0
21	0	0	0	0	0	0	0.619	0	0	0	0	0	0.651	0	0	0	0	0	0
22	0	0	0	0.358	0.232	0	0	0	0.396	0	0	0	0	0	0	0	0	0	0
23	0	0	0	0	0	0	0	0.27	0.778	0	0	0	0	0	0	0	0	0	0
24	0.162	0	0.239	0.385	0.177	0	0.296	0.272	0.714	0	0.275	0	0.963	0	0.386	0	0.198	0.18	0
25	0	0	0	0	0	0	0	0	0.224	0	0	0	0	0	0	0	0	0	0
26	0	0	0	0.173	0	0	0	0	0.307	0	0	0	0.831	0	0.392	0.864	0	0	0
27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
29	0	0	0	0	0	0	0	0.158	0.158	0	0	0	0	0	0	0	0	0	0
30	0	0	0	0	0	0	0	0	0	0	0	0	0.653	0	0	0	0	0	0
31	0	0	0	0	0	0	0	0.294	0.796	0	0	0	0	0	0	0	0	0	0
32	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
33	0	0	0	0	0	0	0	0	0	0	0	0	0	0.626	0	0	0	0	0
34	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
35	0.87	0.156	0.658	0.409	0.425	0.369	0.82	0.817	0.816	0	0.211	0	0.92	0.64	0.645	0	0	0	0
36	0.853	0	0	0.37	0.348	0	0.819	0.817	0.816	0.156	0	0	0.915	0.96	0.638	0	0	0	0
37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.992	0
38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
39	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0.29	0	0.243	0
40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
41	0	0	0.908	0.906	0.906	0	0	0	0	0	0	0	0	0.642	0	0	0.234	0	0
42	0.902	0	0.922	0.908	0.907	0	0	0	0.902	0.208	0.322	0	0	0	0	0	0.198	0	0
43	0	0	0	0.905	0.906	0	0	0	0.902	0	0.276	0	0	0	0	0	0	0	0
44	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
45	0	0	0	0.905	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
46	0	0	0	0.903	0.903	0	0.169	0.169	0	0	0.162	0	0.16	0.161	0.17	0.161	0.163	0	0
47	0.377	0	0.478	0.242	0.169	0	0	0.902	0	0	0.969	0	0	0	0	0	0	0.176	0
48	0	0	0	0.169	0.169	0	0	0	0	0	0	0	0	0	0	0	0	0	0
49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
50	0.208	0	0.208	0	0	0	0	0	0	0	0.173	0	0	0	0	0	0	0	0
51	0	0	0	0.152	0	0	0	0.91	0.947	0	0	0	0	0	0	0	0	0	0
52	0	0	0	0	0	0	0.882	0.914	0.956	0.182	0.908	0	0.317	0	0	0	0	0	0
53	0.243	0	0.324	0.38	0.186	0	0.919	0.808	0.814	0	0.319	0	0	0.904	0.639	0	0	0	0
54	0	0	0	0.164	0.161	0	0	0	0	0	0	0	0	0	0	0	0	0.229	0
55	0.16	0	0.16	0.209	0.17	0	0	0	0	0	0	0	0	0.904	0	0	0.901	0	0
56	0	0	0	0.166	0.298	0	0.801	0.801	0.803	0	0	0	0.903	0.906	0	0	0	0	0
57	0	0	0.357	0.163	0.161	0	0	0	0	0	0	0	0	0.902	0	0	0	0	0
58	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
59	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

From the above table we can see that the highest number of interaction is with gene no. 6 and 24 which is CALM2, GRIN2B and it is associated with 49 no. Of genes, and the lowest number of interaction is with gene no. 15 which is CPEB3 and it is associated with 1 number of genes. Thus, we can obtain the highest and lowest interaction score of a gene, a group of genes.

4.2 K means Clustering

K-means [23] is one of the simplest unsupervised learning algorithms that solve the well known clustering problem. The procedure follows a simple and easy way to classify a given data set through a certain number of clusters (assume k clusters) fixed a priori. The main idea is to define k centroids, one for each cluster.

K means algorithm aims at minimizing an objective function, in this case a squared error function. The objective function

$$J = \sum_{j=1}^{k} \sum_{j=1}^{k} \left\| x_i^{(j)} - c_j \right\|^2$$

where $\left\| x_i^{(j)} - c_j \right\|^2$ is a chosen distance measure between a data point $x_i^{(j)}$ and the cluster centre c_j, is an indicator of the distance of the n data points from their respective cluster centres.

