Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 1, 2021, Pages. 4895 - 4904
Received 15 December 2020; Accepted 05 January 2021.

Exploring Vital Gene in Rheumatoid Arthritis-A Network and Pharmacology Approach

Muhammad Kamil Bin Anas[1], Venkataramanan Swaminathan[2],Mohammed Kaleemullah[3]
[1,2]Faculty of Health and Life Sciences, Management and Science University, Seksyen 13, 40100, Shah Alam, Selangor, Malaysia.
[3]Faculty of Pharmacy, Management and Science University, Seksyen 13, 40100, Shah Alam, Selangor, Malaysia.

ABSTRACT

Rheumatoid arthritis is the most common inflammatory arthritis and is a major effect of disability. This study aimed to explore the vital genes of auto immune diseases(rheumatoid arthritis) in homo sapiens by using bioinformatics analysis and exploration on the genes that causes the rheumatoid arthritis.The gene expression profile GSE 100191 was downloaded from the gene expression omnibus (GEO) database. Pathway enrichment analysis of genes were performed in the database for annotation, visualization and integrated discovery (DAVID) server. Further analysis was carried out by STRING server and Cytoscape to perform functional annotation and protein-protein interactions (PPI), network construction. The genes were separated to do the further analysis in Cytoscape and String website to find the hub gene and confident score. SNP analysis were performed using dbSNP, PolyPhen-2, SNAP2, pMUT, SNPs & GO and Mutpred software by using HDAC1 gene. There were 138 downregulated gene and 112 upregulated gene. HDAC1 was chosen from the 20 hub genes for SNP analysis. 34 SNP was found to be malignant. The gene HDAC1 (Histone Deacetylase 1), which is involved in rheumatoid arthritis, is important. This is the study which used computational approaches to classify significant nsSNP in HDAC1. The discovery of featured genes which were significantly related to cell cycle and DNA replication has potential for use in the clinic for the diagnosis of rheumatoid arthritis in the future.

Keywords
Upregulated, Downregulated, Gene exploration, SNP analysis, Functional annotation

Introduction

Rheumatoid arthritis is the most common inflammatory arthritis and is a major effect of disability. In this study, we will be focusing on identify the key gene marker which cause the rheumatoid arthritis by gene expression analysis and performing SNP analysis. Rheumatoid arthritis is a chronic inflammatory condition which affects more than just the joints. Some people can experience the disease affecting a wide range of body systems including the skin , eyes, lungs, heart and blood vessels. Signs and symptoms of rheumatoid arthritis may include tender, wet, swollen joints, joint pain that typically gets worse in the morning and after inactivity, fatigue , fever and lack of appetite. Rheumatoid arthritis occurs when the synovium affects the immune system and the lining of the membranes that cover the joints(Book: Mayo Clinic on Arthritis). The inflammatory process densifies the synovium which eventually may damage the joint cartilage and bone. Clinical trials show that remission of symptoms is more likely when treatment begins early with medications known as disease-modifying antirheumatic drugs (DMARDs).Rheumatoid arthritis affects about five in 1000 people in Malaysia. This study is focusing on finding hub genes of rheumatoid arthritis and performing SNP analysis through this disease.

Methodology

The Geo Dataset was open in NCBI and rheumatoid arthritis was selected and choose homo sapiens. The gene expression profile was selected which is GSE100191. The 19 samples were separated to healthy and unhealthy group (Meng Zhang et al., 2019) . Next, that sample

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 1, 2021, Pages. 4895 - 4904
Received 15 December 2020; Accepted 05 January 2021.

wereanalyse to choose the top 250 gene which are related to lung cancer. The top 250 gene then were downloaded and separated in excel with first group upregulated and the second group downregulated. The upregulated have a positive value of logFC while downregulated have negative value of logFC.

DAVID server

The DAVID server were open and the functional annotation was selected. The gene list were enter from the file save in upregulated or downregulated in the provided box. The identifier were selected and in this study we uses the UNIGENE id. The last steps were to submit the list which were differentiate between upregulated and downregulated gene. DAVID now offers researchers a robust collection of functional annotation resources to understand the biological significance behind a broad gene list. So, in this server, I try to find the cellular components of genes upregulated and downregulated, biological pathway and molecular function of the genes.

