lncRNA Mediated Hijacking of T-cell Hypoxia Response Pathway by Mycobacterium Tuberculosis Predicts Latent to Active Progression in Humans

Jyotsana Mehra[1#], Vikram Kumar[1#], Priyansh Srivastava[2#], Tavpritesh Sethi[1*]

jyotsana18237@iiitd.ac.in[1], vikram18250@iiitd.ac.in[1], spriyansh29@gmail.com[2], tavpriteshsethi@iiitd.ac.in[1*]

[1]Indraprastha Institute of Information Technology, Delhi, India, [2]Amity University, Noida, Uttar Pradesh, India

[#]Equal Contribution, [*]Corresponding Author

Abstract- **Cytosolic functions of Long non-coding RNAs including mRNA translation masking and sponging are major regulators of biological pathways. Formation of T cell- bounded hypoxic granuloma is a host immune defense for containing infected Mtb-macrophages. Our study exploits the mechanistic pathway of Mtb-induced HIF1A silencing by the antisense lncRNA-HIF1A-AS2 in T cells. Computational analysis of in-vitro T-cell stimulation assays in progressors(n=119) versus latent(n=221) tuberculosis patients revealed the role of lncRNA mediated disruption of hypoxia adaptation pathways in progressors. We found 291 upregulated and 227 downregulated lncRNAs that were correlated at mRNA level with HIF1A and HILPDA which are major players in hypoxia response. We also report novel lncRNA-AC010655 (AC010655.4 and AC010655.2) in cis with HILPDA, both of which contain binding sites for the BARX2 transcription factor, thus indicating a mechanistic role. Detailed comparison of infection with antigenic stimulation showed a non-random enrichment of lncRNAs in the cytoplasmic fraction of the cell in progressors. The lack of this pattern in non-progressors indicates the hijacking of the lncRNA dynamics by Mtb. The in-vitro manifestation of this response in the absence of granuloma indicates pre-programmed host-pathogen interaction between T-cells and Mtb regulated through lncRNAs, thus tipping this balance towards progression or containment of Mtb. Finally, we trained multiple machine learning classifiers for reliable prediction of latent to the active progression of patients, yielding a model to guide aggressive treatment.**

Keywords- ***Mycobacterium tuberculosis*, Long non-coding RNAs, Hypoxia, Granuloma, Differential Expression**

Introduction

Long non-coding RNAs (lncRNAs) are defined as transcripts of lengths exceeding 200 nucleotides that are not translated into protein and form the major part of the non-coding transcriptome. Genome-wide association studies (GWAS) have evaluated their role in disease progression and development. They play a crucial role in gene expression by controlling the translational freedom of protein-coding genetic elements[1]. Previous findings have escalated the role of Nuclear-enriched abundant transcript 1 (lncRNA-NEAT1) as an important player in immune regulation of Tuberculosis (TB) prognosis[2]. 2016's Microarray studies of Mtb infected macrophages eluted out MIR3945HG V1 and MIR3945HG V2, as novel biomarkers for TB that play a vital role in Mtb-macrophage interaction[3]. In the successive year lncRNAs role in the regulation of alpha-beta T cell activation and the T cell receptor signalling pathway have been actively studied making them a potent early-diagnosable biomarker of TB[4]. Experimental studies of the previous year have also documented the effects of lncPCED1B-AS on macrophage apoptosis[5]. Moreover, lncRNA also controls CD8+ Immune Responses[6].

Mycobacterium tuberculosis (Mtb) first infects the Macrophages through vesicle trafficking events by disrupting the antigen processing and presentation pathway which includes

phagosome-lysosome fusion[7]. Once the soluble antigens are presented to the CD4+ T cells by the un-infected dendritic cells the adaptive-immune response is stimulated. As a result of which these virgin T cells become polyfunctional through antigen-driven differentiation[8]. Among all the antigens, diagnostic antigens ESAT and Ag85 play a vital role in disease progression and pathogenesis[9,10]. ESAT-6 inhibits NF-κB activation that restricts myeloid differentiation, whereas the Ag85 prevents the formation of phagolysosome[9,10]. Hypoxia-induced dormancy of Mtb is one of the primary goals of the granuloma formation by T cells and other immune cells[11]. However, adaptive survival of Mtb in hypoxic granulomas has been extensively studied[12]. CD8+ T cells are the major killers of dormant Mtb as they go deep in the granulomas[13]. To withstand the hypoxic environment of granulomas T cells induce HIF-1A which is one of the pioneer genes involved in the hypoxic-homeostasis[14].

