World Inventia Publishers

Journal of Pharma Research

http://www.jprinfo.com/

Vol. 8, Issue 6, 2019 ISSN: 2319-5622 USA CODEN: JPROK3

Research Article

ANALYSIS OF TUBERCULOSIS METAGENOME TOWARDS PATHOGENICITY AND SCREEN OF INTERLOGS IN HUMAN

Tanusree Chaudhuri *, Vidya Niranjan, A.H. Manjunath Reddy

Department of Biotechnology, R. V. College of Engineering, Bengaluru, Karnataka 560059, INDIA.

Received on: 22-04-2019; Revised and Accepted on: 10-06-2019

ABSTRACT

The infection by Mycobacterium tuberculosis (Mtb) depends greatly on how the host responds to the bacteria and its community and how the bacterial community manipulates the host. Thus, to understand this process, it is important to study the bacterial community first and identify the proteins that mediate pathogenicity. Metagenomic analysis is an approach that facilitates us to study these microbial communities. Once we understand these communities, there is a need for elucidating protein interactions between human and Mtb considering the existing microbiota present in the host. This may enable us to characterize specific molecular mechanisms allowing the bacteria to persist and survive under different environmental conditions. These interactions also will shed light on the interplay between Mtb and its human host. This might contribute to the process of repurposing existing drugs with new biological mechanisms of action. In this work, we identified 11 interlogs present in human host. After screening we have reported about 78 diseases, which have either direct or indirect effect on different types of tuberculosis. We believe studying these human interlogs targets and screening potential existing drugs for it, might repurpose some existing medicine towards the cure of tuberculosis.

KEYWORDS: *Tuberculosis, Metagenome, Network, Interlogs, Human diseases.*

INTRODUCTION

Tuberculosis (TB), caused by Mycobacterium tuberculosis (MTB), is a major global health concern according to the World Health Organization (WHO) report, 10.0 million people (range, 9.0–11.1 million) developed TB disease in 2017, where 5.8 million men, 3.2 million women and 1.0 million children. There were cases in all countries and age groups, but overall 90% were adults (aged ≥15 years), 9% were people living with HIV (72% in Africa) and two thirds were in eight countries: India (27%), China (9%), Indonesia (8%), the Philippines (6%), Pakistan (5%), Nigeria (4%), Bangladesh (4%) and South Africa (3%) and overall 1.3 million TB-related deaths in 2017 [1]. So, TB is a disease that can exist as a latent infection through-out the lifetime of a host, but that may be reactivated or re-acquired when host resistance is impaired [2]. In addition to the extremes of life, individuals become susceptible to active disease when suffering from poor nutrition, other serious medical conditions, or severe mental or physical stress. Today, it is estimated by the WHO that one-third of the global population is infected with tubercle bacilli but that only 10% of individuals will develop disease in their lifetime [1]. The implication of the high proportion of latent infections is that there has been a prolonged period of co-evolution of host and pathogen. Chan and colleagues in 1994 describe a mixed-strain infection with two Mycobacterium tuberculosis genotypes in the mummified body of Terézia Hausmann, who died in the late 18th century [3]. This dual infection would not have been detected with the methods of diagnosis in routine use today. The numbers of mixed infections are massively underestimated because the majority of diagnostic techniques rely on cultures grown from single colonies [4]. With the use of molecular biologic techniques or analysis of multiple colonies, 2 to 19% of patients in countries with a moderate-to-high incidence of tuberculosis have been found to be infected with multiple strains [5, 6]. This clearly, increases the number of TB-related deaths in single year is alarmingly higher than the roughly 300,000 deaths [1]. Further, the regimens recommended for the treatment of TB are complex, as Whole-genome sequencing provides comprehensive data on resistance mutations and strain typing for monitoring transmission, but unlike for conventional molecular tests, this has previously been achievable only from cultures of *M. tuberculosis* [7]. Metagenomics (also referred to as environmental and community genomics) is the genomic analysis of microorganisms by direct extraction and cloning of DNA from an assemblage of microorganisms. Metagenome contains vast variety of genome diversity of most microbes. Thus, metagenome data helps in developing computational methods

*** Corresponding author:**

Tanusree Chaudhuri

Department of Biotechnology,

R. V. College of Engineering, Bengaluru,

Karnataka 560059, INDIA.