STRING database

Login to String database and the next part in the study select new gene set were added. The name of the set were identified as upregulated and downregulated. The gene id which is from upregulated and downregulated were paste on the identifier provided. The protein-protein interaction were launch and proceed to the next steps which is clicking the continue. To view the analyse data, go back to stored genes and by clicking show the analysis can be view with an interactive way(Damian Szklarczyk et al., 2017).

Cytoscape

In the Cytoscape, string protein was selected and the upregulated gene and downregulated gene was copy and paste to the search engine. The selected confident score was 0.7 and the maximum additional score was 40 and then continuing to next step which where to import the gene. The apps was selected than which is CentiScaPE 2.2 and degree and betweenness were selected in the implemented centralities.

SNP Dataset Collection

Human HDAC1 gene SNPs were recovered from dbSNP which obtained their amino acid sequence (Uniprot) from UniProt Protein Data-base (http:/www.uniprot.org) in FASTA format.

PolyPhen-2

Polymorphism phenotyping v2 (PolyPhen-2) is a web-based method where it uses a naïve Bayesian classifier to predict by physical and comparative consideration the effect of amino acid substitution on protein function and structure. The FASTA sequence and amino acid variants are the input query for PolyPhen 2 (Muhammad Ramzan et al., 2012).

SNAP2

SNAP2 predicts the practical impact of SNP on a machine learning system based on neutral network. The input question to be entered to know the SNAP2 result is a protein sequence in FASTA format(MiliNaiwal et al . , 2018).

SNPs & GO

Single nucleotide polymorphism database and gene ontology (SNPs&GO) tool is server based on vector control system (SVM). The SNPs&GO input question is a UNIPROT Accession Number,

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 1, 2021, Pages. 4895 - 4904
Received 15 December 2020; Accepted 05 January 2021.

Mutation Location and Wild-type residue as well as mutant-type residue (Muhammad Ramzan et al., 2012).

PMut

PMut2017 is maintained by Molecular Modeling and Bioinformatics, which is developed using the PyMut library along with a 12-function Random Forest classifier. It predicts whether SNP is neutral or causes disease. UniProt ID and mutations are input question for PMut (MiliNailwal et al . , 2018).

Prediction of Harmful Mutations by MutPred

MutPred is a web application tool built to predict the replacement of amino acids as disease-associated or not in humans. The input query method is a series of proteins, a list of variants and an email address.

Results

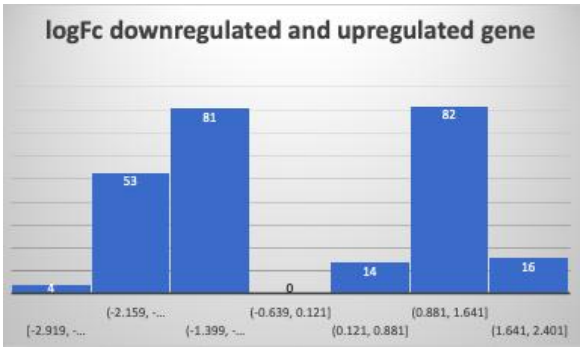

Figure 1: Downregulated with logFc value negatives and upregulated with logFc value positive

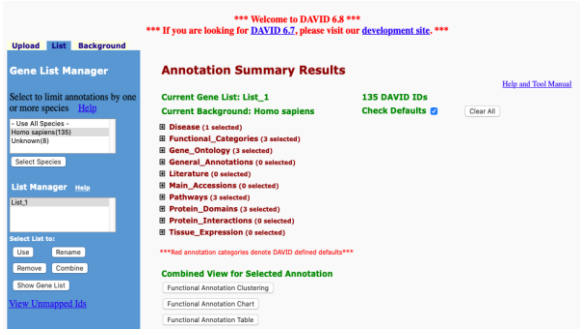

Figure 2:135 gene was selected in DAVID server for downregulated results. Functional annotation table were saved in excel format. There are 117 genes were recorded after that and being saved.

Hub Genes selection by Cytoscape

The biomarkers (hub gene) were selected with any gene that havebetweenness of more than 0.25 and the degree of 2.0 or higher. A total of 20 hub genes with degree more than 2.0 and betweenness of 0.25 were identified.

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 1, 2021, Pages. 4895 - 4904
Received 15 December 2020; Accepted 05 January 2021.

