Infection, Genetics and Evolution 46 (2016) 118–123

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

Novel tuberculosis susceptibility candidate genes revealed by the reconstruction and analysis of associative networks

Elena Yu. Bragina [a,*], Evgeny S. Tiys [b,c], Alexey A. Rudko [a], Vladimir A. Ivanisenko [b], Maxim B. Freidin [a]

[a] Laboratory of Population Genetics, Research Institute of Medical Genetics, Tomsk NRMC, Nabereznaya Ushaiki Str. 10, Tomsk 634050, Russia
[b] Laboratory of Computer-Assisted Proteomics, The Federal Research Centre Institute of Cytology and Genetics of The Siberian Branch of the Russian Academy of Sciences, Lavrentiev Ave. 10, Novosibirsk 630090, Russia
[c] Laboratory of Computer Genomics, Novosibirsk State University, Pirogova Str. 2, Novosibirsk 630090, Russia

ARTICLE INFO

Article history:
Received 2 July 2016
Received in revised form 25 October 2016
Accepted 30 October 2016
Available online 31 October 2016

Keywords:
Tuberculosis
Susceptibility
Genes
Associative network

ABSTRACT

Tuberculosis (TB) is a common infectious disease caused by *M. tuberculosis*. The risk of the disease is dependent on complex interactions between host genetics and environmental factors. Accumulated genomic data, along with novel methodological approaches such as associative networks, facilitate studies into the inherited basis of TB. In the current study, we carried out the reconstruction and analysis of an associative network representing molecular interactions between proteins and genes associated with TB. The network predominantly comprises of well-studied key proteins and genes which are able to govern the immune response against *M. tuberculosis*. However, this approach also allowed us to reveal 12 proteins encoded by genes, the polymorphisms of which have never been studied in relation to *M. tuberculosis* infection. These proteins include surface antigens (*CD4*, *CD69*, *CD79*, *CD80*, *MUC16*) and other important components of the immune response, inflammation, pathogen recognition, cell migration and activation (*HCST*, *ADA*, *CP*, *SPP1*, *CXCR4*, *AGER*, *PACRG*). Thus, the associative network approach enables the discovery of new candidate genes for TB susceptibility.

1. Introduction

The study of the mechanisms of human susceptibility to infections caused by pathogenic microorganisms such as *M. tuberculosis* is an important and productive field of human genetics. It aims to reveal genetic variants predisposing to tuberculosis (TB) in order to facilitate the development of new effective measures to combat the disease. For this reason, both candidate genes and genome-wide association studies are usually carried out. The results of these studies confirm the importance of host genetic components in the risk of contracting TB; however, their reproducibility is poor and the convincing examples of causal genes are solitary.

Contemporary technologies and tools for the analysis of the human genome, allow for detailed estimations of the significance of genetic variability in the development of infectious diseases (Meyer and Thye, 2014). However, the understanding of gene function is impossible without systems biology, which, if applied to human pathology, offers a unique approach for studying multifactorial diseases. Systems biology allows revealing molecular links between diseases and identifies shared genes and pathways to achieve better understanding of the molecular origin of different disease groups. For instance, such the integrative analysis helped discover that leprosy susceptibility genes are also crucially important for the development of autoimmune diseases (Zhang et al., 2016). In systems biology methodology, a network is a central object that represents links between multiple objects (Koonin, 2011). In gene networks, the nodes represent genes and the connections are their interactions (Barabási and Oltvai, 2004). Thus, a gene network is a group of genes functioning in a coordinated fashion, providing control for vital functions in an organism (Kolchanov et al., 2000). Studying the patterns of functioning of such networks may lead to the identification of genes responsible for the development of diseases. For instance, a number of studies using gene network methodologies have revealed novel genes that are important for the development of breast cancer (Pujana et al., 2007), inherited ataxia (Lim et al., 2006), asthma (Hwang et al., 2008), diabetes mellitus (Kussmann et al., 2013; Liu et al., 2007), neurodegenerative diseases (Parikshak et al., 2015), and cardiovascular diseases (Shangguan et al., 2014). Moreover, it is apparent that a breakthrough in the analysis of pathological processes and the development of effective therapeutics for complex diseases (including infectious disorders) is impossible without using the knowledge about functioning of molecular networks.

The wealth of experimental data in molecular biology provides a strong basis for the development of methods for the reconstruction of gene networks based on the automatic analysis of scientific texts and databases (Rzhetsky et al., 2004). One of the most popular softwares

* Corresponding author.
E-mail address: elena.bragina72@gmail.com (E.Y. Bragina).

http://dx.doi.org/10.1016/j.meegid.2016.10.030
1567-1348/

for such analysis is STRING (Szklarczyk et al., 2015) which uses "text-mining" technology for the reconstruction of networks. Another software, ANDSystem, was developed for automatic extraction of knowledge about molecular genetic interactions between proteins, genes, mRNA, metabolites, biological processes and diseases, from the texts of scientific publications, and their presentation in the form of associative semantic networks. Unlike STRING, ANDSystem provides more detailed descriptions of interactions between objects (Ivanisenko et al., 2015). The ANDSystem has been used to reconstruct associative networks for myopia and glaucoma (Podkolodnaya et al., 2011) as well as to describe molecular genetic interactions for co-morbid (Glotov et al., 2015) and inverse co-morbid pathological conditions (Bragina et al., 2014).

