Article

Investigate to find common gene and design a PPI network for vector borne diseases (Malaria, Dengue and Chikungunya) – A bioinformatics approach

Tanjina Akter[1], **Lubna Yasmin Pinky**[1], **Md. Mosaddik Hasan**[1], **Farzana Akter Chowdhury**[2], **Md. Imam Hossain**[3]

[1]Department of Computer Science and Engineering (CSE), MawlanaBhashani Science and Technology University (MBSTU), Santosh, Tangail-1902, Bangladesh

[2]Department of Computer Science and Engineering (CSE), University of South Asia, Banani, Dhaka, Bangladesh

[3]Department of Biotechnology and Genetic Engineering (BGE), MawlanaBhashani Science and Technology University (MBSTU), Santosh, Tangail-1902, Bangladesh

E-mail: lubnaju@yahoo.com

Received 22 March 2018; *Accepted 30 April 2018*; *Published 1 September 2018*

Abstract

Malaria, Dengue and Chikungunya are the most common mosquito-borne viral diseases transmitted to humans by day-biting *Aedes aegypti* and *Aedes albopictus* mosquitoes. Different types of gene are responsible for these viruses. The principal study of this research is to find the relationship between genetic variant for these three diseases and to create a common pathway regulatory or Protein-Protein Interaction (PPI) network. Our investigation goes through preprocessing, filtering, sorting and gene mining on the gathered gene (Malaria, Dengue and Chikungunya) using R to find the common associated genes by the process of reduction. The investigation shows that about 60%of the collected gene from different standard gene database is responsible for animal virus attack. After preprocessing, filtering and sorting using R toolkit, the number of collected gene for three diseases(A=malaria, B=dengue and C=chikunguniya) is reduced to 35%. Gene mining is done by intersection operation on (A, B), (B, C) and (C, A) that reduces the common associated gene from 35% to 5%. Finally, the reduction is done by intersecting AB,BC and CA that reduces the common gene from 5% to less than 1%.We have discovered five (5) common associated genes for these three virus diseases. However a common pathway with the five (5) common associated genes that has been designed for selective diseases.

Keywords malaria; dengue; chikungunya; data mining; PPI network; R toolkit; NCBI;UniHi.

Network Biology
ISSN 2220-8879
URL: http://www.iaees.org/publications/journals/nb/online-version.asp
RSS: http://www.iaees.org/publications/journals/nb/rss.xml
E-mail: networkbiology@iaees.org
Editor-in-Chief: WenJun Zhang
Publisher: International Academy of Ecology and Environmental Sciences

1 Introduction

The Computational Biology & Gene Regulation group aims at developing cutting-edge bioinformatics tools

with immediate application to real-life biological problems. Research are being focused on gene expression regulation and the mechanisms by which it can be disrupted in human diseases. Various Prior research have been held on bioinformatics that includes analyzing genes, finding PPI network and common pathway for associated diseases. Malaria, Dengue and Chikunguniya are vector borne viral disease caused and spread by mosquito's bite that can be lethal for humans. These three diseases may have direct and indirect connection with each other through inter-related associated gene. The presented research is a descendant of previous research which aimed to design a common PPI network with common associated gene for these vector borne diseases.

2 Background

In the era of technological evolution bioinformatics research has a great impact on processing and analyzing different living being's data using computer. Bioinformatics is conceptualizing biology in terms of molecules and then applying informatics techniques to understand and deals with the biological information: genes, genomes, proteins, cells, ecological systems, medical information and so on. Bioinformatics plays an important role in data processing, sequencing, handling large dataset, storing, transformation, DNA compressing, visualizing PPI network and last but not least drug design (Stéphaneet al., 2010). The presented research is a descendant of previous research which aimed to design a common PPI network for vector borne diseases (like malaria, dengue and chikunguniya).

Malaria is a parasitic infection transmitted by the female Anopheles mosquito, infecting humans and insects alternatively. Caused by four Plasmodium species (*P.vivax*, *P. falciparum*, *P.ovale* and *P.malariae*), malaria is a public health problem in 91 countries around the world, affecting 300 million people and responsible directly for about one million deaths annually (Nazrul and ZulKifleet al., 2011). Approximately 40% of the world's population is susceptible to malaria, mostly in the tropical and sub-tropical areas of the world. Africa accounts for 90% of the mortality burden for malaria and South-east Asia accounts for 9% of the burden. Bangladesh is considered as one of the malaria endemic countries in South Asia. Out of 65 districts, 13 districts bordering east and northeast parts of Bangladesh facing Indian states of Assam, Tripura and Meghalaya and part of Myanmar belong to the high risk malaria zone. The three hilly districts of Bangladesh (Khagrachari, Rangamatiand Bandarban) account for 80% of the total burden of malaria in this country.

Dengue is a mosquito-borne viral disease that has rapidly spread in all regions of world. It is caused by four types of viruses (DENV-1, DENV-2, DENV-3, DENV-4) belonging to the *Flaviviridae* family. The global prevalence of dengue has grown dramatically in recent decades.One recent estimate indicates 390 million dengue infections per year (95% credible interval 284–528 million), of which 96 million (67–136 million) manifest clinically (with any severity of disease).Another study, of the prevalence of dengue, estimates that 3.9 billion people, in 128 countries, are at risk of infection with dengue viruses. The number of cases from WHO region reported that increased from 2.2 million in 2010 to 3.2 million in 2015.The disease is now endemic in more than 100 countries in Africa, the Americas, the Eastern Mediterranean, South-East Asia and the Western Pacific. Dengue occurred sporadically in Bangladesh from 1964 until a large epidemic in 2000 established the virus. In 2015, six dengue-related deaths were reported, while a total 2,816 patients were reported, according to the DGHS (Directorate General of Health Services) (Lopamudra and Anirban, 2017).

Chikungunya is also mosquito-borne viral disease first described during an outbreak in southern Tanzania in 1952. It is an RNA virus that belongs to the alpha virus which is spread by female Aedes mosquitoes that cause dengue and characterized by fever, rash, and arthralgia. In 2006, more than 1.25 million cases were reported in India, the majority of which were from the states of Karnataka and Maharashtra. The first

outbreak in Bangladesh (northern Rajshahi and Chapainawabganj districts) was observed in December 2008 when 32 cases were identified. An outbreak of fever with prolonged joint pain was investigated in Bangladesh in 2011.The risk of death is around 1 in 1,000 (Talya and Blanka, 2017). We present seven confirmed cases of chikungunya having different clinical presentations occurring among middle aged males and females from different socio-economic background in Dhaka city, the capital of Bangladesh.