Steps to compute K means Clustering:

1. Place K points into the space represented by the objects that are being clustered. These points represent initial group centroids.
2. Assign each object to the group that has the closest centroid.
3. When all objects have been assigned, recalculate the positions of the K centroids.
4. Repeat Steps 2 and 3 until the centroids no longer move. This produces a separation of the objects into groups from which the metric to be minimized can be calculated.

The number of Clusters is equal to 8 because the number of clusters given by STRING 9.1 is 8. Here, K =8

Steps to calculate Interaction Score:The Pseudo code for calculating the interaction score is given below:

Step 1: Calculate the Interaction Score for each Cluster

Pseudo Code:

```
Load Interaction Score Matrix (59 x 59)
Cluster Interaction Score = 0
Number of Cluster = n
Cluster Element = [c1, c2, ..., cn]
For i = 1 to n
{
        For j = i+1 to n
        {
                Cluster Interaction Score = Cluster Interaction Score + Interaction Score Matrix (Cluster Element[i], Cluster Element[j])
        }
}
```

Step 2: Calculate Total Interaction Score.

$$\text{Total Interaction Score} = \sum (\text{Interaction Score of Each Cluster})$$

The Output Obtained for clustering is given below: in table 5.

Table 5: Clustering Output

K-means output showing clusters and genes belonging to them

Cluster 1: CHRFAM7A, COMT, HTR2A, NRXN1, WWC1

Cluster 2: BDNF, GRIN3A, GRIN3B, GRINA, GRM4, GRM5, PRKCA, PRKCG

Cluster 3: GRIA1, GRIA4, GRIN1, GRIN2A, GRIN2B, GRIN2C, GRIN2D, GRM1, GRM3, PKD2L2

Cluster 4: CAMK2A, CAMK2B, CAMK2D, CAMK2G, PPP1CA, PPP3CA, PPP3CB, PRKAG1

Cluster 5: PDYN, PRKACA, PRKACB, PRKACG, PRKCE

Cluster 6: APOE, CAMTA1, FYN, NCAM1, PPP3CC, PRKCD, PRKCH, PRNP

Cluster 7: ADCY1, ADCY2, ADCY8, CPEB3, CREB1, CREB3, CREBBP, PRKAR1A, PRKAR2B

Cluster 8: CALM2, CLSTN2, MAPK1, MAPK3, PPP1CB, PPP1CC

Here, We obtain Kmeans interaction Score= 73.176

Table 6 shows the comparison of genes in each cluster. It is observed that distribution of genes in each cluster from the k means method is almost uniform in comparison to the STRING 9.1 result. This uniformity of gene cluster would give uniform score in each cluster.

Table 6
Comparison of Results

No. of genes in each Cluster	*Cluster1*	*Cluster2*	*Cluster3*	*Cluster4*	*Cluster5*	*Cluster6*	*Cluster7*	*Cluster8*
K means Clustering by STRING 9.1	4	5	6	6	6	7	12	13
K means Clustering	5	8	10	8	5	8	9	6

5. CONCLUSION

In this paper we have presented biological, cellular, molecular and signalling pathways related to our set of candidate genes of episodic memory. The gene-gene interaction networks and their interaction scores in confidence views are discussed and presented. Based on global interaction score candidate gene are clustered using STRING 9.1 resulting in 8 clusters having 4 , 5, 6, 6, 6, 7, 12, 13 genes in clusters 1, cluster 2, cluster 3, cluster 4, cluster 5, cluster 6, cluster 7 & cluster 8 respectively. We have proposed K means based Clustering for all 8 clusters using global Interaction score to get evenly distributed clusters of. Our algorithm give the distribution of clusters in 5-10 genes range compare to 4-13 genes cluster provided by STRING 9.1. However we can increase the number of clusters and also use hierarchical clustering and SOM based clustering, as a scope for further work.

References

[1] Andreas Papassotiropoulos and Dominique J.-F.deQuervain, “Genetics of human episodic memory: dealing with complexity”, Trends in Cognitive Sciences September 2011, Vol. 15, No. 9.

[2] STRING 9.1 web tool and databases.

[3] Dominique J.-F. deQuervain and Andreas Papassotiropoulos, “Identification of a genetic cluster influencing memory performance and hippocampal activity in humans”, 4270–4274 , PNAS ,2006 ,vol. 103 , no. 11.

