

RESEARCH ARTICLE

Open Access



Strain-specific behavior of *Mycobacterium tuberculosis* in A549 lung cancer cell line

Shima Hadifar^{1,2†}, Shayan Mostafaei^{3,4†}, Ava Behrouzi^{1,2}, Abolfazl Fateh^{1,2}, Parisa Riahi⁵, Seyed Davar Siadat^{1,2} and Farzam Vaziri^{1,2*}

*Correspondence: farzam_vaziri@yahoo.com; f_vaziri@pasteur.ac.ir
†Shima Hadifar and Shayan Mostafaei contributed equally to this article
¹ Department of Mycobacteriology and Pulmonary Research, Pasteur Institute of Iran, Tehran, Iran
Full list of author information is available at the end of the article

Abstract

Background: A growing body of evidence has shown the association between tuberculosis (TB) infection and lung cancer. However, the possible effect of strain-specific behavior of *Mycobacterium tuberculosis* (*M.tb*) population, the etiological agent of TB infection in this association has been neglected. In this context, this study was conducted to investigate this association with consideration of the genetic background of strains in the *M.tb* population.

Results: We employed the elastic net penalized logistic regression model, as a statistical-learning algorithm for gene selection, to evaluate this association in 129 genes involved in TLRs and NF- κ B signaling pathways in response to two different *M.tb* sub-lineage strains (L3-CAS1 and L 4.5). Of the 129 genes, 21 were found to be associated with the two studied *M.tb* sub-lineages. In addition, *MAPK8IP3* gene was identified as a novel gene, which has not been reported in previous lung cancer studies and may have the potential to be recognized as a novel biomarker in lung cancer investigation.

Conclusions: This preliminary study provides new insights into the mechanistic association between TB infection and lung cancer. Further mechanistic investigations of this association with a large number of *M.tb* strains, encompassing the other main *M.tb* lineages and using the whole transcriptome of the host cell are inevitable.

Keywords: *Mycobacterium tuberculosis*, Lung Cancer, Lineage, Elastic net penalized logistic regression

Background

Lung cancer and tuberculosis (TB) are the leading causes of human death and represent a global health concern. According to estimates from the 2018 Global Cancer Observatory (GLOBOCAN), lung cancer is the most frequent cancer, accounting for 11.6% of the 18.1 million newly diagnosed cancer cases [1]. In 2018, the World Health Organization (WHO) reported 10 million new cases of TB and 1.4 million TB-related deaths [2]. Different factors have been identified that might increase the risk of lung cancer [3]. Many risk factors are shared between lung cancer and TB infection [4]. Recent population-based studies have reported that the risk of lung cancer can increase with infection by *Mycobacterium tuberculosis* (*M.tb*), the etiological agent of TB infection [5–10]. An



inflammatory microenvironment driven by cytokines, chemokines, and inflammatory cells during TB infection has been recognized as a process that can induce genetic and host tissue damage and contribute to carcinogenesis in lung tissue [11].

There is increasing evidence of a significant genetic diversity in the *M.tb* population [12, 13]. *M.tb* strain-specific host–pathogen interactions have been demonstrated in previous studies [14–18]. This characteristic may affect the trend of *M.tb* pathogenesis and the molecular mechanism behind the association between the risk of lung cancer and TB infection. However, the association between TB infection and lung cancer has been evaluated previously without considering the bacterial genotype. In the current study, we tried to provide new insight into this association with consideration of the genetic background of strains in *M.tb* population.

In Iran, the presence of remarkable diversity in *M.tb* population structure with predominance of L3-CAS1 and L4.5 (NEW1) sub-lineages has been documented. Epidemiologically, Iran has been identified as the probable origin of L4.5 and the ecological adaptation and national occurring of this subpopulation was not unexpected. L3-CAS1 sub-lineage is almost found in around the Indian Ocean and the influx of Afghan refugees may contribute to ongoing circulation of L3-CAS1 sub-lineage in Iran. However, it seems genetic variability be the main driver of this epidemiological trend and transmission potential in both sub-lineages [13, 19, 20]. In line with, in our previous study, we compared these dominant sub-lineages of *M.tb* strains in interrupting TLRs and NF-κB signaling pathways in alveolar epithelial cell type II (A549 cell line) and observed strain-specific characteristics in interactions with host cells [16]. In the light of these results, we examined the gene expression profile of cancerous cell line in response to two *M.tb* with divergent genetic background by employing penalized statistical model and systems biology methods.

Results

Differential analysis of gene expression data

Based on the statistical comparisons by Limma R package, 37 genes were identified with significant down-regulation (all fold-regulations were less than - 2) compared to the *RPLP0* expression as the housekeeping gene, including *BTK*, *AKT1*, *CD86*, *CD80*, *IFNG*, *BCL3*, *TLR10*, *MAPK8IP3*, etc. In contrast, 25 genes were identified with a significant up-regulation (all fold-regulations were more than 2) compared to *RPLP0*, including *PTGS2(COX-2)*, *FOS*, *HSPA1A*, *NFKBIL1*, *MAP4K4*, *HSPD1*, *HSPA1A*, *FADD*, etc. (Table 1).

Table 1 The list of the differentially expressed genes in comparison with the housekeeping gene

Significantly down-regulated genes (n = 37)	Significantly up-regulated genes (n = 25)
<i>AKT1</i> , <i>BCL2L1</i> , <i>BCL3</i> , <i>BTK</i> , <i>CARD11</i> , <i>CD180</i> , <i>CD80</i> , <i>CD86</i> , <i>ELK1</i> , <i>HMGB1</i> , <i>PSIP1</i> , <i>PPARA</i> , <i>SIGIRR</i> , <i>TAB1</i> , <i>TBK1</i> , <i>TIRAP</i> , <i>TLR10</i> , <i>IFNG</i> , <i>IKBKG</i> , <i>IL1R1</i> , <i>IL2</i> , <i>IL10</i> , <i>IRF1</i> , <i>LTA</i> , <i>MAPK8IP3</i> , <i>TLR2</i> , <i>TLR4</i> , <i>TLR6</i> , <i>TLR7</i> , <i>TLR8</i> , <i>TNFRSF10A</i> , <i>TNFRSF1A</i> , <i>TRADD</i> , <i>TRAF2</i> , <i>TOLLIP</i> , <i>TRAF3</i> , <i>TNFAIP2</i>	<i>BCL2A1</i> , <i>CCL2</i> , <i>CCL5</i> , <i>CXCL2</i> , <i>CXCL8</i> , <i>EIF2AK2</i> , <i>FADD</i> , <i>FOS</i> , <i>PELI1</i> , <i>PTGS2</i> , <i>RELA</i> , <i>SARM1</i> , <i>HRAS</i> , <i>HSPA1A</i> , <i>HSPD1</i> , <i>HMOX1</i> , <i>IRF3</i> , <i>JUN</i> , <i>MAP2K4</i> , <i>MAP3K1</i> , <i>MAP4K4</i> , <i>NFK-</i> <i>BIA</i> , <i>NFKBIL1</i> , <i>NFRKB</i> , <i>NR2C2</i>