** E-Mail: tanusree.ch3@gmail.com*

DOI: https://doi.org/10.5281/zenodo.3265321

that maximize the understanding of the genetic composition and the activities of the communities [8]. So, in case one want to analyse tuberculosis infection, should analyse the metagenomic data collected from tuberculosis infected patients.

In our present study, we have tried to analyze the long co evolutionary relationship between host and pathogen, through host pathogen interaction. To understand this interaction well, we have collected the metagenome of Mycobacterium tuberculosis complex for Indian Mycobacterium tuberculosis strains with Alr mutations. Our study revealed the evidences of the occurrence of recurrent TB was associated with a higher prevalence of mixed-strain infections. In our analysis, we have followed recurrent Host pathogen interaction through homology approach and identified 11 potential human transcripts. Our further analysis reveals that there 78 diseases that are direct or indirect relation with tuberculosis.

MATERIALS AND METHODS

1. ***Analysis of Metagenomic data:*** is downloaded from the NCBI-SRA TOOL KIT by giving the search term as "Mycobacterium tuberculosis and Human". We downloaded the sequence for the id ERR1394200 as it was having a good spot intensity served for our purpose using the command "./fastq-dump ERR1394280. And mapped it against H37RV as reference genome. We used Trim Galore to remove poor quality and adapter sequences in metagenome and Galaxy workflow for data integration, and data processing for screening of the metagenome against reference Mycobacterium genome as well as mapping against bacterial genome [9].

   Bam viewer tools are being used to visualise the sequence information and karaken is used for taxonomic ID conversion and pie chart mapping.

2. ***Pathogenic protein screening:*** MP3 is being used for screening of Pathogenic Proteins in the meta-genome. We used MP3 Software manually for the FASTA file of the aligned Metagenome file, to screen for pathogenic and non-pathogenic proteins is Mycobacterium. The pathogenic proteins and separated out of the file and put up with in an Excel sheet and with further reference with M.Tb pathogenic proteins (H37RV genome) [10].

3. ***Creation of Protein-protein interaction network:*** We analysed the protein-protein interaction network through String database which contains information from numerous sources, including experimental data, computational prediction methods and public text collections. We constructed the network by proving the protein along with specified organism as input. We screened the network through clustering approach. Here, in our study, we have followed MCL clustering strategy [11].

4. ***Screening of interlogs:*** Interlogs are the proteins that mediate interaction between Host and pathogen proteins. In this study, to identify the inetrlogs, we first identified the bacterial orthologs for M.Tb and mapped that with mouse to identify the human orthologs. We used TBLASTN and BLASTN for the screening purpose. The genome annotation data has been visualised by Integrative Genomics Viewer (IGV) [12].

5. ***Screening of human diseases:*** We screened human diseases through DisGeNET web server, which is a discovery platform containing one of the largest publicly available collections of genes and variants associated to human diseases. We uploaded our gene name for human and collected the disease names from the server [13].

RESULTS AND DISCUSSION

In our present study, at first, we tried to align *Mycobacterium tuberculosis* metagenome with M.Tb H37Rv full genome as the reference genome. Here, we have analyzed ERR1394280 metagenomic sample.

To perform this operation, we have done data preprocessing trim-galore. The trim report generated shows, our data contain good quality. So, the trimmed file generated from this tool is used for next analysis. To align our metagenome against reference genome, we used Bowtie 2 tool present in the GALAXY server. Bowtie 2 generated two separate files for aligned as well as unaligned reads. The aligned reads are from mycobacterium species and the unaligned reads are from some other species. This result clearly indicates that the metagenome does not entirely from the *Mycobacterium*, but it contains different types of other species.