The present study found the mRNA silencing potential of lncRNA-HIF1A-AS2 in T cells which disrupts the hypoxia adaptation pathways in progressors. According to our proposed hypotheses, lncRNA-HIF1A-AS2 silences its anti-sense mRNA-HIF-1A and is induced by Mtb during latent to active TB progression. The adaptation mechanism is also assisted with HILPDA which is responsible for lipid accumulation as in low oxygen environments. We applied statistical procedures to find out differentially expressed genes in Mtb-infected T cells samples against Antigen-stimulated T cells samples, to decipher the Mtb T-cell interaction during active TB progression. Statistical inferences from our study show that 291 upregulated lncRNAs that were correlated at mRNA level with HIF1A and HILPDA which are major players in hypoxia response. Computational analysis of in-vitro T-cell stimulation assays in progressors versus latent tuberculosis patients also shows major differences in the chromosomal distribution of genes and sub-cellular localization of the DE lncRNAs. Additionally, we trained a machine learning classifier using Random forest statistics to predict the latent samples with potential to become active.

Methods

2.1 Data Collection and Curation

The data for analysis was taken from the expression profiling study carried on 150 adolescents (12-18 years), with a gap cycle of 6 months[15]. The reads were recorded using Illumina HiSeq 2000 (*Homo sapiens*) (GPL11154) and are available at GEO with Accession ID "GSE103147"[15,18]. 106 blood samples of non-progressors (Mtb infected controls) and 44 blood samples of progressors (developed TB during two years of followup) constituted the initially reported data set. A total of 1650 raw RNASeq reads of all 150 blood samples were systematically retrieved from Gene Expression Omnibus through GEOquery using Bioconductor (R.3.6)[16-18]. We proceeded with 340 samples which included 119 progressors and 221 non-progressors under the umbrella of "0 Day Collection", having the four distinctive conditional classes as "Mtb Infected", "Ag85 Stimulated", "ESAT Stimulated" and "Unstimulated" as control. We only analyzed the RNASeq data from T cells variants as it plays a vital role in the adaptive immune response against Mtb invasion.

2.2 Construction of alignment maps

All spliced reads in FASTQ files after retrieval were first aligned with the human genome to produce the alignment maps. Hierarchical Indexing for Spliced Alignment of Transcript (HISAT) was used for the alignment of the reads with the GRCh37-hg19 database[19]. All the alignment maps (.SAM files) were converted into binary aligned maps (.BAM files) using Sam tools[20].

2.3 Extraction of lncRNAs and mRNAs

Ultrafast Comprehensive Long Non-Coding RNA (UClncR) pipeline was implemented on the binary aligned maps files that use GENCODE to detect the lncRNA[21,22]. Since the expression values of lncRNAs are very low, therefore, Reads per Kilobase per Million Mapped Reads (RPKM) normalized expression matrix was taken for further analysis after UClncR[21]. 68,000

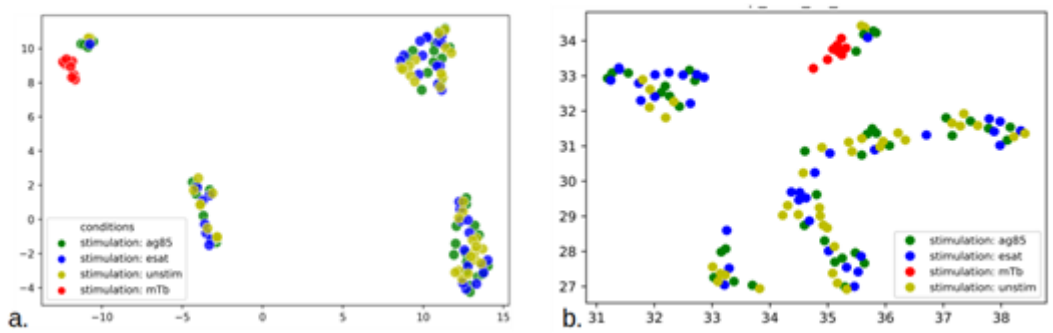

Figure 1 (a). Clustering of the expression data classes of the lncRNAs derived through the UClncR pipeline, classes include Stimulated by Ag85 in red, Stimulated by ESAT in green, Infected by Mtb in violet, Unstimulated in blue, clusters of Mtb are isolated against Stimulated sets of ESAT and Ag85; (b) Clustering of the expression data classes of the mRNA derived through the Ht-Seq count pipeline, classes include Stimulated by Ag85 in red, Stimulated by ESAT in green, Infected by Mtb in violet, Unstimulated in blue, clusters of Mtb are isolated against Stimulated sets of ESAT and Ag85