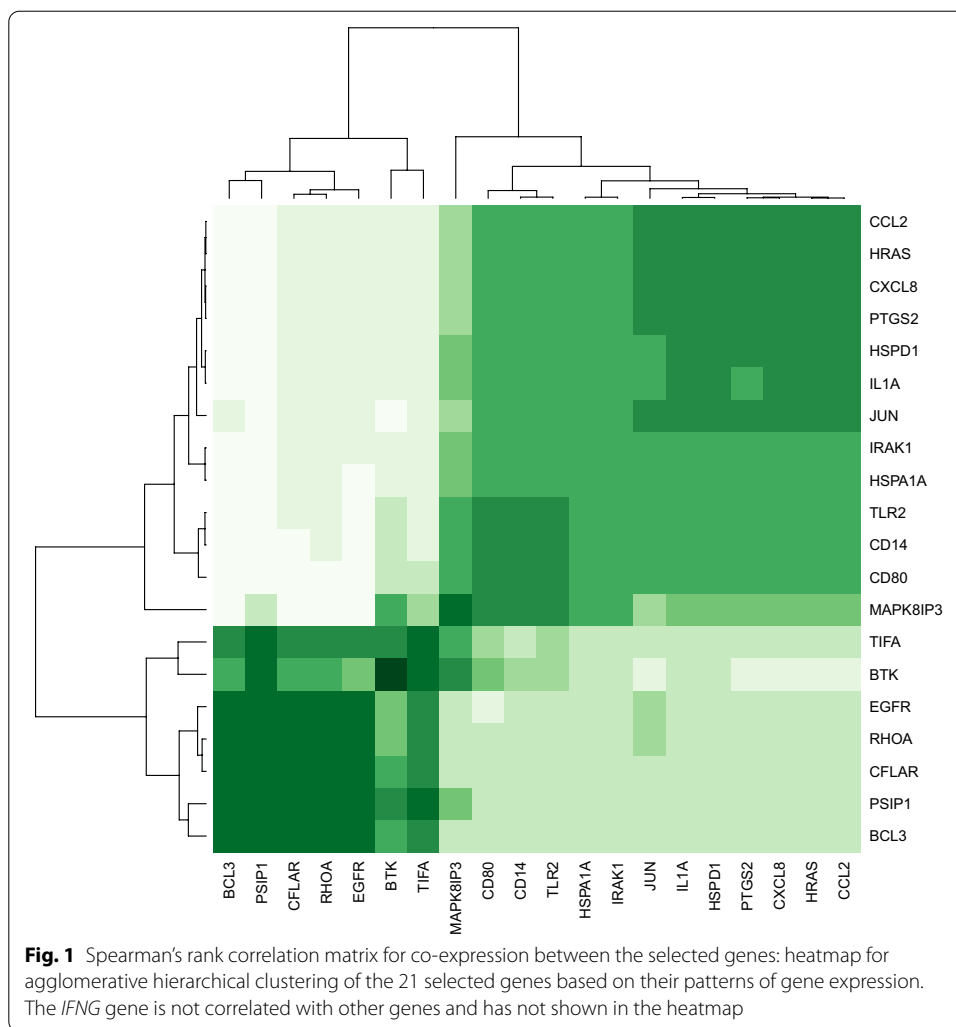
Table 2 Genes selected by elastic-net regularized multinomial logistic regression for the association between the genes and the dominant genotypes of *M.tb* strains

	Gene Symbol	Elastic net	Fold Regulation (L3-CAS1 vs. Con.)	Adjusted P-value (L3-CAS1 vs. Con.)	Fold Regulation (L4.5 vs. Con.)	Adjusted <i>p</i> value (L4.5 vs. Con.)	Fold Regulation (L4.5 vs. L3-CAS1)	Adjusted <i>p</i> value (L4.5 vs. L3-CAS1)
1	<i>CFLAR</i>	100%	- 1.544	0.082	1.376	0.117	2.125	0.044
2	<i>TIFA</i>	93.9%	- 1.392	0.199	- 1.206	0.399	1.154	0.256
3	<i>HRAS</i>	91.8%	3.279	0.0001	6.681	0.0001	2.038	0.001
4	<i>IRAK1</i>	82.36%	1.395	0.082	1.136	0.503	- 1.228	0.256
5	<i>CD14</i>	77.93%	4.724	0.0001	- 1.217	0.171	- 5.749	0.0001
6	<i>BTK</i>	67.28%	- 4.258	0.055	- 440.602	0.001	- 103.476	0.0001
7	<i>MAPK8IP3</i>	59.63%	- 13.117	0.0002	- 12.323	0.0003	1.064	0.114
8	<i>IFNG</i>	55.51%	- 24.028	0.0001	- 9.426	0.0002	2.549	0.001
9	<i>JUN</i>	41.91%	3.146	0.0001	12.582	0.0002	3.999	0.001
10	<i>IL1A</i>	39.77%	2.204	0.0001	3.160	0.0001	1.434	0.081
11	<i>CD80</i>	33.64%	- 3.630	0.0001	- 5.856	0.0002	- 1.613	0.044
12	<i>PSIP1</i>	31.03%	- 3.972	0.045	- 2.075	0.126	1.914	0.011
13	<i>PTGS2</i>	29.73%	40.880	0.0001	84.449	0.0001	2.066	0.001
14	<i>BCL3</i>	26.79%	- 12.098	0.035	- 1.617	0.188	7.482	0.0001
15	<i>HSPA1A</i>	26.63%	10.581	0.0001	4.009	0.0002	- 2.639	0.001
16	<i>RHOA</i>	26.27%	- 1.245	0.055	1.526	0.003	1.900	0.011
17	<i>CCL2</i>	26.23%	2.023	0.045	5.993	0.0002	2.962	0.001
18	<i>CXCL8</i>	24.92%	6.364	0.0001	20.966	0.0001	3.294	0.0001
19	<i>HSPD1</i>	23.79%	7.345	0.0001	9.759	0.0002	1.329	0.114
20	<i>TLR2</i>	22.52%	- 8.980	0.0001	1.651	0.0017	14.826	0.0001
21	<i>EGFR</i>	21.83%	- 3.547	0.035	1.286	0.433	4.561	0.0001
	Misclassification error rate	14.8%	-	-	-	-	-	-
	lambda	0.00849	-	-	-	-	-	-
	Alpha	0.50	-	-	-	-	-	-