In this study, we set out to construct an associative network for TB using the ANDSystem software to reveal key molecules driving the disease pathogenesis.

2. Material and methods

We used the ANDSystem software to reconstruct an associative network for pulmonary TB. Extraction of the data by ANDSystem is carried out by applying a text-mining algorithm to the ANDCell knowledge base, followed by the visualisation of the results by the ANDVisio software (Ivanisenko et al., 2015).

We used a "tuberculosis pulmonary" search query to build the network, followed by a re-assessment through the expert review to remove the network redundancy caused by the incomplete formalization of the texts of research articles. The following parameters were taken into account for the expert review: 1) participation of a protein in TB pathogenesis; 2) correct protein/gene name recognition; 3) correct context (protein/gene mentioned in the abstract is really associated with the disease); 4) nominal statistical significance ($p < 0.05$) for the association between the protein and TB. After the review, the associative network was shrunk to contain only the most essential objects and their interactions.

We used the public databases Ensembl (http://www.ensembl.org/index.html), NCBI (http://www.ncbi.nlm.nih.gov/), and HuGE Navigator (https://phgkb.cdc.gov/HuGENavigator) to find out whether there are known polymorphisms of the genes in the network associated with TB or immune-mediated diseases.

Gene ontology-based overrepresentation analysis was conducted using the BiNGO program with the following parameters: hypergeometric test; Benjamini-Hochberg FDR-correction; significance level = 0.05; ontology = go-basic.obo (data-version: releases/2016-03-23); annotation = UniProt-GOA (Submission Date: 3/16/2016).

3. Results and discussion

Using the ANDSystem software we compiled a list of 131 different proteins and genes associated with TB according to data published in scientific literature. After the expert review, the list was reduced to the 40 genes/proteins most essential in TB pathogenesis (Supplementary Table 1; Fig. 1).

In the associative network, genes and proteins are separate components (Fig. 1). The network comprises 40 genes and their 45 protein products. The links between genes and the proteins they encode are defined by the type "expression". This link type is assessed in case of the presence of a Gene Id link in SwissProt card for the protein. The total number of proteins exceeds the total number of genes in the network because for *HLA-A* and *HLA-DRB1* genes there were 7 proteins; MHC