Table 1.Collection of hub gene of upregulated gene in rheumatoid arthritis

Gene	**Betweenness**	**Degree**
IL22	0.25	3.0
PVRL1	0.547	4.0
LPHN2	2.0	2.0
ETS1	2.395	6.0
PVRL3	3.515	6.0
PVR	5.812	6.0
ITGA2	5.993	6.0
GGH	28.681	6.0

Table 2.Collection of hub gene of downregulated gene in rheumatoid arthritis

Gene	Betweenness	Degree
MAML3	1.67	4.0
SPHK2	2.295	2.0
CENPT	2.531	7.0
DNAH2	3.625	4.0
TNK2	5.822	3.0
AMBRA1	6.756	4.0
LMO4	7.403	3.0
COPZ2	14.029	7.0
AP1S1	15.719	7.0
ADAM15	19.118	3.0
RUNX2	33.132	5.0
HDAC1	38.636	7.0

Protein-protein interaction (PPI) between the hub genes in STRING database

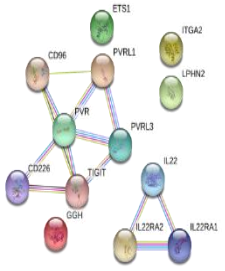

Figure 3 :**DEGs PPI network was developed using STRING database. PPI, protein – protein interaction; STRING, Search Method for Interacting Genes Retrieving. This is upregulated gene hub gene.**

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 1, 2021, Pages. 4895 - 4904
Received 15 December 2020; Accepted 05 January 2021.

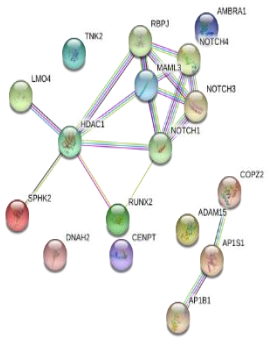

Figure 4: The PPI network of DEGs was constructed using STRING database. PPI, protein–protein interaction; STRING, Search Tool for the Retrieval of Interacting Genes. This is the hub gene of downregulated gene.

It can be concluded that HDAC1 may play important roles in the carcinogenesis or progression of rheumatoid arthritis from the network of the hub genes and their co-expression genes.

SNP Retrieval from dbSNP Database

The HDAC1 gene SNPs data were recovered from dbSNP database (https:/www.ncbi.nlm.nih.gov / snp/?term = DBY). The human gene HDAC1 contains a total of 355 SNPs, of which 201 contain missense and contiguous reference mutation.

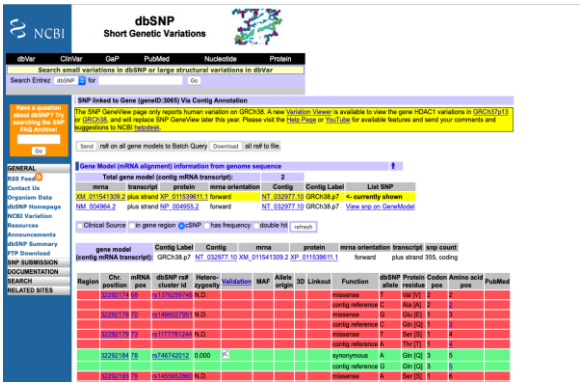

Figure 5: The 355 sample of SNP from HDAC1 gene.

Prediction of deleterious nsSNPs of HDAC1 gene

Using PolyPhen2, Total 201 missense and contiguous reference SNPs of the dbSNP HDAC1 gene were analysed. Of the 201 SNPs, 15 have been projected to be harmful. Among the SNPs predicted by SNAP2, all SNPs showed effect on protein function.

Table 3:Functional consequences of nsSNPs in human HDAC1 gene.