These three discussed vector borne diseases have individual genes. Our research aim is to find the gene-gene interconnection among these diseases. To do this it is necessary to reduce the number of connected genes. In this research, we use a bioinformatic approach for investigation to find the common gene among malaria, dengue and chikunguniya. We want to design a common gene regulatory network path that shows the interconnection of common genes of these three vector borne diseases.

3 Proposed Methodology

Different kinds of bioinformatics tools are being used to show the interaction or to find the common genes (Zhang, 2012, 2015, 2016a, 2016b, 2017, 2018; Zhang and Li, 2015). The article of Klingstrom and Plwczynski (2010) provides information about the bio-informatics tool that can be used to showing interaction, finding common genes and representing the PPI network among genes as well as proteins. The UNIHI tool that is used to predict PPI network, common metabolic pathway and drug design is referred by Kalathuret (Nahida, 2016). The role of gene reduction for creating gene network evolution has been investigated in the paper (Zhang and Li, 2015b). The research methodology consists of designing a common PPI network for vector borne diseases. Different steps are performed in this research to attain the desired goal. We use R toolkit for performing some specified steps in our methodology to trace out the common genes. Then, a PPI network will be generated with those common genes that show a random network. In this regard, the steps are intended to take is shown in the following block diagram.

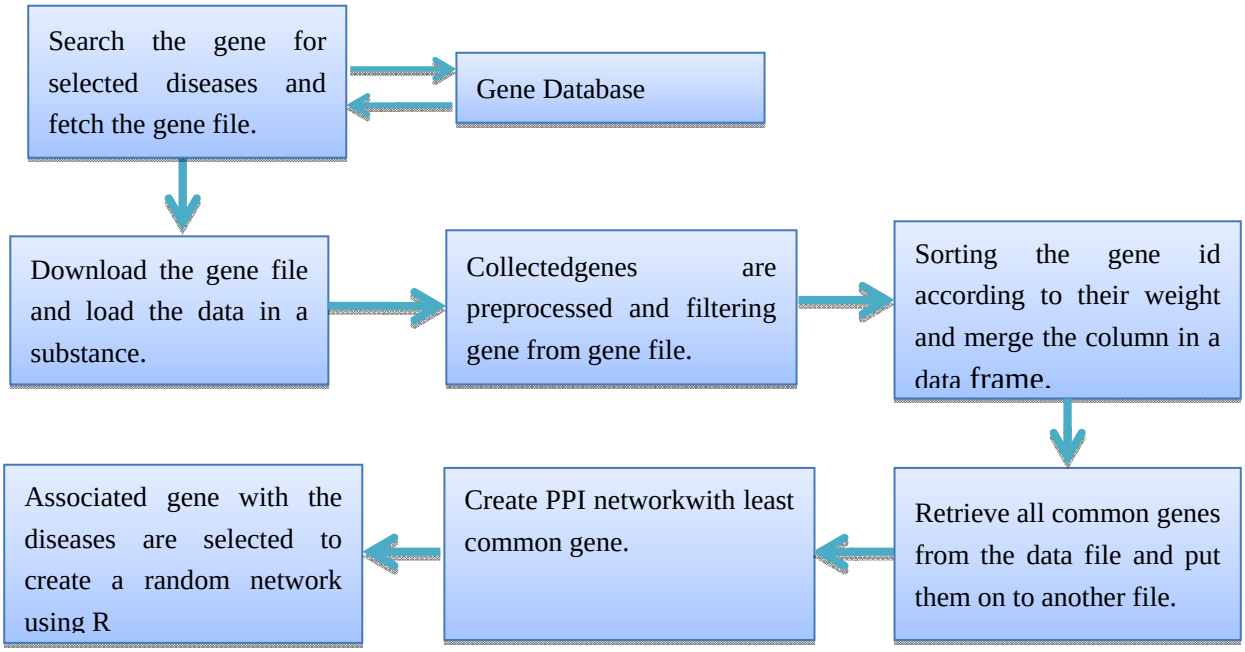

Fig. 1 Block diagram of proposed methodology.

3.1 Collection of gene

The NCBI (National Center for Biotechnology Information) includes a series of databases relevant to biotechnology and is an important resource for bioinformatics tool. The NCBI is the most common on-line downloadable gene database. Major databases include Gene Bank, PubMed are used for DNA sequences and the biomedical literature. Other databases include the NCBI Epigenomics database. All these databases are available online through the Entrez search engine. For this research genes associated with Malaria, Dengue and Chikungunya virus are gathered from gene database using R. The gene collecting code is given in Fig. 2.

```
1. library  (rentrez)
2. entrez _ dbs()
3. entrez_db_searchable("gene")
4. r_search<- entrez_search(db="gene",  term="(disease_name[ALL])",retmax=700)
5. r_search
6. r_search$ids
7. r_seqs<-entrez_fetch(db="gene", id=r_search$ids,    rettype="txt",retmode="text")
```

Fig. 2 Process of gene collection (Code Segment).

3.2 Preprocessing and filtering

The collected gene is required to modify that is called preprocessing. In this step we have filtered genes which are responsible only for Homo sapiens from the total collected genes of each disease. The gene processing and filtering are using R query is given in Fig.3.

```
1. r_search<-entrez_search(db="gene",term="(disease_name[ALL]) AND homo
sapiens[ORGN]", retmax=700)
2. r_search
```

Fig. 3 Genefiltering process (Code Segment).

3.3 Sorting and filtering the gene

The downloaded genes are kept in the text file separately and then imported in the Microsoft Excel for sorting genes in increasing order by their weight. Then the sorted gene files of each disease are imported in R toolkit.

3.4 Gene mining using R

Gene mining is another important part of this research. The responsible gene for human is mined and kept in several columns. Larger number of gene can create difficulty. The sorted gene files that are mined are linked to find out interrelated genes among diseases. The associated mined genes are kept in sorted order in three column of a list which is imported by R for further processing.

3.5 Linkage of gene

In this step, the interrelated genes among diseases are recognized. To find the linkage among diseases enter the diseases name in the query as follow in below Fig. 4(a) and 4(b).

```
1. r_search<-entrez_search(db="gene",term="(disease_name[ALL] AND
Disease_name[ALL])", retmax=700)
2. r_search
3. r_search$ids
4. r_seqs<-entrez_fetch(db="gene",id=r_search$ids,rettype="txt",retmode="text")
```

Fig. 4(a) Gene linkage procedure(Code Segment).