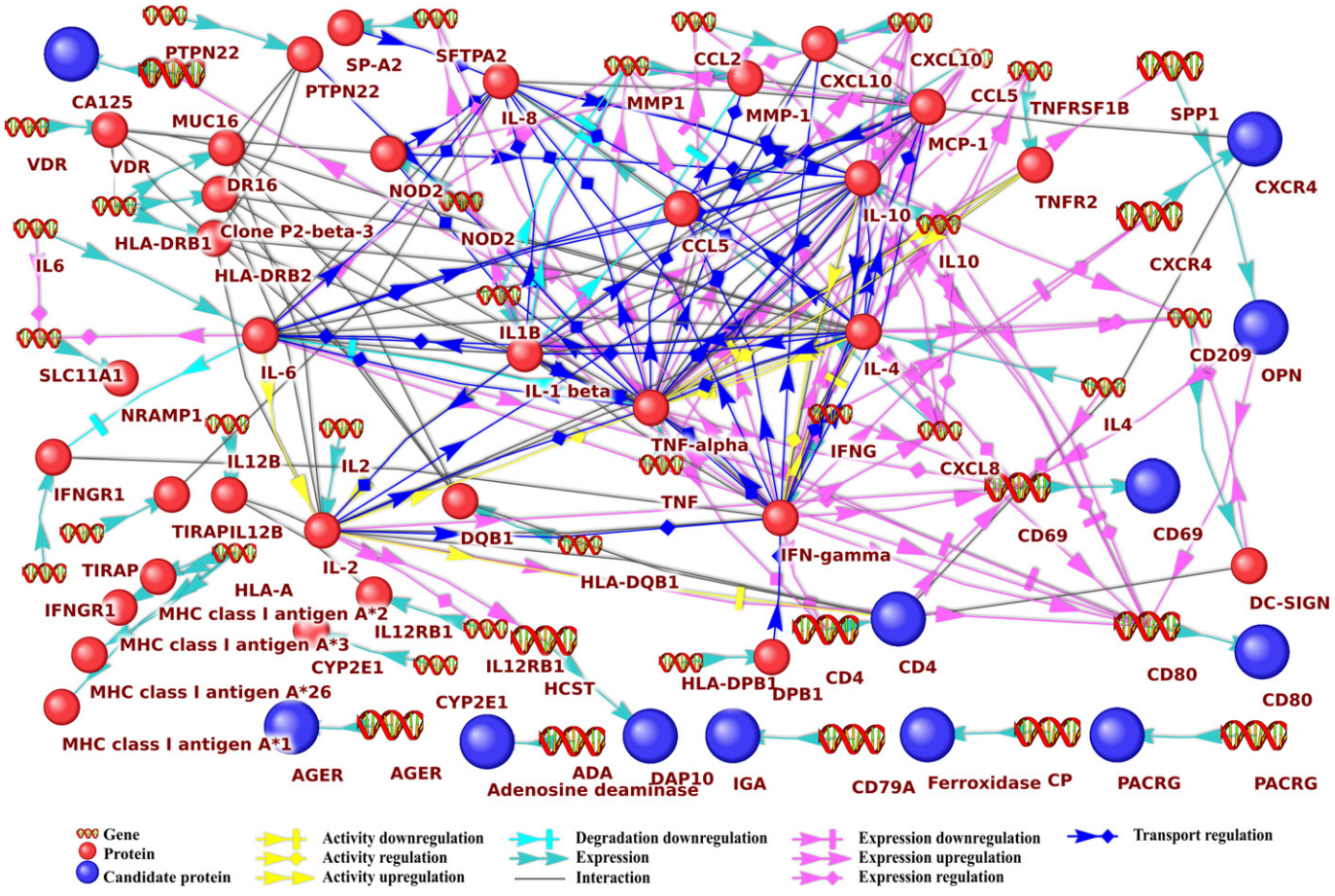

Fig. 1. Molecular genetic network of functional interactions between human genes/proteins associated with tuberculosis (associative network).

class I antigens A*1, A*2, A*3, and A*26 and MHC class II antigens DRB1*15, DRB1*16, DRB1*3.

The network comprises of 274 interactions of many different types. The most abundant interaction type is "expression regulation" (n = 60). This connection type describes activation (upregulation) or inactivation (downregulation) of a gene under the control of a regulatory protein. For instance, tumour necrosis factor (TNF) protein and *MMP-1* gene are linked by "expression upregulation" connection type, while TNF protein and *COL1A1* and *COL3A1* are linked by "expression downregulation". The connections have been automatically extracted by the ANDSystem from PubMed (Zeldich et al., 2010; Sato et al., 1998).

The second most abundant type of interaction in the network is protein-protein interaction (PPI) with a total of 54 links in the tuberculosis associative network. For 46 of these interactions, ANDSystem specified PubMed database alone as the source of information, while for the rest HPRD, DIP and innatedb were additionally used.

"Transport/release regulation" connection type was also relatively common in the network, with 45 protein pairs linked via this type. This connection type is used by the ANDSystem for the description of regulation of transport or release of a protein under the action of another protein. For example, the connection established between TNF-α and MCP-1 according to PubMed publications (Bouchelouche et al., 2006; Kim et al., 2010).

In the network, there are 6 objects comprised of genes/proteins with no direct connection with other network participants. They include *HLA-A*, *CD79A*, *CYP2E1*, *CP*, *PACRG*, and *AGER* genes. Despite such "isolation" of these objects, they exhibit functional interactions with other members of the network indirectly, via intermediates for which no information about association with TB was known. For instance, the IGA protein encoded by *CD79A* gene interacts with 36 intermediary proteins which, in turn, interact with 23 proteins in the tuberculosis associative network. According to the ANDSystem, IGA protein participates in a direct interaction between HS1 (hematopoietic cell–specific Lyn substrate 1) which is involved in regulation of TNF-α and also interacts with IL-2 protein. The total number of intermediary participants connecting *HLA-A*, *CD79A*, *CYP2E1*, *CP*, *PACRG*, and *AGER* with the associative network was comparable with the total number of the main network members (>100). These mediators are hidden to facilitate the visualisation of the network. Thus, we can assume that all the participants of the tuberculosis associative network are functionally tightly inter-connected.

The vast majority of genes in the associative network have already been studied in terms of associations with the risk of TB; however, for 12 genes, such studies have never been conducted according to the publicly available data (Supplementary Table S1). These genes can therefore be defined as novel candidate genes for tuberculosis susceptibility studies.

Proteins and genes in the network can be classified into functional categories such as immune response, regulation of transcription, xenobiotics metabolism/detoxification.

Immune response genes and proteins in the network are currently the most well studied with regards to TB and other infectious diseases, due to the predomination of the candidate genes approach in the study of these disorders. Antigen recognition, macrophage activation, granuloma formation, inflammation, and overall efficacy of immune response against *M. tuberculosis* are crucially important in the pathogenesis of TB, and are dependent on the genetic make-up of the host. Numerous genes involved in the immune response to TB have been studied, which is reflected in the obtained associative network.

Among the immune response genes presented in the network, genes encoding human leukocyte antigens (*HLA*) are known to have the strongest association with TB. Another group of immune response genes encode cytokines and their corresponding receptors, such as *IL1B*, *IL2*, *IL4*, *IL6*, *IL8*, *IL10*, *IL12B*, *TNFA*, *IFNG*, *IFNGR1*, *IL12RB1*, and *TNFRSF1B*, regulating many aspects of the immune response and inflammation, are known to be associated with TB, too. Also in the associative network, there are genes encoding pattern recognition receptors of macrophages and dendritic cells, such as Toll-like receptors and NOD-receptors (*TIRAP* and *NOD2*), which activate signals inducing cytokine production.