lncRNAs were generated by the UClncR. The 54,837 un-annotated Ensemble Ids of 68,000 lncRNAs were computationally removed and 13,870 were taken for further processing. Out of 13,870 Ensemble Ids the low expression (0 expression count) i.e. 2,382 were systematically removed. Similarly, 55,190 mRNAs were generated using the HtSeq-count[23]. After mapping with RefSeq db through Biomart, only 17,469 protein-coding mRNAs were obtained[24,25]. Eventually, 16,969 mRNAs were taken for differential gene expression studies after removal of 0 expression counts.

2.4 Differential Expression Analysis

DE lncRNAs were statistically filtered out by applying analysis of variance (ANOVA)[26]. ANOVA was applied on three classes ("Mtb Infected", "Ag85 Stimulated", and "ESAT Stimulated"), with all the 340 samples with the false discovery rate (FDR) of <0.01[26]. Differential expression analysis for lncRNA is done using lncDIFF, as the lncRNAs show very low expression as compared to mRNA[27]. Similarly, ANOVA was applied on all the 340 samples of mRNA with three classes with the FDR cut-off of <0.0.1[26]. Differentially expressed genes were calculated using EdgeR[28]. And similarly, differentially expressed mRNAs were further reduced on the basis of the FDR cut-off of <0.01.

2.5 Gene ontology and pathway enrichment analysis

ShinyGO was used for gene ontology enrichment and pathway enrichment analysis [29]. In the case of differential mRNAs, we selected the Gene Ontology database (GO) and pathway enrichment analysis was done vis Kyoto Encyclopedia of Gene and Genome (KEGG)[30,31]. The most significant GO terms and pathways were filtered using the p-value cut-off of <0.05. Since lncRNAs have low enrichment for GO and KEGG pathways, therefore, Pearson-correlation test among differentially expressed mRNAs (DE mRNAs) and differentially expressed lncRNAs (DE lncRNAs) was performed with the filter of -0.5 to +0.5. The correlated mRNAs were then enriched using the ShinyGO with the p-value cut-off of <0.05 with GO and KEGG[29-31].

2.6 Chromosomal Mapping and Subcellular localization

Biomart service was used to map the chromosomal location of the differentially expressed lncRNAs and mRNA that were having correlation[25]. The correlated pairs with the heterogeneous chromosomal numbers were systematically dropped. Filter criteria (>1000 bp) was applied to lncRNA-mRNA pairs on the same chromosome to study co-expression relationships. JASPER database was used to retrieve the profiles of the transcription factor binding sites in the

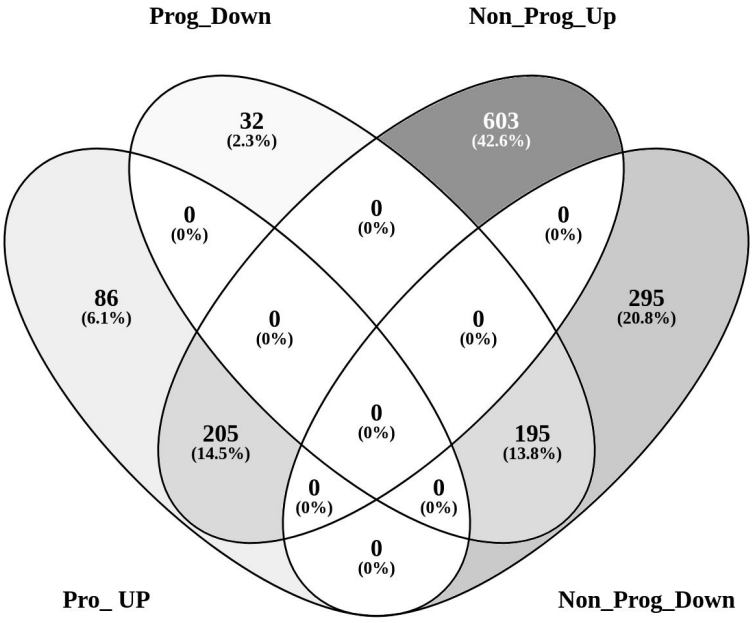

Figure 2. Venn diagram for comparative analysis of the differential expression the lncRNAs in the progressive and non-progressive samples of tuberculosis.

transcripts of *Homo sapiens*[32]. To map each transcription site to their respective chromosome location we used Ciiider with GRCh38.p12 assembly[33,34]. For subcellular localization analysis, we specifically used the raw data of B-cells from LncATLAS database and programmatically mapped them to the respective DE lncRNAs[35]. For deriving the coding potential of nuclear elements we used Coding Potential Calculator[36].