The correlative genes between each pair of selected diseases ((Disease_1,Disease_2), (Disease_2, Disease_3) and (Disease_3,Disease_1)) will be identified that will help to find out the associative gene.

```
<- Reduce(intersect, list(disease _1,disease_2))
<- Reduce(intersect, list(disease _2, disease_3))
<- Reduce(intersect, list(disease _3, disease_1))
```

Fig. 4(b) Linkage procedure using R(Code Segment).

3.6 Finding the common genes

We have alreadymade linkage between the pair of (Disease_1, Disease_2), (Disease_2, Disease_3) and (Disease_3, Disease_1) to find the linkage among the genes. The final linkage is done by intersecting the pair of ((Disease_1,Disease_2), (Disease_2,Disease_3), (Disease_3, Disease_1)) to get the common associated genes.common_gene <-Reduce(intersect,list ((Disease_1, Disease_2), (Disease_2, Disease_3), (Disease_3, Disease_1))).

3.7 Design PPI network

PPI or protein-protein interaction is a biological network referring to intentional physical contact established between two or more proteins as a result of biochemical events (Zhang, 2012, 2017, 2018). Unified Human Interactome or UniHi is a tool that is used for visualization of PPI. We will generate the PPI network of common genes using this tool.

The PPI network consists of common genes which are responsible for all of the concerned diseases. We generate a random network with desired common genes that will be found by R is given in Fig .5.

```
1. library(GGally)
2. library(ggnet)
3. library(network)
4. library(sna)
5. library(ggplot2)
6. library(grid)
7. net = rgraph(5, mode = "graph", tprob = 0.5)
8. net = network(net, directed = FALSE)
9.ggnet2(net,mode="circle",size=17,label=c("IL1B","IL6","MIF","CCL5","Ifng"),color=rep(c("red","ora
nge","yellow","palevioletred","goldenrod")),label.color="white",edge.color="blue") +
theme(panel.background = element_rect(fill
+ = "darksalmon"))
```

Fig. 5 Procedure of random network creation(Code Segment).

4 Results and Discussion

4.1 Collecting the genes

We have collected corresponding genes from the gene database. The collected primary numbers of genes without preprocessing & filtering are calculated as 712 for malaria, 225 for dengue and 33 for chikungunya. The number of genes responsible for chikungunya is very few. The selected genes have been stored in increasing order by their weight.

4.2 Preprocessing, filtering and sorting the genes

After completing preprocessing and filtering the responsible genes for *Homo sapiens* are 329 for malaria,214 for dengue and 28 for chikungunya. The selected number of resultant genes is kept in Table 1.

Table 1 Gene selected for each diseases.

Name of Diseases	Initial Number of Gene	Responsible Gene for Human
Malaria	712	329
Dengue	225	214
Chikungunya	33	28

The number of filtered genes of each disease is not equal. The number of total genes is listed in Fig. 6 and all of the genes are sorted in ascending order with their weight.

Malaria

[1] "ABCA1" ,"ABO" ,"ADORA2A" ,"ADRB2" [5] "AGTR1", "ANGPT1" , "ANGPT2" , "APOE" [9] "FAS" ,"FASLG" , "ATP2A3", "ATP2B4" [13] "BSG" , "C1QBP" , "SERPING1","C5AR1" [17] "CAPN1", "CAPN2", "CASP1" ,"CASP3" [21] "MS4A1" "CD36" "SCARB1" "CD40" [25] "CD40LG" "CD59" "CD81" "CDH13" [29] "CHGA" "CHI3L1" "CHIT1" "CISH" [33] "CNTFR" "CR1" "CRP" "CTGF"[37] "CYBB" "CD55" "ACE" "DDC" [41] "EDN1" "ENG" "EPO" "EPOR" [45] "FCGR2A" "FCGR2B" "FCGR3B" "FCN2" [49] "FOXO3" "FLT1" "ACKR1" "G6P" [53] "GNAS" "GRK5" "GSR" "GSTM1" [57] "GSTP1" "GSTT1" "GYPA" "GYPB" [61] "GYPC" "HBB" "HBE1" "HLA-A" [65] "HLA-B" "HLA-DRB1" "HLA-G" "HMOX1" [69] "HP" "HPRT1" "ICAM1" "IFNAR1" [73] "IFNB1" "IFNG" "IFNGR1" "IGF1" [77] "IL1B" "IL1RN" "IL2" "IL2RA" [81] "IL3" "IL4" "IL4R" "IL6" [85] "IL6ST" "CXCL8" "IL10" "IL10RA" [89] "IL10RB" "IL12A" "IL12B" "IL12RB1" [93] "IL18" "CXCL10" "IRF1" "KIR2DL3" [97] "KIR2DS3" "KIR2DS5""LAIR1" "LEPR" [101] "LGALS2" "LMAN1" "LTA" "MBL2" [105] "MIF" "CXCL9" "MMP1" "MMP9" [109] "MYD88" "NFE2L2" "NOS1" "NOS2" [113] "NOS3" "TNFRSF11B" "PDCD1" "PF4" [117] "ABCB1" "PI4KB" "PKLR" "PTGS2"[121] "RNASE3" "S100A8" "SAA1" "SCO1" [125] "CCL3" "CCL3L1" "CCL5" "CCL18" [129] "CX3CL1" "SLC4A1" "SMN1" "SPTA1" [133] "STAT6" "SYN1" "TCN2" "TRD" [137] "TRG" "TGFB1" "THBD" "TLR1" [141] "TLR2" "TLR4" "TNF" "TNFRSF1A" [145] "TNFRSF1B" "TP53" "TPI1" "TNFRSF4" [149] "VCAM1" "VEGFA" "VWF" "YWHAE" [153] "PRRC2A" "DDX39B" "RAB11A" "VNN1" [157] "CD163" "APOBEC3B" "TLR6" "PROCR" [161] "GNLY" "TNFSF13B" "FUT9" "MASP2" [165] "ADAMTS13" "Abca1" "Ahr" "Ahsg" [169] "Ank1" "Aqp4" "Bdnf" "C2" [173] "C5ar1" "Casp1" "Cd1d1" "Cd28" [177] "Cd36" "Cd4" "Cd6" "Chil1"[181] "Socs3" "Cxcr3" "Ccr5" "Cnr2" [185] "Cr2" "Ackr1" "Edn1" "Epo" [189] "Fcgr2b" "Flt3l" "Fth1" "Fyn" [193] "Gfap" "Hba-a1" "Hc" "Hmox1" [197] "Icam1" "Irf8" "Ido1" "Cxcl10" [201] "Ifnar1" "Ifng" "Ikbkb" "Il10" [205] "Il12rb2" "Il15" "Il17a" "Il18" [209] "Il2" "Il2ra" "Il2rg" "Il6" [213] "Itgam" "Itgax" "Cd47" "Jak3" [217] "Klf4" "Psmb8" "Lta" "Ltbr" [221] "Il1rl1" "Mif" "Cxcl9" "Myd88" [225] "Nfe2l2" "Nos2" "Pdcd1" "Prf1" [229] "Prkcq" "Pklr" "Pla2g2a" "Pparg" [233] "Ptgs2" "Sirpa" "Ccl17" "Foxp3" [237] "Stat3" "Stat6" "Tcrb" "Tgfb1" [241] "Tlr6" "Tnf" "Tnfrsf1b" "Tnfsf11" [245] "Vnn1" "Vwf" "IL17RA" "Tlr2" [249] "Tnfsf13b" "Cd24" "Mapk9" "HAVCR1" [253] "Clec4a2" "Cd36" "Tnfsf14" "Il27ra" [257] "FOXP3" "IL23A" "TLR9" "Irf7" [261] "Icos" "TREM1" "SLC38A2" "KIR2DL5A" [265] "AICDA" "Tbx21" "HAMP" "IL21" [269] "ACE2" "COL18A1" "SLC38A1"