So, the effect of *Mycobacterium* on human health can be understood well if we understand how infection with mycobacterium changes the microbiota present in human host [14].

For this process binning and search against all bacterial and viral genomes has been done through galaxy server. After that, we performed taxonomical analysis through KRAKEN tool, which provided a graphical output containing the details of abundance of the species. This file is converted to produce the taxonomic names, through KRONA PIE Chart. The abundance of the bacteria was found to be 24%, eukaryotes 60% and virus 14% in the meta-genome (Figure 1).

The above data (Fig. 1) clearly indicates that tuberculosis infection is an infection of immunocompromised host. It is quite evident from the above data that, consistent infection with Influenza A virus, increases the chances of infected with *M.tuberculosis*.

So, our study, suggested that, to identify the host protein interaction between M.Tb and host human, we should map the interaction between *M.Tuberculosis* and its orthologs. For this purpose, we decided to narrow down our search for only pathogenic proteins of *M.Tuberculosis*.

To satisfy this goal, we collected fasta sequences of the aligned reads from the Bowtie2 server and run in MP3 software. The software predicted 31 novel pathogenic proteins present in the M.Tb metagenome which belongs to Mycobacterium spp only (Table 1).

We constructed protein-protein interaction network with these proteins through String database, to identify the most essential proteins in *m.tb*. The protein-protein interaction network is shown in the following figure (Figure 2).

We used K-means clustering to cluster the proteins in the network. The k- means clustering identified 5 clusters in the networks, which are as follows (Table 2).

Through statistical analysis, we could identify 11 pathogenetic proteins in *Mycobacterium* which are very important for pathogenesis in tuberculosis. Hence, our study aimed to predict novel host pathogen interacting proteins which are related to these 11 proteins.

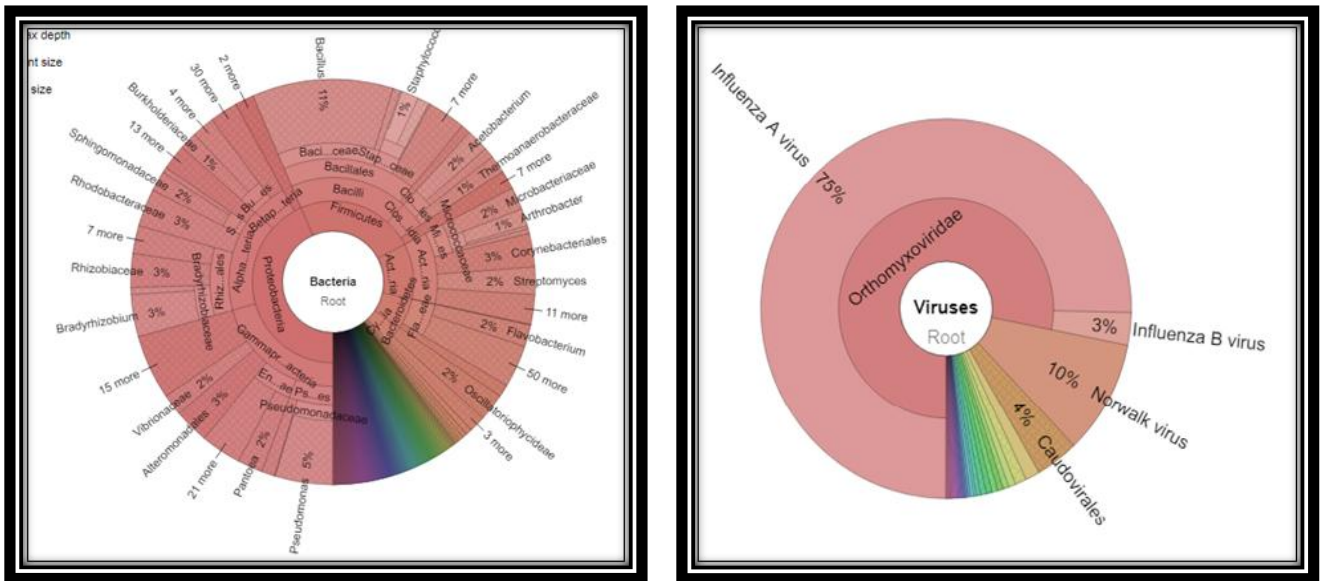
We decided to predict the host pathogen interaction through Interlog method. Here, we have considered these 11 proteins and run a genome-wide TBLASTN searches against all bacteria except Mycobacterium. Orthologous with more than 80% sequence identity have been forwarded to the next blast search against mouse. In order to find the human orthologs of M.Tb, we first mapped the proteins obtained from the blast search of mycobacterium against all bacteria and then mapped the output file against mouse, which gradually mapped to human.