At the same time, some immune response genes/proteins in the associative network have never been studied in terms of their association with TB, despite their obvious importance in the pathogenesis of the disease. For instance, the *HCST* (*DAP10*) gene encodes a transmembrane signalling adaptor containing the YxxM motif in its cytoplasmic domain. The DAP10-signal-subunit-deficient mice infected by *M. tuberculosis* exhibit decreased cytotoxicity of CD8 + T-cells due to an impaired release of cytotoxic granules (Hessmann et al., 2011). Thus, this gene is of interest not only for the study of susceptibility to TB, but also for the study of the mechanisms of the development of specific clinical features of the disease.

Another understudied gene *ADA* encodes adenosine deaminase enzyme, which converts adenosine to inosine. Even though adenosine does not influence TB pathogenesis directly, it is involved in the IL-6 signalling cascade important for TB pathogenesis (Law et al., 1996). High levels of this enzyme were detected in the most severe TB patients with remarkable destruction and infiltration (Balasaniants et al., 2001). The genomic region in which the *ADA* gene resides (20q13.12) was found to be associated with TB (Stein et al., 2008).

The *CP* gene encodes ceruloplasmin, a metalloprotein which binds blood cuprum and participates in the oxidation of the Fe(II)-transferrin to Fe(III)-transferrin. Possibly, antibacterial function of ceruloplasmin is based on its participation in the homeostasis of iron and copper; important elements of antibacterial defense (Neifakh et al., 1969; Karyadi et al., 2000). Ceruloplasmin is an acute phase protein; its concentration, along with copper ion levels is elevated in lung TB patients (Cernat et al., 2011). *CP* gene expression is significantly different between TB patients with low and high content of iron and transferrin saturation index in blood serum (Cowie, 2014). Similar to SLC11A1, a protein responsible for divalent metal cations transport (Cellier et al., 2001), defects in the production or function of *CP* may result in impairment of its binding and transport functions; thus leading to elevated sensitivity to intracellular pathogens, such as mycobacteria. Binding of iron ions is important for the suppression of replication of mycobacteria; therefore, genes involved in this biological process are of interest as potential drug targets for TB treatment. Again, no studies of association between this gene variants and TB are available.

The *CD69* gene encodes type II transmembrane glycoprotein. In invariant NKT type I cells in TB patients and HIV-infected patients, an increased expression of the *CD69* gene is observed (Montoya et al., 2008). To date, no studies of the polymorphisms of *CD69* gene in relation to TB have been published.

In the reconstructed associative network, the *SFTPA2B* gene is known to be associated with TB in India (Madan et al., 2002) and Ethiopia (Malik et al., 2006); however, there are no data so far on its association with the disease and its clinical forms in Europeans. The studies of these molecules are promising taking into account the wide spectrum of immune-modulating effects of cell surface molecules.

The *AGER* gene product is a member of a superfamily of cell surface immunoglobulin molecules, which acts as receptors for various molecules participating in inflammation, development and homeostasis. Enhanced expression of *AGER* was reported in lung TB patients (Arce-Mendoza et al., 2008), but there are no studies to date which have shown polymorphisms of the *AGER* gene to be associated with TB and its clinic features.

Osteopontin encoded by the *SPP1* gene is a pro-inflammatory cytokine predominantly involved in Th1-type immune response. Differential expression of the *SPP1* gene is found in cells infected by *Mycobacteria* and non-infected cells (Wang et al., 2003), thus suggesting that *SPP1* can be considered as a promising candidate gene for the development of TB.

The analysis of the associative network revealed several clusters of differentiation molecules such as *CD4*, *CD79A*, *CD80*, *MUC16*, involved

in TB pathogenesis coded for by genes which have not previously been studied in relation to the disease.

The *CD4* gene encodes the membrane glycoprotein of T-lymphocytes which interacts with *HLA* class II antigens and also serves as a receptor for HIV. The gene is expressed in T- and B-lymphocytes, macrophages and granulocytes and plays an important role in Th-mediated immunity. It is important for preventing the growth of *M. tuberculosis* and in the development of granulomatous inflammation in the lungs of infected knockout mice (Saunders et al., 2002).

The *CD79A* gene encodes the Ig-α protein of the B-cell antigenic complex and is associated with immune-deficient pathological conditions (Wang et al., 2002). The protein is essential for immunopathogenesis of TB and is an important marker of the course of TB, as its local levels in the lungs are elevated in patients with active disease (Goyal et al., 1990).

B-lymphocyte activation antigen B7-1 encoded by the *CD80* gene is one of the co-stimulatory molecules which, when its expression is decreased, results in the impairment of immunological reactivity of T-lymphocytes. *CD80* was shown to be important for anti-tuberculosis immunity (Bhatt et al., 2009).

Glycoprotein *MUC16* is a mucin specifically expressed in uterus epithelia (Hattrup and Gendler, 2008). Its levels are significantly increased in TB patients (Fortún et al., 2009).