Lambda is the tuning parameter and it is for controlling the overall strength of the penalty. Alpha is a controlled parameter between LASSO (L1) and ridge (L2) penalties. Alpha = 0.5 corresponds to elastic-net regularization. Importance value for each selected gene was reported, *p* values adjusted by Benjamini-Hochberg-FDR correction at $\alpha = 0.05$

Gene selection and literature search

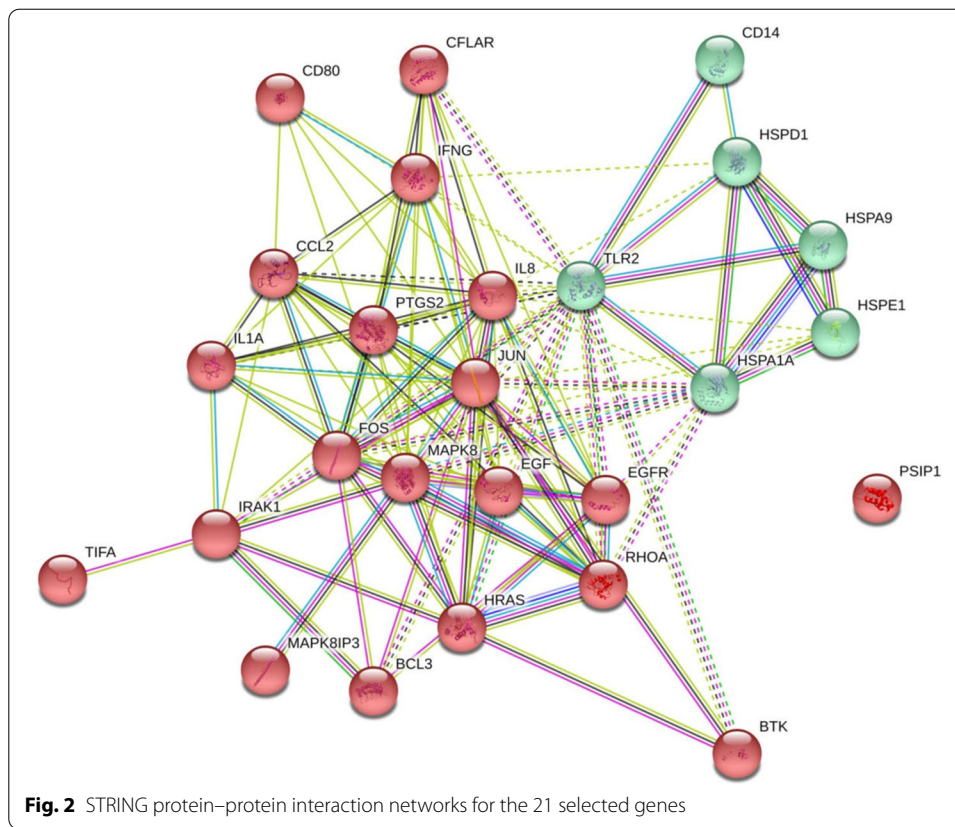
By using elastic net regularized logistic regression as a statistical-learning algorithm for gene selection, 21 genes were selected all in association with either of the two *M.tb* sub-lineages (Table 2). The most important selected genes included *CFLAR*, *TIFA*, *HRAS*, *IRAK1*, *CD14*, *BTK*, *MAPK8IP3*, *IFNG*, *JUN*, *IL1A*, *CD80*, *PSIP1*, *PTGS2*, *BCL3*, *HSPA1A*, *RHOA*, *CCL2(MCP-1)*, *CXCL8*, *HSPD1*, *TLR2*, and *EGFR*. The importance values of all the selected genes were more than 20%. Co-expression hierarchical heatmap between the selected genes was presented using Spearman's rank correlation matrix (Fig. 1). As shown in Fig. 1, the selected genes were clustered in the two main clusters based on the co-expression pattern. According to the co-expression patterns of the selected genes, *TIFA*, *BTK*, *EGFR*, *RHOA*, *CFLAR*, *PSIP1*, and *BCL3* had a similar expression pattern while the other selected genes formed another group with the same expression patterns (e.g. *CCL2*, *HRAS*, *CXCL8*, *PTGS2*, *HSPD1*, *IL1A*, and *JUN*). Figure 2 displays the functional protein-association networks for the 21 selected genes.



Furthermore, results of the literature search demonstrated 20 out of the 21 genes had been previously reported in the literature to be associated with lung cancer or lung function diseases (Table 3). We identified *MAPK8IP3* as a novel gene, which has not been previously detected in lung function studies. These selected genes may deserve special attention for the future development of therapeutics intervention.

Comparison of selected genes expression between *M.tb* L4.5 and L3-CAS1 sub-lineages groups

We investigated the differential expression of the 21 selected genes between *M.tb* L4.5 and L3-CAS1 sub-lineages groups. A comparison of *M.tb* L4.5 to L3-CAS1 sub-lineages showed significant differences in the expression of 13/21 (61.9%) genes. Among these 13 genes, *CFLAR*, *HRAS*, *IFNG*, *JUN*, *PTGS2*, *BCL3*, *CCL2*, *CXCL8*, *TLR2*, and *EGFR* were significantly up-regulated in the L4.5 sub-lineage group compared to the L3-CAS1 sub-lineage group, while *CD14*, *BTK*, and *HSPA1A* genes were significantly down-regulated in the *M.tb* L4.5 sub-lineage group compared to the *M.tb* L3-CAS1 sub-lineage group (Table 2). In



addition, *MAPK8IP3* was not changed in the *M.tb* L4.5 sub-lineage group compared to the *M.tb* L3-CAS1 sub-lineage group (adj. $P=0.114$).

