Genes Genom (2015) 37:679–691
DOI 10.1007/s13258-015-0297-y

RESEARCH ARTICLE

Identification of hub genes and their SNP analysis in West Nile virus infection for designing therapeutic methodologies using RNA-Seq data

Iftikhar Aslam Tayubi[1] · Ahmad Firoz[2,3] · Omar M. Barukab[1] · Adeel Malik[4]

Received: 9 March 2015 / Accepted: 19 May 2015 / Published online: 24 June 2015

Abstract The West Nile virus (WNV) infections are generally asymptomatic and are considered as immediate concerns of biodefense due to the lack of any therapeutic remedies. In this work, we created an interaction network of 1159 differentially expressed genes to detect potential hub genes from WNV infected primary human macrophages. We go on to explore the genetic variations that can alter the expression and function of identified hub genes (*HCLS1, SLC15A3, HCK,* and *LY96*) using the PROVEAN Protein Batch tool and PolyPhen-2. Community analysis of the network revealed that these clusters were enriched in GO terms such as inflammatory response and regulation of proliferation. Analysis of hub genes can aid in determining their degree of conservation and may help us in understanding their functional roles in biological systems. The nsSNPs proposed in this work may be further targeted through experimental methods for improving treatment towards the infection of WNV.

Keywords West Nile virus · Infection · Interaction network · Hub genes · Community analysis · nsSNPs

Introduction

West Nile virus (WNV) is a mosquito-borne neurotropic viruses and belongs to a member of Japanese encephalitis virus (JEV) serogroup in the family Flaviviridae (Heinz et al. 2000), which comprises viruses for yellow fever and dengue (Brinton 2002). Originally first detected in Uganda, it has been endemically spreading in various regions across the world, including the Middle East, some regions of Africa, Europe and United States at different periods (Suthar et al. 2013). However, serologically it was first tested in Assam, India in 2006 (Chowdhury et al. 2014). Belonging to the Flaviviridae family, it is encoded by a ~11 kb positive-sense, single-stranded RNA (ssRNA) genome (Suthar et al. 2013). It follows a replication life cycle, which is involved in binding to cell surface receptor, leading to fusion with the membrane and final delivery of infectious RNA genome into the cytoplasm (Suthar et al. 2013). The genome injected is further translated as a single polyprotein, which undergoes subsequent cleavage by viral and host proteases to generate structural and non-structural proteins. The structural proteins form the virion encapsidating the viral RNA, whereas, the non-structural proteins form the replication complex, which synthesize the negative- and positive-sense viral RNA. Hence, targeting the replication pathway and primarily the infectious genome can aid in curbing the expression of WNV (Suthar et al.

Electronic supplementary material The online version of this article (doi:10.1007/s13258-015-0297-y) contains supplementary material, which is available to authorized users.

✉ Iftikhar Aslam Tayubi
Iftikhar.tayubi@gmail.com

[1] Faculty of Computing and Information Technology, King Abdulaziz University, Rabigh 21911, Kingdom of Saudi Arabia

[2] School of Chemistry and Biochemistry, Thapar University, Patiala 147004, Punjab, India

[3] Biomedical Informatics Center of ICMR, Post Graduate Institute of Medical Education and Research (PGIMER), Chandigarh 160012, India

[4] Perdana University Centre for Bioinformatics, MARDI Complex, Jalan MAEPS Perdana, 43400 Serdang, Selangor, Malaysia

2013). Additionally, targeting at gene level will also enhance the effectiveness of therapy involved. Identification of the prime regions of interest and further the candidate genes for gene therapy can be determined via RNA sequencing (RNA-Seq) methods (Mercer et al. 2014).

The RNA-Seq method comprises of the direct sequencing of cDNA for monitoring quantitative gene expression (Nagalakshmi et al. 2010). It involves direct sequencing of cDNAs using high-throughput DNA sequencing technologies followed by mapping of reads to the reference genome (Nagalakshmi et al. 2010). The methodology allows identifications of introns and exons and further helps in recognizing intronic and exonic boundaries and boundary ends of the gene of interest (Nagalakshmi et al. 2010). RNA-Seq is presently suggested as cost-effective new tool, which can be applied for deciphering the genetic basis of diseases, and traits that could not be detected based on previous conventional gene-discovery methods (Bamshad et al. 2011). RNA-Seq has been employed to investigate biological phenomena in diverse areas such as viral infections (Jones et al. 2014), cancer (Young et al. 2014) and cardiomyopathy (Christodoulou et al. 2014). Recently, we have also used RNA-Seq analysis to explore the role of glycogenes in skeletal muscle development in $MYOG_{kd}$ cells (Lee et al. 2014) and also used precomputed expression values from mouse RNA-Seq data to understand the role glycogenes in various biological processes that are involved in the development of brain, muscle, and liver tissues (Firoz et al. 2014).

RNA-Seq has also been applied to identify 1514 differentially expressed (DE) transcripts from primary human macrophages, in response to infection with WNV (Qian et al. 2013). In vivo studies of WNV infection using high-throughput digital gene expression analysis have reported common genes that are differentially expressed in multiple tissues upon WNV infections as well as genes that are differentially expressed in specific tissues following WNV infection (Peña et al. 2014). These studies have helped in developing a framework that can aid in the discovery of genes and their associated biological functions and in turn, decipher their impact on causing such infectious diseases. Therefore, in this work, we have performed a network analysis to identify the key hubs from the list of candidates differentially expressed transcripts that were recently detected from RNA-Seq analysis of primary human macrophages infected with WNV (Qian et al. 2013). We also explored the Single Nucleotide Polymorphisms (SNPs) within these potential hubs that can be suggested as the prime regions for susceptibility towards WNV infection. Additionally, we also explored functionally important modules in the network of WNV infected DE genes by using community analysis. The regions suggested in this study can be further targeted via experimental therapeutic technologies like gene knockout technique, gene targeting, etc., for developing treatment towards suppressing the pathogenic action of WNV.

Materials and methods

Datasets

In this work, 1514 differentially expressed (DE) transcripts were used for our computational analysis and were downloaded from a recent study on WNV infected macrophages (Qian et al. 2013). These DE transcripts represent the RNA-Seq analysis originally carried out by Qian et al. (2013) using Illumina Genome Analyzer 2 and expression levels for each transcript estimated by using maximum likelihood based method employed in the Cufflinks program (Trapnell et al. 2010). The dataset was filtered by removing any duplicate transcripts and represent the final set of differentially expressed genes (DEGs) used in this study.

Network analysis

GeneMANIA [www.genemania.org/] cytoscape plugin (Montojo et al. 2010) was utilized to determine the functional interactions between DEGs based on the GO term "biological process" and *Homo sapiens* as reference species. The predicted relationships between the genes in the network comprises of co-expression, physical and genetic interactions, pathways, co-localization, protein domain similarity, and predicted interactions.

Identification of hub genes

Biological networks exhibit the scale-free property (Albert 2005) with hubs representing nodes with many connections in the network. NetworkAnalyzer plugin of Cytoscape (Smoot et al. 2011) was utilized to determine the hubs by calculating the values of the node degree distribution. Three genes with highest node degree distribution were identified as hubs in the current network.

Community analysis

The modules with functional property were determined by applying greedy community-structure detection algorithm via GLay [http://brainarray.mbni.med.umich.edu/sugang/glay] (Su et al. 2010) plugin in Cytoscape. For identifying the over-represented biological functions within each cluster, the clusters were subjected to a functional enrichment analysis by focusing only on communities with at least 10 nodes. The functional enrichment analysis was carried out using the DAVID functional analysis tool.

