Genomics 111 (2019) 958–965

Contents lists available at ScienceDirect

Genomics

journal homepage: www.elsevier.com/locate/ygeno

Methods Paper

# Exploring the multi-drug resistance in *Escherichia coli* O157:H7 by gene interaction network: A systems biology approach

Sravan Kumar Miryala, Sudha Ramaiah*

*Medical and Biological Computing Laboratory, School of Biosciences and Technology, Vellore Institute of Technology (VIT), Vellore 632014, Tamil Nadu, India.*

ARTICLE INFO

*Keywords:*
Gene network
*Escherichia coli* O157:H7
Multidrug resistance
Antimicrobial resistance gene
Clustering analysis

ABSTRACT

In the present study, we have constructed an interaction network of 29 antibiotic resistant genes along with 777 interactions in *E. coli* O157:H7. Gene ontology analysis reveals that 94, 89 and 67 genes have roles in the cellular process, biological process and molecular function respectively. Gene complexes related to tripartite efflux pumps *mdtEF-tolC* and ABC family efflux pump *macAB-tolC* play key roles in multidrug efflux systems. It is noteworthy to mention that, 19 genes are involved in multi-efflux pumps and they play a significant role in multidrug resistance (MDR); while 18 genes are vital for fatty acid synthesis. Interestingly, we found that the four genes *arnABCD* are involved in both MDR and in fatty acid synthesis. Hence these genes could be targeted for new drug discovery. On the whole, our results provide a detailed understanding of the mode of MDR mechanisms in *E.coli* O157:H7.

## 1. Introduction

*Escherichia coli (E. coli)* is a Gram-negative, rod-shaped bacterium and a well-known gut microbial flora. The pathogenicity attained by the non-pathogenic *E. coli* strains is due to the gene transfers or external plasmid insertions [1, 2]. The *E. coli* O157:H7, a Shiga-toxin producing Entero-hemorrhagic strain is the causative pathogen of the hemolytic uremic syndrome (HUS) in humans. It transmits mainly through the contamination of meat while butchering and packaging. The HUS symptoms include bloody diarrhoea and severe abdominal pain that leads to the formation of clots in capillaries and causes hemolytic anemia, thrombocytopenia and renal failure. The treatment is limited to rehydration, medication for fever and pain [3–5]. *E. coli* O157:H7 strain has much higher intrinsic levels of resistance to different antibiotics due to the effectiveness of the outer membrane (OM) as a barrier and through multidrug efflux pumps. Efflux pumps (EP's) helps in reducing OM permeability and thus reduces antibiotic uptake which results in drug resistance [6]. EP's affects almost all classes of antibiotics including Macrolides, Tetracyclines, and Fluoroquinolones. These antibiotics act on inhibition of DNA or protein biosynthesis and hence they need to penetrate into the cell to exert the effect. Some EP's are drug specific but most of them are multidrug transporters and capable of expelling a huge spectrum of structurally unrelated drugs [7]. Unlike the targeting of single pathway, a multi-targeted approach is necessary to establish successful treatment for HUS caused by *E. coli* O157:H7. Thus the better understanding of the pathogenesis and drug resistance mechanisms is important to identify better targets and to develop novel drugs to treat HUS [8].

The gene network based approaches becomes essential to determine the impact of genes or proteins on biological functions [9]. The interaction network analysis helps to explore the useful biological information of AMR mechanisms which in return helps to identify the key target genes or proteins in cascade and to design novel drugs to control the infections caused by the AMR pathogenic strains [10–13]. Previous work using computational approaches from our lab has reported the regulation of genes on *ampC* beta-lactamase synthesis, multidrug-resistant gene interaction networks in *A. baumannii* and *S. aureus* [14–16]. The present study includes the extraction of AMR genes, construction and analysis of interactome in pathogenic *E. coli* O157:H7 strain (Fig. 1). The molecular interactions and the molecular mechanisms of the AMR genes are presented which will be very much-needed for the discovery of the novel and potent drugs for the successful treatment of the disease. We have used the clustering approach and topology approaches to identify the biologically important genes which play an important role in drug resistance mechanisms. The gene interaction network approach is one of the promising ways of studying the

*Abbreviations:* AMR, Antimicrobial resistance; ARDB, Antibiotic resistant genes Database; BP, Biological processes; CC, Cellular components; COG, Clusters of Orthologous Groups; EP, Efflux pump; GM, GeneMania; GO, Gene Ontology; HUS, Hemolytic Uremic Syndrome; KEGG, Kyoto Encyclopaedia of Genes and Genomics; MCODE, Molecular Complex Detection; MDR, Multidrug Resistance; MF, Molecular Functions; OM, Outer membrane; RND-pump, Resistance-nodulation-division pump; STRING, Search Tool for the Retrieval of Interacting Genes/Proteins

* Corresponding author at: Department of Bio-Sciences, Vellore Institute of Technology (VIT), Vellore, India.
*E-mail address:* sudhaanand@vit.ac.in (S. Ramaiah).

https://doi.org/10.1016/j.ygeno.2018.06.002
Received 17 November 2017; Received in revised form 8 June 2018; Accepted 11 June 2018
Available online 13 June 2018
0888-7543/ 