We obtained a total set of 14 interlogs, which are human proteins act as mediator with TB proteins. These 14 proteins are as follows (Table3).

To narrow down our search once more, we tried to map these 14 proteins through the protein-protein interaction network construction (Figure 3).

So, we have taken the connected nodes MSX1,LBX1,SIX2,TLX1 for our further analysis.

So, we tried to map the diseases associated with these proteins through DisGenet tool and we got a total set of 78 diseases having direct or indirect contact with the prevalence of tuberculosis. These are as follows (Table 4).



a **b**

Fig. 1a: Presence of a) Bacteria, b) virus in tuberculosis metagenome

Table No. 1: Pathogenetic proteins in tuberculosis metagenome

Symbol	Genomic nucleotide accession. version	Start position on the genomic accession	End position on the genomic accession	Aliases	Function
esxD	NC_000962.3	4374049	4374372	Rv3891c	ESAT-6 like protein EsxD
moxR2	NC_000962.3	4133516	4134592	Rv3692	methanol dehydrogenase transcriptional regulator MoxR
cyp137	NC_000962.3	4127295	4128725	Rv3685c	cytochrome P450 Cyp137
Rv3636	NC_000962.3	4075752	4076984	Rv3636	pseudo
PE_PGRS58	NC_000962.3	4031404	4033158	Rv3590c	PE-PGRS family protein PE_PGRS58
Rv3528c	NC_000962.3	3964479	3965192	Rv3528c	hypothetical protein
PE_PGRS54	NC_000962.3	3931005	3936710	Rv3508	PE-PGRS family protein PE_PGRS54
yrbE4B	NC_000962.3	3919220	3920062	Rv3500c, supB	integral membrane protein
Rv3486	NC_000962.3	3905772	3906221	Rv3486	hypothetical protein
choD	NC_000962.3	3826991	3828727	Rv3409c	cholesterol oxidase