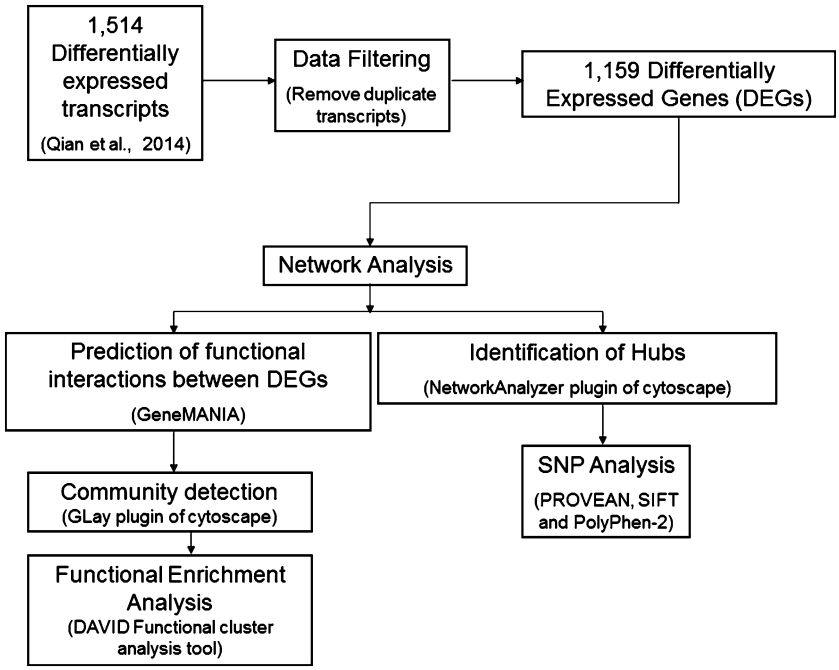

Fig. 1 Flowchart showing the overall methodology implemented in this study

Computational analysis of non-synonymous SNPs (nsSNPs) of predicted hub genes

Data mining the SNP information for hub genes identified in humans in this study were retrieved from National Centre for Biotechnology Information (NCBI) database dbSNP [http://www.ncbi.nlm.nih.gov/projects/SNP]. Three different programs were used to predict the damaged or deleterious coding nsSNPs.

Prediction of deleterious or damaging coding nsSNPs using the PROVEAN protein batch tool

The PROVEAN Protein Batch tool [http://provean.jcvi.org/index.php] was used to provide PROVEAN (Protein Variation Effect Analyzer) and SIFT (Sorting Intolerant From Tolerant) predictions for a list of protein variants. Both PROVEAN and SIFT are software tools that predict whether an amino acid substitution has an impact on the biological function of a protein if the score, lies below a certain threshold value. In PROVEAN, the clustering of BLAST hits is carried out, and the best 30 clusters of closely linked sequences of the supporting sequence set that is further employed to make the predictions. For every supporting sequence, a delta alignment score is computed and then averaged within and across clusters generating a final PROVEAN score (Choi et al. 2012). A default score threshold of −2.5 or above is considered deleterious whereas anything less than this cut-off score has a neutral effect. On the other hand, SIFT is a multi-step algorithm and uses sequence homology based method to classify amino acid substitutions (Kumar et al. 2009). The PROVEAN Protein Batch tool accepts a list of protein sequence variants as input for the predictions.

Prediction of functional modification of coding nsSNPs by polymorphism phenotyping v2 (PolyPhen-2)

PolyPhen-2 [http://genetics.bwh.harvard.edu/pph2/] tool was used to study the possible consequence of nsSNPs on protein structure and function. The web server predicts the potential effect of amino acid substitution on the stability and function of human proteins by employing structural as well as comparative evolutionary considerations. For each specified amino acid residue substitution in a protein, PolyPhen-2 mines a variety of sequence and structural characteristics of the replacement position and supplies them to a probabilistic classifier. For this work, we have used the provided batch query option and selected the default advance option for generating the predictions (Adzhubei et al. 2010).

Results

We downloaded a list of 1514 transcripts that were determined to be differentially expressed between control unaffected and WNV-infected samples (Qian et al. 2013). The data were further filtered by removing any duplicate transcripts so that each transcript represents a unique DEG. As a result, a non-redundant set of 1159 DEGs was

Table 1 Top 30 significantly enriched gene ontology (GO) terms detected by GeneMANIA for differentially expressed genes and additional related genes

GO ID	Description	Q-value
GO:0019221	Cytokine-mediated signaling pathway	1.55E−39
GO:0060337	Type I interferon-mediated signaling pathway	3.75E−30
GO:0071357	Cellular response to type I interferon	3.75E−30
GO:0034340	Response to type I interferon	4.83E−30
GO:0034341	Response to interferon-gamma	4.83E−30
GO:0071346	Cellular response to interferon-gamma	6.73E−27
GO:0060333	Interferon-gamma-mediated signaling pathway	1.85E−23
GO:0051707	Response to other organism	3.10E−21
GO:0009607	Response to biotic stimulus	1.00E−20
GO:0001817	Regulation of cytokine production	1.00E−20
GO:0031347	Regulation of defense response	2.62E−20
GO:0001816	Cytokine production	7.21E−20
GO:0006954	Inflammatory response	7.52E−17
GO:0002694	Regulation of leukocyte activation	7.96E−17
GO:0050865	Regulation of cell activation	7.96E−17
GO:0051249	Regulation of lymphocyte activation	1.85E−16
GO:0046649	Lymphocyte activation	1.18E−15
GO:0045088	Regulation of innate immune response	3.63E−15
GO:0002252	Immune effector process	4.69E−15
GO:0050867	Positive regulation of cell activation	4.69E−15
GO:0002696	Positive regulation of leukocyte activation	8.86E−15
GO:0009615	Response to virus	2.20E−14
GO:0051251	Positive regulation of lymphocyte activation	2.68E−14
GO:0042110	T cell activation	3.73E−14
GO:0050778	Positive regulation of immune response	3.37E−13
GO:0031349	Positive regulation of defense response	5.70E−13
GO:0050863	Regulation of T cell activation	1.76E−12
GO:0001819	Positive regulation of cytokine production	2.05E−12
GO:0002237	Response to molecule of bacterial origin	2.25E−12
GO:0007249	I-kappaB kinase/NF-kappaB cascade	3.05E−12

obtained that consisted of 253 down-regulated and 906 up-regulated genes (Supplementary Table S1). The complete workflow of computational methodology adopted in this study is diagrammatically represented in (Fig. 1).

Network construction and identification of hub genes

How DEGs interact with each other and additional related genes in the network was investigated by GeneMANIA Cytoscape plugin (Montojo et al. 2010). A GeneMANIA network analysis of DEGs points to enrichment of signaling pathways such as cytokine-mediated and type I interferon-mediated signaling pathways. The other over-represented GO terms includes inflammatory response, regulation of leukocyte and lymphocyte activation, and regulation of cell activation (Table 1).

The interaction network deduced via GeneMANIA was further analyzed in Cytoscape 2.8.2 (Smoot et al. 2011). The initial network comprising of 952 nodes and 48,263 edges was filtered to 952 nodes and 44,785 edges by removing duplicate edges. All the genes in the network are represented by circles and the interactions between them are represented as edges. The up-regulated genes are shown in *green* whereas; *red* nodes represent down-regulated genes. Moreover, the additional related genes predicted by GeneMANIA are shown in *cyan* (Fig. 2).