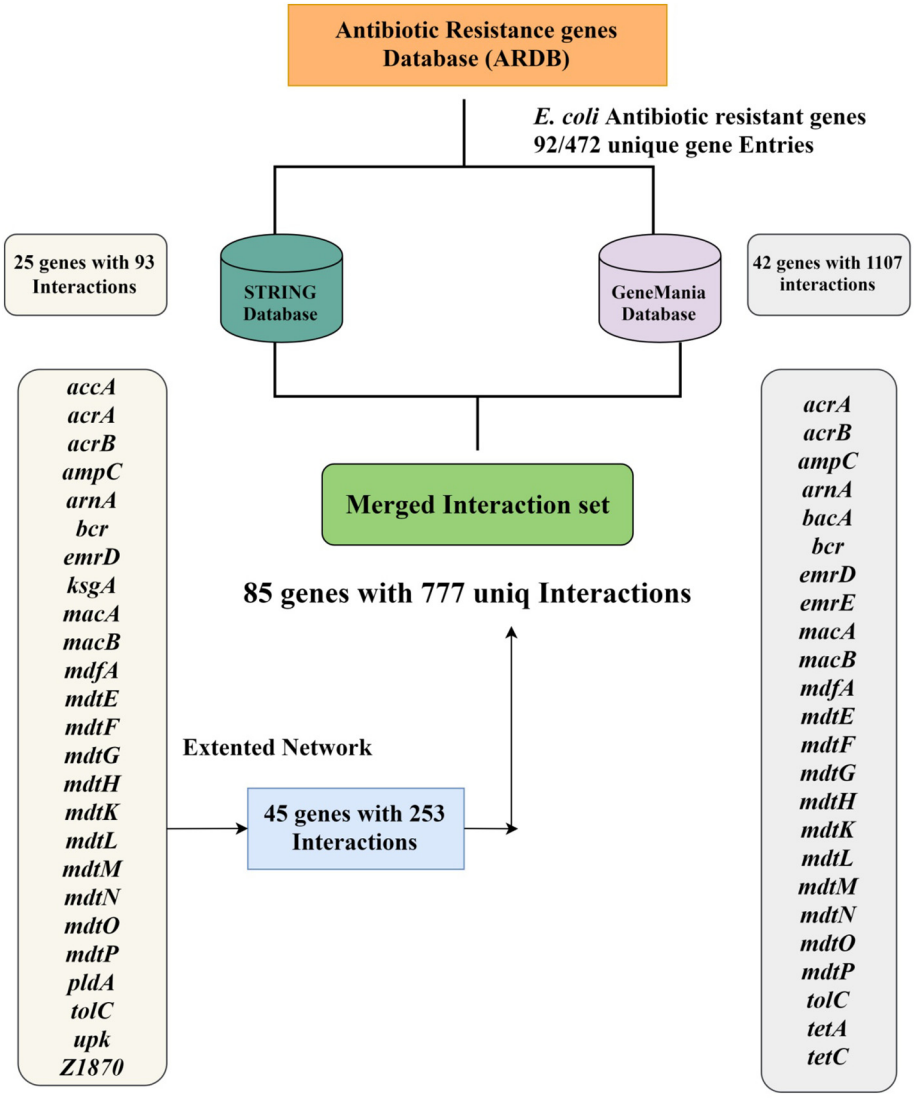


**Fig. 1. A systematic workflow of AMR gene interaction network analysis of *E. coli* O157:H7 strain.** The list of AMR genes collected from the interaction databases GeneMania and STRING. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

functions of gene or protein and the associated partners. Highly interconnected genes are clustered and share a similar functional annotations and pathways. By using the topological approach we have measured the centrality in the network to study the global structure. The shortest path length describes the length of edge between the nodes. Simultaneously we have calculated the shortest path length between antibiotic resistant genes or proteins to find the correlation in gene expression levels. Similarly, the connecting edges of genes or proteins within the clusters which are exhibiting high edge betweenness centrality are considered as connecting bridges of the network [17–19].

## 2. Results

### 2.1. Data collection and network construction

We have collected resistance genes in *E. coli* from Antibiotic Resistance gene Database (ARDB). Out of 472 collected resistance genes 92 entries are found to be unique. We have used two well-known databases for interaction data, STRING and GeneMania (GM) for collecting interaction data. STRING and GM databases have returned interactions for 25 AMR genes out of 92 unique genes. We further verified for the redundancy in obtained interacting pairs from both the databases. Out of 50 genes only 29 genes are unique and the interactions were collected from the two databases. We have collected interactions from each database separately. STRING database has given hits related to various *E. coli* strains and we have chosen *E. coli* O157:H7 strain for the study because of its severe impact on humans. The STRING database has 90 interactions between 25 AMR genes. The genes include *accA*, *acrAB, ampC, arnA, bcr, emrD, macAB, mdfAEFGHKLMNOP, pldA, tolC, upk* and *Z1870*. As we have found there is less connectivity between the AMR genes found in STRING database, we have modified our search criteria from medium confidence (0.4–0.6) to low confidence (0.1–0.3) to capture a maximum number of interactions. We further extended the network with the interactors of AMR genes and returned with 45 nodes and 254 edges. GeneMania has 1226 interactions between 42 genes, out of which 25 genes are AMR genes. The set of 25 AMR genes collected from GeneMania includes *acrAB, ampC, arnA, bacA, bcr, emrDE, macAB, mdfAEFGHKLMNOP, tolC* and *tetAC*. The interaction data collected from both databases has many repetitions, to reduce the redundancy we have combined the interaction data and removed the duplicate values. The final interaction data is left over with 777 interactions between 85 interactors. We have used Cytoscape 3.5.1 to construct the interaction network of 29 AMR genes.

**Table 1**
Clustering analysis using Cytoscape-MCODE tool.

| Cluster | Score (density) | Nodes | Edges | Gene name | AMR genes |
|---|---|---|---|---|---|
| 1 | 25.517 | 30 | 397 | *aaeB, acrA, acrD, acrE, acrF, ampC, arnA, bacA, bcr, cusC, emrA, emrD, emrK, macA, macB, mdfA, mdtB, mdtC, mdtE, mdtF, mdtH mdtI, mdtJ, mdtL, mdtN, mdtP, sugE, tolC, ydhJ, ydhK* | *mdtN, mdtE, emrD, mdtL, mdfA, macA, macB, ampC, mdtF, acrA, bacA, mdtP, bcr, mdtH, arnA, tolC* |
| 2 | 12.833 | 13 | 77 | *Z4863, Z4866, accB, fabD, accD, accA, fabH, fabF, ECs3207, Z1549, ECs1472, fabG, accC* | *accA* |
| 3 | 10 | 10 | 45 | *betI, envR, slmA, ybjK, ycfQ, acrR, ybiH, uidR, tetC, bdcR* | *tetC* |
| 4 | 4 | 4 | 6 | *arnC, arnD, arnT, arnB* | – |
| 5 | 3 | 3 | 3 | *acrB, cusA, mdtK* | *acrB, mdtK* |

List of genes involved in clusters along with the MCODE scores. AMR genes from each cluster are given in separate column.

