ATLANTIS PRESS

The Gene Expression Profile of AIDS Patients with Tuberculosis Based on Bioinformatics Analysis

Yong-hong Ma[1,2, §], Jiao Tan[1, §,a,*], Ming-juan Shi[1], Jing Lei[1] and Rong-qiang Zhang[3,b,*]

[1]Department of Epidemiology and Health Statistics, College of Public Health, Xi'an Medical University,

Xi'an, Shaanxi, China

[2]Department of Health Statistics, School of Public Health, Fourth Military Medical University, Xi'an, Shaanxi,

China

[3]Department of Epidemiology and Health Statistics, College of Public Health, Shaanxi University of Chinese

Medicine, Xianyang, Shaanxi, China

[a]tanjiao2009@hotmail.com
[b]zhangrqxianyang@163.com

*corresponding author

[§]co-first author

Keywords: AIDS, Tuberculosis, Differentially Expressed Genes, Bioinformatics

Abstract: Objective To analyze the differentially expressed genes in peripheral blood of AIDS patients with TB and explore the biological regulatory mechanism and network of key proteins, providing new evidence for early diagnosis and clinical treatment of AIDS patients combined with tuberculosis. Methods Microarray gene chip data of AIDS patients with TB was downloaded from public databases GEO ans imported into the analysis software STRING, PANTHER, and GenClip. The gene expression profiles, protein interaction networks, the process of molecular biology, gene function were analyzed. Results The expression profiles of 15,529 genes between the two groups of patients were similar; gene expression profiles from 44 subjects were highly correlated; 251 differentially expressed genes had a good diagnostic capabilities in the diagnosis of HIV/TB infection; CDC5L(cell division cycle 5 like)，EIF4E(eukaryotic translation initiation factor 4E)，RBBP7(RB binding protein 7, chromatin remodeling factor) might be the key genes in the diagnosis of HIV/TB infection; differentially expressed genes and positive regulators showed some functions, including cellular function , RNA binding and cellular metabolic process. Conclusions The incidence of HIV/TB infection might closely related to CDC5L, EIF4E and RBBP7, which mainly related to cellular function and RNA binding, cellular metabolic process and other biological functions could play a role in its development, which provide some reference in diagnosis of HIV/TB infection.

1. Introduction

AIDS is an infectious disease which caused by human immunodeficiency virus (HIV) infection. It can severely damage the human immune system and has a high mortality rate [1].The Tuberculosis is a chronic infectious disease caused by tuberculosis bacilli (TB) infection. It mainly affects the lungs and also has a high mortality rate. The early diagnoses of these two infectious diseases have great significance for the active use of reasonable treatment options [2]. However, at the current stage of medical technology, the early diagnosis of TB in HIV-infected individuals is extremely challenging. In recent years, genomics technology has made great progress, which providing an important reference for the diagnosis of TB infection, especially HIV combined TB infection (HIV/TB) [3, 4].

In this study, bioinformatics analysis was performed on a set of microarray data on differentially expressed peripheral blood in HIV- and HIV/TB-infected persons which collected by the Gene Expression Omnibus (GEO), and investigated the molecular network changes in the peripheral blood

of HIV/TB-infected individuals and their possible molecular biological functions, in order to further understand the possible biological relevance of some core genes and HIV, and to help clinical prevention, early diagnosis and treatment of HIV/TB infection.

2. Material and methods

2.1 Data source

Searching in PubMed's GEO database, then using "HIV" as a search term to obtain the GSE50834 genechip data which Dawany N and Showe LC submitted in June 2014. The GSE50834 data was obtained by using the human HumanHT-12 V4.0 gene expression chip platform GPL10558, provided by Illumina Company. The microarray data were derived from a peripheral blood mononuclear cell RNA sample which from a population cohort in South Africa (22 HIV-infected and 22 HIV/TB-infected). The GSE50834 dataset contains a total of 15,529 gene expression data.