Discussion

In the present study, we investigated lung cancer-related genes that differentially were regulated by different genotypes of *M.tb* in lung adenocarcinoma cell line using the statistical penalized algorithm. In our analyses, we identified 21 potentially lung cancer-related genes during infection with *M.tb* L3-CAS1 and L4.5 sub-lineages.

Various inflammatory processes and functional pathway-associated genes, which are involved in carcinogenesis, have been investigated in different studies. Chemokines secretion is one of the main ways for recruitment of host cell and inhibition of antitumor immune responses in cancerous cells [21, 22]. MCP-1 is one of the chemotactic stimuli that is secreted from cancerous cells and induces immunosuppressive microenvironments [23]. There are some controversies about the role of this chemokine in lung cancer pathogenesis [24, 25]. However, Fridlender et al. found that the blockade of *MCP-1* could inhibit lung tumorigenesis and could be proposed as a promising approach to lung cancer treatment [26]. Besides, the role of IL8 as another chemokine, which typically plays a role in the induction of angiogenesis and its overexpression, has been reported in lung cancer. In addition, the overexpression of *COX-2*, an inflammation-associated gene, has been found in different stages of lung

Table 3 Confirmation of the associations of 21 selected genes with lung cancer/or lung function by literature reviewing in PubMed with keywords (“Lung Cancer” OR “Lung Function”) AND “name of each selected gene”) in the title and abstract fields

Gene Symbol	Number of studies	References (PMIDs)
<i>CFLAR</i>	8	29,258,089, 29,165,042, 27,186,395, 24,387,758, 21,430,781, 20,713,531, 19,860,666, 17,450,141
<i>TIFA</i>	1	29,975,933
<i>HRAS</i>	37	30,196,239, 29,641,535, 29,636,358, and etc
<i>IRAK1</i>	4	25,550,857, 29,667,303, 26,262,504, 24,144,839
<i>CD14</i>	100	30,325,558, 30,112,108, 29,343,695, 29,448,000, and etc
<i>BTK</i>	5	28,990,652, 28,734,581, 27,437,104, 27,111,445, 17,327,079
<i>MAPK8IP3</i>	0	-
<i>IFNG</i>	12	29,516,506, 24,244,422, and etc
<i>JUN</i>	200	29,262,633, 29,858,032, and etc
<i>IL1A</i>	3	25,554,695, 12,752,325, 12,737,276
<i>CD80</i>	30	11,422,893, 26,702,740, and etc
<i>PSIP1</i>	1	24,670,920
<i>PTGS2</i>	22	22,761,909, 22,464,751, and etc
<i>BCL3</i>	2	20,420,878, 26,122,346
<i>HSPA1A</i>	5	22,037,874, 29,137,250, 28,331,811, 26,093,302, 20,404,511
<i>RHOA</i>	100	29,995,590, 29,337,059, 28,888,686, and etc
<i>CCL2</i>	66	30,061,946, 29,892,522, 29,722,145, and etc
<i>CXCL8</i>	55	27,578,214, 29,636,079, 28,836,545, and etc
<i>HSPD1</i>	5	28,978,099, 28,698,656, and etc
<i>TLR2</i>	32	27,097,965, 22,952,638, 24,630,931, and etc
<i>EGFR</i>	3692	30,350,450, 30,349,313, 30,343,004, and etc

cancer [27]. This up-regulation may be explained the functional role of this gene in lung tumorigenesis and prediction of patient outcomes. In line with lung cancer studies, in our analysis, the expression of *MCP-1*, *IL-8*, and *COX-2* was up-regulated in the cancerous cell line in response to the infection with *M.tb* L4.5 sub-lineage when compared to the infection with *M.tb* L3-CAS1 sub-lineage ($p < 0.001$).

IFN- γ is generally considered as a cytokine with antitumor activity. However, there are significant controversies about the role of this cytokine. Increasing evidence suggested that IFN- γ may have dual aspects in its function and act as both an anti-tumorigenic and a pro-tumorigenic cytokine [28]. The pro-tumorigenic property of IFN- γ is based on the upregulation of immunosuppressive cells such as Treg cells and Th17 [29]. The pivotal role of this cytokine in regulating the Programmed death-ligand 1 (*PD-L1*) gene expression, as a factor that has an inhibitory role in cancer immunity, and promoting the immune evade has been found in tumor cells [30]. Besides, the upregulation of PD-L1 expression and induction of lung carcinoma by IFN- γ have been revealed [30]. Similarly, in vitro and in vivo concomitant H37Rv infection in non-small cell lung cancer showed that lung cancer progression facilitated by enhancing Treg cells proportion and the upregulation of PD-L1 expression that induced by H37Rv as a part of *M.tb* lineage 4 [31]. In the current study, based on the expression of IFN- γ in response to the infection with *M.tb* L4.5 sub-lineage, compared to L3-CAS1 sub-lineage, this expression profile might be favor for better control of *M.tb* L4.5 sub-lineage strain compared to L3-CAS1

sub-lineage in infected host cell, in addition may contribute to the pathogenesis and deterioration of lung cancer during infection with *M.tb* L4.5 sub-lineage strain. Evaluating the expression level of *IL-8*, *MCP-1* and *IFN- γ* genes in A549 cell line in response to infection with F15/LAM4/KZN(LAM sub-lineage), F11(LAM sub-lineage), F28 (S sub-lineage) and Beijing (L2-Beijing sub-lineage) genotypes showed that the higher level of upregulation in the both genes are induced in response to infection with *M.tb* LAM sub-lineages compared to the other sub-lineages[32],While they failed to detect the expression of *IFN- γ* in response to all strains. However, the *M.tb* L4.5, LAM and S strains are members of lineage 4, differentially induced host response. The strain-specific characteristic of *M.tb* population may have the potential to be considering in lung cancer cells studies.