In the associative network, there are also some genes that have been tested for an association with TB only in a very limited number of published studies. Among them, there is *PTPN22* which encodes a tyrosine phosphatase specific for immune cells. A polymorphism of Parkin Co-Regulated Gene Protein (*PACRG*) is associated with the development of leprosy, an infectious disease caused by *M. leprae*, another representative of *Mycobacterium* genus (Mira et al., 2004). The protein *PACRG* also participates in ubiquitin-mediated protein degradation, a biochemical pathway which so far has not been well studied with regard to infectious disease.

Mutations in the gene encoding Chemokine (C-X-C Motif) Receptor (*CXCR4*) are associated with WHIM syndrome (MIM #193670). The expression levels of *CXCR4* in peripheral blood are significantly different in TB patients and healthy individuals (Zhao et al., 2015).

Little is known about the impact of variants of the *CTLA4* gene encoding cytotoxic T-lymphocyte associated antigen on the predisposition to TB. The respective protein inhibits T-cells and leads to a bias of the immune response towards Th1-type (Nasta et al., 2006). A haplotype of three polymorphisms of the *CTLA4* gene (+49A>G, +6230G>A, and 11430G>A) has been found to be associated with the severity of TB (Wang et al., 2012); however, no further data have been published so far.

Genes encoding xenobiotic-biotransformation enzymes, represented in the associative network, are of significant interest for studies into the sensitivity and toxicity of anti-tuberculosis drugs (Sheng et al., 2014). However, they are totally understudied in this capacity. Meanwhile, for some of these genes, e.g. *CYP2E1*, an association with infiltrative TB has been shown (Bikmaeva et al., 2004).

For the purpose of ranking the understudied genes as candidates for associative studies, we carried out an assessment of the level of their functional connection with genes with known polymorphisms associated with TB (called reference nodes). The levels of functional connection were assessed using two statistics—centrality (p1) and specificity (p2). Centrality parameter is estimated as the total number of edges in a network between candidate and reference nodes. Specificity is estimated as a ratio between centrality and the total number of edges between a candidate gene/protein and all protein/genes presented in global ANDSystem network (Table 1). In other words, the centrality parameter characterises the quantity of connections with reference genes, while the specificity parameter shows a proportion of these connections among all links of the analysed candidate. We assume that the higher p1 and p2, the greater the probability of discovering associations between polymorphisms of the candidate gene and TB.

Based on this analysis, according to p1 value, the most plausible candidate for subsequent investigation was *CD80*. This gene was connected with six reference proteins (IL-2, IL-4, IL-6, IL-10, IFN-γ, and TNF-α) which can be involved in regulation of its expression, according to the ANDSystem data. Moreover, the p2 value for these genes exceeded the figures for all other candidates except for *HCST*.

Table 1
Centrality and specificity parameters for candidate genes.

Gene/protein	Centrality (p1)	Global degree	Specificity (p2)
CD80	6	312	1,92E-02
HCST	1	52	1,92E-02
CD69	5	285	1,75E-02
MUC16	1	228	4,39E-03
CD4	5	1333	3,75E-03
SPP1	1	722	1,39E-03
CXCR4	1	738	1,36E-03

ADA, *AGER*, *CD79A*, *CP*, and *PACRG* are not shown as they have 0 centrality due to the lack of direct connections with the associative network members.

Gene Ontology analysis was carried out for genes that have not yet been studied for association with TB, using the BiNGO tool. Overall, 529 biological processes were detected (Supplementary Table S2). Among these processes, the following groups could be revealed: adaptive immune response (*ADA*, *CD4*, *AGER*, *CD80*), response to nutrient (*ADA*, *CP*, *CD4*, *SPP1*, *AGER*), respiratory system development (*ADA*, *CP*, *AGER*), apoptosis and cell death (*ADA*, *AGER*, *PACRG*, *CXCR4*), regulation of protein phosphorylation (*CD4*, *AGER*, *CD80*, *CXCR4*), and neurogenesis (*SPP1*, *AGER*, *CXCR4*). Interestingly, in addition to the immune response which has been actively studied, other specific processes were found. It is known that not only chemotherapy but also therapeutic diets are used for tuberculosis therapy, since poor nutrition leads to decreased immunity and, consequently, to the development of tuberculosis (Gupta et al., 2009). Identification of the "response to nutrient" group of processes among the overrepresented ones seems to be associated with diet therapy of patients with tuberculosis and is not directly related to the disease itself. Overrepresentation of the development process of the respiratory system was probably due to the active recovery of the latter after damage by *M. tuberculosis*. Of particular note is the regulation of apoptosis in various cell types. On the one hand, apoptosis can be an attacking mechanism of the pathogen or a defense mechanism of the host (Zychlinsky and Sansonetti, 1997); on the other hand, it serves for selecting T-lymphocytes in the thymus (Starr et al., 2003). In addition to the positive regulation of apoptosis, its downregulation was also found to be carried out by *M. tuberculosis* to prevent death of the infected host cells (Velmurugan et al., 2007).