The genes determined as hubs with high node degree distribution encode, HCLS1 (Hematopoietic lineage cell-specific protein or Hematopoietic cell-specific LYN substrate 1), (Supplementary Fig. 1a) a top hub with the node degree of 342 whereas, SLC15A3 (Solute carrier family 15 member 3) (Supplementary Fig. 1b) was the second hub having a node degree of 323. The genes with third highest node degree are represented by HCK (Tyrosine-protein

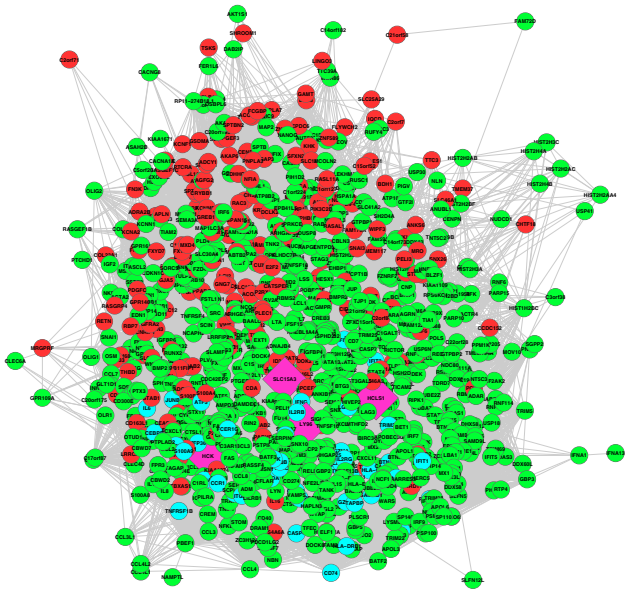

Fig. 2 Interaction network between DEGs and related additional genes. In this network, *green* nodes represent up-regulated genes whereas *red* nodes represent down-regulated genes and GeneMANIA predicted genes are shown in *cyan*. Hub genes are shown as *purple* diamonds and represent the genes with highest node degree

kinase) (Supplementary Fig. 1c) and LY96 (Lymphocyte antigen 96) (Supplementary Fig. 1d) because both have a node degree of 318.

Community analysis and functional annotation of detected modules

Five biologically-related clusters were identified using Fast Greedy community-structure identification algorithm (Fig. 3). Among all the detected clusters, Cluster 1 is the largest with 425 genes and also consists of hubs HCK and LY96 (Fig. 3a). Cluster 2 and cluster 3 consists of 262 (Fig. 3b) and 254 (Fig. 3c) genes, respectively. Cluster 3 also consists of two hubs, SLC15A3 and HCLS1 (Fig. 3c). Only 9 genes were observed in cluster 4 whereas cluster 5 has 2 genes (Fig. 3d). Therefore, only those communities were selected for enrichment analyses that have at least 10 nodes. Based on this criterion, only 3 communities (Cluster 1, 2 and 3) were finally analyzed for over-representation of GO terms.

To biologically categorize these clusters, DAVID functional analysis tool was used to classify the genes in each module, and observed the enrichment of GO term "Biological Process" in three selected modules. The top 20 statistically significant enriched GO terms for DEGs in top 3 clusters for community analysis are summarized in (Table 2). The three most statistically significant GO terms that were enriched in cluster 1 are response to wounding, defense response and inflammatory response. Other significant GO terms in this cluster include regulation of apoptosis or programmed cell death as well as positive regulation of response to stimulus. Cluster 2 shows higher enrichment for GO terms related to phosphorous metabolic processes as well as regulation of proliferation. The GO terms of cluster 3 were mostly related to antigen processing and presentation of peptide antigen, regulation of I-kappaB kinase/NF-kappaB cascade and positive regulation of T cell activation.

SNP analysis

Non-synonymous SNPs (nsSNPs) are positioned predominantly in coding regions and their consequences are observed in phenotypic characteristics of translated protein products (Ramensky et al. 2002). On the other hand, influence of synonymous SNPs is not determined easily on translated proteins probably owing to the uncertainty of genetic code (Hunt et al. 2009). In the current study we have focussed on deciphering the role of variants on proteins and their related pathways in successive steps, therefore we confined and selected nsSNPs for our computational analysis. Figure 4 shows the distribution of nsSNPs and total SNPs in four hub genes. From the figure it can be observed that in spite of a large number of SNPs identified in each hub gene, only a small percent represent the nsSNPs. The top hub *HCLS1* has 689 SNPs and 68 nsSNPs, whereas, *SLC15A3* has 443 SNPs and 60 nsSNPs. *HCK* has 1138 SNPs, which is highest among all four genes, but has only 48 nsSNPs. Finally, *LY96* has 796 SNPs and only 13 nsSNPs. The estimation of ratio of synonymous and nonsynonymous substitution rates helps in determining the evolutionary effect of substitutions on protein coding genes (Yang and Nielsen 1998).

Deleterious or damaged nsSNPs by PROVEAN protein batch tool

The PROVEAN Human protein batch tool was used for the prediction of damaged or deleterious nsSNPs. The program provides both PROVEAN as well as SIFT prediction scores. The input to the PROVEAN protein batch tool consists of detected coding nsSNPs for four hubs, that is, 68 nsSNPs for HCLS1 (UniProt ID: P14317), 60 for SLC15A3 (UniProt ID: Q8IY34), 48 for HCK (UniProt ID: A8K4G3), and 13 for LY96 (UniProt ID: Q9Y6Y9). The input for each hub gene was submitted independently to PROVEAN Human protein batch tool. Among 68 nsSNPs for HCLS1, 29 were identified to be deleterious by PROVEAN whereas 36 were identified to be damaging by SIFT (Supplementary Table S2). For SLC15A3, 28 and 36 out of 60 nsSNPs were predicted as deleterious or damaging by PROVEAN and SIFT, respectively (Supplementary