2.7 Progression Classifier Construction

We estimated the differentially expressed mRNA between the latent and progressor samples of TB using the EdgeR[28]. Since the lncRNA profiles showed exclusive enrichment in the cytosol, we leveraged the signal by combining the DE lncRNA enriched in the cytoplasm with DE m RNAs. The dataset was divided into training and testing sets with 80:20 ratio. We implemented the least absolute shrinkage and selection operator (LASSO), K-Nearest Neighbor, Random Forest (RF) and Decision Tree classifiers to check the accuracy, sensitivity and specificity of the model. Hyperparameters of these models such as lambda (0.017, lasso) and number of trees (10000, RF classifier) were optimized for model selection.

Results

3.1 Expression Data

Visualization of 4 classes ("Mtb Infected", "Ag85 Stimulated", "ESAT Stimulated" and "Unstimulated"), showed the isolated clusters of "Mtb infected" against the other 3 classes, in case of both mRNAs and lncRNAs (Figure 1). Unstimulated sets acted as the control for the other three classes. To generate insights on the underlying mechanisms of the key lncRNAs we dropped the set of "Unstimulated T cells" and proceeded with the "Infection Set" vs "Stimulated Set" for different gene analysis in both the progressors and non-progressor samples.

3.2 Differential Expression analysis

For lncRNA, after lncDIFF, a total of 11,488 differentially expressed lncRNAs were isolated in the stimulus set of progressors. 10,964 lncRNAs were dropped from a total of 1,148 based on the FDR cut-off of <0.01. Similarly, for mRNAs, a total of 1,148 based on the FDR cut-off of <0.01 were taken for further analysis. In the case of non-progressors, 1595 lncRNAs and 5859 mRNAs were eluted out (Fig. 2).

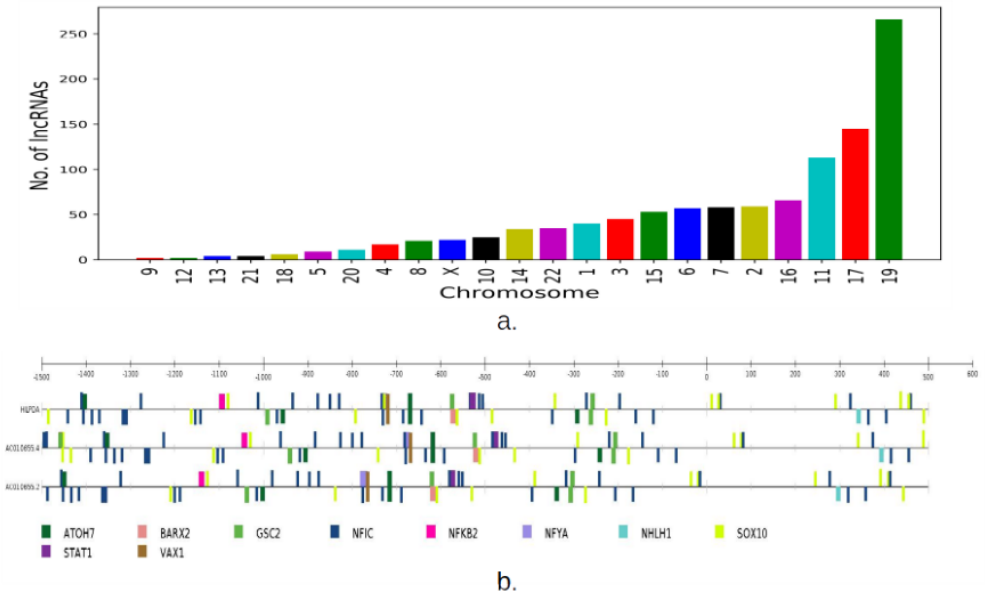

Figure 3 ((a) lncRNA distribution across the human genome; Chromosome 19 being the smaller chromosome holds a higher number of lncRNA as compared to the chromosome 1,2 and 3; (b) Transcription factor binding site of lncRNA AC010655.4 and lncRNA AC010655.2 in close proximity of mRNA HILPDA on the same chromosome location. Designed using Ciiider and JASPER profiles)