Qlucore Omics Ex-plorer 3.1 (QOE) is a new type of bioinformatics analysis software system which analyzed the gene expression pedigrees in both groups that enables comprehensive and rapid gene expression, high-throughput gene chips, real-time PCR (RT-PCR), DNA methylation and other primitive biological databases analysis. In this study, the dataset GSE50834 was imported into the QOE software, and the gene chip data in the data set GSE50834 was normalized (mean=0, S.D=1); after filtered ($P<0.01, q<0.01$) then performed two independent sample t-tests to screen for differentially expressed genes between individuals infected with HIV and those infected with HIV/TB. Then the principal component analysis and hierarchical cluster analysis were performed on the genechip data of data set GSE50834.

The protein-protein interaction map analysis of differentially expressed genes uploads the protein names corresponding to the 200 genes which differentially expressed in HIV-infected and HIV/TB-infected individuals to STRING 9.05 protein-protein interaction online analysis software. And according to the specific analysis results, the credibility and additional node parameters were appropriately adjusted, and the protein-protein interaction maps of two differentially expressed genes were drawn. The top 40 differentially expressed genes in two groups were shown in table 1.

Table 1 Names of the top 40 differentially expressed genes in two groups

No	name of genes	abbreviation	No	name of genes	abbreviation
1	C-C motif chemokine receptor 2	CCR2	21	septin 7	7-Sep
2	methionine sulfoxide reductase B2	MSRB2	22	centromere protein C	CENPC
3	acyl-CoA synthetase short-chain family member 2	ACSS2	23	Heterogeneous nuclear ribonucleo protein H1 (H)	HNRNPH1
4	BR serine/threonine kinase 1	BRSK1	24	dual specificity phosphatase 2	DUSP2
5	acyl-CoA synthetase short-chain family member 2	ACSS2	25	coiled-coil domain containing 91	CCDC91
6	myelin protein zero like 2	MPZL2	26	phosphoinositide-3-kinase regulatory subunit 1	PIK3R1
7	solute carrier family 46 member 2	SLC46A2	27	serine and arginine rich splicing factor 11	SRSF11
8	TBC1 domain family member 2	TBC1D2	28	REL proto-oncogene, NF-kB subunit	REL
9	transmembrane protein 164	TMEM164	29	ATPase phospholipid transporting 11B (putative)	ATP11B
10	purinergic receptor P2X 7	P2RX7	30	DnaJ heat shock protein family (Hsp40) member C2	DNAJC2
11	sorting nexin family member 27	SNX27	31	SUN domain containing ossification factor	SUCO
12	apolipoprotein L3	APOL3	32	CD69 molecule	CD69
13	syntaxin 3	STX3	33	BTG anti-proliferation factor 3	BTG3
14	nuclear receptor subfamily 1 group H member 3	NR1H3	34	REL proto-oncogene, NF-kB subunit	REL
15	methylenetetrahydrofolate reductase (NAD(P)H)	MTHFR	35	protein phosphatase 1 regulatory subunit 15B	PPP1R15B
16	guanylate binding protein 2	GBP2	36	peptidylprolyl isomerase G	PPIG
17	major histocompatibility complex, class II, DM beta	HLA-DMB	37	sphingosine-1-phosphate receptor 1	S1PR1
18	RAB24, member RAS oncogene family	RAB24	38	deoxynucleotidyltransferase terminal interacting protein 2	DNTTIP2
19	glucosamine-6-phosphate deaminase 1	GNPDA1	39	UDP-GlcNAc:betaGal beta-1,3-N-acetylglucosaminyltransferase 5	B3GNT5
20	branched chain ketoacid dehydrogenase kinase	BCKDK	40	G protein-coupled receptor 183	GPR183

2.2 Go functional enrichment analysis of the differentially expressed genes

Upload the list of the most significant differential genes to human gene function and network analysis software GenCliP 2.0. The main parameters of each type of analysis were adjusted according to need, GO function enrichment status of the differentially expressed genes involved in the two groups of patients were analyzed.