"Tlr9" [273] "FCRL5" "Il23a" "Trem2" "FAM234A" [277] "Hamp" "IL33" "NLRP12" "ZNF804A" [281] "MARVELD3" "Cpn1" "Kmo" "Hrh3" [285] "Nod1" "IL17F" "NLRP3" "TIRAP" [289] "P2ry6" "SPPL3""Tlr7""Tlr8" [293] "Btla" "Nlrp3" "Erfe" "Tnfrsf14" [297] "Clec9a" "Il27" "OR51F1" "Nod2" [301] "NCR3" "MIF-AS1" "Cd68" "Card9" [305] "C5" "C5" "Nlrp12" "NCF1" 309] "IL10" "G6PD" "PF10_0001" "PF10_0225" [313] "PF10_0268" "PF10_0322" "AMA1" "PfOPRT"[317] "ETRAMP5" "PFI0215c" "MSP1" "SP21"[321] "PF13_0248""PFC0210c" "NEWENTRY "NEWENTRY" [325] "PB000298.03.0" "NEWENTRY" "NEWENTRY" "MAL13P1.540"

Dengue

[1] "ABCF1" "ACTA1" [3] "ADAR" "AKT1" [5] "ANGPT1" "ANGPT2"[7] "ARF4" "ARF5"[9] "ARF6" "BAK1"[11] "BCL2" "CALR"[13] "CASP1" "CAV1"[15] "CCNT1" "CD27"[17] "CDK9" "CMA1"[19] "CCR1""CRP"[21] "MAPK14" "CSF2"[23] "DAXX" "DDX6" [25] "ENO1" "F3"[27] "FASN" "FCGR1A"[29] "FCGR2A" "FLT1" [31] "CXCR3" "PDIA3"[33] "GSK3B" "CFH"[35] "HGF" "HLA-A"[37] "HLA-B" "HLA-DRB1"[39] "HMGB1" "HMGCR"[41] "HMOX1" "HNRNPC"[43] "HNRNPH1" "HNRNPK"[45] "HSPA8" "HSPA9"[47] "HSPD1" "ICAM1"[49] "IFIT3" "IFNA1"[51] "IFNB1" "IL1B"[53] "IL1RN" "IL2RG" [55] "IL4""IL6"[57] "CXCL8" "IL10" [59] "IL11" "TNFRSF9"[61] "ILF3" "CXCL10" [63] "ITGB1" "ITGB3" [65] "KDR" "KIR3DL1"[67] "KPNB1" "KPNA2"[69] "RPSA" "STMN1"[71] "LGALS3BP" "LGALS9"[73] "LTA" "CAPRIN1" [75] "MBL2""RAB8A"[77] "MICB" "MIF"[79] "MMP2""MMP9"[81] "MRC1" "NCL"

[83] "NFKB1" "OAS1"[85] "OAS2" "OAS3"[87] "FURIN""SERPINE1"[89] "PCBP2" "PIK3CA"[91] "PLAT" "PLG"[93] "PML""EIF2AK2"[95] "PROC" "PTBP1" [97] "PTGIS" "PTGS2" [99] "RAB5A" "RPL18"[101] "CCL2" "CCL3" [103] "CCL5""CXCL11" [105] "SDC2" "SSB" [107] "STAT2" "STAT3" [109] "ADAM17" "TAP1" [111] "TAP2" "TGFB1" [113] "THPO" "TLR2"[115] "TLR3" "TLR4" [117] "TNF" "TP53"[119] "TPSAB1" "TRAF6" [121] "VCAM1" "VDR" [123] "VEGFA" "VIM" [125] "YWHAE" "IFITM1" [127] "GBF1" "TNFSF10" [129] "SQSTM1" "CLDN1" [131] "IL1RL1" "DDX21"[133] "IKBKE" "G3BP2" [135] "DNM1L" "G3BP1"[137] "TLR6" "IFITM3"[139] "IFITM2""LILRB1" [141] "CD300A" "Cxcr3" [3] "Ccr1" "Ccr2"[145] "Ccr4""Fcgr3" [147] "Cxcl10" "Ifnb1" [149] "Ifng" "Ifngr1" [151] "Il12a" "Il18" [153] "Il2rg" "Il6" [155] "Kpnb1" "Mif" [157] "Nfkb1" "Nfkbia" [159] "Nfkbib" "Nos2" [161] "P4hb" "Plg" [163] "Sphk1" "Stat2"

[165] "Tek""Tlr6" [167] "Tnf" "RAB18" [169] "SCAP""UBR4" [171] "DDX58" "CLEC5A" [173] "Clec5a" "Tlr2" [175] "Mapk1""Mapk3" [177] "SND1" "DLL1" [179] "NRBP1""CD209" [181] "Tl" "Il22" [183] "PLCE1" "TREM1" [185] "DLL4""RBFOX1" [187] "C19orf66" "EXOC1" [189] "ADAP2" "MYO5C" [191] "MAVS" "IFIH1" [193] "PDIA2" "Ifi30" [195] "TRIM56" "IL33" [197] "RSAD2" "Hnrnpk" [199] "Nlrp3" "TMEM173"

[201] "MIRLET7C" "MIR12" [203] "MIR146A" "MIR150" [205] "MIR21" "MIR223" [207] "RPSA" "flaviviruspolyprotein gene"[209] "flaviviruspolyprotein gene" "flaviviruspolyprotein gene"[211] "NEWENTRY" "Isg15" [213] "MIR548G" "MICA".