4. Conclusion

The mechanisms which underlie the pathogenesis of TB are complex and not fully understood, thus preventing and complicating the development of effective strategies for treatment and prophylaxis of the disease. Normal functions of macrophages and Th1-immunity are important to deter the mycobacterial infection, but are not enough to eliminate the pathogen from the host organism. Mycobacteria can remain latent in granulomas for decades. Adaptive immunity induced by BCG vaccination is not sufficient to prevent latent TB activation, and the mechanisms of the activation are not fully understood (Lin and Ottenhoff, 2008).

Using the reconstruction of the associative network allowed us, from the body of available published data, to pick out proteins and genes most significant for the development of the infectious process. Of note, the reconstruction helps identify not only proteins which genes are well studied regarding to the development of TB, but also proteins which are important for the disease development and coded by genes which have never been studied in terms of associations with TB. Thus, the obtained results give the reference point for focused experimental and associative studies of the molecular basis of the development of TB and different clinical forms of the disease.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.meegid.2016.10.030.

Conflict of interest

The authors declare no conflict of interest.

Acknowledgments

The analysis of the associative network of TB was financially supported by the grant from Russian Science Foundation "Genetic factors of susceptibility to different forms of tuberculosis infection" (grant number 15-15-00074).

The analysis of relation between TB pathogenesis and apoptosis was financially supported from Russian Science Foundation grant "Programmed cell death induced via death receptors: Delineating molecular mechanisms of apoptosis initiation via molecular modeling" (grant number 14-44-00011).

References

Arce-Mendoza, A., Rodriguez-de Ita, J., Salinas-Carmonam, M.C., Rosas-Taraco, A.G., 2008. Expression of CD64, CD206, and RAGE in adherent cells of diabetic patients infected with *Mycobacterium tuberculosis*. Arch. Med. Res. 39, 306–311.

Balasaniants, G.S., Titarenko, O.T., D'iakova, M.N., 2001. Diagnostic and prognostic significance of adenosine deaminase in acutely progressive pulmonary tuberculosis. Probl. Tuberk. 8, 46–49.

Barabási, A.L., Oltvai, Z.N., 2004. Network biology: understanding the cell's functional organization. Nat. Rev. Genet. 5, 101–113.

Bhatt, K., Uzelac, A., Mathur, S., McBride, A., Potian, J., Salgame, P., 2009. B7 costimulation is critical for host control of chronic *Mycobacterium tuberculosis* infection. J. Immunol. 182, 3793–3800.

Bikmaeva, A.R., Sibiriak, S.V., Khusnutdinova, E.K., 2004. Insertional polymorphism of the *CYP2E1* gene in infiltrative pulmonary tuberculosis in populations of Bashkortostan Republic. Mol. Biol. 38, 196–199.

Bouchelouche, K., Andresen, L., Alvarez, S., Nordling, J., Nielsen, O.H., Bouchelouche, P., 2006. Interleukin-4 and 13 induce the expression and release of monocyte chemoattractant protein 1, interleukin-6 and stem cell factor from human detrusor smooth muscle cells: synergy with interleukin-1beta and tumor necrosis factor-alpha. J. Urol. 175, 760–765.

Bragina, E.Y., Tiys, E.S., Freidin, M.B., Koneva, L.A., Demenkov, P.S., Ivanisenko, V.A., et al., 2014. Insights into pathophysiology of dystropy through the analysis of gene networks: an example of bronchial asthma and tuberculosis. Immunogenetics 66, 457–465.

Cellier, M.F., Bergevin, I., Boyer, E., Richer, E., 2001. Polyphyletic origins of bacterial Nramp transporters. Trends Genet. 17, 365–370.

Cernat, R.I., Mihaescu, T., Vornicu, M., Vione, D., Olariu, R.I., Arsene, C., 2011. Serum trace metal and ceruloplasmin variability in individuals treated for pulmonary tuberculosis. Int. J. Tuberc. Lung Dis. 15, 1239–1245.

Cowie, D., 2014. Iron and Tuberculosis (TB) Pathogenesis (Thesis presented in partial fulfilment of the requirements for the degree of Master in Genetics in the Faculty of AgriSciences at Stellenbosch University).

Fortún, J., Martín-Dávila, P., Méndez, R., Martínez, A., Norman, F., Rubi, J., et al., 2009. Ca-125: a useful marker to distinguish pulmonary tuberculosis from other pulmonary infections. Open Respir. Med. J. 3, 123–127.

Glotov, A.S., Tiys, E.S., Vashukova, E.S., Pakin, V.S., Demenkov, P.S., Saik, O.V., et al., 2015. Molecular association of pathogenetic contributors to pre-eclampsia (pre-eclampsia associome). BMC Syst. Biol. 9 (Suppl. 2), S4.

Goyal, M., Dhand, R., Gupta, N., Ganguly, N.K., Dang, N., 1990. Monocyte/macrophage functions & humoral response in blood & bronchoalveolar lavage fluid of pulmonary tuberculosis patients. Indian J. Med. Res. 91, 349–354.

Gupta, K.B., Gupta, R., Atreja, A., Verma, M., Vishvkarma, S., 2009. Tuberculosis and nutrition. Lung India 26, 9.

Hattrup, C.L., Gendler, S.J., 2008. Structure and function of the cell surface (tethered) mucins. Annu. Rev. Physiol. 70, 431–457.