**Table 2**
Network analysis by using NetworkAnalyzer.

| Gene name | MCODE cluster | Average shortest path length | Cluster coefficient | Closeness centrality |
|---|---|---|---|---|
| *fabG* | Cluster 2 | 1 | 0.492424 | 1 |
| *fabD* | Cluster 2 | 1 | 0.461538 | 1 |
| *fabH* | Cluster 2 | 1 | 0.406593 | 1 |
| *Z1549* | Cluster 2 | 1 | 0.352564 | 1 |
| *mdtG* | NIC | 1 | 0.30303 | 1 |
| *msyB* | NIC | 1.5 | 0.3 | 0.666667 |
| *ydhJ* | Cluster 1 | 1.555556 | 0.421371 | 0.642857 |
| *aaeB* | Cluster 1 | 1.629032 | 0.366667 | 0.613861 |
| *arnB* | Cluster 4 | 1.666667 | 0.238095 | 0.6 |
| *sugE* | Cluster 1 | 1.83871 | 0.398551 | 0.54386 |
| *mdtI* | Cluster 1 | 1.967742 | 0.353846 | 0.508197 |
| *acrE* | Cluster 1 | 2.016129 | 0.413306 | 0.496 |
| *ydhK* | Cluster 1 | 2.145161 | 0.438095 | 0.466165 |
| *acrF* | Cluster 1 | 2.209677 | 0.375223 | 0.452555 |
| *acrD* | Cluster 1 | 2.258065 | 0.393048 | 0.442857 |
| *yeeO* | NIC | 2.274194 | 0.428947 | 0.439716 |
| *tolC* | Cluster 1 | 2.274194 | 0.296154 | 0.439716 |
| *ampC* | Cluster 1 | 2.306452 | 0.362698 | 0.433566 |
| *cusC* | Cluster 1 | 2.322581 | 0.414773 | 0.430556 |
| *mdtF* | Cluster 1 | 2.322581 | 0.330014 | 0.430556 |

The top 20 genes with the shortest path length and high closeness centrality are listed here. The average shortest path length gives the expected distance between the two connected nodes and genes with shortest path length and high closeness centrality are considered as the controlling points of molecular communication. The cluster coefficient value lies in between 0 and 1. The value 0 indicates the node with less than 2 neighbours. (Note: NIC -Not in cluster).

### 2.2. Clustering analysis

We have used a Cytoscape plug-in called MCODE for clustering the genes based on the connectivity scores and marked the five clusters (C1–C5). The clustering analysis was performed using default parameter of MCODE tool to ensure the efficiency of functional partners towards AMR genes. We have identified five efficient clusters based on the interactions between the genes. Around 60 functional partners were found to be present in five clusters. Among the clusters, the first cluster possessed 30 nodes with a score 25.517 while the second cluster contained 13 nodes with score 12.833. Subsequently, the third, fourth and fifth clusters contained 10 nodes, 4 nodes and 3 nodes with the scores 10, 6 and 3 respectively (Table 1).

### 2.3. Functional enrichment analysis

Functional enrichment analysis using STRING tool has resulted the enriched GO terms such as cellular process (CC), molecular function (MF) and biological process (BP) for 29 resistance genes and for their 777 interacting functional partners. Further, the significant GO terms were selected based on the *p*-value less than 0.05. Out of 777 functional interactions, the corresponding GO terms were extracted, in which 89 genes are found to have role in biological processes, 67 genes in molecular function and 94 genes for cellular process from the STRING.

### 2.4. Shortest path length and closeness centrality analysis

The shortest path lengths and closeness centrality of the nodes in clusters are calculated by using NetworkAnalyzer. We have computed the pairwise betweenness of each functional partner in the network. The genes along with the calculated average shortest path length, cluster coefficient and high closeness centrality of the genes has been provided as Supplementary file 1.

## 3. Discussion

AMR genes in pathogenic bacteria play a very crucial role in attaining resistance to various drugs used against the disease. There are multiple ways adopted in order to resist the entry of antibiotics inside the cell by different efflux mechanisms. In our study the gene network, clustering analysis and functional enrichment analysis of AMR genes and their functional partners has provided valuable information on drug efflux pumps and tripartite efflux systems. The enriched GO terms such as cellular components, biological processes and molecular functions, gene complexes, signalling pathways related to the multidrug efflux systems have provided the complete profile of AMR genes interactions, functional partners and the role of their association towards the drug resistance.

For each pair of functional partners in clusters the average shortest path length, cluster coefficient and closeness centralities are calculated by using NetworkAnalyzer. The average shortest path length gives the expected distance between the two connected nodes. Genes with shortest path length and high closeness centrality are considered as the controlling point of molecular communication. The clustering coefficient of each is calculated based on the number of neighbours it is connected. The nodes with less than two neighbours are assumed to have a clustering coefficient of 0. The clustering coefficient values normally lie in between 0 and 1 (Table 2). The interaction network of the genes obtained from STRING and GeneMania has shown the dense interactions between AMR genes and the functional interactors. We further carried out the functional enrichment analysis of all the interactors using STRING tool, it has shown that the genes in network play a major role in biological processes related to lipid metabolism and fatty acid synthesis (Supplementary file 2; GO enrichment). The processes lipid metabolic process (GO.0006629), lipopolysaccharide biosynthetic process (GO.0009103), biosynthetic process (GO.0008610), fatty acid biosynthetic process (GO.0006633), cellular lipid metabolism process (GO.0044255) are highly enriched in our results. Along with the processes related to drug response and drug transportation such as response to the chemical (GO.0042221), response to an antibiotic (GO.0046677), drug transport (GO.0015893), response to the drug (GO.0042493) has been enriched. Molecular functions such as transporter activity (GO.0005215), drug transporter activity (GO.0090484), drug transmembrane transporter activity (GO.0015238), fatty acid

**Table 3**
Functional assessment of genes within the five clusters.

| AMR genes | Genes related to fatty acid synthesis | Genes responsible for Multi Drug resistance |
|---|---|---|
| *accA* | *accA* | *acrA* |
| *acrA* | *accB* | *arnA* |
| *acrB* | *accC* | *arnB* |
| *ampC* | *accD* | *arnC* |
| *arnA* | *arnA* | *arnD* |
| *bacA* | *arnB* | *macA* |
| *bcr* | *arnC* | *macB* |
| *emrD* | *arnD* | *mdfA* |
| *emrE* | *arnT* | *mdtE* |
| *ksgA* | *ECs1472* | *mdtF* |
| *macA* | *ECs3207* | *mdtG* |
| *macB* | *fabD* | *mdtH* |
| *mdfA* | *fabF* | *mdtK* |
| *mdtE* | *fabG* | *mdtL* |
| *mdtF* | *fabH* | *mdtM* |
| *mdtG* | *pldA* | *mdtN* |
| *mdtH* | *Z4863* | *mdtO* |
| *mdtK* | *Z4866* | *mdtP* |
| *mdtL* | | *Upk* |
| *mdtM* | | |
| *mdtN* | | |
| *mdtO* | | |
| *mdtP* | | |
| *pldA* | | |
| *tetA* | | |
| *tetC* | | |
| *tolC* | | |
| *upk* | | |
| *Z1870* | | |

List of AMR genes and functional partners responsible for fatty acid synthesis and multi drug resistance in *E.coli* O157:H7. The genes *arnABCD* are involved in both fatty acid synthesis and MDR.

synthase activity (GO.0004312). Enriched KEGG pathways include fatty acid biosynthesis, fatty acid metabolism, biotin metabolism, pyruvate metabolism and beta-lactam resistance.