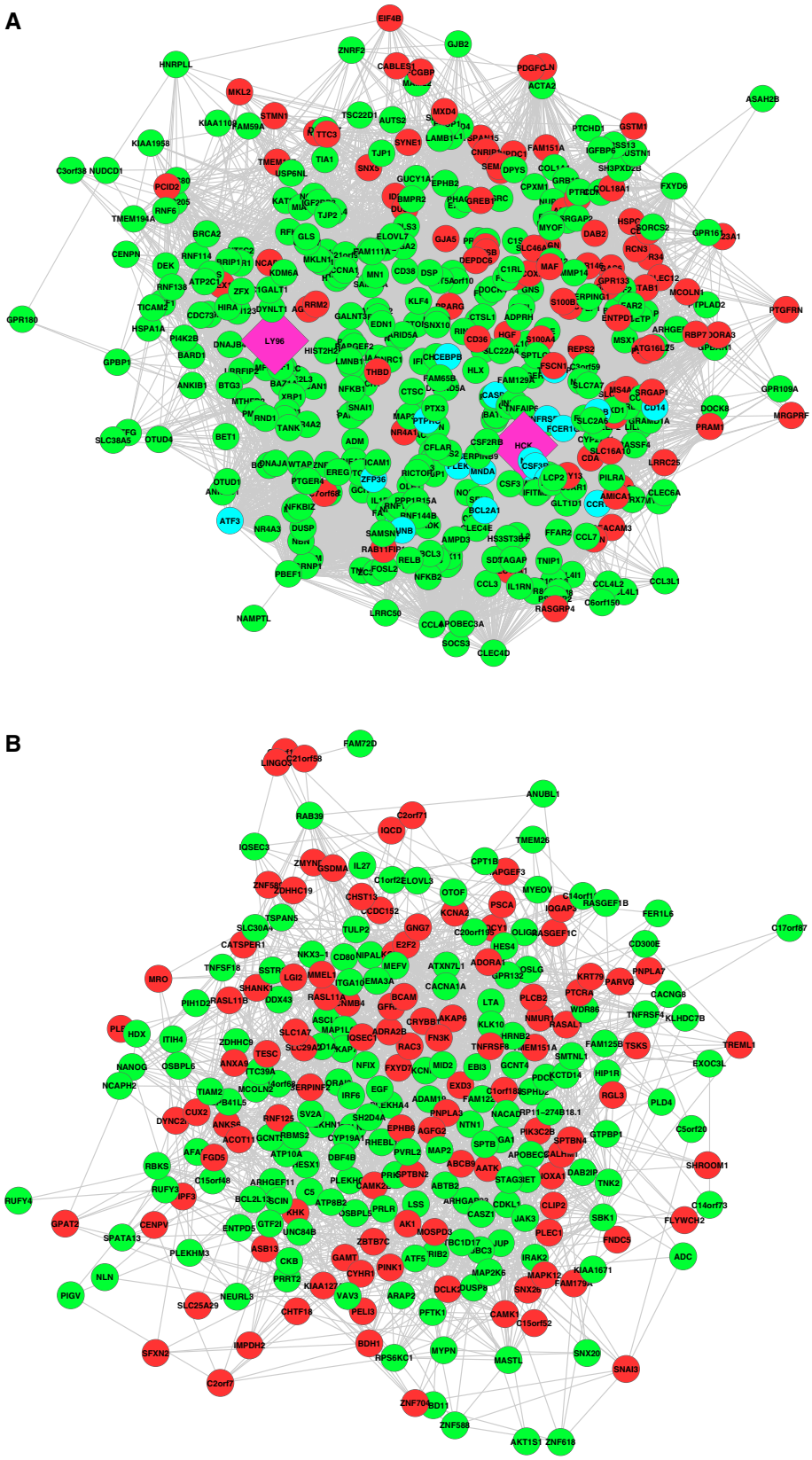

Fig. 3 Communities detected by fast greedy (GLay) clustering algorithm are shown. In each community **a** Cluster 1 **b** Cluster 2 **c** Cluster 3 **d** Cluster 4 **e** Cluster 5, up- and down-regulated genes are shown as *green* and *red* nodes respectively, whereas GeneMANIA predicted genes are represented as *cyan*. Nodes in *purple* diamond shape represent hub genes

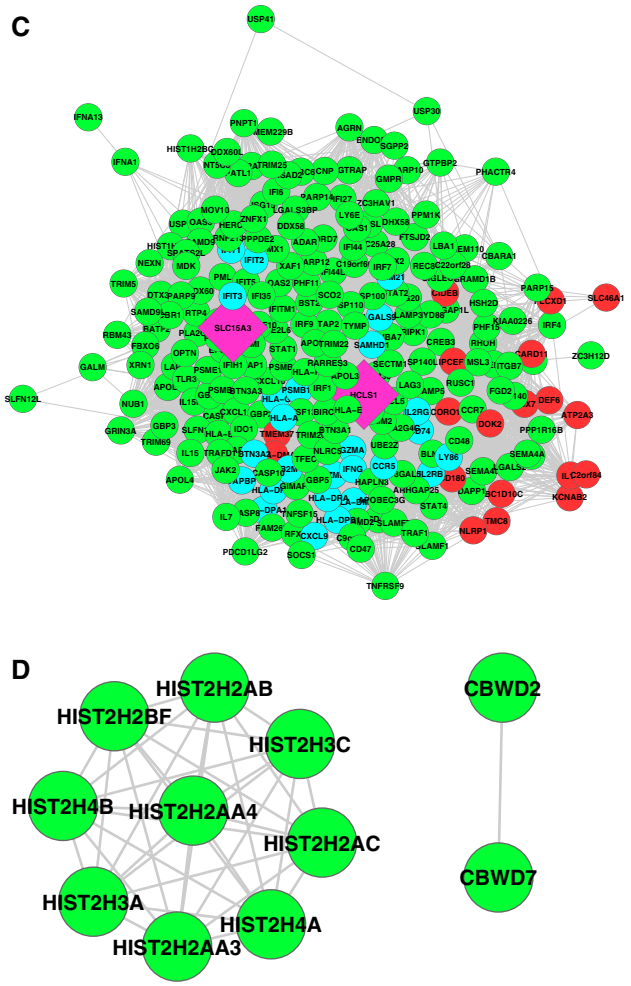

Fig. 3 continued

Table S3). PROVEAN identified 21 nsSNPs as deleterious while SIFT predicted 24 as damaging out of total 48 nsSNPs for HCK (Supplementary Table S4). Similarly, 6 and 5 nsSNPs were detected by PROVEAN and SIFT for 13 nsSNPs of LY96 gene (Supplementary Table S5).

Damaged nsSNPs by PolyPhen-2 web server

The protein sequences of each gene and their SNP substitution were submitted independently to the Polyphen-2 web server. Among 68 nsSNPs for HCLS1, 47 were identified to be damaging (Supplementary Table S2) whereas, 37 nsSNPs were predicted to have a damaging effect for SLC15A3 (Supplementary Table S3). Similarly, the predicted number of damaging nsSNPs observed for HCK was 21 out of 48 (Supplementary Table S4) and 8 nsSNPs were identified to be damaging for LY96 out of 13 nsSNPs (Supplementary Table S5). Overall, the number of nsSNPs that are commonly predicted to be deleterious or damaging by PROVEAN, SIFT and Polyphen-2, which may affect protein function, is 26 for HCLS1 and SLC15A3, 16 for HCK, and 5 for LY96 (Table 3), respectively.

Discussion

Discovery of molecular targets and targeted therapeutics have turned out to be a vital remedial treatment for diseases, especially with the advancement of bioinformatics in the last few decades (Dinh et al. 2007). In order to explore the role of DEGs that were recently identified by RNA-Seq analysis (Qian et al. 2013), and additionally related genes involved in WNV infection, an interaction network was constructed and node degree for each gene in the network was calculated. *HCLS1*, *SLC15A3*, *HCK* and *LY96* were the genes with the highest degree and considered as hubs in the network created from DEGs of primary human macrophages and additional related genes predicted by GeneMANIA. The most significant enriched GO terms within the interaction network identified by GeneMANIA include cytokine-mediated signaling and type I interferon-mediated signaling pathways, processes related to regulation of defense and immune response. Additional over-represented GO terms in the network represent processes related to immune cell activation (Table 1). This analysis is in consistency with the recent study on WNV infection in which DAVID functional annotation cluster analysis highlighted the enrichment of processes involved in immune or defense responses (Qian et al. 2013).