3. Results

3.1 The differentially expressed genes for the diagnosis of HIV/TB infection

After the QOE software standardized the original data, the two independent t-test were used to found that there were a total of 193 differentially expressed genes (P<0.05); it showed that differentially expressed genes have a good diagnostic ability to diagnose HIV/TB infection. The result of the preprocessing of the original data was shown in Figure 1. The GSE50834 original data was standard normalization in the top 40 genes of two groups, and we can see after normalization, the average value of the gene expression profile of each sample was a straight line, which indicating that the data was stable and can provide reliable conclusions for follow-up analysis.
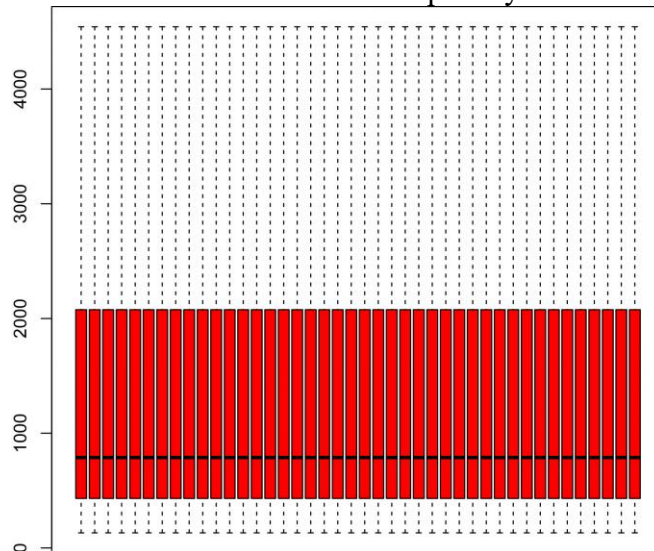

Figure 1 Standard normalized box map for GSE50834 original data

3.2 The Differentially expressed genes for the two groups of HIV/TB infection

The differentially expressed genes analysis results of volcanic maps of the two infection groups showed in figure 2. In the end, a total of 193 differentially expressed genes were found. Compared with the simple HIV infection group, there were 122 genes which expressed decreased in the HIV-infected combine tuberculosis group and 71 genes expressions increased.
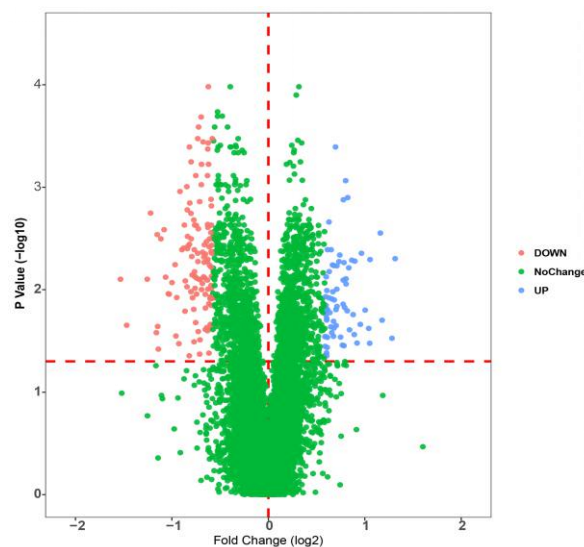