From the enriched GO terms we have identified the genes responsible for fatty acid synthesis and MDR efflux related genes. The AMR genes *acrA, arnA, arnB, arnC, arnD, macA, macB, mdfA, mdtE, mdtF, mdtG, mdtH, mdtK, mdtL, mdtM, mdtN, mdtO, mdtP* and *Upk* are found to be related to MDR and the genes *accA, accB, accC, accD, arnA, arnB, arnC, arnD, arnT, ECs1472, ECs3207, fabD, fabF, fabG, fabH, pldA, Z4863* and *Z4866* are responsible for fatty acid synthesis or lipid metabolism (Table 3). The genes *arnA, arnB, arnC* and *arnD* have roles in both fatty acid synthesis and also in MDR (Fig. 2).

The clustering analysis between 85 interactors and 777 unique interactions by using MCODE has given the local dense graph of the gene complexes. Based on the connectivity score obtained from MCODE we have selected 5 clusters (C1-C5) with 60 functional partners (Fig. 3). Among the five clusters the C1 and C5 has enriched with GO terms related to response to antibiotics and drug transport activity, whereas the C2, C4 clusters are enriched with the lipid and fatty acid synthesis (Fig. 4). The functional enrichment analysis of 30 interacting genes in C1 cluster has shown the top enriched biological process, molecular function, cellular components and KEGG pathways. Cluster C2 has genes responsible for fatty acid synthesis and lipid metabolism and it is also shown to have highly dense interactions with C4, whereas the cluster C3 has no significant interactions with other clusters. The genes in cluster 3 are observed that they have the role in biological processes such as negative regulation of cellular process and transcription. Genes in cluster 3 contain PFAM protein domains belonging to *tetR* family related bacterial regulatory proteins (Supplementary file 2; MCODE cluster analysis).

In the cluster C1, out of 30 genes 12 genes (*acrA, arnA, macA, macB, mdfA, mdtE, mdtF, mdtH, mdtL, mdtN, mdtP* and *upk*) are associated to the antibiotics response and drug transport. The transport systems or efflux pumps in bacteria involves in transporting various compounds into or outside of the cell. The genes *acrA, macB, mdfA* and *mdtN* are responsible for biological processes such as drug transport (GO.0015893), drug response (GO.0042493), drug transporter mechanism (GO.0090484) and molecular function drug transporter activity (GO.0090484), drug transmembrane transporter activity (GO.0015238). The cluster C5 has densely connected with cluster C1 and enriched with the molecular functions like drug transmembrane transport and response to the drug. Enrichment of response to

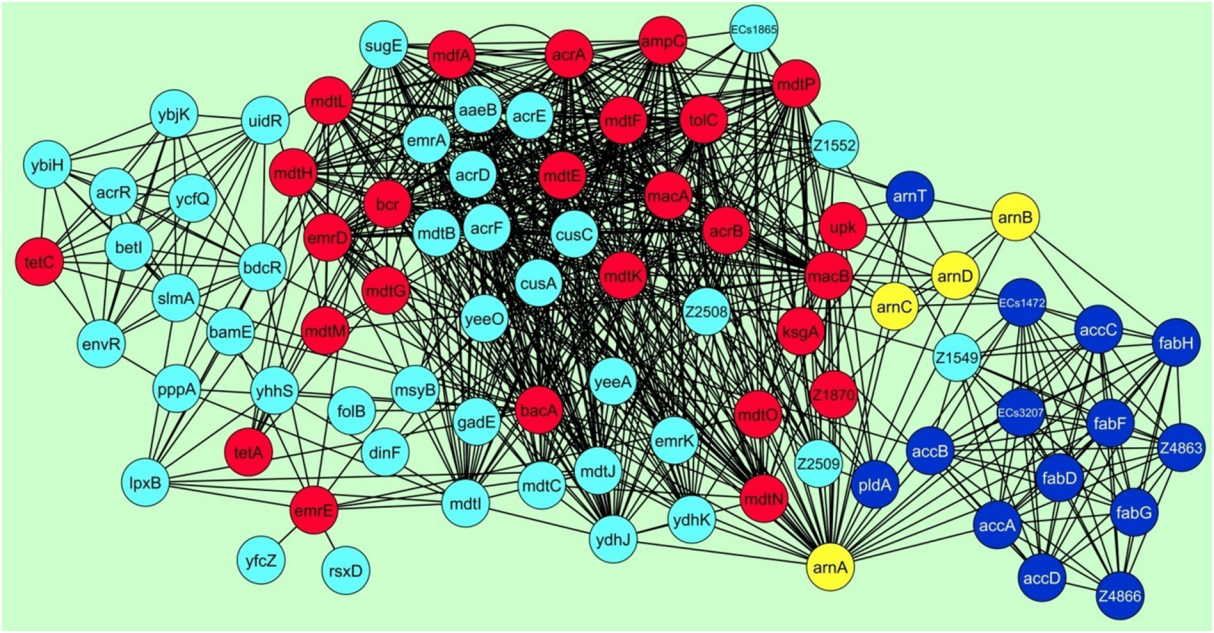


**Fig. 2. Functional assessment analysis of AMR genes.** The AMR genes along with their functional partners are mainly involved in fatty acid synthesis and multi drug resistance (MDR). The MDR related genes are highlighted with red colour. The genes related to fatty acid synthesis are highlighted with blue colour. Genes that share both the functionality (MDR and fatty acid synthesis) are given yellow colour. The other functional partners which are not involving in any of these functions are given in light blue colour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