Among *HCLS1*, *SLC15A3*, *HCK* and *LY96*, only *HCK* was detected as one of the hubs from a list of DEGs whereas GeneMANIA predicted *HCLS1, SLC15A3,* and *LY96* genes to be a part of interaction network. These genes may possibly play a key role in regulating the biological process of WNV resistance and control viral infection. *HCLS1* gene codes hematopoietic cell-specific Lyn substrate 1 protein (HCLS1 or HS1) that consists of an SH3 (Src homology 3) adapter domain and can trigger activation of receptor-coupled tyrosine kinases (Kitamura et al. 1989; Van Rossum et al. 2005; Huang and Burkhardt 2007). Elevated levels of HCLS1 are related with chronic lymphoblastic leukemia (Huang et al. 2008; Scielzo et al. 2005), whereas lymphocyte precursors deficient in HCLS1 tends towards defective proliferation and differentiation of B-lymphocytes following B cell receptor activation (Skokowa et al. 2012). The members of the SLC15 protein family have been verified to be significant drug targets at the level of drug transport (Sasawatari et al. 2011). SLC15A3 and SLC15A4 are the two types of histidine transporters that have been identified in the lysosomes of immune cells (Sakata et al. 2001). Previous studies have established the fact that amino acids are essential in the

Table 2 Enrichment of GO terms for DEGs in each cluster

Cluster number	Term	*P* value
1	GO:0009611, response to wounding	3.3E−28
	GO:0006952, defense response	3.63E−25
	GO:0006954, inflammatory response	5.64E−25
	GO:0010033, response to organic substance	1.81E−15
	GO:0002237, response to molecule of bacterial origin	5.35E−10
	GO:0042330, taxis	5.6E−09
	GO:0006935, chemotaxis	5.6E−09
	GO:0042981, regulation of apoptosis	6.16E−09
	GO:0043067, regulation of programmed cell death	8.52E−09
	GO:0032496, response to lipopolysaccharide	8.93E−09
	GO:0001817, regulation of cytokine production	8.97E−09
	GO:0010941, regulation of cell death	9.62E−09
	GO:0048584, positive regulation of response to stimulus	1.07E−08
	GO:0048584, positive regulation of response to stimulus	1.07E−08
	GO:0009617, response to bacterium	2.81E−08
	GO:0001775, cell activation	2.89E−08
	GO:0002684, positive regulation of immune system process	5.7E−08
	GO:0002252, immune effector process	3.64E−07
	GO:0002252, immune effector process	3.64E−07
	GO:0002252, immune effector process	3.64E−07
2	GO:0006468, protein amino acid phosphorylation	6.19E−05
	GO:0016310, phosphorylation	0.000353
	GO:0046578, regulation of Ras protein signal transduction	0.000769
	GO:0006793, phosphorus metabolic process	0.002414
	GO:0006796, phosphate metabolic process	0.002414
	GO:0042325, regulation of phosphorylation	0.002418
	GO:0043242, negative regulation of protein complex disassembly	0.002544
	GO:0051693, actin filament capping	0.00345
	GO:0051174, regulation of phosphorus metabolic process	0.003524
	GO:0019220, regulation of phosphate metabolic process	0.003524
	GO:0051129, negative regulation of cellular component organization	0.003904
	GO:0030835, negative regulation of actin filament depolymerization	0.004442
	GO:0043244, regulation of protein complex disassembly	0.005631
	GO:0050670, regulation of lymphocyte proliferation	0.006219
	GO:0030834, regulation of actin filament depolymerization	0.006224
	GO:0032944, regulation of mononuclear cell proliferation	0.006539
	GO:0070663, regulation of leukocyte proliferation	0.006539
	GO:0030837, negative regulation of actin filament polymerization	0.0069
	GO:0051494, negative regulation of cytoskeleton organization	0.00736
	GO:0050671, positive regulation of lymphocyte proliferation	0.00736
3	GO:0019882, antigen processing and presentation	1.35E−16
	GO:0048002, antigen processing and presentation of peptide antigen	3.75E−12
	GO:0002474, antigen processing and presentation of peptide antigen via MHC class I	2.22E−09
	GO:0043122, regulation of I-kappaB kinase/NF-kappaB cascade	5.18E−09
	GO:0002684, positive regulation of immune system process	1.27E−08
	GO:0002684, positive regulation of immune system process	1.27E−08
	GO:0050863, regulation of T cell activation	1.56E−08
	GO:0043123, positive regulation of I-kappaB kinase/NF-kappaB cascade	1.66E−08

Table 2 continued

Cluster number	Term	*P* value
	GO:0010740, positive regulation of protein kinase cascade	2.22E−08
	GO:0019884, antigen processing and presentation of exogenous antigen	2.48E−08
	GO:0019884, antigen processing and presentation of exogenous antigen	2.48E−08
	GO:0051249, regulation of lymphocyte activation	3.45E−08
	GO:0050865, regulation of cell activation	4.18E−08
	GO:0050867, positive regulation of cell activation	7.7E−08
	GO:0010627, regulation of protein kinase cascade	1.47E−07
	GO:0002694, regulation of leukocyte activation	1.47E−07
	GO:0050870, positive regulation of T cell activation	1.48E−07
	GO:0050870, positive regulation of T cell activation	1.48E−07
	GO:0051251, positive regulation of lymphocyte activation	1.66E−07
	GO:0046649, lymphocyte activation	2.28E−07

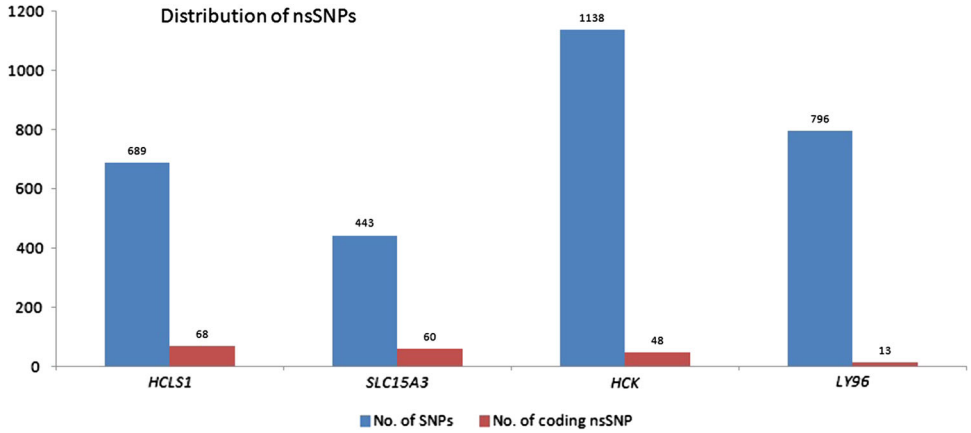

Fig. 4 Distribution of total SNPs and nsSNPs in four hub genes

regulation of immune responses (Sasawatari et al. 2011). *HCK* gene is mainly expressed in cells having monocyte or macrophage lineage (Ziegler et al. 1987) that are important HIV-1 target cells and viral reservoirs (Martín and Bandrés 1999; Orenstein 2001; Crowe et al. 2003). *HCK* encodes a 505-residue polypeptide, which is closely related to $pp56^{lck}$, a lymphocyte-specific protein-tyrosine kinase (Ziegler et al. 1987). HCK also one of the members of Src kinase family, is known to interact with NEF (HIV-1 virulence factor) and promotes viral pathogenesis of HIV/AIDS (Trible et al. 2006). *LY96* gene encodes a glycoprotein lymphocyte antigen 96 (LY96; also denoted as ESOP-1 or MD-2) that is involved in endotoxin recognition and offers a foundation for antiseptic drug development (Ohto et al. 2007). MD-2 interacts with the extracellular domain of Toll-like receptor 4 (TLR4) and is necessary for the activation of TLR4 by lipopolysaccharide (LPS) present on the outer membranes of Gram-negative bacterial cell walls (Mullen et al. 2003; Dziarski and Gupta 2000).