Figure 2 Volcanic maps of differentially expressed genes

3.3 Results of GO function enrichment analysis

The GenClip online software was used to analyze the GO function enrichment status of the biofunctional histograms. Through GO enrichment analysis of differentially expressed genes, it was found that most key genes involved in cellular component function, include intracellular, many kinds of organelle. The main process involved in the biological process is RNA binding; the molecular function was associated with primary metabolic process, cellular metabolic process, and macromolecule metabolic process. It can be seen that differential genes may be mainly related to cellular function and RNA binding, cellular metabolic process. The changes in cellular function have the greatest impact on the group of AIDS combine tuberculosis. The biofunctional histogram was showed in Figure 3.
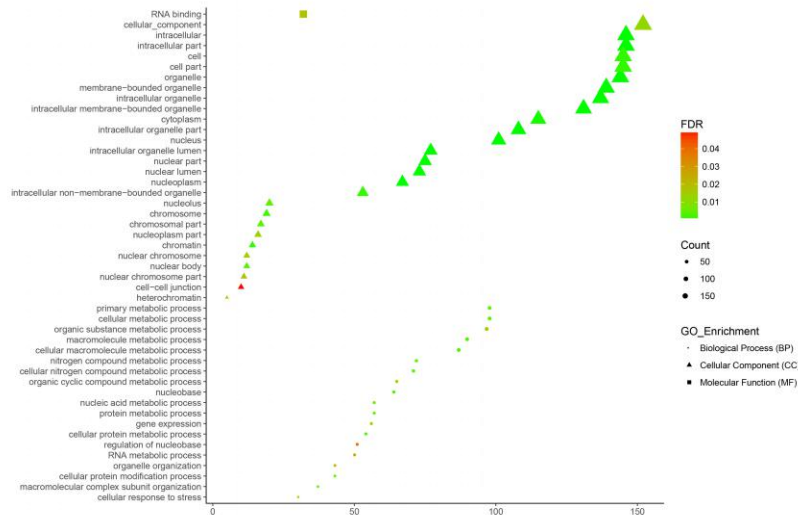

Figure 3 Biofunctional histograms of GO enrichment analysis

3.4 Protein-protein interaction results of top 200 differentially expressed genes

In this study, the first 200 differentially expressed genes were selected for protein-protein interaction analysis. The protein names corresponding to the 200 genes were imported into the STRING 9.05 software to analyze the protein-protein interaction network and predict the protein inter-actions. The results are presented in Fig. 4. It was showed that PPI network was widened by adding 53 partner proteins. The interaction of nodes of >5 was considered as the key central node in the network in Fig. 4. Among them, there were three core genes which connect those proteins, CDC5L（cell division cycle 5 like），EIF4E（eukaryotic translation initiation factor 4E），RBBP7（RB binding protein 7, chromatin remodeling factor）。Through the PPI analysis network, four related genes can be predicted：TAF9（TATA-box binding protein associated factor 9），CENPC1（centromere protein C1），KIAA0947（KIAA0947 ortholog），RPL7A（ribosomal protein L7a）.The genes in the PPI network were mainly associated with cell division and protein binding.
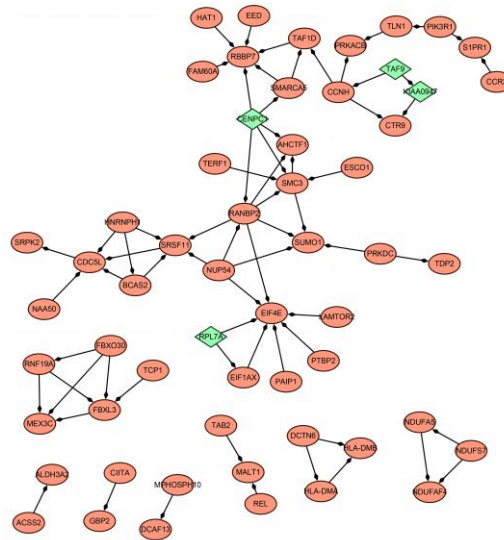
Figure 4 Network of protein-protein interactions for Top 200 differentially expressed genes

4. Discussion

Tuberculosis is the most common contagious infectious disease among HIV-infected persons (including AIDS patients) and also is an ordinary cause of death in this type of patients. HIV and TB can interact in patient's body and have an important impact on the patients, manifesting as: the patient can rapidly develop into an active tuberculosis patient; the chance of re-infecting TB is increasing because of the destruction of the patient's immune system, and definitive diagnosis is getting more and more difficult. Epidemiological studies have shown that HIV is a driving force for the promotion of TB epidemic, and TB is the most common opportunistic infection in AIDS patients [5-8]. At present, AIDS is prevalent in the world, and HIV/TB infection has become an increasingly prominent problem especially in Africa. The prevention, diagnosis and treatment of HIV/TB have attracted more and more attention of medical personnel.