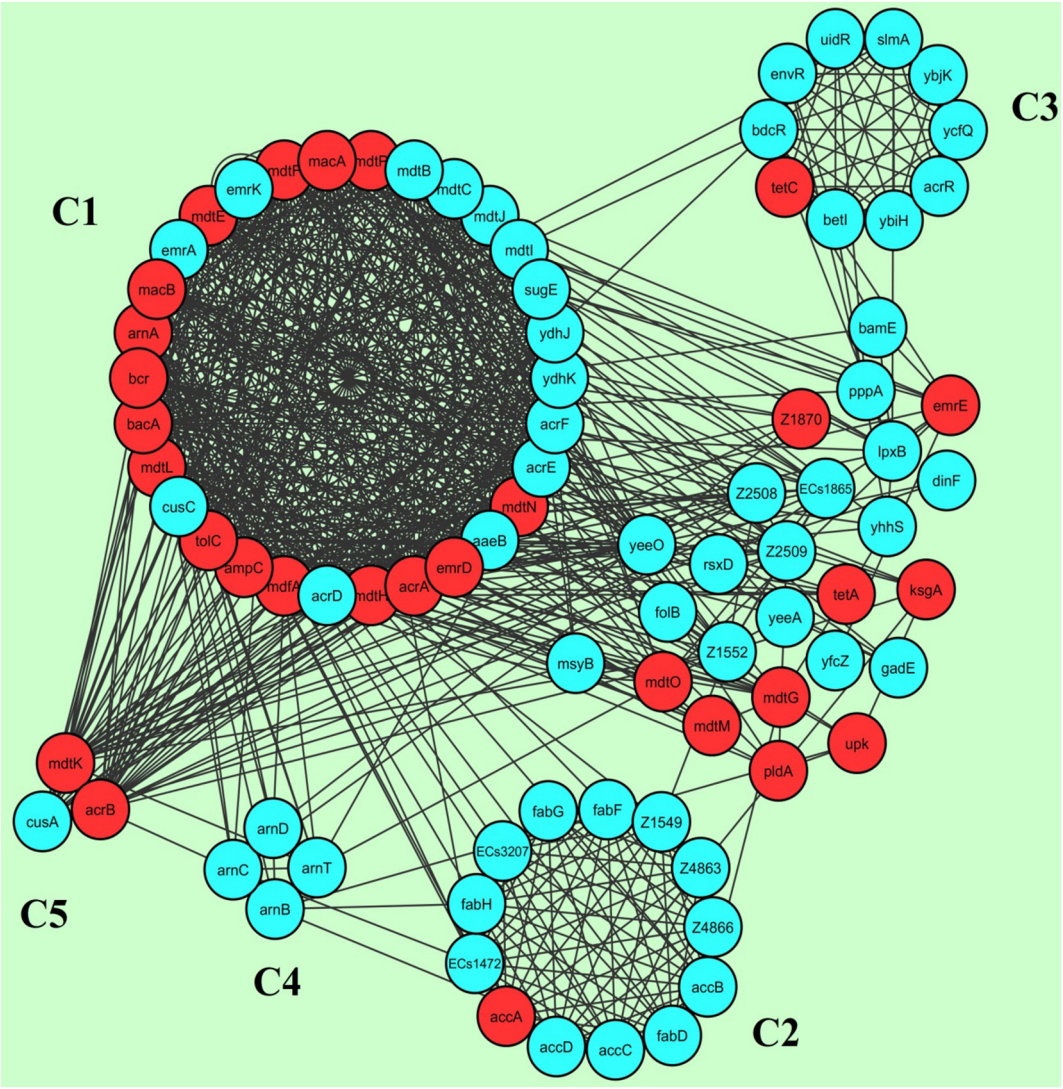


**Fig. 3. Clustering analysis of AMR gene interaction network:** Gene clusters obtained from MCODE are isolated and marked as C1 - C5. For easy identification AMR genes in the network are highlighted with red colour and the other functional partners are given light blue colour. The cluster C1 has 16 AMR genes and the cluster C5 has 2 AMR genes. The clusters C2, C3 has only one AMR gene each and there are no AMR genes in Cluster C4. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

antibiotics and transport of drugs shows the role of AMR genes in MDR mechanisms. The genes in clusters C2 (*Ecs1472, accA, accB, accC, accD, fabD, fabF* and *fabH*) and C4 (*arnB, arnC, arnD* and *arnT*) has been involved in enriched biological processes such as fatty acid biosynthesis, lipid-A biosynthetic process and lipo-polysaccharides biosynthesis process; molecular functions like fatty acid synthase activity, acetyl-CoA carboxylase activity and pathways related to fatty acid biosynthesis, fatty acid metabolism. Our enrichment results show that the AMR genes and associated functional partners play a major role in fatty acid biosynthesis which is considered to be one of the significant potential targets for the development of antibacterial compounds [19]. Our work may support the earlier studies carried out on Gram-negative bacteria [20, 21] and it has been reported the need to concentrate on bacterial fatty acid synthesis mechanism for novel antibacterial targets.

Another notable result includes the enrichment of gene complexes related to tripartite efflux pumps. The outer membrane of the Gram-negative bacteria acts as a barrier and will not permit substances like antibiotics, antiseptics, dyes, and detergents to enter into the cell thus the Gram-negative bacteria shows much intrinsic levels of resistance than Gram-positive bacteria [22, 23]. The porin channels permit the small molecules like nutrients and slow down the influx of antibiotics or effluxes the large compounds. Usually, the antibiotics are larger than the nutrients. Along with the outer membrane, certain Resistance-Nodulation-Division (RND) efflux pumps contribute in generating intrinsic resistance by forming a tripartite complex with periplasmic proteins and outer membrane channels. All the three components are important for RND pumps function; the absence of a single component may cause the complex non-functional [24]. In our results, we have noticed the enrichment of multidrug efflux pumps, along with the fatty acid synthesis pathways which infer the contribution of the AMR genes in the tripartite complex.

In our gene network, out of 85 interacting genes, 22 genes are specifically related to various drug efflux pumps. We have calculated the number of interactions for each gene (Supplementary file 2; Functional annotation) and the top five genes which have the highest number of interactions are *macA, tolC, mdtF, acrA* and *mdtE* with 45, 45, 45, 44 and 42 edges respectively. *macA* gene is a macrolide transporter subunit and it is a part of *macAB-tolC* system. The genes *mdtE, mdtF* are the components of *mdtEF-tolC* tripartite efflux system [25]. *AcrA* gene plays important role in *acrABZ-tolC* drug efflux system [26, 27]. The *macB* gene with 37 edges is a macrolide transporter ATP binding permease and part of tripartite efflux system *macAB-tolC*, ABC family efflux