Chikungunya

[1] "ATP1B1" "BST2" "C1QBP" "HLA-DQB1" "HSP90AA1" "NDST1" [7] "IL1B" "IL6" "MIF" "OAS1" "OAS3" "PHB" [13] "CCL5" "TLR3" "TM9SF2" "G3BP2" "G3BP1" "Ifng" [19] "DDX58" "Clec4a2" "CD209" "Irf7" "Irf3" "MAVS" [25] "Bst2" "RSAD2" "Tlr3" "CHIKVgp2"

Fig. 6 Sorted and filtered genes.

4.3 Linkage of gene

In this step we have recognized the interrelated genes among diseases. The correlative genes between each pair of selected diseases (Malaria-Dengue, Dengue-Chikungunya and Chikungunya-Malaria) are identified. To find the linkage among diseases enter the diseases name in the query as follow in below Fig.7(a) and 7(b).

```
1. r_search<-entrez_search(db="gene",term="(malaria[ALL] AND
dengue[ALL])", retmax=700)
2. r_search
3. r_search$ids
```

Fig. 7(a) Gene linkage procedure (Code Segment).

```
AD <- Reduce(intersect, list(Malaria, Dengue))
> AD
 [1] "ANGPT1"  "ANGPT2"  "CASP1"   "CRP"     "FCGR2A"  "FLT1"  [7] "HLA-A"   "HLA-B"
"HLA-DRB1" "HMOX1"  "ICAM1"  "IFNB1"   [13] "IL1B"  "IL1RN" "IL4"  "IL6" "CXCL8" "IL10"
[19] "CXCL10"  "LTA"  "MBL2" "MIF" "MMP9" "PTGS2" [25] "CCL3"  "CCL5" "TGFB1" "TLR2"
"TLR4"  "TNF"  [31] "TP53" "VCAM1" "VEGFA"  "YWHAE"  "TLR6"  "Cxcr3" [37] "Cxcl10"
"Ifng" "Il18" "Il2rg" "Il6" "Mif"  [43] "Nos2" "Tlr6" "Tnf" "Tlr2" "TREM1"   "IL33"
[49] "Nlrp3"   "NEWENTRY"
BD <- Reduce(intersect, list(Dengue, Chikungunya))
> BD
 [1] "IL1B"  "IL6"  "MIF"  "OAS1"  "OAS3"  "CCL5"  "TLR3"  "G3BP2" "G3BP1"
[10] "Ifng"  "DDX58" "CD209" "MAVS"  "RSAD2"
CD <- Reduce(intersect, list(Malaria, Chikungunya))
> CD
[1] "C1QBP"  "IL1B"  "IL6"  "MIF"  "CCL5"  "Ifng"  "Clec4a2" [8] "Irf7"
```

Fig. 7(b) Linkage procedure using R (Code Segment).

Here AD,BD, CD are the linkage variable of common gene identification procedure. AB indicates (Malaria and Dengue), BD indicates (Dengue and Chikungunya), CD indicates (Malaria and Chikungunya).

4.4 Common gene finding after linkage and mining

We have already made linkage between the pair of (malaria, dengue), (dengue, chikungunya) and (malaria, chikungunya) to find the linkage among the genes. The final linkage is done by intersecting the pair of ((malaria, dengue), (dengue, chikungunya), (malaria, chikungunya)) to get the common associated genes. We have found five common associated genes among the diseases malaria, dengue and chikunguniya. Now the common genes are display in the Fig. 8.

```
common_gene<-Reduce(intersect,list(AD,BD,CD))
>common_gene
[1] "IL1B" "IL6"  "MIF"  "CCL5" "Ifng"
```

Fig. 8 Finding common genes using R (Code Segment).

The gene share kept in a mined form and there exist two extra gene in PPI network . The final 5 common gene are IL6, IL1B, Ifng, MIF and CCL5. One gene has not directly connected name as MIF. The ultimate cross linkage genes are shown in Table 2.

Table 2 Number of cross linkage gene.

Cross Linkage	No. of Gene
Malaria and Dengue	49
Dengue and Chikungunya	14
Chikungunya and Malaria	8
Common gene	5

4.5 Finding common gene regulatory pathway or PPI network

The 5 common genes have been used to create a PPI network using UniHI tool. The output of the network is shown in Fig.9.