Hessmann, M., Rausch, A., Rückerl, D., Adams, P.S., Simon, M., Gilfillan, S., et al., 2011. DAP10 contributes to CD8(+) T cell-mediated cytotoxic effector mechanisms during *Mycobacterium tuberculosis* infection. Immunobiology 216, 639–647.

Hwang, S., Son, S.W., Kim, S.C., Kim, Y.J., Jeong, H., Lee, D., 2008. A protein interaction network associated with asthma. J. Theor. Biol. 252, 722–731.

Ivanisenko, V.A., Saik, O.V., Ivanisenko, N.V., Tiys, E.S., Ivanisenko, T.V., Demenkov, P.S., et al., 2015. ANDSystem: an Associative Network Discovery System for automated literature mining in the field of biology. BMC Syst. Biol. 9 (Suppl. 2), S2.

Karyadi, E., Schultink, W., Nelwan, R.H.H., Gross, R., Amin, Z., Dolmans, W.M.V., et al., 2000. Poor micronutrient status of active pulmonary tuberculosis patients in Indonesia. J. Nutr. 130, 2935–2958.

Kim, E.J., Park, W.H., Ahn, S.G., Yoon, J.H., Kim, S.W., Kim, S.A., 2010. 5′-nitroindirubinoxime inhibits inflammatory response in TNF-alpha stimulated human umbilical vein endothelial cells. Atherosclerosis 211:77–83. http://dx.doi.org/10.1016/j.atherosclerosis.2010.01.040.

Kolchanov, N.A., Anan'ko, E.A., Kolpakov, F.A., Podkolodnaya, O.A., Ignat'eva, E.V., Goryachkovskaya, T.N., et al., 2000. Gene networks. Mol. Biol. 34, 449–460.

Koonin, E.V., 2011. The Logic of Chance: the Nature and Origin of Biological Evolution. FT Press, Upper Saddle River .

Kussmann, M., Morine, M.J., Hager, J., Sonderegger, B., Kaput, J., 2013. Perspective: a systems approach to diabetes research. Front. Genet. 4, 205.

Law, K., Weiden, M., Harkin, T., Tchou-Wong, K., Chi, C., Rom, W.N., 1996. Increased release of interleukin-1 beta, interleukin-6, and tumor necrosis factor-alpha by bronchoalveolar cells lavaged from involved sites in pulmonary tuberculosis. Am. J. Respir. Crit. Care Med. 153, 799–804.

Lim, J., Hao, T., Shaw, C., Patel, A.J., Szabó, G., Rual, J.F., et al., 2006. A protein-protein interaction network for human inherited ataxias and disorders of Purkinje cell degeneration. Cell 125, 801–814.

Lin, M.Y., Ottenhoff, T.H., 2008. Host-pathogen interactions in latent *Mycobacterium tuberculosis* infection: identification of new targets for tuberculosis intervention. Endocr. Metab. Immune Disord. Drug Targets 8, 15–29.

Liu, M., Liberzon, A., Kong, S.W., Lai, W.R., Park, P.J., Kohane, I.S., et al., 2007. Network-based analysis of affected biological processes in type 2 diabetes models. PLoS Genet. 3, e96.

Madan, T., Saxena, S., Murthy, K.J., Muralidhar, K., Sarma, P.U., 2002. Association of polymorphisms in the collagen region of human *SP-A1* and *SP-A2* genes with pulmonary tuberculosis in Indian population. Clin. Chem. Lab. Med. 40, 1002–1008.

Malik, S., Greenwood, C.M., Eguale, T., Kifle, A., Beyene, J., Habte, A., et al., 2006. Variants of the *SFTPA1* and *SFTPA2* genes and susceptibility to tuberculosis in Ethiopia. Hum. Genet. 118, 752–759.

Meyer, C.G., Thye, T., 2014. Host genetic studies in adult pulmonary tuberculosis. Semin. Immunol. 26, 445–453.

Mira, M.T., Alcaïs, A., Nguyen, V.T., Moraes, M.O., Di Flumeri, C., Vu, H.T., et al., 2004. Susceptibility to leprosy is associated with PARK2 and PACRG. Nature 427, 636–640.

Montoya, C.J., Cataño, J.C., Ramirez, Z., Rugeles, M.T., Wilson, S.B., Landay, A.L., 2008. Invariant NKT cells from HIV-1 or *Mycobacterium tuberculosis*-infected patients express an activated phenotype. Clin. Immunol. 127, 1–6.

Nasta, F., Ubaldi, V., Pace, L., Doria, G., Pioli, C., 2006. Cytotoxic T-lymphocyte antigen-4 inhibits GATA-3 but not T-bet mRNA expression during T helper cell differentiation. Immunology 117, 358–367.

Neifakh, S.A., Monakhov, N.K., Shapshnikov, A.M., Zubzhitski, Y.N., 1969. Localization of ceruloplasmin biosynthesis in human and monkey liver cells and its copper regulation. Experientia 25, 337–344.