Deregulation of apoptosis and cell proliferation pathways are the key mechanisms playing important roles in cancer pathogenesis [33]. The blockade of apoptosis can be mediated by the overexpression of anti-apoptotic proteins such as FLIP. The upregulation of FLIP also has been detected in lung carcinoma [34, 35]. In line with lung cancer study, the expression of *FLIP* was upregulated in response to L4.5 sub-lineage when compared with L3-CAS1 sub-lineage in our study ($p < 0.05$). The overexpression of FLIP by *M. tb* L4.5 sub-lineage may contribute to exacerbation of lung cancer during infection with this strain. This overexpression and inhibiting of the apoptosis is also favor for *M.tb* pathogenesis.

In addition, it has been demonstrated the blockade of the Rho/Rho-kinase pathway, which is involved in cancer proliferation and invasion, inhibited tumor migration and invasion [36, 37]. The knockdown of *RhoA* as the member of the Rho family inhibits lung cancer cell proliferation and induces apoptosis [38]. In our study, the expression of *RhoA* did not change in response to L4.5 sub-lineage when compared with L3-CAS1 sub-lineage. In addition, the deregulation of the *HRAS* gene, which is involved in the proliferation of different cancers, has been reported [39, 40]. Overexpression of the *Ras* oncogene family member was identified in response to L4.5 sub-lineage when compared with L3-CAS1 sub-lineage. This upregulation can contribute to the progression of cancerous cells.

Among the 21 selected genes, *BCL3* has an inhibitory function. The deregulation of this gene as an atypical member of the I κ B family has been shown in different solid tumors[41]. In addition, Dimitrakopoulos et al. described the role of this gene in lung carcinogenesis [42]. They reported an increase in *BCL3* expression in lung cancer. This overexpression could be directly related to the increased level of *EGFR* expression. Aberrant *EGFR* expression are implicated in the progression of malignant cells manner [43, 44]. *EGFR* can promote angiogenesis by upregulation of main angiogenesis mediators such as Vascular endothelial growth factor (*VEGF*). Angiogenesis plays important role in the solid tumors growth and metastasis spreading [45]. Moreover, the expression of *HDM2* as a negative regulator of p53 that is a tumor suppressor gene was induced by the upregulation of *BCL3*[46]. In our analysis, based on the expression of *BCL3* and *EGFR* in response to L4.5 sub-lineage when compared to L3-CAS1 sub-lineage, we hypothesize that infection by the L4.5 sub-lineage strain may be potent to deteriorate lung carcinoma by promoting tumor growth and angiogenesis.

Besides, it has been proposed that *HSPA1A* (*HSP70*), a chaperone molecule, is strongly involved in promoting and development of different tumor cells and overexpression of this heat-shock protein has been shown to be associated with the progression of several tumors such as lung cancer [47]. The current findings were consistent with the previous studies. However, level of *HSP70* expression in response to infection with *M.tb* L4.5 sub-lineage was lower in compared to L3-CAS1 sub-lineage.

It is noteworthy that some inconsistent results were found in the current study. It has been shown TLRs pathway molecules such as *IRAK1* and *TLR2* have important roles in neoplasm diseases [48] and the significant upregulation of *IRAK1* and its involvement in the development of solid tumors including lung cancer have been reported [49, 50]. Besides, high expression of *TIFA* [51] and *IL1A* as a gene, which regulates tumor growth, angiogenesis, and metastasis in lung carcinoma cell has been reported. [52]. Contrary, no changes in the expression of all aforementioned genes were observed during infection with *M.tb* L4.5 sub-lineage compared to L3-CAS1 sub-lineage. Although the expression of the genes is controversial, the expression profile of other genes suggested the possibility that infection with the *M.tb* L4.5 sub-lineage strain drive cancer cell to progression. In the other word, the risk of progression might have promoted in lung cancer patients with lung that infected by *M.tb* L4.5 sub-lineage strain compared to *M.tb* L3-CAS1 sub-lineage strain. These patients also are more potent to secondary infections.

In contrast to our results, Mvubu et al. [32] showed that the expression level of *IRAK1* and *IL1A* were increased in response to infection with LAM sub-lineages (F15/LAM4/KZN, F11), S sub-lineage (F28) and Beijing sub-lineage. Level of this increase was higher in response to LAM sub-lineages compared to the other sub-lineages. It is possible that infection with LAM sub-lineages similar to L4.5 sub-lineages is more potent to drive cancer cell to progression.

In our analysis, we also identified *MAPK8IP3* as a novel and potent target that has not been reported in previous lung cancer studies. *MAPK8IP3* is a scaffold gene, also known as JIP3, that exhibits function in the JNK pathway [53]. The overexpression of this gene has been shown in different tumor cells [53, 54]. Therefore, *MAPK8IP3* may have the potential to be recognized as a novel biomarker in lung cancer investigation.

Based on the results of previous studies that demonstrated elastic net penalized logistic regression frequently performed better than Ridge, LASSO, and some statistical-based learning algorithms for model selection consistency and prediction accuracy [55], the use of this modern and accepted computational method in high dimensional gene expression data is a strength of the current study. The validation of all the results by literature review, the use of an appropriate cross-validation method (repeated 5-CV), the address of potential sources of bias and the use of STRING networks are the other strengths of the present study. However, the main limitations of our study are that the cell line selection was confined to adenocarcinoma of lung cell line and protein levels of selected genes were not assessed.

Conclusions

The evidence of epidemiological association between TB infection and lung cancer is well established. This preliminary study provides new insights into the mechanistic association between TB infection and lung cancer. The two studied *M.tb* sub-lineages

promoted cancer development by creating an inflammatory environment through differentially down/up-regulation of gene involved in TLRs and NF- κ B signaling pathways. This environment has crucial impact on cell proliferation, apoptosis and angiogenesis. Based on significant strain-specific behavior of *M.tb* population in host–pathogen interactions and according to our findings, investigation of linking TB infection to lung cancer in the context of the genetic background of *M.tb* strains might be more effective to gain a better understanding of this association, identification of *M.tb* strain-specific behavior and therapeutic intervention. Further investigations with a large number of *M.tb* strains, encompassing the other main *M.tb* lineages and using the whole transcriptome of the host cell are inevitable. However, providing further information to fully understand of significant *M.tb* strain-specific behavior related to lung cancer progression and minimizing bias are needed by means of high throughput methods.