Recognizing the structure and function of biological networks is indispensable for the investigation of biological processes. Many recent studies have used network based studies to investigate various biological problems (Malik et al. 2014; Lee and Lee 2013). In this work, we identified 5 functional modules or communities in the interaction network using fast greedy algorithm implemented as GLAY (Su et al. 2010) plugin for Cytoscape. Furthermore, the functional enrichment of only 3 modules (based on the criteria described in the methods section) was explored by using functional annotation tool DAVID. Our analysis shows that the modules are enriched in functions related to defense response, inflammatory response, phosphorous metabolic processes and regulation of I-kappaB kinase/NF-kappaB cascade. According to previous studies, subjects with severe infection exhibit high expression levels of cytokine and chemokine genes (IL-8, TNF), antiviral signaling genes such as TMEM158 (involved in antiviral immune signaling) (Ishikawa and Barber 2008),

Table 3 List of nsSNPs that are commonly predicted to be damaging/deleterious/affect protein function by PROVEAN, SIFT as well as PolyPhen-2

Gene (No. of predicted nsSNPS)	SNP ID	AA change	PPH2_Prob	PPH2	PROVEAN score	PROVEAN (cutoff = −2.5)	SIFT score	SIFT (cutoff = 0.05)
LY96 (5)	rs141631661	12 S C	0.997	Probably damaging	−3.04	Deleterious	0.003	Damaging
	rs199978698	29 D N	1	Probably damaging	−2.54	Deleterious	0	Damaging
	rs142571384	70 D Y	1	Probably damaging	−7.36	Deleterious	0.002	Damaging
	rs367973351	99 D H	0.999	Probably damaging	−4.5	Deleterious	0.004	Damaging
	rs202107994	143 E A	0.962	Probably damaging	−2.85	Deleterious	0	Damaging
HCK (16)	rs199815006	125 F C	1	Probably damaging	−6.34	Deleterious	0	Damaging
	rs372761182	168 D N	0.999	Probably damaging	−3.61	Deleterious	0.002	Damaging
	rs200896933	186 G R	1	Probably damaging	−7.13	Deleterious	0	Damaging
	rs202001086	193 R Q	0.936	Possibly damaging	−3.13	Deleterious	0.001	Damaging
	rs199531584	212 G R	0.996	Probably damaging	−7.01	Deleterious	0	Damaging
	rs189947820	239 R W	1	Probably damaging	−6.86	Deleterious	0	Damaging
	rs200867413	257 W R	1	Probably damaging	−12.37	Deleterious	0	Damaging
	rs200178250	271 T M	0.973	Probably damaging	−4.89	Deleterious	0	Damaging
	rs151054795	339 I T	0.981	Probably damaging	−3.86	Deleterious	0.029	Damaging
	rs114505022	350 A T	0.983	Probably damaging	−3.52	Deleterious	0.001	Damaging
	rs145632103	384 R W	1	Probably damaging	−7.46	Deleterious	0	Damaging
	rs377468533	392 T M	0.554	Possibly damaging	−3.83	Deleterious	0.022	Damaging
	rs137871205	435 R W	1	Probably damaging	−6.94	Deleterious	0	Damaging
	rs374360915	435 R Q	0.993	Probably damaging	−3.2	Deleterious	0.001	Damaging
	rs142491977	470 M V	0.773	Possibly damaging	−3.13	Deleterious	0.001	Damaging
	rs17093828	482 P Q	1	Probably damaging	−7.15	Deleterious	0	Damaging
HCLS1 (26)	rs150386736	7 G A	1	Probably damaging	−4.789	Deleterious	−4.789	Damaging
	rs372797954	19 D E	0.999	Probably damaging	−3.437	Deleterious	−3.437	Damaging
	rs148254045	21 W C	1	Probably damaging	−11.321	Deleterious	−11.321	Damaging
	rs146539124	54 I T	0.986	Probably damaging	−3.533	Deleterious	−3.533	Damaging
	rs371758156	155 R C	0.999	Probably damaging	−2.749	Deleterious	−2.749	Damaging
	rs372191150	163 V L	0.971	Probably damaging	−2.684	Deleterious	−2.684	Damaging
	rs34767273	166 D H	1	Probably damaging	−5.588	Deleterious	−5.588	Damaging
	rs144275149	181 T M	1	Probably damaging	−2.589	Deleterious	−2.589	Damaging
	rs147129061	184 H Y	1	Probably damaging	−5.782	Deleterious	−5.782	Damaging
	rs368893881	184 H Q	1	Probably damaging	−7.69	Deleterious	−7.69	Damaging
	rs36104070	200 I T	0.908	Possibly damaging	−3.216	Deleterious	−3.216	Damaging
	rs143408084	203 D E	0.891	Possibly damaging	−3.417	Deleterious	−3.417	Damaging
	rs202020296	211 G S	0.977	Probably damaging	−2.55	Deleterious	−2.55	Damaging
	rs61749596	218 P L	0.97	Probably damaging	−2.595	Deleterious	−2.595	Damaging
	rs144589441	225 T M	1	Probably damaging	−3.399	Deleterious	−3.399	Damaging
	rs150713543	240 A V	0.998	Probably damaging	−3.167	Deleterious	−3.167	Damaging
	rs145562548	291 P L	0.984	Probably damaging	−2.73	Deleterious	−2.73	Damaging
	rs368506277	430 G W	1	Probably damaging	−5.551	Deleterious	−5.551	Damaging
	rs376245237	441 G V	1	Probably damaging	−2.85	Deleterious	−2.85	Damaging
	rs201252242	451 P L	1	Probably damaging	−9.08	Deleterious	−9.08	Damaging
	rs202175021	452 D N	0.997	Probably damaging	−3.99	Deleterious	−3.99	Damaging
	rs142545513	453 D N	1	Probably damaging	−4.19	Deleterious	−4.19	Damaging
	rs377000996	464 G A	1	Probably damaging	−5.148	Deleterious	−5.148	Damaging
	rs78618042	467 R Q	0.999	Probably damaging	−3.049	Deleterious	−3.049	Damaging
	rs374286282	476 L P	1	Probably damaging	−6.108	Deleterious	−6.108	Damaging
	rs142693789	479 A E	0.998	Probably damaging	−4.51	Deleterious	−4.51	Damaging

Table 3 continued

Gene (No. of predicted nsSNPS)	SNP ID	AA change	PPH2_Prob	PPH2	PROVEAN score	PROVEAN (cutoff = −2.5)	SIFT score	SIFT (cutoff = 0.05)
SLC15A3 (26)	rs377584250	51 G R	0.991	Probably damaging	−4.34	Deleterious	0.002	Damaging
	rs199743083	93 L P	0.994	Probably damaging	−4.74	Deleterious	0.001	Damaging
	rs201623464	187 V G	1	Probably damaging	−6.04	Deleterious	0	Damaging
	rs140918009	191 G C	0.992	Probably damaging	−7.49	Deleterious	0	Damaging
	rs142952480	197 R H	1	Probably damaging	−3.35	Deleterious	0.007	Damaging
	rs377073448	213 S L	1	Probably damaging	−4.21	Deleterious	0.002	Damaging
	rs140942899	256 P L	0.999	Probably damaging	−5.93	Deleterious	0.002	Damaging
	rs147368970	258 G D	1	Probably damaging	−5.28	Deleterious	0	Damaging
	rs373404805	318 P L	1	Probably damaging	−6.64	Deleterious	0.003	Damaging
	rs372172597	327 W R	1	Probably damaging	−11.75	Deleterious	0.004	Damaging
	rs34738190	337 Y C	0.998	Probably damaging	−7.53	Deleterious	0	Damaging
	rs188010920	369 T M	0.983	Probably damaging	−3.17	Deleterious	0.021	Damaging
	rs201019779	371 P L	1	Probably damaging	−8.64	Deleterious	0.001	Damaging
	rs140596697	392 R C	0.99	Probably Damaging	−4.74	Deleterious	0.005	Damaging
	rs140945002	400 R W	0.518	Possibly damaging	−4.65	Deleterious	0.027	Damaging
	rs199509067	405 P S	1	Probably damaging	−4.34	Deleterious	0.025	Damaging
	rs142111240	464 Q R	1	Probably damaging	−3.54	Deleterious	0	Damaging
	rs377527454	471 S G	1	Probably damaging	−3.54	Deleterious	0	Damaging
	rs148648672	473 I T	0.998	Probably damaging	−3.85	Deleterious	0.002	Damaging
	rs377256065	481 E Q	0.999	Probably damaging	−2.63	Deleterious	0	Damaging
	rs143864458	488 P L	1	Probably damaging	−9.1	Deleterious	0	Damaging
	rs367924777	489 R C	1	Probably damaging	−4.81	Deleterious	0.001	Damaging
	rs369243035	531 G E	1	Probably damaging	−5.26	Deleterious	0	Damaging
	rs374216372	537 R W	0.996	Probably damaging	−3.22	Deleterious	0.023	Damaging
	rs199844407	552 T M	0.999	Probably damaging	−3.56	Deleterious	0.002	Damaging
	rs371994103	562 R C	1	Probably damaging	−3.57	Deleterious	0.001	Damaging