3.3 GO and pathway enrichment analysis

KEGG pathway analysis of the DE mRNA-DE lncRNAs correlation pairs were enriched with Thermogenesis (20 gene hits), Parkinson's Disease (14 gene hits) and Systemic lupus erythematosus (13 gene hits) (Supp. Table 1, Appendix II). However, for DE mRNA, only Parkinson's Disease pathway with 34 gene hits and Thermogenesis with 46 gene hits were enriched (Supp. Table 1, Appendix II). GO term enrichment analysis upvoted 190 for molecular functions as Nucleic-acid binding, 184 for cellular compartments as Nucleosome and 116 for Phosphorylation as biological processes in case of DE mRNA-DE lncRNAs correlation pairs (Supp. Table 1, Appendix II). In the case of DE mRNAs, 27 for cellular compartment "Nucleosome" and 340 for biological processes as "Phosphorylation" were enriched using GO (Supp. Table 1, Appendix II). Surprisingly, the gene set enrichment analysis of the non-progressive DE mRNA-DE lncRNAs correlation pairs yielded only basic cellular pathways like "RNA transport" (155 hits) and "Metabolic pathways" (298 hits), that clearly shows the cell is trying to maintain its natural metabolism.

3.4 Chromosomal Mapping

Upon chromosomal mapping of all the progressive DE mRNA-DE lncRNAs correlation pairs, we found out the pairs that were located in the close proximity of 1000 base pairs. Interestingly the mRNA-PBX2 and lncRNA-AL662884.1 reside on the same location on chromosome 6. Hypoxia-Inducible Lipid Droplet Associated (HILPDA) was mapped in close proximity to AC010655.4 and AC010655.2 on chromosome 5 (Supp. Table 2). It is also interesting to see that chromosome 19, being the smaller in size, holds more lncRNA than larger chromosomes like 1, 2, and 3 [Figure 3(a)]. After the transcription factor binding site prediction on the HILPDA- AC010655.4 and AC010655.2 trio we found 10 (ATOH7, BARX2, GSC2, NFIC, NFKB2, NFYA, NHLHI, SOX10, VAX1) common transcription factors among them [Figure 3(b)]. Comparatively, the chromosomal mapping of all the non-progressive DE mRNA-DE lncRNAs correlation pairs found no pairs that were located in the close proximity of 1000 base pairs. Moreover, the chromosomal distribution of the non-progressive DE lncRNA was significantly different in comparison to the progressive DE lncRNA's [Suppp. Figure 3]. Chromosome 1 is the largest chromosome that holds a higher number of lncRNA as expected.

Table 1. Model Performance Indicators for Prediction of Progression

	LASSO Classifier	KNN Classifier	Random Forest	Decision Tree
Accuracy	72.46%	63.76%	69.56%	60.86%
Sensitivity	45.45%	38.46%	57.14%	36.84%
Specificity	85.11%	69.64%	70.96%	70.0%
F 1 Score	0.50	0.61	0.64	0.60

3.5 Subcellular localization

Localization signal analysis revealed the location of 175 lncRNAs from our set of upregulated lncRNAs in the Mtb-infected set. Out of 175 lncRNA, 125 lncRNA were enriched in the nucleus with random expression value distribution i.e. 63 upregulated and 62 downregulated (Appendix III). However, in the case of cytosol 40 lncRNAs were upregulated in the cytosol in comparison to only 11 being downregulated (Appendix III). This major upregulation of lncRNAs clearly shows the evidence for cytosolic activity. Interestingly our finding enriched lncRNA-HIF1A-AS2 in the cytosol which is an anti-sense to HIF1A. The subcellular localization of the DE lncRNAs from non-progressive samples showed random distribution both in nucleus and cytosol. We also compared the DE genes (both of progressors and non-progressors) with the human housekeeping gene data set and found that the number of housekeeping genes in non-progressor DE (4 genes) expression set was more in comparison to progressor DE expression set (1 gene). This shows that in case of progression normal functions of T cells are downgraded as they move towards hypoxic stress.

3.6 Classifier Predictions

Random forest and lasso classifiers were found to be independently useful on the basis of model performance indicators (Table 1), While RF had a highest F1 score(64%) and sensitivity(57.14%) for predicting progressors, lasso had the best specificity(85.11%) and overall accuracy(72.46%). Therefore, the lasso model can be used for decisions to aggressively treat with a low false positive rate whereas Random Forest can be used to triage for treatment escalation.