Methods

Study design

The study was designed in accordance with our previous study [16] which investigated the gene expression profile of infected A549 cell line (ATCC CCL-185) in response to dominant genotypes of *M.tb*. Briefly, the dominant genotypes of *M.tb* (L3-CAS1 and L4.5 strains) in the capital of Iran were identified based on 24 loci MIRU-VNTR and Spoligotyping[19] and confirmed by whole genome sequencing method. Then, the A549 cell line (maintained in antibiotic-free media) was infected in triplicates with the dominant genotypes an multiplicity of infection (MOI) of ~50:1 (50 bacteria:cell) for 72 h supplemented Dulbecco's modified Eagle medium (DMEM) and After the time, cellular response involved in TLRs and NF- κ B signaling pathways was evaluated by qRT-PCR. RT² Profiler™ PCR Array kits (QIAGEN), which include RT² Profiler™ PCR Array Human Toll-Like Receptor Signaling Pathway (QIAGEN, Cat.No. PAHS-018ZF-2) and RT² Profiler™ PCR Array Human NF- κ B Signaling Pathway (QIAGEN, Cat.No. PAHS-025YF-2) according to the manufacturer's instructions was used to perform qRT-PCR. The expression of 168 pathway-specific genes was evaluated and 39 genes were shared between these pathways. Secretion level of 12 cytokines/chemokines was assessed by ELISA arrays kit (QIAGEN). Viability of infected and mock cells was evaluated by the trypan blue exclusion test based on the manufacturer's instructions (Sigma Aldrich, Germany). In addition, intracellular growth assay and intracellular internalization index were carried out [16].

Gene expression analysis

The comparative cycle threshold (Ct) method ($2^{-\Delta Ct} \times 10^3$) was used to demonstrate the relative gene expression across the samples and the fold change was calculated using the $2^{-\Delta\Delta Ct}$ method [56]. Next, the primary gene expression data were qualified and normalized. Linear modeling for statistical comparison was applied by “limma” R package [57]. The cutoff of the false discovery rate for statistical comparison between the control and TB groups was considered at the level of 0.10.

Gene selection model

Elastic net regularization produced a sparse model with good prediction accuracy and good grouping capability. Elastic net frequently has served better than the Ridge, LASSO, and many other statistical learning algorithms in gene selection consistency and prediction accuracy in gene datasets [55, 58]. Elastic net is introduced as a compromise between these two techniques, combining strength between the Ridge and LASSO penalized regression [59]. The elastic net penalized logistic regression was performed by “glmnet” R package (<https://cran.r-project.org/web/packages/glmnet>). The two *M.tb* sub-lineages were considered as dependent variable and expression level of the 129 genes were considered as independent/ or predictive variables in the elastic net regularized logistic regression for gene selection. The importance value of each selected gene was calculated using “varImp” function in “Caret” R package. Interactive agglomerative hierarchical clustering heatmap was applied by “heatmaply” R package in order to draw the co-expression heatmap between the selected genes (<https://cran.r-project.org/web/packages/heatmaply>). Statistical significance was considered at the level of 0.05 in the all of statistical methods.

Cross-validation and literature validation

In order to validate the performance of the elastic net penalized regression, the repeated fivefold cross-validation was used. The model split the dataset by using repeated random sub-sampling with 100 repetitions in the fivefold cross-validation, permuting the sample labels every time. The cross-validated performance was summarized by observed misclassification error rate. In addition, to assess the literature validation for any result, a literature mining was used in PubMed by the search strategy of (“Lung Cancer” OR “Lung Function”) AND (“name of each selected gene”) and related MeSH terms in title and abstract fields.

Abbreviations

Ct: Cycle threshold; GLOBOCAN: Global Cancer Observatory; *M.tb*: *Mycobacterium tuberculosis*; PD-L1: Programmed Death-Ligand 1; TB: Tuberculosis.

Acknowledgements

We thank all the personnel of Mycobacteriology and Pulmonary Research Department, Pasteur Institute of Iran for their assistance in this project. We also thank Dr. Maha Reda Farhat for reviewing the manuscript.

Authors' contributions

SH, FV and SM wrote the manuscript. FV supervised the project. FV, SDS and AF designed the project. SH and AB performed laboratory work. SM and PR performed the statistical analysis. All authors have read and approved the final manuscript.

Funding

This work was supported by a grant [Number 927] from Pasteur Institute of Iran. The funding body had no role in the design of the study and collection, analysis, and interpretation of data and in writing the manuscript.

Availability of data and materials

The datasets used and/or analyzed during the current study could become available through the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study was approved by Ethical Committee of the Pasteur Institute of Iran (Tehran, Iran). Under the Ethics code of 927.

Consent to publish

This section is not applicable to the current study.

Competing interests

The authors declare that they have no conflicting interests.

Author details

¹ Department of Mycobacteriology and Pulmonary Research, Pasteur Institute of Iran, Tehran, Iran. ² Microbiology Research Center (MRC), Pasteur Institute of Iran, Tehran, Iran. ³ Department of Biostatistics, School of Health, Kermanshah University of Medical Sciences, Kermanshah, Iran. ⁴ Epidemiology and Biostatistics Unit, Rheumatology Research Center, Tehran University of Medical Sciences, Tehran, Iran. ⁵ Department of Biostatistics, Faculty of Medical Sciences, Tarbiat Modares University, Tehran, Iran.