TB9.4	NC_000962.3	3585677	3585949	Rv3208A	hypothetical protein
Rv3191c	NC_000962.3	3557311	3558345	Rv3191c	transposase
PPE43	NC_000962.3	3076894	3078078	Rv2768c	PPE family protein PPE43
PPE43	NC_000962.3	3076894	3078078	Rv2768c	PPE family protein PPE43
speE	NC_000962.3	2928388	2929959	Rv2601	spermidine synthase
plsB2	NC_000962.3	2786914	2789283	Rv2482c	glycerol-3-phosphate acyltransferase
PPE38	NC_000962.3	2632923	2634098	Rv2352c	PPE family protein PPE38
PE_PGRS37	NC_000962.3	2387202	2387972	Rv2126c	PE-PGRS family protein PE_PGRS37
acg	NC_000962.3	2279129	2280124	Rv2032	NAD(P)H nitroreductase
PPE24	NC_000962.3	1981614	1984775	Rv1753c	PPE family protein PPE24
PPE21	NC_000962.3	1751297	1753333	Rv1548c	PPE family protein PPE21
Rv1490	NC_000962.3	1679322	1680629	Rv1490	membrane protein
PE_PGRS28	NC_000962.3	1636004	1638229	Rv1452c	PE-PGRS family protein PE_PGRS28
glpX	NC_000962.3	1227596	1228684	Rv1099c	fructose 1,6-bisphosphatase
PE_PGRS11	NC_000962.3	846159	847913	Rv0754	PE-PGRS family protein PE_PGRS11
fadB2	NC_000962.3	558895	559755	Rv0468	3-hydroxybutyryl-CoA dehydrogenase
Rv0435c	NC_000962.3	522347	524533	Rv0435c	ATPase
PPE6	NC_000962.3	372820	375711	Rv3873	PPE family protein PPE68
esxG	NC_000962.3	351525	351818	Rv0287, TB9.8	ESAT-6 like protein EsxG
mtn	NC_000962.3	99684	100451	Rv0091, pfs	5'-methylthioadenosine/S-adenosylhomocysteine nucleosidase
Rv0076c	NC_000962.3	85183	85572	Rv0076c	membrane protein

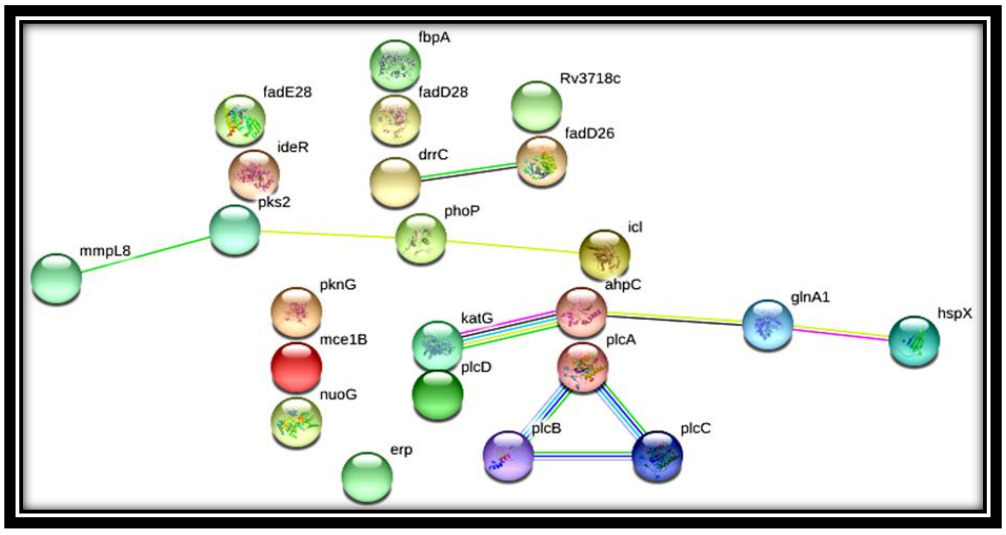


Fig. 2: Protein Protein interaction of the pathogenetic proteins in tuberculosis metagenome

Table No. 2: No of Pathogenetic Protein clusters present in Tuberculosis metagenome

Cluster Number	Cluster Colour	Gene Count	Protein Name	Protein Identifier
1	Red	3	plcC	83332.Rv2349c
1	Red	3	plcA	83332.Rv2351c
1	Red	3	plcB	83332.Rv2350c
4	Cyan	2	katG	83332.Rv1908c
4	Cyan	2	ahpC	83332.Rv2428
3	Green	2	pks2	83332.Rv3825c
3	Green	2	mmpL8	83332.Rv3823c
2	Yellow	2	drrC	83332.Rv2938
2	Yellow	2	fadD26	83332.Rv2930
5	Blue	2	icl	83332.Rv0467
5	Blue	2	phoP	83332.Rv0757