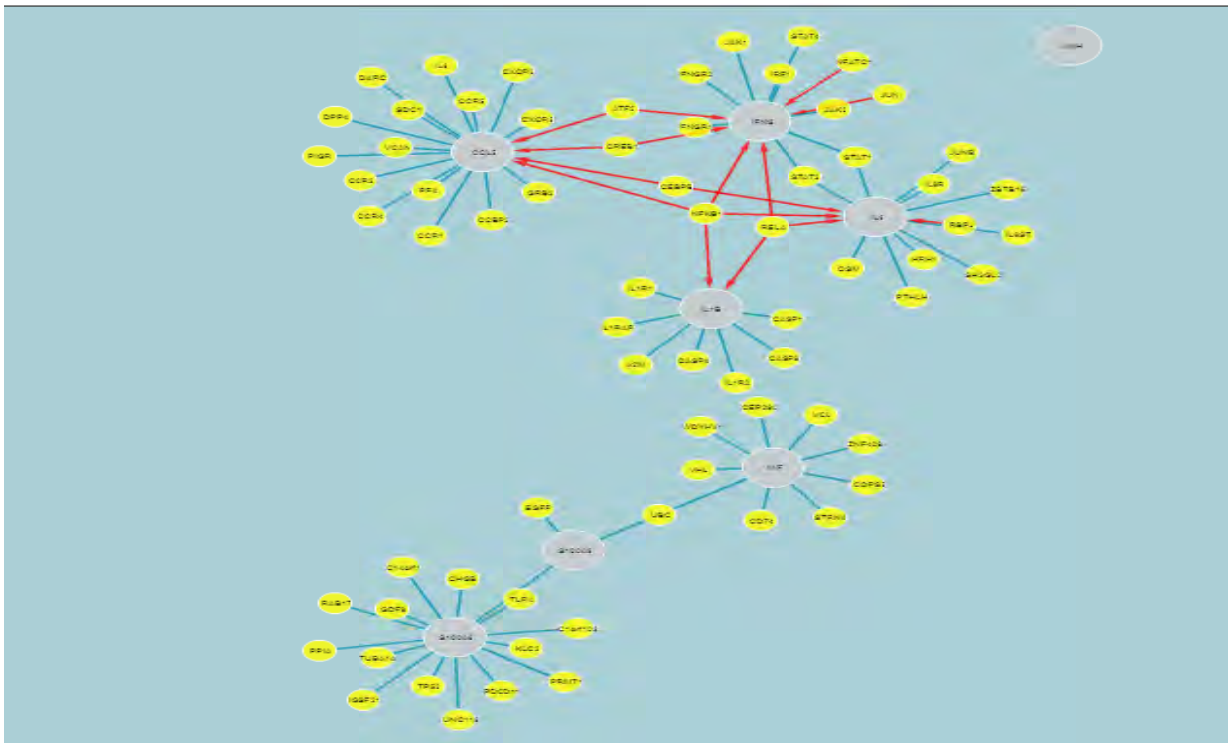

Fig. 9 PPI network with 5 common genes of selected diseases.

The protein-protein interaction or PPI network represents the interrelation among the genes and protein, including those genes some of them are connected directly and some are connected indirectly. But in last it is recognized that only 4 genes are interconnected directly and 1 gene is connected indirectly with each other in PPI network. Now we have been observed the directly connected 4 common gene PPI network in Fig. 10.

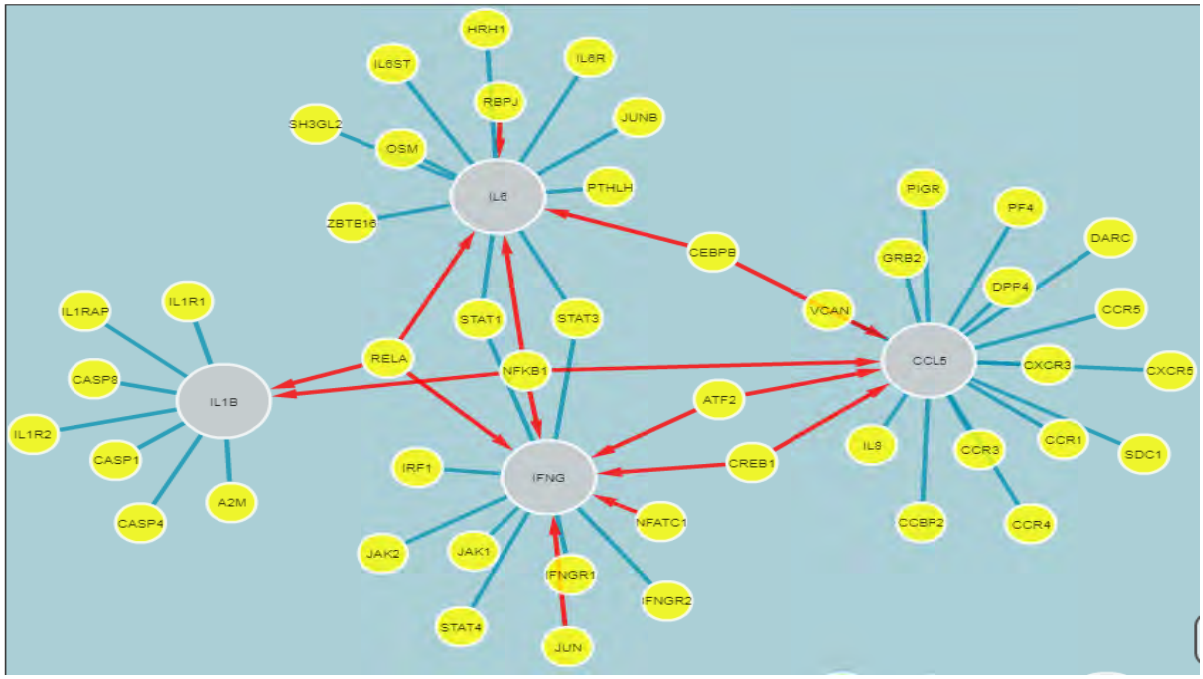

Fig. 10 PPI Network of interconnected 4 common genes.

The four common gene are connected each other, but one other common gene named “MIF” is not connected directly with other four genes. This gene is responsible for connecting two calcium binding gene named “S100A8” and “S100A9”.Now we have been observed it in Fig. 11.

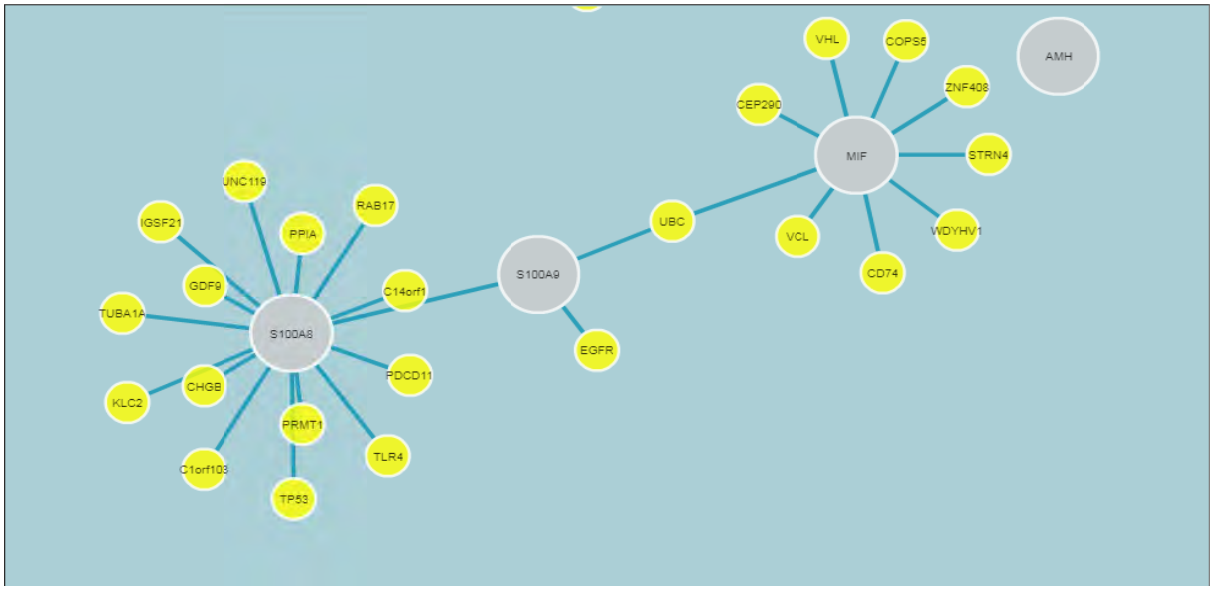

Fig. 11 Calcium binding gene.