Parikshak, N.N., Gandal, M.J., Geschwind, D.H., 2015. Systems biology and gene networks in neurodevelopmental and neurodegenerative disorders. Nat. Rev. Genet. 16, 441–458.

Podkolodnaya, O.A., Yarkova, E.E., Demenkov, P.S., Konovalova, O.S., Ivanisenko, V.A., Kolchanov, N.A., 2011. Application of the ANDCell computer system to reconstruction and analysis of associative networks describing potential relationships between myopia and glaucoma. Russ. J. Genet. 1, 21–28.

Pujana, M.A., Han, J.D., Starita, L.M., Stevens, K.N., Tewari, M., Ahn, J.S., et al., 2007. Network modeling links breast cancer susceptibility and centrosome dysfunction. Nat. Genet. 39, 1338–1349.

Rzhetsky, A., Iossifov, I., Koike, T., Krauthammer, M., Kra, P., Morris, M., et al., 2004. GeneWays: a system for extracting, analyzing, visualizing, and integrating molecular pathway data. J. Biomed. Inform. 37, 43–53.

Sato, M., Ishikawa, O., Abe, M., Miyachi, Y., 1998. Opposite effects of tumour necrosis factor-alpha on type I and III collagen gene expression by human dermal fibroblasts in monolayer and three-dimensional cultures. Br. J. Dermatol. 138, 118–121.

Saunders, B.M., Frank, A.A., Orme, I.M., Cooper, A.M., 2002. CD4 is required for the development of a protective granulomatous response to pulmonary tuberculosis. Cell. Immunol. 216, 65–72.

Shangguan, W.J., Tang, L.M., Li, H., 2014. Screening of genes related to acute coronary syndrome and identification of functional modules in the PPI network. Eur. Rev. Med. Pharmacol. Sci. 18, 52–57.

Sheng, Y.J., Wu, G., He, H.Y., Chen, W., Zou, Y.S., Li, Q., et al., 2014. The association between *CYP2E1* polymorphisms and hepatotoxicity due to anti-tuberculosis drugs: a meta-analysis. Infect. Genet. Evol. 24, 34–40.

Starr, T.K., Jameson, S.C., Hogquist, K.A., 2003. Positive and negative selection of T cells. Annu. Rev. Immunol. 21, 139–176.

Stein, C.M., Zalwango, S., Malone, L.L., Won, S., Mayanja-Kizza, H., Mugerwa, R.D., et al., 2008. Genome scan of *M. tuberculosis* infection and disease in Ugandans. PLoS One 3, e4094. http://dx.doi.org/10.1371/journal.pone.0004094.

Szklarczyk, D., Franceschini, A., Wyder, S., Forslund, K., Heller, D., Huerta-Cepas, et al., 2015. STRING v10: protein-protein interaction networks, integrated over the tree of life. Nucleic Acids Res. 43 (Database issue), D447–D452.

Velmurugan, K., Chen, B., Miller, J.L., Azogue, S., Gurses, S., Hsu, T., et al., 2007. *Mycobacterium tuberculosis* nuoG is a virulence gene that inhibits apoptosis of infected host cells. PLoS Pathog. 3, e110.

Wang, Y., Kanegane, H., Sanal, O., Tezcan, I., Ersoy, F., Futatani, T., et al., 2002. Novel Igalpha (CD79a) gene mutation in a Turkish patient with B cell-deficient agammaglobulinemia. Am. J. Med. Genet. 108, 333–336.

Wang, J.P., Rought, S.E., Corbeil, J., Guiney, D.G., 2003. Gene expression profiling detects patterns of human macrophage responses following *Mycobacterium tuberculosis* infection. FEMS Immunol. Med. Microbiol. 39, 163–172.

Wang, C., Jiang, T., Wei, L., Li, F., Sun, X., Fan, D., et al., 2012. Association of *CTLA4* gene polymorphisms with susceptibility and pathology correlation to pulmonary tuberculosis in southern Han Chinese. Int. J. Biol. Sci. 8, 945–952.

Zeldich, E., Koren, R., Dard, M., Weinberg, E., Weinreb, M., Nemcovsky, C.E., 2010. Enamel matrix derivative induces the expression of tissue inhibitor of matrix metalloproteinase-3

in human gingival fibroblasts via extracellular signal-regulated kinase. J. Periodontal Res. 45:200–206. http://dx.doi.org/10.1111/j.1600-0765.2009.01218.x.

Zhang, D.F., Wang, D., Li, Y.Y., Yao, Y.G., 2016. Integrative analyses of leprosy susceptibility genes indicate a common autoimmune profile. J. Dermatol. Sci. 82:18–27. http://dx.doi.org/10.1016/j.jdermsci.2016.01.001.

Zhao, N., Xiao, J., Zheng, Z., Fei, G., Zhang, F., Jin, L., et al., 2015. Single-nucleotide polymorphisms and haplotypes of non-coding area in the CP gene are correlated with Parkinson's disease. Neurosci. Bull. 31, 245–256.

Zychlinsky, A., Sansonetti, P., 1997. Perspectives series: host/pathogen interactions. Apoptosis in bacterial pathogenesis. J. Clin. Invest. 100, 493.