Table No. 3: Potential Human interlogs between tuberculosis infections

S. No.	Name
1	ARX
2	CTCF
3	DLX3
4	EN2
5	FAM198B
6	HPSE2
7	LBX1
8	MSX1
9	NAGLU
10	NEUROG3
11	SIX2
12	SLC8A1
13	TLX1
14	VAX1

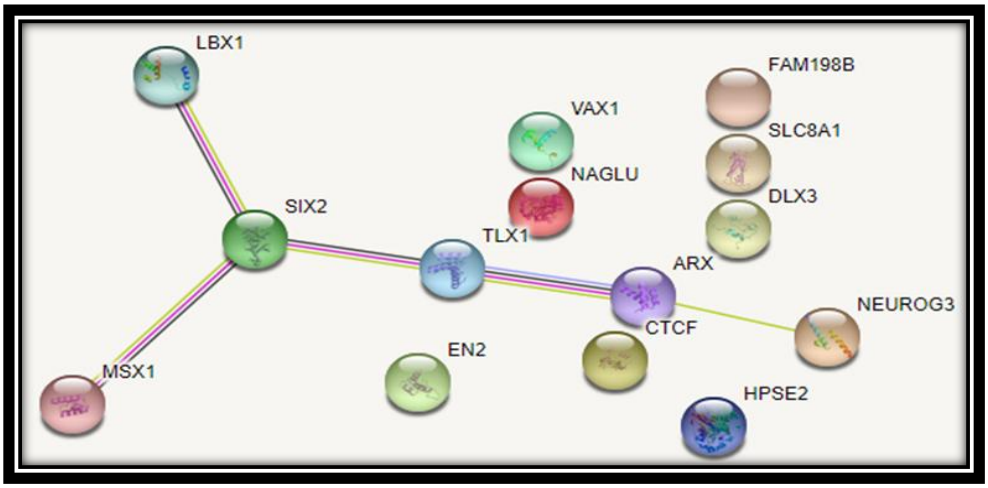


Fig. 3: Protein protein interaction between potential interlogs

Table No. 4: Potential diseases that have direct or indirect effect in tuberculosis

Gene Name	Disease ID	Disease Name
TLX1	C0023493	Adult T-Cell Lymphoma/Leukemia
SIX2	C1862939	AMYOTROPHIC LATERAL SCLEROSIS 1
TLX1	C0002938	Aneuploidy
MSX1	C1290511	Anodontia of Permanent Dentition
LBX1	C0005586	Bipolar Disorder
MSX1	C0008924	Cleft Lip
MSX1	C1298692	Cleft lip and alveolus