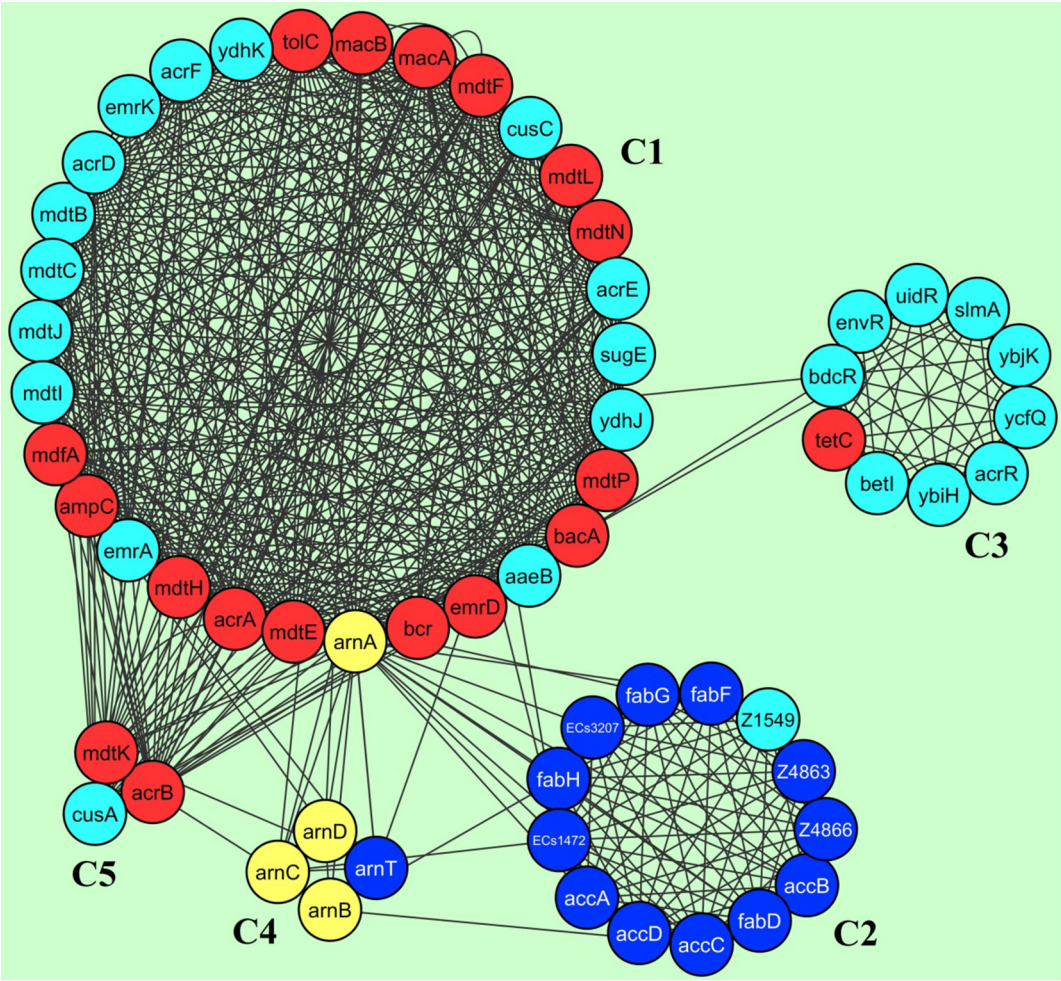


**Fig. 4. Functional similarity between the clusters.** The five gene clusters are compared using the functional enrichment analysis. The clusters C1 and cluster C5 share the genes that are involved in MDR by efflux pumps. Clusters C2 and C4 has genes related to fatty acid synthesis. Whereas the cluster C3 has no significant enrichment of fatty acid synthesis or MDR. As shown in the figure cluster C3 has very few interactions with other clusters. The MDR related genes are highlighted with red colour. The genes related to fatty acid synthesis are highlighted with blue colour. Genes that share both the functionality (MDR and fatty acid synthesis) are given yellow colour and the other functional partners are given light blue colour. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

pump. The gene *tolC*, responsible for outer membrane channel protein has been found to have a key role in many well known multidrug efflux systems [28,29]. Our results provide a clear understanding of the molecular mechanism works in attaining MDR in *E. coli* O157:H7 strain. The interaction network of AMR genes has shown functional partners which are responsible for drug efflux system, the functional annotations of all the interacting partners are provided. Thus our results show the complete profile of drug resistance mechanisms through the tripartite efflux systems by AMR genes along with the functional partners involved in the interaction network.

## 4. Conclusion

MDR in pathogenic bacteria is a major concern worldwide. Therefore there is an urgent need to explore the interactions between AMR genes and their functional partners for better understanding of the AMR mechanisms. We have presented an approach to identify MDR mechanisms along with the functional enrichment of GO terms. The resistance attained by the pathogenic bacteria *E. coli* O157:H7 strain is mainly through drug efflux mechanisms by the active efflux pumps and exploited to improve the treatments by developing new potent drugs. Our results provide a detailed information on the interactions of major efflux pumps which plays an active role in drug efflux in *E. coli* O157:H7. The interaction network generated in the present study helps us to understand the various mechanisms and their associated gene/protein interactions of AMR genes. These observations will be helpful for better understanding the AMR mechanisms in *E. coli* O157:H7 and the interactions between the resistance genes and their functional partners.

## 5. Materials and methods

### 5.1. Antibiotic resistant genes database (ARDB)

ARDB (http://ardb.cbcb.umd.edu/) is a manually curated database of publically available information on antibiotic resistance genes. The annotation of each AMR gene includes resistance profile, mechanism of action, ontology, and Cluster of Orthologous Groups (COG) along with the external links to sequence and protein databases. The database has information of 23,137 genes from 632 genomes of 1737 species [30].

### 5.2. STRING network analysis

STRING (https://string-db.org/) is a protein interaction database which consists of pre-computed gene or protein interactions. The interactions includes direct (through physical contact) and indirect

(functional similarity). Data can be retrieved by using the raw amino acid sequence or the unique protein identifier. The functional partners involved in the interactions are ranked by estimated confidence scores obtained from prediction algorithms based on the genomic information. The interactions are categorised in four classes based on the confidence scores, such as highest confidence (0.9–1), high confidence (0.7–0.8), medium confidence (0.4–0.6) and low confidence (0.1–0.3). These probabilistic confidence scores were scored based on a special scoring framework by STRING and it is based on the association in a common reference set. The interactions are integrated with the sources like genomic context, high throughput experimental data, literature survey, co-expressed gene analysis and database data mining. The current STRING version 10.5 consists of 1,380,838,440 interactions of 9,643,763 proteins from 2031 organisms [31–33].