SNPs	AA variant	PolyPhen2	SNAP2
rs1376259746	A2V	Probably Damaging	Effect
rs749855844	Y23C	Probably Damaging	Effect

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 1, 2021, Pages. 4895 - 4904
Received 15 December 2020; Accepted 05 January 2021.

rs1318049482	R36H	Probably Damaging	Effect
rs1420901500	L43P	Probably Damaging	Effect
rs767396990	A59D	Probably Damaging	Effect
rs777501351	H68R	Probably Damaging	Effect
rs748444579	N83K	Probably Damaging	Effect
rs1302971614	G105S	Probably Damaging	Effect
rs1281325781	G116D	Probably Damaging	Effect
rs756177951	A132T	Probably Damaging	Effect
rs371689839	V183M	Probably Damaging	Effect
rs768748943	T190M	Probably Damaging	Effect
rs773477997	G231R	Probably Damaging	Effect
rs1391570201	G430D	Probably Damaging	Effect
rs1412932910	R431C	Probably Damaging	Effect

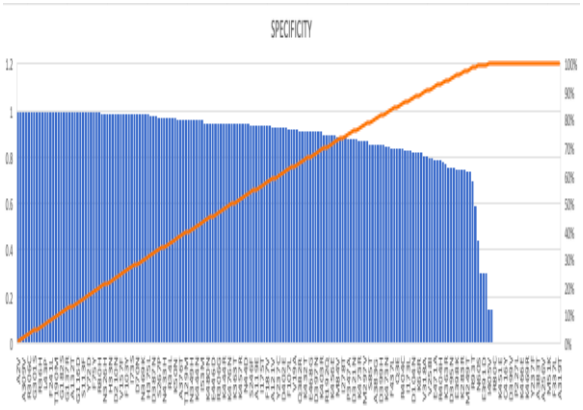

Figure 6: The horizontal line shows the amino acid variant and the vertical line shows that the specificity to be malignant from PolyPhen2 analysis.

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 1, 2021, Pages. 4895 - 4904
Received 15 December 2020; Accepted 05 January 2021.

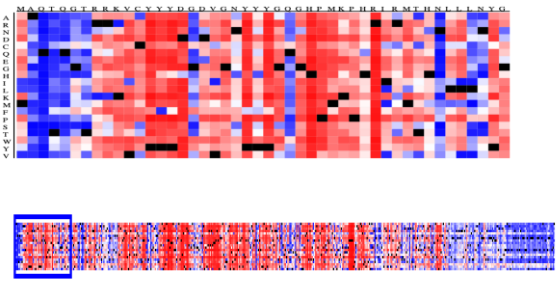

Figure 7: This is the whole sequence of HDAC1 gene and it is differentiate between normal and effect based on SNAP2 analysis.

Table 4:Prediction of disease associate nsSNPs in human HDAC1 gene.

SNPs	AA variant	SNPs and GO	pMUT
rs1376259746	A2V	Neutral	Disease
rs749855844	Y23C	Disease	Disease
rs1318049482	R36H	Disease	Disease
rs1420901500	L43P	-	Neutral
rs767396990	A59D	Neutral	Neutral
rs777501351	H68R	Disease	Neutral
rs748444579	N83K	Neutral	Neutral
rs1302971614	G105S	Neutral	Neutral
rs1281325781	G116D	Disease	Disease
rs756177951	A132T	Disease	Disease
rs371689839	V183M	Disease	Neutral
rs768748943	T190M	Disease	Neutral
rs773477997	G231R	Disease	Disease
rs1391570201	G430D	Neutral	Disease
rs1412932910	R431C	Disease	Disease

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 1, 2021, Pages. 4895 - 4904
Received 15 December 2020; Accepted 05 January 2021.

Prediction of diseases associated gene

All the SNPs were further analysed for their association with disease using SNPs&GO and PMut.

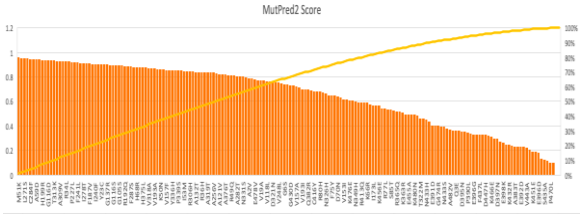

Figure 8: Prediction of nsSNPs effect in HDAC1 using MutPred.

From **Figure 9**, MutPred is used to predict structural character changes as well as functional locations between mutant and wild-type sequences. The modifications are defined as structure and role benefit or loss along with insight into the molecular mechanism associated with disease-related mutations. This database offers precise and robust disease hypothesis based on molecular data for around 11 percent of identified mutations that cause genetic disease.