AA Change amino acid substitution, *PPH2_Prob* a classifier probability of the variation being damaging as predicted by PolyPhen-2, *PPH2* PolyPhen-2, an automatic tool for the prediction of potential effect of an amino acid substitution on the structure and function of human proteins by using physical and evolutionary comparative considerations, *PROVEAN Score* score that correlates with biological activity level and might be used as an indicator for the amount of functional effect of a protein variation. If the PROVEAN score is $\leq$ to a given threshold, the variation is predicted as deleterious, whereas the variant is predicted to have a neutral effect if the score is above the threshold, *PROVEAN* Protein Variation Effect Analyzer (PROVEAN) is a tool that predicts whether an amino acid substitution or indel effects the biological function of a protein, *SIFT Score* scaled probability of an amino acid substitution being tolerated. Substitutions with scores <0.05 are predicted to affect protein function, *SIFT* Sorting Tolerant From Intolerant (SIFT) is a tool to predict whether an amino acid substitution affects protein function

and CD69, a C-type lectin that plays a vital role in T cell activation, proliferation and signaling (Martín and Sánchez-Madrid 2011). Moreover, it has been observed in subjects with history of severe infection displayed higher expression levels of anti-inflammatory proteins such as PI3, a protease inhibitor that acts to reduce inflammation (Verrier et al. 2012), and TNFAIP3 that restrains NF-kappa B activation (Song et al. 1996).

One of the prime objectives of human genetics lies in deciphering the functional consequences of inherited variations between individuals. Even though millions of SNPs are said to be present in the human population, the small percentage of this set will be posing potential disease risk. Among the genomic variants, non-synonymous SNPs are of prime interest as they lead to amino acid change in translated protein products, and hence, these variants are primarily studied for their relevance in inheritable diseases and drug sensitivity (Yue and Moult 2006). It has been further termed as one of the promising arena for research in the field of genomics (Karchin et al. 2005). The effect of majority of nsSNPs is non-detrimental as they are deleted by natural selection. Assessment of lethality of detrimental nsSNPS is traced primarily based on phylogenetic information (i.e. correlation with residue conservation) and structural approaches (PolyPhen). However, studies have also revealed many human disease genes as a resultant of

exonic or noncoding mutations affecting regulatory regions (Hudson 2003; Yan et al. 2002).

The findings from this work propose that the application of computational algorithms, such as PROVEAN, SIFT, and PolyPhen-2 analysis may offer a substitute approach to pick target SNPs by understanding the effect of SNPs on the functional attributes or molecular phenotype of a protein. In future studies, structural models of the mutants will be built that may be applicable for predicting the deleterious nsSNPs, which in turn will be useful for additional genotype-phenotype research as well as pharmacogenetics studies.

Conclusion

Identification of hub genes that possess potential to cause disease-related symptoms will lead to a comprehensive approach to devise a therapeutic method for curing complex diseases. This study determined interactions between hub genes *HCLS1*, *SLC15A3, HCK* and *LY96* from DEGs of primary human macrophages. These genes act in signal transduction, cell differentiation, peptide transport and related important cellular pathways. Moreover, we shortlisted three functional modules related to defense and inflammatory responses in WNV infected macrophages. Hence, these clusters can be targeted in further studies to devise a therapeutic target that stabilizes their gene expression and aids in curbing symptoms related to WNV-linked infection. In addition to targeting multiple targets, gene-based approach aids in devising more personalized approach, which further enhances its efficacy.

Acknowledgments The Deanship of Scientific Research (DSR), King Abdulaziz University, Jeddah, funded this work under Grant No. (830-011-D1434). The authors, therefore, acknowledge with thanks DSR for technical and financial support.

Conflict of interest Authors declare no conflict of interest.

References

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR (2010) A method and server for predicting damaging missense mutations. Nat Method 7:248–249

Albert R (2005) Scale-free networks in cell biology. J Cell Sci 118:4947–4957

Bamshad MJ, Ng SB, Bigham AW, Tabor HK, Emond MJ, Nickerson DA, Shendure J (2011) Exome sequencing as a tool for Mendelian disease gene discovery. Nat Rev Genet 12:745–755

Brinton MA (2002) The molecular biology of West Nile virus: a new invader of the western hemisphere. Annu Rev Microbiol 56:371–402

Choi Y, Sims GE, Murphy S, Miller JR, Chan AP (2012) Predicting the functional effect of amino acid substitutions and indels. PLoS One 7:e46688

Chowdhury P, Khan SA, Dutta P, Topno R, Mahanta J (2014) Characterization of West Nile virus (WNV) isolates from Assam, India: insights into the circulating WNV in northeastern India. Comp Immunol Microb 37:39–47

Christodoulou DC, Wakimoto H, Onoue K, Eminaga S, Gorham JM, DePalma SR, Herman DS, Teekakirikul P, Conner DA, McKean DM et al (2014) 5′RNA-Seq identifies Fhl1 as a genetic modifier in cardiomyopathy. J Clin Invest 124:1364–1370

Crowe S, Zhu T, Muller WA (2003) The contribution of monocyte infection and trafficking to viral persistence, and maintenance of the viral reservoir in HIV infection. J Leukoc Biol 74:635–641

Dinh P, Sotiriou C, Piccart MJ (2007) The evolution of treatment strategies: aiming at the target. Breast 16:10–16

Dziarski R, Gupta D (2000) Role of MD-2 in TLR2- and TLR4-mediated recognition of Gram-negative and Gram-positive bacteria and activation of chemokine genes. J Endotoxin Res 6:401–405

Firoz A, Malik A, Singh SK, Jha V, Ali A (2014) Comparative analysis of glycogene expression in different mouse tissues using RNA-Seq data. Int J Genomics 2014:837365

Heinz FX, Collett MS, Purcell RH, Gould EA, Howard CR, Houghton M, Moormann RJM, Rice CM, Thiel HJ (2000) Family *Flaviviridae*. In: van Regenmortel MHV, Fauquet CM, Bishop DHL, Carstens EB, Estes MK, Lemon SM et al (eds) Virus taxonomy: classification and nomenclature of viruses. Academic Press, San Diego, pp 859–878

Huang Y, Burkhardt JK (2007) T-cell-receptor–dependent actin regulatory mechanisms. J Cell Sci Rev 120:723–730

Huang Y, Comiskey EO, Dupree RS, Li S, Koleske AJ, Burkhardt JK (2008) The c-Abl tyrosine kinase regulates actin remodeling at the immune synapse. Blood 112:111–119

Hudson TJ (2003) Wanted: regulatory SNPs. Nat Genet 33:439–440

Hunt R, Sauna ZE, Ambudkar SV, Gottesman MM, Kimchi-Sarfaty C (2009) Silent (synonymous) SNPs: should we care about them? Methods Mol Biol 578:23–39