4.6 Generating random network using R

In final step, we have investigated 5 common genes that make a random network using R. Fig. 12 shows the figure of random network is given.

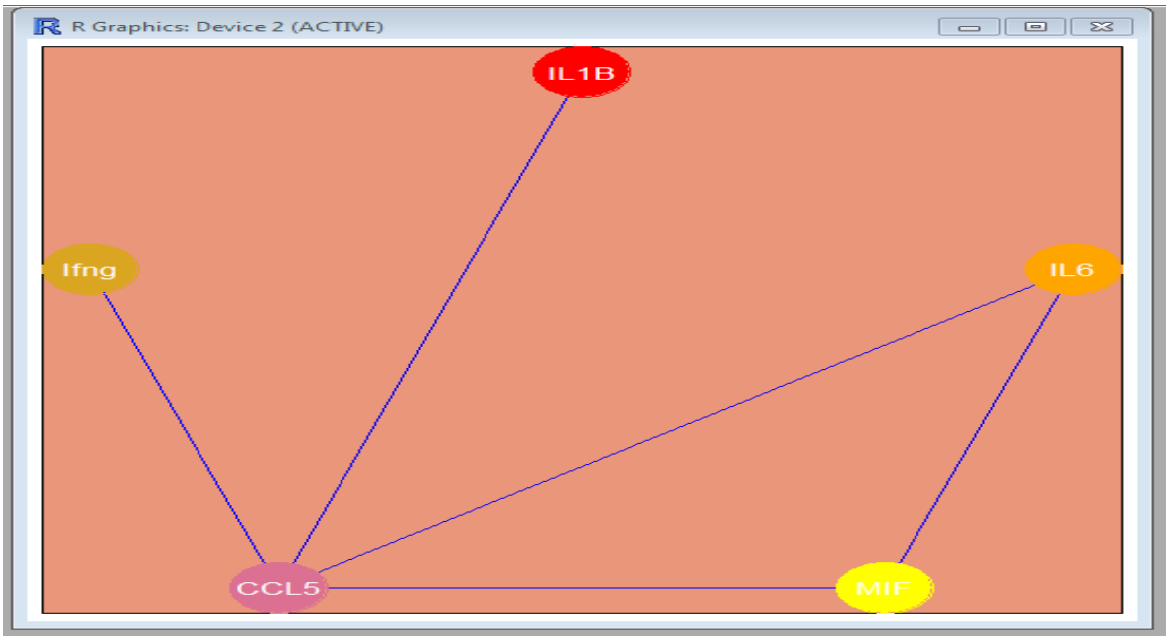

Fig. 12 Prepare a random network by R.

5 Discussion

The three mosquitoes borne diseases (malaria, dengue and chikunguniya) has five common genes that are directly or indirectly connected to each other. The gene database access, uses of R, uses of bioinformatics tools like UniHi, makes it easy to find common gene and generation of PPI network, which shows interrelation among genes. The linkage and mined on collected gene have been used to identify the resulting common genes for associated diseases and generate a protein-protein interaction or PPI network. In final process the number of collected gene are reduced up to 60%, 35%, 5%, 1% and finally less than 1% respectively. Then the calculation is shown in a Pie Chart Fig.13.

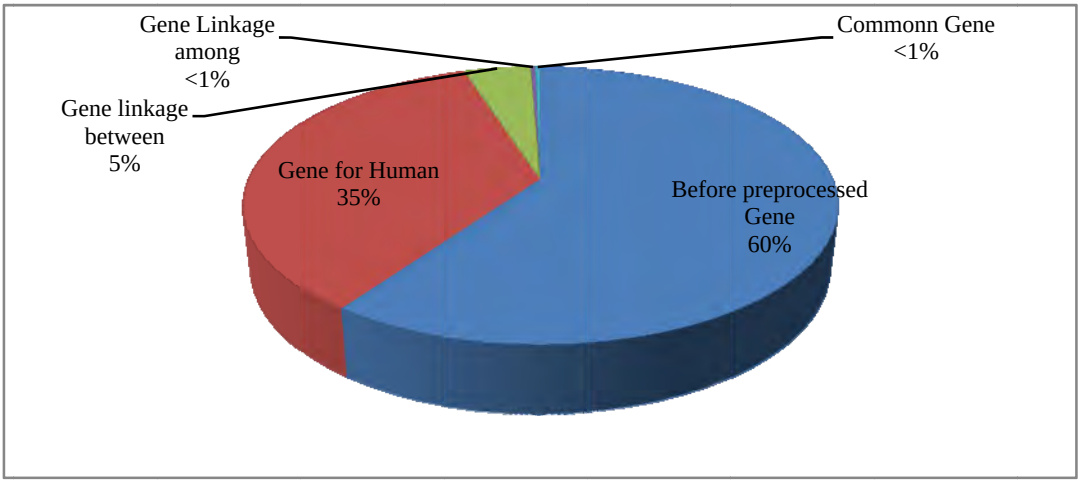

Fig. 13 Reduced collected gene after linkage and mine.

The genes are reduced gradually after processing and the percentage of the gene reduction value are 44%, 20%, 14%, 13%, 5%, 2% and finally less than 1% given in a pie chart in Fig. 14.