Clinically, the diagnosis of HIV/TB infection mainly relies on clinical comprehensive diagnosis, but there are great difficulties. Bacterial culture is the "gold standard" for the diagnosis of tuberculosis including HIV/TB infection, but it takes a long time to test and the positive rate is low. In recent years, the diagnostic technology of molecular biology has made great progress. However, in the actual clinical work, the equipment, technical capabilities, and reagent consumables have high requirements and cannot be widely promoted in practical work. Therefore, the definitive diagnosis of HIV/TB infection needs to be further studied. Finding reliable diagnostic techniques and widely applicable in various hospitals are the common aspiration of medical workers. With the development of genomics research methods, many research results and data are also available in the field of HIV/TB infection. Using bioinformatics technology to dig deeper into these data may have important clinical value for the early diagnosis of HIV/TB infection [9-11].

In this study, molecular biology techniques were used to conduct a detailed excavation of microarray data from a panel of HIV-infected (n=22) and HIV/TB-infected (n=22) from an African cohort. The results of gene profile suggested that the 15 529 gene expression patterns in the two groups were highly consistent, which may be the genetic basis that HIV/TB infection individuals were difficult to diagnostic. And it also suggested that it was necessary to perform deeper excavation of the two genes expression data. Using QOE online analysis software to perform the principal component analysis of gene expression patterns of the two infection groups, it showed that the gene expression patterns of the two groups were actually significantly different. A total of 200 differentially expressed genes were found, and the distinction between HIV infection and HIV/TB infection could be well distinguished. The reason may be that after infect HIV and combine with TB bacteria, the patient's immune system was inhibited, which affects the synthesis of certain important

genes or the key link of protein degradation, and thus affects the expression level of DNA [12-14]. The protein-protein interaction network of the first 40 genes differentially expressed in the two groups was mainly involved in protein synthesis. CDC5L, EIF4E and RBBP7 are the core nodes of this protein network, which suggested that they may have some clinical value in distinguishing between two types of patients. The genes in the PPI network were mainly associated with cell division and protein binding. The pathogenesis of AIDS is not yet clear, but it will have similar changes in the immune function of patients with SLE, such as T cell abnormalities. Currently, there was evidence that HIV can induce autoimmune diseases such as AIDS and SLE. The results of the above studies indicate that it may play a role in HIV retrovirus-induced autoimmune diseases such as SLE and AIDS, but further exploration is needed [15]. The relationship between CDC5L, EIF4E and RBBP7 genes and pulmonary tuberculosis has not been reported yet, however the results of this study suggest that: CDC5L, EIF4E and RBBP7 gene may has closely association with HIV co-infected TB bacteria.

The results of gene biological function analysis showed that: The first 40 genes differentially expressed in the two infection groups were mainly involved in the formation of organelles, transduction of signal pathways, protection of the immune system, and other biological processes; and involved in free radical-related apoptosis pathways, inflammatory Reaction pathways and other signal pathways. It suggested that after HIV infection combine with TB pathogens infected, the key factors of the body genetic regulatory network activate the apoptosis-related signaling pathways, inflammation-related signaling pathways, and affect the body's signal transduction ability, immune system status, then presenting different clinical symptoms meanwhile affecting the prognosis. Through the literature excavation, most of the analysis results of this study can find similar evidence in the currently published literature.

In summary, in our study, bioinformatics methods were used to analyze the expression of 15,529 genes in peripheral blood mononuclear cells from 22 HIV-infected and 22 HIV/TB infected individuals in South Africa. A total of 251 differentially expressed genes were found which can be distinguished between HIV infection and HIV/TB infection individuals. The CDC5L, EIF4E and RBBP7 genes contained in the protein-protein interaction network of the differentially expressed 40 genes may be closely associated with HIV-infected TB; the differential genes of the two groups are mainly involved in cell division and protein binding, which provide a scientific basis for the early diagnosis of HIV combine TB infection. In further studies, the gene expression profile of peripheral blood cells should be added to the normal population. And more detailed bioinformatics analysis may yield even more comprehensive and thorough analysis results.