MSX1	C0008925	Cleft Palate
MSX1	C0158646	Cleft palate with cleft lip
CTCF	C1864897	Cognitive delay
MSX1	C0376634	Craniofacial Abnormalities
CTCF	C0010417	Cryptorchidism
MSX1	C4280643	Decreased projection of maxilla
MSX1	C4280611	Decreased size of teeth
MSX1	C4280612	Decreased width of tooth
MSX1	C4280642	Deficiency of upper jaw bones
CTCF	C0262444	Dental abnormalities
MSX1	C0399352	Developmental absence of tooth
CTCF	C4020876	Dull intelligence
MSX1	C1853246	Everted lower lip vermilion
CTCF	C0232466	Feeding difficulties
MSX1	C0423867	Fine hair
MSX1	C1856963	Fragile nails
SIX2	C1876203	Frontonasal dysplasia
CTCF	C0557874	Global developmental delay
CTCF	C0020490	Hyperopia
MSX1	C0020608	Hypodontia
MSX1	C0406716	Hypodontia and nail dysgenesis
MSX1	C0240310	Hypoplasia of the maxilla
MSX1	C1855694	Hypoplastic deciduous teeth
MSX1	C0406735	Hypoplastic enamel-onycholysis-hypohidrosis syndrome
MSX1	C1856786	Hypoplastic fingernails
MSX1	C1857130	Hypoplastic mandible condyle
MSX1	C1837279	Hypoplastic toenails
MSX1	C4280641	Hypotrophic maxilla
CTCF	C3714756	Intellectual Disability
SIX2	C0022658	Kidney Diseases
MSX1	C0221261	Koilonychia
MSX1	C0239174	Late tooth eruption
CTCF	C1860789	Leukemia, Megakaryoblastic, of Down Syndrome
CTCF	C0423903	Low intelligence
CTCF	C0079773	Lymphoma, T-Cell, Cutaneous
LBX1	C1458155	Mammary Neoplasms
MSX1	C0240295	Mandibular hypoplasia
MSX1	C4082243	Maxillary retrognathia
MSX1	C0240340	Microdontia (disorder)
MSX1	C1856203	Microdontia of primary teeth
MSX1	C0025990	Micrognathism
MSX1	C0263523	Micronychia (disorder)
CTCF	C0026827	Muscle hypotonia
TLX1	C0007621	Neoplastic Cell Transformation
MSX1	C4082304	Oligodontia
MSX1	C1837210	OROFACIAL CLEFT 5
MSX1	C0266037	Peg-shaped teeth
MSX1	C0150993	Pitting of nails
MSX1	C4012359	Pointed teeth
CTCF	C1843367	Poor school performance
TLX1	C1961099	Precursor T-Cell Lymphoblastic Leukemia-Lymphoma
MSX1	C1866234	Protruding lower lip
MSX1	C4280640	Retrusion of upper jaw bones
MSX1	C1849392	Ridged fingernails
MSX1	C0423820	Ridged nails
CTCF	C0036341	Schizophrenia
CTCF	C0349588	Short stature
CTCF	C0424688	Small head
MSX1	C1837770	Sparse hair
TLX1	C0037997	Splenic Diseases

MSX1	C0038356	Stomach Neoplasms
CTCF	C0038379	Strabismus
CTCF	C0578038	Thin lips
MSX1	C3554113	Thin toenails
MSX1	C1860844	Thin, sparse hair
CTCF	C0040427	Tooth Abnormalities
MSX1	C0040427	Tooth Abnormalities
MSX1	C0457756	Tooth absent
MSX1	C3489529	Tooth Agenesis, Familial
MSX1	C2981150	Uranostaphyloschisis
MSX1	C1956097	Wolf-Hirschhorn Syndrome

CONCLUSION

Our study clearly indicates that, tuberculosis is predominant in the immunocompromised host. And since these diseases show either direct or indirect interaction with tuberculosis, the potential drugs for these diseases can be the potential target for drug repurposing to cure tuberculosis.

REFERENCES:

1. Global tuberculosis report **2018**, https://www.who.int/tb/publications/global_report/en/
2. Madhukar Pai, Marcel A. Behr, David Dowdy, Keertan Dheda, Maziar Divangahi, Catharina C. Boehme, Ann Ginsberg, Soumya Swaminathan, Melvin Spigelman, Haileyesus Getahun, Dick Menzies & Mario Raviglione. Tuberculosis. Nature Reviews Disease Primers **2016**;2: Article number: 16076.
3. Chan JZ, Sergeant MJ, Lee OY, Minnikin DE, Besra GS, Pap I, Spigelman M, Donoghue HD, Pallen MJ. Metagenomic analysis of tuberculosis in a mummy. N Engl J Med **2013**; 369(3):289-90.
4. Suzanne M. Hingley-Wilson. Metagenomic Analysis of Tuberculosis - Current Limitations. N Engl J Med **2013**; 369:1572.
5. Tran N. Buu, Dick van Soolingen, Mai NT. Huyen, Nguyen TN. Lan, Hoang T. Quy, Edine W. Tiemersma, Kristin Kremer, Martien W. Borgdorff, and Frank GJ. Cobelens. Increased Transmission of Mycobacterium tuberculosis Beijing Genotype Strains Associated with Resistance to Streptomycin: A Population-Based Study. PLoS One **2012**; 7(8):e42323.
6. Juan Garberi, Jorge Labrador, Federico Garberi, Juan Ezequiel Garberi, Julian Peneipil, Miguel Garberi, Luis Scigliano and Alcides Troncoso. Diagnosis of Mycobacterium tuberculosis using molecular biology technology. Asian Pac J Trop Biomed **2011**;1(2):89–93.
7. Amanda C. Brown, Josephine M. Bryant, Katja Einer-Jensen, Jolyon Holdstock, Darren T. Houniet, Jacqueline Z. M. Chan, Daniel P. Depledge, Vladyslav Nikolayevskyy, Agnieszka Broda, Madeline J. Stone, Mette T. Christiansen, Rachel Williams, Michael B. McAndrew, Helena Tutill, B Julianne Brown, Mark Melzer, Caryn Rosmarin, Timothy D. McHugh, Robert J. Shorten, Francis Drobniewski,d Graham Speight and Judith Breuerb. Rapid Whole-Genome Sequencing of Mycobacterium tuberculosis Isolates Directly from Clinical Samples. J Clin Microbiol **2015**;53(7):2230–2237.
8. The New Science of Metagenomics: Revealing the Secrets of Our Microbial Planet. National Research Council (US) Committee on Metagenomics: Challenges and Functional Applications. Washington (DC): National Academies Press (US); **2007**.
9. Kozich JJ, Westcott SL, Baxter NT, Highlander SK. Schloss PD.: Development of a dual-index sequencing strategy and curation pipeline for analyzing amplicon sequence data on the MiSeq Illumina sequencing platform. Applied and Environmental Microbiology. Appl Environ Microbiol **2013**;79(17):5112-5120.
10. Gupta A, Kapil R, Dhakan DB, Sharma VK. MP3: a software tool for the prediction of pathogenic proteins in genomic and metagenomic data. PLoS One **2014**; 9(4):e93907.
11. Damian Szklarczyk, John H Morris, Helen Cook, Michael Kuhn, Stefan Wyder, Milan Simonovic, Alberto Santos, Nadezhda T Doncheva, Alexander Roth, Peer Bork, Lars J. Jensen, and Christian von Mering. The STRING database in 2017: quality-controlled protein–protein association networks, made broadly accessible. Nucleic Acids Res **2017**;45(Database issue):D362–D368.
12. Suyu Mei, Erik K. Flemington, and Kun Zhang, Transferring knowledge of bacterial protein interaction networks to predict pathogen targeted human genes and immune signaling pathways: a case study on M. tuberculosis, BMC Genomics **2018**;19:505.
13. Janet Piñero, Núria Queralt-Rosinach, Àlex Bravo, Jordi Deu-Pons, Anna Bauer-Mehren, Martin Baron, Ferran Sanz, and Laura I. Furlong. DisGeNET: a discovery platform for the dynamical exploration of human diseases and their genes, Database (Oxford). **2015**; 2015:bav028.
14. Bo-Young Hong, Nancy Paula Maulén, Alexander J. Adami, Hector Granados, María Elvira Balcells, and Jorge Cervantes. Microbiome Changes during Tuberculosis and Antituberculous Therapy. Clin Microbiol Rev **2016**;29(4): 914–925.

How to cite this article:

Tanusree CH, et al. ANALYSIS OF TUBERCULOSIS METAGENOME TOWARDS PATHOGENICITY AND SCREEN OF INTERLOGS IN HUMAN. J Pharm Res 2019;8(6):395-401. **DOI: https://doi.org/10.5281/zenodo.3265321**

Conflict of interest: The authors have declared that no conflict of interest exists.

Source of support: Nils