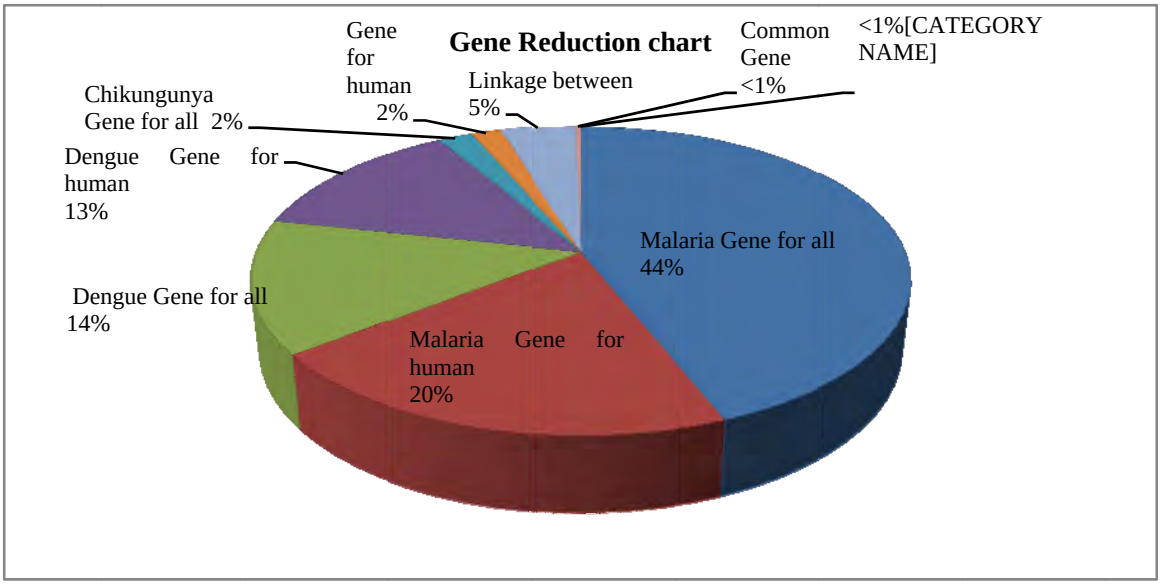

Fig. 14 Gene optimization chart.

6 Conclusion

Study of bioinformatics includes the identification of candidate genes and nucleotides. It can conclude that, malaria, dengue and chikunguniya may be directly, indirectly and genetically associated with each other. To design a drug for a disease, it is required to know the affected genes associated with the diseases. Findings of the interconnected genes, linkage and common genes among associated diseases have a great impact upon the analysis of the diseases for drug design correctly. After processing we have found the common genes which are inter-related each other. The common genes will be maintained a gene regulatory network pathway known as protein-protein interaction or PPI network which can be designed to show the association and interaction among the proteins of these diseases which leads to the way to drug design. We have successfully reduced the number of gene for the associated diseases. Moreover, designing a common PPI for three associated diseases will help to design common drug for these diseases.

Abbreviation

DENV-1=Dengue Virus-1; DENV-2=Dengue Virus-2; DENV-3=Dengue Virus-3; DENV-4=Dengue Virus-4; DGHS = Directorate General of Health Services; PPI= Protein-Protein Interaction; NCBI=National Center for Biotechnology Information.

References

DumrongMairiang, Huamei Zhang, Ann Sodja,et al. 2013. Identification of New Protein Interactions between Dengue Fever Virus and Its Hosts, Human and Mosquito. PLoS One, 8(1): e53535

Javad Zahiri, Joseph Hannon Bozorgmehr, Ali Masoudi-Nejad. 2013. Computational Prediction of Protein–Protein Interaction Networks: Algo-rithms and Resources. Current Genomics, 14(6): 397-414

Klingstrom T, Plwczynski D. 2010. Protein-protein interaction and pathway databases, a graphical review.

Briefings in Bioinformatics, 12(6): 702-713

Lopamudra Dey, Anirban Mukhopadhyay. 2017. DenvInt: A database of protein–protein interactions between dengue virus and its hosts. PLOS Neglected Tropical Diseases, 11(10): e0005879

MihaelaAngelova, Slobodan Kalajdziski, LjupcoKocarev. 2010. Computational Methods for Gene Finding in Prokaryotes. ICT Innovations 2010, Web Proceedings

Mohammad Nazrul Islam, MD Mohammad ZulKifle, Arish Mohammad Khan Sherwani, et al. 2011. Prevalence of malaria, dengue, and Chikungunya significantly associated with mosquito breeding sites. JIMA, 43(2): 58–67

Nahida Habib, Kawsar Ahmed, IffatJabin, Mohammad Motiur Rahman, 2016.Application of R to investigate common gene regulatory network pathway among bipolar disorder and associate diseases. Network Biology, 2016, 6(4): 86-100

Oussema Souiai, Fatma Guerfali, Slimane Ben Miled, Christine Brun, Alia Benkahla. 2014. In silico prediction of protein-protein interactions in human macrophages. BMC Research Notes, 7: 157

Stéphane Tchankouo-Nguetcheu, Huot Khun, Laurence Pincet, Pascal Roux, et al. 2010. Differential Protein Modulation in Midgutsof Aedesaegypti Infected with Chikungunya and Dengue 2 Viruses. PLoS One. 2010; 5(10): e13149

Talya Shragai, Blanka Tesla, Courtney Murdock, Laura C. Harrington. 2017. Zika and chikungunya: mosquito-borne viruses in a changing world. Annals of the New York Academy of Sciences, 1399(1): 61-77

Teichmann SA, Babu MM, Teichmann SA. 2014. Gene regulatory network growth by duplication. Nature Genetics, 36(5): 492-496

Zhang WJ. 2012. Computational Ecology: Graphs, Networks and Agent-based Modeling. World Scientific, Singapore

Zhang WJ. 2015. A hierarchical method for finding interactions: Jointly using linear correlation and rank correlation analysis. Network Biology, 5(4): 137-145

Zhang WJ. 2016a. A node degree dependent random perturbation method for prediction of missing links in the network. Network Biology, 6(1): 1-11

Zhang WJ. 2016b. A random network based, node attraction facilitated network evolution method. Selforganizology, 3(1): 1-9

Zhang WJ. 2018. Fundamentals of Network Biology. World Scientific Europe, London, UK

Zhang WJ, Feng YT. 2017. Metabolic pathway of non-alcoholic fatty liver disease: Network properties and robustness. Network Pharmacology, 2(1): 1-12

Zhang WJ, Li X. 2015a. General correlation and partial correlation analysis in finding interactions: with Spearman rank correlation and proportion correlation as correlation measures. Network Biology, 5(4): 163-168

Zhang WJ, Li X. 2015b. Linear correlation analysis in finding interactions: Half of predicted interactions are undeterministic and one-third of candidate direct interactions are missed. Selforganizology, 2(3): 39-45