### 5.3. GeneMANIA

GeneMANIA (http://www.genemania.org) is a user-friendly web interface used for generating hypothesis about gene function, analyzing gene lists and prioritizing genes for functional assays. A query of gene list results with the functionally similar genes which are identified using genomics and proteomics data. The weights assigned to the interactions indicate the prediction value of each selected data set. Currently the database supports with nine organisms including *A. thaliana, C. elegans, D. rerio, D. melanogaster, E. coli, M. musculus, H. sapiens, R. novergicus,* and *S. cerevisiae*. The data is collected from different databases such as GEO, BioGRID, Pathway commons, I2D and the organism-specific functional genomics datasets [34–36].

### 5.4. Gene interaction network construction

Cytoscape (http://www.cytoscape.org/) is one of the popular tools available as open source software and is used for visualization, analysis of molecular and genetic interaction networks. Cytoscape is compatible with all operating systems such as Windows, Linux and Mac OS [37, 38].

### 5.5. Clustering analysis

MCODE is a Cytoscape app used for clustering analysis. It is based on MCODE (Molecular Complex Detection) algorithm. It is used to identify clusters of nodes which are highly interconnected. The algorithm operates mainly in three stages, vertex weighting, complex prediction and optionally post-processing. The network of interacting molecules is modelled as a graph where every vertex is a molecule and the edge as a molecular interaction. Directed graph is used for known cell signalling and known pathways, otherwise undirected graph is used [39].

### 5.6. Shortest path length and closeness centrality analysis

NetworkAnalyzer is a user friendly and versatile Cytoscape plugin, used to compute and display topological parameters such as number of nodes, connecting edges, the network diameter, density, radius, centralization, heterogeneity, clustering coefficient, the characteristic path length, the distribution of node degrees, neighbourhood connectivity, average clustering coefficients and the shortest path lengths. It can be used to analyze both directed and undirected networks also allow constructing the intersection or union of two networks [17].

Supplementary data to this article can be found online at https://doi.org/10.1016/j.ygeno.2018.06.002.

## Conflict of interest statement

The authors declare that there is no conflict of interest.

## Acknowledgements

SR gratefully acknowledges the Indian Council of Medical Research (ICMR), Government of India agency for the research grant (IRIS ID: 2014-0099). MSK thanks ICMR for the research fellowship. The authors would like to thank the management of VIT for providing the necessary facilities to carry out this research work.

## References

[1] M.J. Blaser, J.M. Musser, M.S. Donnenberg, T.S. Whittam, Bacterial polymorphisms pathogenesis and evolution of virulence in enteropathogenic and enterohemorrhagic Escherichia coli, J. Clin. Invest. 107 (2001) 539–548, http://dx.doi.org/10.1172/JCI12404.
[2] J.Y. Lim, J. Yoon, C.J. Hovde, A brief overview of Escherichia coli O157:H7 and its plasmid O157, J. Microbiol. Biotechnol. 20 (2010) 5–14, http://dx.doi.org/10.4014/jmb.0908.08007.
[3] Wong, The risk of hemolytic-uremic syndrome, New Engl. J. Med. 342 (2000) 1930–1936, http://dx.doi.org/10.1056/NEJM200006293422601.
[4] J.B. Kaper, M.A. Karmali, The continuing evolution of a bacterial pathogen, Proc. Natl. Acad. Sci. U. S. A. 105 (2008) 4535–4536, http://dx.doi.org/10.1073/pnas.0801435105.
[5] Z.D. Blount, The unexhausted potential of *E. coli*, Elife 4 (2015) 1–12, http://dx.doi.org/10.7554/eLife.05826.
[6] L. Amaral, A. Martins, G. Spengler, J. Molnar, Efflux pumps of Gram-negative bacteria: What they do, how they do it, with what and how to deal with them, Front. Pharmacol. (2014) 1–11, http://dx.doi.org/10.3389/fphar.2013.00168.
[7] S. Dzidic, J. Suskovic, B. Kos, Antibiotic resistance mechanisms in Bacteria: biochemical and genetic aspects, Food Technol. Biotechnol. 46 (2008) 11–21.
[8] P.N. Goldwater, K.A. Bettelheim, Treatment of enterohemorrhagic Escherichia coli (EHEC) infection and hemolytic uremic syndrome (HUS), BMC Med. 10 (2012) 12, http://dx.doi.org/10.1186/1741-7015-10-12.
[9] S.K. Miryala, A. Anbarasu, S. Ramaiah, Discerning molecular interactions: A comprehensive review on biomolecular interaction databases and network analysis tools, Gene 642 (2018), http://dx.doi.org/10.1016/j.gene.2017.11.028.
[10] M.W. Gonzalez, M.G. Kann, Chapter 4: Protein Interactions and Disease, 8 (2012), http://dx.doi.org/10.1371/journal.pcbi.1002819.
[11] I. Journal, H. Genetics, Network and Polymorphic Analysis of Obesity Candidate Gene-plin1: A Bioinformatics Approach, (2015).
[12] S. Bag, A. Anbarasu, Revealing the strong functional association of adipor2 and cdh13 with adipoq: a gene network study, Cell Biochem. Biophys. 71 (2015) 1445–1456, http://dx.doi.org/10.1007/s12013-014-0367-9.
[13] S. Bag, S. Ramaiah, A. Anbarasu, Fabp4 is central to eight obesity associated genes: a functional gene network-based polymorphic study, J. Theor. Biol. 364 (2015) 344–354, http://dx.doi.org/10.1016/j.jtbi.2014.09.034.
[14] P. Anitha, A. Anbarasu, S. Ramaiah, Computational gene network study on antibiotic resistance genes of Acinetobacter baumannii, Comput. Biol. Med. 48 (2014) 17–27, http://dx.doi.org/10.1016/j.compbiomed.2014.02.009.
[15] P. Anitha, S. Bag, A. Anbarasu, S. Ramaiah, Gene and Protein network analysis of AmpC beta lactamase, Cell Biochem. Biophys. 71 (2015) 1553–1567, http://dx.doi.org/10.1007/s12013-014-0379-5.
[16] P. Anitha, A. Anbarasu, S. Ramaiah, Gene network analysis reveals the association of important functional partners involved in antibiotic resistance: a report on an important pathogenic bacterium Staphylococcus aureus, Gene 575 (2016) 253–263, http://dx.doi.org/10.1016/j.gene.2015.08.068.
[17] A. Fiannaca, M. La Rosa, A. Urso, R. Rizzo, S. Gaglio, A knowledge-based decision support system in bioinformatics: an application to protein complex extraction, BMC Bioinformatics 14 (Suppl. 1) (2013) S5, http://dx.doi.org/10.1186/1471-2105-14-S1-S5.
[18] Y. Assenov, F. Ramírez, S.E.S.E. Schelhorn, T. Lengauer, M. Albrecht, Computing topological parameters of biological networks, Bioinformatics 24 (2008) 282–284, http://dx.doi.org/10.1093/bioinformatics/btm554.
[19] J. Yoon, A. Blumer, K. Lee, An algorithm for modularity analysis of directed and weighted biological networks based on edge-betweenness centrality, Bioinformatics 22 (2006) 3106–3108, http://dx.doi.org/10.1093/bioinformatics/btl533.
[20] J. Yao, C.O. Rock, How bacterial pathogens eat host lipids: implications for the development of fatty acid synthesis therapeutics, J. Biol. Chem. 290 (2015) 5940–5946, http://dx.doi.org/10.1074/jbc.R114.636241.
[21] C.R.H. Raetz, C.M. Reynolds, M.S. Trent, R.E. Bishop, Lipid a modification systems in gram-negative bacteria, Annu. Rev. Biochem. 76 (2007) 295–329, http://dx.doi.org/10.1146/annurev.biochem.76.010307.145803.
[22] J.B. Parsons, C.O. Rock, Is bacterial fatty acid synthesis a valid target for antibacterial drug discovery? Curr. Opin. Microbiol. 14 (2011) 544–549, http://dx.doi.org/10.1016/j.mib.2011.07.029.
[23] H. Nikaido, Multidrug efflux pumps of gram-negative bacteria, J. Bacteriol. 178 (1996) 5853–5859, http://dx.doi.org/10.1128/jb.178.20.5853-5859.1996.
[24] J. Anes, M.P. McCusker, S. Fanning, M. Martins, The ins and outs of RND efflux pumps in Escherichia coli, Front. Microbiol. 6 (2015) 1–14, http://dx.doi.org/10.3389/fmicb.2015.00587.
[25] H. Nikaido, Y. Takatsuka, Mechanisms of RND multidrug efflux pumps, Biochim. Biophys. Acta, Proteins Proteomics 1794 (2009) 769–781, http://dx.doi.org/10.1016/j.bbapap.2008.10.004.
[26] T. Horiyama, K. Nishino, AcrB, AcrD, and MdtABC multidrug efflux systems are