[4] Neurochemistry and Molecular Neurobiology of Memory P. Dash . A.N. Moore # Springer-Verlag Berlin Heidelberg 2007

[5] M. Mayadevi*, G.M. Archana*,Ramya R. Prabhu* and R.V. Omkumar, Molecular Mechanisms in Synaptic Plasticity, Neuroscience – Dealing with Frontiers, www.intechopen.com,

[6] Citri, A. and R. Malenka (2008). "Mechanisms of plasticity in excitatory synapses."Neuropsychopharmacology Rev 1.

[7] Malinow, R., H. Schulman, et al. (1989). "Inhibition of postsynaptic PKC or CaMKII blocks induction but not expression of LTP." Science 245(4920): 862-866.

[8] Klann, E., S.-J. Chen, et al. (1993). "Mechanism of protein kinase C activation during the induction and maintenance of long-term potentiation probed using a selective peptide substrate." Proceedings of the National Academy of Sciences 90(18): 8337-8341.

[9] Pang, P. T., H. K. Teng, et al. (2004). "Cleavage of proBDNF by tPA/plasmin is essential for long-term hippocampal plasticity." Science 306(5695): 487-491.

[10] Yao, Y., M. T. Kelly, et al. (2008). "PKMζ maintains late long-term potentiation by N-ethylmaleimide-sensitive factor/GluR2-dependent trafficking of postsynaptic AMPA receptors." The Journal of Neuroscience 28(31): 7820-7827.

[11] Gerdeman, G. L. and D. M. Lovinger (2003). "Emerging roles for endocannabinoids in long-term synaptic plasticity." British journal of pharmacology 140(5): 781-789.

[12] Yamauchi, T., J. Kamon, et al. (2002). "Adiponectin stimulates glucose utilization and fatty-acid oxidation by activating AMP-activated protein kinase." Nature medicine 8(11): 1288-1295.

[13] R. C. Malenka and R. A. Nicoll.Long-term potentiation–a decade of progress? Science, 285(5435):1870-4, 1999.

[14] N. Otmakhov, J. H. Tao-Cheng, S. Carpenter, B. Asrican, A. Dosemeci, T. S. Reese, and J. Lisman. Persistent accumulation of calcium/calmodulin-dependent protein kinase II in dendritic spines after induction of NMDA receptor-dependent chemical long-term potentiation. J Neurosci, 24(42):9324-31, 2004.

[15] Ito T, Chiba T, Ozawa R, Yoshida M, Hattori M, Sakaki Y: A comprehensive two-hybrid analysis to explore the yeast protein interactome. ProcNatlAcadSci USA 2001, 98(8):4569-74. 3.

[16] Miller JP, Lo RS, Ben-Hur A, Desmarais C, Stagljar I, Noble WS, Fields S: Large-scale identification of yeast integral membrane protein interactions. ProcNatlAcadSci USA 2005, 102(34):12123-8.]

[17] Ho Y, Gruhler A, Heilbut A, Bader GD, Moore L, et.al.: Systematic identification of protein complexes in Saccharomyces cerevisiae by mass spectrometry. Nature 2002, 415(6868):180-3. 6.

[18] Uetz P, Giot L, Cagney G, Mansfield TA, et.al.: A comprehensive analysis of protein- protein interactions in Saccharomyces cerevisiae. Nature 2000, 403(6770):623-7. 2.

[19] Gavin AC, Bosche M, Krause R, Grandi P, Marzioch M, Bauer A, Schultz J, et.al.: Functional organization of the yeast proteome by systematic analysis of protein complexes. Nature 2002, 415(6868):141-7. 5.

[20] Gavin AC, Aloy P, Grandi P, Krause R, Boesche M, Marzioch M, Rau C, et.al.: Proteome survey reveals modularity of the yeast cell machinery. Nature 2006, 440(7084):631-636. 7.

[21] Krogan NJ, Cagney G, Yu H, Zhong G, Guo X, Ignatchenko A, Li J, Pu S, et.al.: Global landscape of protein complexes in the yeast Saccharomyces cerevisiae. Nature 2006, 440(7084):637-643.].

[22] Sylvain Brohée* and Jacques van Helden, “Evaluation of clustering algorithms for protein-protein interaction Networks”, BMC Bioinformatics 2006, 7:488.

[23] J. MacQueen, “Some methods for classification and analysis of multivariate observations” University of California, Los Angeles, 1967.