DISCUSSIONS

In the Geo database, association of gene expression with rheumatoid arthritis was evaluated by the minimum logFC value approach with downregulated will be the negative and the logFC value in upregulated is the positive. At all possible cut-off points, they were then divided into normal and abnormal groups, and the threat differences of any 2 groups were calculated by the log-rank method in the Geo database (okayama et al., 2011). In DAVID server, the functional annotation was used to determine the biological pathway and enrichment score of the genes that were involved from upregulated gene and downregulated gene. The enrichment score indicates that the genes or the protein are over-represented in a large set of genes or proteins and in this case the rheumatoid arthritis. By using Cystoscape, 20 hub genes were found by using CytoscapestringAPP and the steps involved required centiscape2.2 in the apps provided by Cytoscape. The confidence score was set to 0.7 and the maximum additional score is set to 40. Protein-protein interaction were performed by using the STRING server. It is an important ingredient for the system level understanding of cellular process. Network analysis and analysis of the functional enrichment are complementary approaches for obtaining an overview of a long list of genes or proteins. Proteins (singletons) are removed from the simulation without any interaction partners within the network. Only one gene were be selected for SNP analysis which is HDAC1.The human *HDAC1* gene contains a total of 355 SNPs, with 201 of them have the mutation of missense and contig reference were chosen for further analysis. Out of the total 201 nsSNPs screened by and PolyPhen-2, 34 nsSNPs were found to be malignant with the PSIC score ranging from 0.75 to 1.00 (Ramensky et al., 2002). PMut is used to evaluate the missense and contiguous reference mutation as neutral or disease along with score and percentage estimation and MutPred is used to predict changes in structural character as well as functional sites between mutant and wild sequences.

CONCLUSIONS

Future use in the clinic for the diagnosis of rheumatoid arthritis is potential for the discovery of featured genes which were substantially linked to the cell cycle and DNA replication. HDCA1 gene plays an important role in rheumatoid arthritis. In the SNP analysis, it shows that most of the mutation position happen in the gene of HDAC1 have the probability 0.9

Annals of R.S.C.B., ISSN: 1583-6258, Vol. 25, Issue 1, 2021, Pages. 4895 - 4904
Received 15 December 2020; Accepted 05 January 2021.

or higher. Thus, there are about 15 mutation position from the SNP analysis in HDAC1 gene which have maximum impact on HDAC1 structure, stability and functions, which may increase the susceptibility to rheumatoid arthritis.

ACKNOWLEDGEMENTS

The researchers would like to thank Management and Science University for allowing this research.

References

[1] Angelo G., Iker S., Enrique C., Marı́a Teresa A., Rocı́o L., Esther Dı́az 1, Emilio C., Manuel N., Rosario M., Enrique E. , Juan Antonio L. andJuan Angel F. (2013) PTRF/Cavin-1 and MIF Proteins Are Identified as Non-Small Cell Lung Cancer Biomarkers by Label-Free Proteomics, PLoS ONE, doi:10.1371/journal.pone.0033752

[2] Centola M, Cavet G, Shen Y, Ramanujan S, Knowlton N, et al. (2013) Development of a Multi-Biomarker Disease Activity Test for Rheumatoid Arthritis. PLoS ONE 8(4): e60635. doi:10.1371/journal.pone.0060635

[3] Gámez-Pozo A, Sánchez-Navarro I, Calvo E, Agulló-Ortuño MT, López-Vacas R, Díaz E, et al. (2012) PTRF/Cavin-1 and MIF Proteins Are Identified as Non-Small Cell Lung Cancer Biomarkers by Label-Free Proteomics. PLoS ONE 7(3): e33752. https://doi.org/10.1371/journal.pone.0033752

[4] Gary S. Firestein (2003) Evolving concepts of Rheumatoid Arthritis. doi:10.1038/nature01661

[5] H. Okayama, Takashi K., Yuko I., Yoko Shimada, Kouya S., Reika I., K. Furuta, Koji Tsuta, Tatsuhiro S., S. Yamamoto, S. Watanabe, H. Sakamoto, K. Kumamoto, S. Takenoshita, N. Gotoh, H. Mizuno, A. Sarai, S. Kawano, Rui Yamaguchi, Satoru Miyano, and Jun Yokota,(11 November 2011) Identification of Genes Upregulated in ALK-Positive and EGFR/KRAS/ALK-Negative Lung Adenocarcinomas, 10.1158/0008-5472.CAN-11-1403