Conflicts of Interest
The authors declare no competing financial interests.

References
[1] APIDECHKUL T. A 20-year retrospective cohort study of TB infection among the Hill-tribe HIV/AIDS populations, Thailand [J]. Bmc Infectious Diseases, 2015, 16(1): 1-14.
[2] KASSA D, DE J W, GEBREMICHAEL G, et al. The effect of HIV coinfection, HAART and TB treatment on cytokine/chemokine responses to Mycobacterium tuberculosis (Mtb) antigens in active TB patients and latently Mtb infected individuals [J]. Tuberculosis, 2016, 96:131-40.
[3] Montales M T, Chaudhury A, Beebe A, et al. HIV-Associated TB Syndemic: A Growing Clinical Challenge Worldwide[J]. Front Public Health, 2015,3:281

[4] LAWN S D, KERKHOFF A D. Rapid diagnosis of TB in HIV-positive in-patients with M. tuberculosis bacteraemia in sub-Saharan Africa [J]. International Journal of Tuberculosis & Lung Disease the Official Journal of the International Union Against Tuberculosis & Lung Disease, 2015, 19(12): 1557-1559.

[5] MATABANE M M Z, ISMAIL F, STRYDOM K A, et al. Performance evaluation of three commercial molecular assays for the detection of Mycobacterium tuberculosis from clinical specimens in a high TB-HIV-burden setting [J]. Bmc Infectious Diseases, 2015, 15(1): 15:508..

[6] NDAGIJIMANA A, RUGIGANA E, UWIZEYE C B, et al. One-stop TB-HIV services evaluation in Rwanda: comparison of the 2001-2005 and 2006-2010 cohorts [J]. Public Health Action, 2015, 5(4): 209-213.

[7] SKOGMAR S, SCH N T, BALCHA T T, et al. Plasma Levels of Neopterin and C-Reactive Protein (CRP) in Tuberculosis (TB) with and without HIV Coinfection in Relation to CD4 Cell Count [J]. Plos One, 2015, 10(12): e0144292.

[8] D'ELIA A, EVANS D, MCNAMARA L, et al. Predictive and prognostic properties of TB-LAM among HIV-positive patients initiating ART in Johannesburg, South Africa [J]. Pan African Medical Journal, 2015, 22:4

[9] FORD N, GETAHUN H. Service integration to reduce HIV-associated TB mortality [J]. Public Health Action, 2015, 5(4): 204.

[10] MEKONNEN D, DERBIE A, DESALEGN E. TB/HIV co-infections and associated factors among patients on directly observed treatment short course in Northeastern Ethiopia: a 4 years retrospective study [J]. Bmc Research Notes, 2015, 8(1): 1-6.

[11] NGLAZI M D, BEKKER L G, WOOD R, et al. The impact of HIV status and antiretroviral treatment on TB treatment outcomes of new tuberculosis patients attending co-located TB and ART services in South Africa: a retrospective cohort study [J]. Bmc Infectious Diseases, 2015, 15(1): 536.

[12] PIGGOTT D A, KARAKOUSIS P C. Timing of Antiretroviral Therapy for HIV in the Setting of TB Treatment [J]. Clinical & Developmental Immunology, 2011:1740-2522.

[13] FISEHA T G T. Tuberculosis Treatment Outcome among HIV Co-infected Patients at Mizan-Aman General Hospital, Southwest Ethiopia: A Retrospective Study [J]. Journal of Bioengineering & Biomedical Science, 2015, 05(1):1

[14] FORD N, MATTEELLI A, SHUBBER Z, et al. TB as a cause of hospitalization and in-hospital mortality among people living with HIV worldwide: a systematic review and meta-analysis [J]. Journal of the International Aids Society, 2016, 19(1): 207-214.

[15] GLAZIOU P, FLOYD K, WEIL D, et al. TB deaths rank alongside HIV deaths as top infectious killer [J]. International Journal of Tuberculosis & Lung Disease the Official Journal of the International Union Against Tuberculosis & Lung Disease, 2016, 20(2): 143-144.