involved in enterobactin export in *Escherichia coli*, PLoS One 9 (2014), http://dx.doi.org/10.1371/journal.pone.0108642.
[27] H.I. Zgurskaya, H. Nikaido, AcrA is a highly asymmetric protein capable of spanning the periplasm, J. Mol. Biol. 285 (1999) 409–420, http://dx.doi.org/10.1006/jmbi.1998.2313.
[28] E.B. Tikhonova, H.I. Zgurskaya, AcrA, AcrB, and TolC of Escherichia coli form a stable intermembrane multidrug efflux complex, J. Biol. Chem. 279 (2004) 32116–32124, http://dx.doi.org/10.1074/jbc.M402230200.
[29] K.N.W. Deininger, A. Horikawa, R.D. Kitko, R. Tatsumi, J.L. Rosner, M. Wachi, J.L. Slonczewski, A requirement of tolc and MDR efflux pumps for acid adaptation and GadAB induction in escherichia coli, PLoS One 6 (2011) 1–7, http://dx.doi.org/10.1371/journal.pone.0018960.
[30] B. Liu, M. Pop, ARDB - Antibiotic resistance genes database, Nucleic Acids Res. 37 (2009) 443–447, http://dx.doi.org/10.1093/nar/gkn656.
[31] D. Szklarczyk, A. Franceschini, M. Kuhn, M. Simonovic, A. Roth, P. Minguez, T. Doerks, M. Stark, J. Muller, P. Bork, L.J. Jensen, C. Von Mering, The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored, Nucleic Acids Res. 39 (2011) 561–568, http://dx.doi.org/10.1093/nar/gkq973.
[32] A. Franceschini, D. Szklarczyk, S. Frankild, M. Kuhn, M. Simonovic, A. Roth, J. Lin, P. Minguez, P. Bork, C. Von Mering, L.J. Jensen, STRING v9.1: Protein-protein interaction networks, with increased coverage and integration, Nucleic Acids Res. 41 (2013) 808–815, http://dx.doi.org/10.1093/nar/gks1094.
[33] D. Szklarczyk, A. Franceschini, S. Wyder, K. Forslund, D. Heller, J. Huerta-Cepas, M. Simonovic, A. Roth, A. Santos, K.P. Tsafou, M. Kuhn, P. Bork, L.J. Jensen, C. Von Mering, STRING v10: protein-protein interaction networks, integrated over the tree of life, Nucleic Acids Res. 43 (2015) D447–D452, http://dx.doi.org/10.1093/nar/gku1003.
[34] S. Mostafavi, D. Ray, D. Warde-Farley, C. Grouios, Q. Morris, GeneMANIA: a real-time multiple association network integration algorithm for predicting gene function, Genome Biol. 9 (2008) S4, http://dx.doi.org/10.1186/gb-2008-9-s1-s4.
[35] D. Warde-Farley, S.L. Donaldson, O. Comes, K. Zuberi, R. Badrawi, P. Chao, M. Franz, C. Grouios, F. Kazi, C.T. Lopes, A. Maitland, S. Mostafavi, J. Montojo, Q. Shao, G. Wright, G.D. Bader, Q. Morris, The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function, Nucleic Acids Res. 38 (2010) 214–220, http://dx.doi.org/10.1093/nar/gkq537.
[36] J. Montojo, K. Zuberi, H. Rodriguez, G.D. Bader, Q. Morris, GeneMANIA: Fast gene network construction and function prediction for Cytoscape, F1000Research 153 (2014) 1–7, http://dx.doi.org/10.12688/f1000research.4572.1.
[37] P. Shannon, A. Markiel, Z. Owen Ozier, N.S. Baliga, J.T. Wang, D. Ramage, N. Amin, B. Schwikowski, T. Ideker, Cytoscape: a software environment for integrated models of biomolecular interaction networks, Genome Res. (2003) 2498–2504, http://dx.doi.org/10.1101/gr.1239303.metabolite.
[38] B. Demchak, T. Hull, M. Reich, T. Liefeld, M. Smoot, T. Ideker, J.P. Mesirov, Cytoscape: the network visualization tool for GenomeSpace workflows, F1000Research 2014 (2014) 1–12, http://dx.doi.org/10.12688/f1000research.4492.2.
[39] G.D. Bader, C.W. Hogue, An automated method for finding molecular complexes in large protein interaction networks, BMC Bioinformatics 4 (2003) 2, http://dx.doi.org/10.1186/1471-2105-4-2.