[6] Hecht M, Bromberg Y, Rost B (2013) News from the protein mutability landscape. J MolBiol 425:3937–3948. https://doi. org/10.1016/j.jmb.2013.07.028

[7] HongyuDiao, Xinxing Li, Sheng Hu, Yunhui Liu (2012) Gene Expression Profiling Combined with Bioinformatics Analysis Identify Biomarkers for Parkinson Disease. PLoS ONE 7(12): e52319. doi:10.1371/journal.pone.0052319

[8] Lu Z.,Song Q., Yang J. , Zhao X., Zhang X., · Yang P., · Kang J., Comparative Proteomic Analysis of Anti-Cancer Mechanism by Periplocin Treatment in Lung Cancer Cells, Cell PhysiolBiochem 2014;33:859-868, https://doi.org/10.1159/000358658

[9] Mansi Desai and J.B. Chauhan (2018), Computational analysis for the determination of deleterious nsSNPs in human MTHFR gene, from journal Computational Biology and Chemistry, https://doi.org/10.1016/j.compbiolchem.2018.02.022

[10] Mansi Desai and JenabhaiChauhan(2016), In silico analysis of nsSNPs in human methyl CpG binding protein 2, from journal Meta Gene,http://dx.doi.org/10.1016/j.mgene.2016.09.004

[11] MiliNailwa and Jenabhai B. Chauhan (2018) Computational Analysis of High-Risk SNPs in Human DBY Gene Responsible for Male Infertility: A Functional and Structural Impact , https://doi.org/10.1007/s12539-018-0290-7

[12] Nadezhda T. Doncheva, John Morris, Jan Gorodkin, and Lars J. Jensen, (19 November 2018) CytoscapeStringApp: Network Analysis and Visualization of Proteomics Data, Journal of Proteome Research, 10.1021/acs.jproteome.8b00702

[13] Peiheng He, Ziji Zhang, Weiming Liao, DongliangXu, Ming Fu and Yan Kang (2016) Screening of gene signatures for rheumatoid arthritis and osteoarthritis based on bioinformatics analysis MOLECULAR MEDICINE REPORTS 14: 1587-1593 , doi: 10.3892/mmr.2016.5423

[14] Screening and Identification of Key Biomarkers in Pancreatic Cancer: Evidence from Bioinformatic Analysis, Journal of Computational Biology,(23 October 2019) DOI: 10.1089/cmb.2019.0189

[15] Screening of gene signatures for rheumatoid arthritis and osteoarthritis based on bioinformatics analysis, (1 August 2016) Molecular Medicine Reports, https://www.spandidos-publications.com/mmr/14/2/1587

[16] Szklarczyk, D., Morris, J. H., Cook, H., Kuhn, M.; Wyder, S., Simonovic, M., Santos, A., Doncheva, N. T., Roth, A., Bork, P.,(2017), The STRING Database in 2017: Quality-Controlled Protein-Protein Association Networks, Made Broadly Accessible. Nucleic Acids Res. 2017, 45 (D1), D362–D368

[17] ToobaYousefi, SeyedMostafa Mir, JahanbakhshAsadi, Mehdi Tourani, AnsarKarimian, MahmoodManiat and SadraSamavarchiTehrani (2019) In silico analysis of non-synonymous single nucleotide polymorphism in a human KLK-2 gene associated with prostate cancer from journal Meta Gene, https://doi.org/10.1016/j.mgene.2019.100578

[18] W.-G. GUO, Y. ZHANG, D. GE, Y.-X. ZHANG, C.-L. LU, Q. WANG, H. FAN , Bioinformatics analyses combined microarray identify the desregulated MicroRNAs in lung cancer ,Departments of Thoracic Surgery, Zhongshan Hospital, Fudan University, Shanghai, China , 2013; 17 ,1509-1516

[19] WeiyingZheng and ShaoqiRao (2015) Knowledge-based analysis of genetic associations of rheumatoid arthritis to inform studies searching for pleiotropic genes: a literature review and network analysis, doi: 10.1186/s13075-015-0715-1

[20] Wenmei Zhang, Wei Gong, Hua Ai, Jie Tang, CeShen, MD, Gene Expression Analysis of Lung Adenocarcinoma and Matched Adjacent Non-tumor Lung Tissue, January 5, 2014, https://doi.org/10.1700/1578.17222