Received: 12 December 2020 Accepted: 23 March 2021

Published online: 25 March 2021

References

- Bray F, Ferlay J, Soerjomataram I, Siegel RL, Torre LA, Jemal A. Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA: a cancer journal for clinicians*. 2018.
- WHO. WHO Global Report, Global Tuberculosis Report 2019. Geneva: World Health Organization. 2019.
- Malhotra J, Malvezzi M, Negri E, La Vecchia C, Boffetta P. Risk factors for lung cancer worldwide. *Eur Respirat J*. 2016;48(3):889–902.
- Skowroński M, Iwanik K, Halicka A, Barinow-Wojewódzki A. Squamous cell lung cancer in a male with pulmonary tuberculosis. *Adv Respirat Med*. 2015;83(4):298–302.
- Bae J-M, Li Z-M, Shin M-H, Kim D-H, Lee M-S, Ahn Y-O. Pulmonary tuberculosis and lung cancer risk in current smokers: the Seoul Male Cancer Cohort Study. *J Korean Med Sci*. 2013;28(6):896–900.
- Wu CY, Hu HY, Pu CY, Huang N, Shen HC, Li CP, et al. Pulmonary tuberculosis increases the risk of lung cancer: a population-based cohort study. *Cancer*. 2011;117(3):618–24.
- Yu Y-H, Liao C-C, Hsu W-H, Chen H-J, Liao W-C, Muo C-H, et al. Increased lung cancer risk among patients with pulmonary tuberculosis: a population cohort study. *J Thoracic Oncol*. 2011;6(1):32–7.
- Wong JY, Zhang H, Hsiung CA, Shiraishi K, Yu K, Matsuo K, et al. Tuberculosis infection and lung adenocarcinoma: Mendelian randomization and pathway analysis of genome-wide association study data from never-smoking Asian women. *Genomics*. 2020;112(2):1223–32.
- Liang H, Guan P, Yin Z, Li X, He Q, Zhou B. Risk of lung cancer following nonmalignant respiratory conditions among nonsmoking women living in Shenyang, Northeast China. *J Womens Health (Larchmt)*. 2009;18(12):1989–95.
- Kreuzer M, Heinrich J, Kreienbrock L, Schaffrath Rosario A, Gerken M, Wichmann HE. Risk factors for lung cancer among nonsmoking women. *Int J Cancer*. 2002;100(6):706–13.
- Su Y-Y, Yen Y-F, Pan S-W, Chuang P-H, Feng J-Y, Chou K-T, et al. Latent tuberculosis infection and the risk of subsequent cancer. *Medicine*. 2016;95(4).
- Gagneux S, Small PM. Global phylogeography of *Mycobacterium tuberculosis* and implications for tuberculosis product development. *Lancet Infect Diseases*. 2007;7(5):328–37.
- Gagneux S. Ecology and evolution of *Mycobacterium tuberculosis*. *Nat Rev Microbiol*. 2018;16(4):202.
- Sarkar R, Lenders L, Wilkinson KA, Wilkinson RJ, Nicol MP. Modern lineages of *Mycobacterium tuberculosis* exhibit lineage-specific patterns of growth and cytokine induction in human monocyte-derived macrophages. *PLoS ONE*. 2012;7(8):e43170.
- Carmona J, Cruz A, Moreira-Teixeira L, Sousa C, Sousa J, Osorio NS, et al. *Mycobacterium tuberculosis* strains are differentially recognized by TLRs with an impact on the immune response. *PLoS ONE*. 2013;8(6):e67277.
- Hadifar S, Behrouzi A, Fateh A, Khatami S, Rahimi Jamnani F, Siadat SD, et al. Comparative study of interruption of signaling pathways in lung epithelial cell by two different *Mycobacterium tuberculosis* lineages. *J Cell Physiol*. 2019;234(4):4739–53.
- Coscolla M, Gagneux S. Does *M. tuberculosis* genomic diversity explain disease diversity? *Drug Discov Today: Disease Mech*. 2010;7(1):e43–e59.
- Mvubu NE, Pillay B, McKinnon LR, Pillay M. *Mycobacterium tuberculosis* strains induce strain-specific cytokine and chemokine response in pulmonary epithelial cells. *Cytokine*. 2018;104:53–64.
- Hadifar S, Shamkhali L, Kamakoli MK, Mostafaei S, Khanipour S, Mansoori N, et al. Genetic diversity of *Mycobacterium tuberculosis* isolates causing pulmonary and extrapulmonary tuberculosis in the capital of Iran. *Mol Phylogenet Evol*. 2019;132:46–52.
- Hadifar S, Fateh A, Pourbarkhordar V, Siadat SD, Mostafaei S, Vaziri F. Variation in *Mycobacterium tuberculosis* population structure in Iran: a systemic review and meta-analysis. *BMC Infect Dis*. 2021;21(1):2.
- Rabinovich GA, Gabrilovich D, Sotomayor EM. Immunosuppressive strategies that are mediated by tumor cells. *Annu Rev Immunol*. 2007;25:267–96.
- Raman D, Baugher PJ, Thu YM, Richmond A. Role of chemokines in tumor growth. *Cancer Lett*. 2007;256(2):137–65.
- Conti I, Rollins BJ, editors. CCL2 (monocyte chemoattractant protein-1) and cancer. *Seminars in cancer biology*; 2004: Elsevier.
- Brown CE, Vishwanath RP, Aguilar B, Starr R, Najbauer J, Aboody KS, et al. Tumor-derived chemokine MCP-1/CCL2 is sufficient for mediating tumor tropism of adoptively transferred T cells. *J Immunol*. 2007;179(5):3332–41.
- Wang T, Dai H, Wan N, Moore Y, Dai Z. The role for monocyte chemoattractant protein-1 in the generation and function of memory CD8+ T cells. *J Immunol*. 2008;180(5):2886–93.
- Fridlender ZG, Kapoor V, Buchlis G, Cheng G, Sun J, Wang L-CS, et al. Monocyte chemoattractant protein-1 blockade inhibits lung cancer tumor growth by altering macrophage phenotype and activating CD8+ cells. *Am J Respirat Cell Mol Biol*. 2011;44(2):230–7.