Ishikawa H, Barber GN (2008) STING is an endoplasmic reticulum adaptor that facilitates innate immune signaling. Nature 455:674–678

Jones M, Dry IR, Frampton D, Singh M, Kanda RK, Yee MB, Kellam P, Hollinshead M, Kinchington PR, Otoole EA et al (2014) RNA-seq analysis of host and viral gene expression highlights interaction between varicella zoster virus and keratinocyte differentiation. PLoS Pathog 10:e1003896

Karchin R, Kelly L, Sali A (2005) Improving functional annotation of non-synonymous SNPs with information theory. Pac Symp Biocomput 10:397–408

Kitamura D, Kaneko H, Miyagoe Y, Ariyasu T, Watanabe T (1989) Isolation and characterization of a novel human gene expressed specifically in the cells of hematopoietic lineage. Nucl Acids Res 17:9367–9379

Kumar P, Henikoff S, Ng PC (2009) Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm. Nat Protoc 4:1073–1081

Lee J, Lee J (2013) Hidden information revealed by optimal community structure from a protein-complex bipartite network improves protein function prediction. PLoS One 8:e60372

Lee EJ, Malik A, Pokharel S, Ahmad S, Mir BA, Cho KH, Kim J, Kong JC, Lee DM, Chung KY et al (2014) Identification of genes differentially expressed in myogenin knock-down bovine muscle satellite cells during differentiation through RNA sequencing analysis. PLoS One 9:e92447

Malik A, Lee J, Lee J (2014) Community-based network study of protein-carbohydrate interactions in plant lectins using glycan array data. PLoS One 9:e95480

Martín JC, Bandrés JC (1999) Cells of the monocyte-macrophage lineage and pathogenesis of HIV-1 infection. J Acquir Immune Defic Syndr 22:413–429

Martín P, Sánchez-Madrid F (2011) CD69: an unexpected regulator of TH17 cell-driven inflammatory responses. Sci Signal 4:pe14

Mercer TR, Clark MB, Crawford J, Brunck ME, Gerhardt DJ, Taft RJ, Nielsen LK, Dinger ME, Mattick JS (2014) Targeted sequencing for gene discovery and quantification using RNA CaptureSeq. Nat Protoc 5:989–1009

Montojo J, Zuberi K, Rodriguez H, Kazi F, Wright G, Donaldson SL, Morris Q, Bader GD (2010) GeneMANIA Cytoscape plugin: fast gene function predictions on the desktop. Bioinformatics 26:2927–2928

Mullen GE, Kennedy MN, Visintin A, Mazzoni A, Leifer CA, Davies DR, Segal DM (2003) The role of disulfide bonds in the assembly and function of MD-2. Proc Natl Acad Sci USA 100:3919–3924

Nagalakshmi U, Waern K, Snyder M (2010) RNA-Seq: a method for comprehensive transcriptome analysis. Curr Protoc Mol Biol 11:1–13

Ohto U, Fukase K, Miyake K, Satow Y (2007) Crystal structures of human MD-2 and its complex with antiendotoxic lipid IVa. Science 316:1632–1634

Orenstein JM (2001) The macrophage in HIV infection. Immunobiology 204:598–602

Peña J, Plante JA, Carillo AC, Roberts KK, Smith JK, Juelich TL, Beasley DWC, Freiberg AN, Labute MX, Arani PN et al (2014) Multiplexed digital mRNA profiling of the inflammatory response in the west nile swiss webster mouse model. PLoS Negl Trop Dis 8:e3216

Qian F, Chung L, Zheng W, Bruno V, Alexander RP, Montgomery RR (2013) Identification of genes critical for resistance to infection by West Nile virus using RNA-Seq analysis. Viruses 5:1664–1681

Ramensky V, Bork P, Sunyaev S (2002) Human non-synonymous SNPs: server and survey. Nucl Acids Res 17:3894–3900

Sakata K, Yamashita T, Maeda M, Moriyama Y, Shimada S, Tohyama M (2001) Cloning of a lymphatic peptide/histidine transporter. Biochem J 356:53–60

Sasawatari S, Okamura T, Kasumi E, Tanaka-Furuyama K, Yanobu-Takanashi R, Shirasawa S, Kato N, Toyama-Sorimachi N (2011) The solute carrier family 15A4 regulates TLR9 and NOD1 functions in the innate immune system and promotes colitis in mice. Gastroenterology 140:1513–1525

Scielzo C, Ghia P, Conti A, Bachi A, Guida G, Geuna M, Alessio M, Cappio FC (2005) HS1 protein is differentially expressed in chronic lymphocytic leukemia patient subsets with good or poor prognoses. J Clin Invest 115:1644–1650

Skokowa J, Klimiankou M, Klimenkova O, Lan D, Gupta K, Hussein K, Carrizosa E, Kusnetsova I, Li Z, Sustmann C et al (2012) Interactions among HCLS1, HAX1 and LEF-1 proteins are essential for G-CSF-triggered granulopoiesis. Nat Med 18:1550–1559

Smoot ME, Ono K, Ruscheinski J, Wang PL, Ideker T (2011) Cytoscape 2.8: new features for data integration and network visualization. Bioinformatics 27:431–432

Song HY, Rothe M, Goeddel DV (1996) The tumor necrosis factor-inducible zinc finger protein A20 interacts with TRAF1/TRAF2 and inhibits NF-kappa B activation. Proc Natl Acad Sci USA 93:6721–6725

Su G, Kuchinsky A, Morris JH, States DJ, Meng F (2010) GLay: community structure analysis of biological networks. Bioinformatics 26:3135–3137

Suthar MS, Diamond MS, Gale M Jr (2013) West Nile virus infection and immunity. Nat Rev Microbiol 11:115–128

Trapnell C, Williams BA, Pertea G, Mortazavi A, Kwan G, van Baren MJ, Salzberg SL, Wold BJ, Pachter L (2010) Transcript assembly and quantification by RNA-Seq reveals unannotated transcripts and isoform switching during cell differentiation. Nat Biotechnol 28:511–515

Trible RP, Emert-Sedlak L, Smithgall TE (2006) HIV-1 Nef selectively activates Src family kinases Hck, Lyn, and c-Src through direct SH3 domain interaction. J Biol Chem 281:27029–27038

Van Rossum AG, Schuuring-Scholtes E, Seggelen V, Kluin PM, Schuuring E (2005) Comparative genome analysis of cortactin and HS1: the significance of the F-actin binding repeat domain. BMC Genom 6:15

Verrier T, Solhonne B, Sallenave JM, Garcia-Verdugo I (2012) The WAP protein Trappin-2/Elafin: a handyman in the regulation of inflammatory and immune responses. Int J Biochem Cell Biol 44:1377–1380

Yan H, Yuan W, Velculescu VE, Vogelstein B, Kinzler KW (2002) Allelic variation in human gene expression. Science 297:1143

Yang Z, Nielsen R (1998) Synonymous and nonsynonymous rate variation in nuclear genes of mammals. J Mol Evol 46:409–418

Young A, Berry R, Holloway AF, Blackburn NB, Dickinson JL, Sakala M, Philips JL, Moore KHB (2014) RNA-seq profiling of a radiation resistant and radiation sensitive prostate cancer cell line highlights opposing regulation of DNA repair and targets for radiosensitization. BMC Cancer 14:808

Yue P, Moult J (2006) Identification and analysis of deleterious human SNPs. J Mol Biol 356:1263–1274

Ziegler SF, Marth JD, Lewis DB, Perlmutter RM (1987) Novel protein-tyrosine kinase gene (hck) preferentially expressed in cells of hematopoietic origin. Mol Cell Biol 7:2276–2285