27. Moraes JL, Moraes AB, Aran V, Alves MR, Schluckbier L, Duarte M, et al. Functional analysis of polymorphisms in the COX-2 gene and risk of lung cancer. *Mol Clin Oncol*. 2017;6(4):494–502.
28. Mandai M, Hamanishi J, Abiko K, Matsumura N, Baba T, Konishi I. Dual faces of IFN γ in cancer progression: a role of PD-L1 induction in the determination of pro-and antitumor immunity. *Clin Cancer Res*. 2016;22(10):2329–34.
29. Irmiler IM, Gajda M, Bräuer R. Exacerbation of antigen-induced arthritis in IFN- γ -deficient mice as a result of unrestricted IL-17 response. *J Immunol*. 2007;179(9):6228–36.
30. Zhang X, Zeng Y, Qu Q, Zhu J, Liu Z, Ning W, et al. PD-L1 induced by IFN- γ from tumor-associated macrophages via the JAK/STAT3 and PI3K/AKT signaling pathways promoted progression of lung cancer. *Int J Clin Oncol*. 2017;22(6):1026–33.
31. Zhou Y, Hu Z, Cao S, Yan B, Qian J, Zhong H. Concomitant Mycobacterium tuberculosis infection promotes lung tumor growth through enhancing Treg development. *Oncol Rep*. 2017;38(2):685–92.
32. Mvubu NE, Pillay B, Gamielidien J, Bishai W, Pillay M. Canonical pathways, networks and transcriptional factor regulation by clinical strains of Mycobacterium tuberculosis in pulmonary alveolar epithelial cells. *Tuberculosis (Edinb)*. 2016;97:73–85.
33. Morgensztern D, Campo MJ, Dahlberg SE, Doebele RC, Garon E, Gerber DE, et al. Molecularly targeted therapies in non-small-cell lung cancer annual update 2014. *J Thoracic Oncol*. 2015;10(1 Suppl 1):S1–63.
34. Shirley S, Micheau O. Targeting c-FLIP in cancer. *Cancer Lett*. 2013;332(2):141–50.
35. Riley J, Hutchinson R, McArt D, Crawford N, Holohan C, Paul I, et al. Prognostic and therapeutic relevance of FLIP and procaspase-8 overexpression in non-small cell lung cancer. *Cell Death Disease*. 2013;4(12):e951.
36. Yang X, Liu Y, Zong Z, Tian D. The Rho kinase inhibitor fasudil inhibits the migratory behaviour of 95-D lung carcinoma cells. *Biomed Pharmacother*. 2010;64(1):58–62.
37. Porter AP, Papaioannou A, Malliri A. Deregulation of Rho GTPases in cancer. *Small GTPases*. 2016;7(3):123–38.
38. Liu D, Mei X, Wang L, Yang X. RhoA inhibits apoptosis and increases proliferation of cultured SPCA1 lung cancer cells. *Mol Med Rep*. 2017;15(6):3963–8.
39. Wu XY, Liu WT, Wu ZF, Chen C, Liu JY, Wu GN, et al. Identification of HRAS as cancer-promoting gene in gastric carcinoma cell aggressiveness. *Am J Cancer Res*. 2016;6(9):1935–48.
40. Miglietta G, Gouda AS, Cogoi S, Pedersen EB, Xodo LE. Nucleic Acid Targeted Therapy: G4 Oligonucleotides Down-regulate HRAS in Bladder Cancer Cells through a Decoy Mechanism. *ACS Med Chem Lett*. 2015;6(12):1179–83.
41. Maldonado V, Melendez-Zajgla J. Role of Bcl-3 in solid tumors. *Mol Cancer*. 2011;10(1):152.
42. Dimitrakopoulos F-ID, Antonacopoulou AG, Kottorou A, Marousi S, Koukourikou I, Kalofonou M, et al. Variant of BCL3 gene is strongly associated with five-year survival of non-small-cell lung cancer patients. *Lung Cancer*. 2015;89(3):311–9.
43. Scagliotti GV, Selvaggi G, Novello S, Hirsch FR. The biology of epidermal growth factor receptor in lung cancer. *Clin Cancer Res*. 2004;10(12):4227s–s4232.
44. Ellis LM. Epidermal growth factor receptor in tumor angiogenesis. *Hematol/Oncol Clin*. 2004;18(5):1007–21.
45. Bergers G, Benjamin LE. Tumorigenesis and the angiogenic switch. *Nat Rev Cancer*. 2003;3(6):401–10.
46. Kashatus D, Cogswell P, Baldwin AS. Expression of the Bcl-3 proto-oncogene suppresses p53 activation. *Genes Develop*. 2006;20(2):225–35.
47. Wu J, Liu T, Rios Z, Mei Q, Lin X, Cao S. Heat shock proteins and cancer. *Trends Pharmacol Sci*. 2017;38(3):226–56.
48. Shatz M, Menendez D, Resnick MA. The human TLR innate immune gene family is differentially influenced by DNA stress and p53 status in cancer cells. *Can Res*. 2012;72:3948.
49. Pilarsky C, Wenzig M, Specht T, Saeger HD, Grützmann R. Identification and validation of commonly overexpressed genes in solid tumors by comparison of microarray data. *Neoplasia*. 2004;6(6):744–50.
50. Zhang X, Dang Y, Li P, Rong M, Chen G. Expression of IRAK1 in lung cancer tissues and its clinicopathological significance: a microarray study. *Int J Clin Exp Pathol*. 2014;7(11):8096.
51. Men W, Li W, Zhao J, Li Y. TIFA promotes cell survival and migration in lung adenocarcinoma. *Cell Physiol Biochem*. 2018;47(5):2097–108.
52. Li Y, Lian H, Jia Q, Wan Y. Proteome screening of pleural effusions identifies IL1A as a diagnostic biomarker for non-small cell lung cancer. *Biochem Biophys Res Commun*. 2015;457(2):177–82.
53. Sun T, Yu N, Zhai L-K, Li N, Zhang C, Zhou L, et al. c-Jun NH2-terminal Kinase (JNK)-interacting protein-3 (JIP3) regulates neuronal axon elongation in a kinesin-and JNK-dependent manner. *J Biol Chem*. 2013;288(20):14531.
54. Standen CL, Kennedy NJ, Flavell RA, Davis RJ. Signal transduction cross talk mediated by Jun N-terminal kinase-interacting protein and insulin receptor substrate scaffold protein complexes. *Mol Cell Biol*. 2009;29(17):4831–40.
55. Mostafaei S, Kazemnejad A, Jamalkandi SA, Amirhashchi S, Donnelly SC, Armstrong ME, et al. Identification of novel genes in human airway epithelial cells associated with chronic obstructive pulmonary disease (COPD) using machine-based learning algorithms. *Sci Rep*. 2018;8(1):15775.
56. Schmittgen TD, Livak KJ. Analyzing real-time PCR data by the comparative C_T method. *Nat Protoc*. 2008;3(6):1101.
57. Smyth GK. Limma: linear models for microarray data. *Bioinformatics and computational biology solutions using R and Bioconductor*: Springer; 2005. p. 397–420.
58. Zou H, Hastie T. Regularization and variable selection via the elastic net. *J R Stat Soc: Ser B (Stat Methodol)*. 2005;67(2):301–20.
59. Friedman J, Hastie T, Tibshirani R. The elements of statistical learning. volume 1. Springer; New York, NY, USA: 2001. Springer Series in Statistics) [Google Scholar].

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.