Discussion

Our study shows that 291 upregulated and 227 downregulated lncRNAs in Mtb infected T cells progressive samples are the major key players. Upon validating with KEGG pathway enrichment analysis, 34 of the DE lncRNA were enriched with Parkinson's Disease pathway and Thermosgeneis pathway combined. What is more interesting to see is that both pathways have DE lncRNA sets enriched within the lumen of Mitochondria whereas non-progressive showed no such shreds of evidence of mitochondria. Concrete evidence for many mitochondria-associated lncRNAs in the regulation of mitochondrial bioenergetics and cross-talk with nuclei has already existed in literature[37]. Our experimental design also enriched NEAT1 as an upregulated gene in the antigen-stimulated set of progressive samples. Role of NEAT1 is widely studied for its paraspeckle formation and paraspeckle dependent cell differentiation[38]. Upregulated levels of NEAT1 could manipulate the cytokine expression through the JNK/ERK MAPK signalling pathway in Macrophages[39]. It has also been reported as a potential biomarker in the previous studies[40]. Our study proposes 118 downregulated DE mRNAs which are highly correlated with the low expressions of NEAT1 in the Mtb-infection set. Since NEAT1 is involved in nuclear retention of A-I mRNAs, therefore, it might be responsible architecturally blocking the expression of 118 DE mRNAs in the nucleus.

After the formation of hypoxic granuloma by the host immune system, T cells surround the infected macrophages to contain the infection. CD8+ cells being cytotoxic digs deep inside the granuloma to kill the infected macrophages[13]. For effective cytotoxic functions, T cells undergo hypoxic reprogramming which results in the

secretion of hypoxic genes such as HILPDA and HI1FA. Our study on non-progressors samples shows the overexpression of lncRNA-HIF1A-AS2 in the cytosol as well, which is antisense to HIF1A. This clearly shows that T cells undergo hypoxic stress even during non-progression. Interestingly there was no expression of HILPDA which is mainly responsible for lipid accumulation during the hypoxic condition. To withstand the hypoxic environment HIF1A is induced by the T cells. Our finding enriched lncRNA-HIF1A-AS2 in the cytosol which is an anti-sense to HIF1A. Coding potential of lncRNA-HIF1A-AS2 was very low, which clearly indicated it's non-protein coding nature. Thus, expression and localization of lncRNA-HIF1A-AS2 in the cytosol are clearly due to the regulation of cytosolic functions of lncRNAs such as RNAi. In our expression set, the levels of HIF1A fail to pass the cut-off criteria of P-value but were majorly upregulated across the samples. Therefore, it is very much evident from our data that lncRNA-HIF1A-AS2 could mask the translation of HIF1A in the cytosol through RNAi.

We also report novel lncRNA-AC010655 (AC010655.2 and AC010655.4) for their consensus binding to HILPDA linked transcription factors in the nucleus. This competitive binding to transcription factors of HILPDA such as BARX2 to lncRNA-AC010655. HILPDA is responsible for lipid accumulation as the cell slowly changes its metabolism towards low oxygen environments. This change in metabolism is reflected in mitochondrial pathways, which are actively enriched in our data set of differentially expressed lncRNAs. It is clear that non-progressors of TB also show the formation of granuloma as well, which contain infection inside the macrophage. Therefore, we report HILPDA as one of the major biomarkers of progressive TB.

In order to exploit these mechanistic insights, we chose to construct machine learning models that leverage the exclusive presence of a cytosolic signal in lncRNAs in combination with the differentially expressed mRNAs. Mechanistically enhanced machine learning models have not been applied to predict tuberculosis progression to the best of our knowledge and may avoid the pitfalls of black-box predictions which may not be actionable in the real world settings. Our models also revealed an interesting conflict between the decision to choose between the overall best model and the potential to pivot clinical decision making and therapeutic implications. RF, though the best performing model based upon F1 score was more sensitive, but not high enough to change decisions. On the other hand, LASSO classifier had a high sensitivity (85%) but a marginally lower F1 score. We propose the use of LASSO classifier, which has low false positive rate for informing decisions to escalate treatment in predicted progressors. On the other hand, RF may be useful for triaging predicted progressors for treatment escalation as it has higher sensitivity. This study has several limitations. The mechanistic insights are derived from the correlation between lncRNA and mRNAs and using mRNA enrichment in pathways as a surrogate for lncRNA function. We have attempted to mitigate some of the false associations by restricting ourselves to the lncRNAs that are present in cis with the correlated mRNA. However this may have led to missing out on the trans acting lncRNAs. We also enriched for this effect by investigating the transcription factor binding sites that are common to the lncRNA-mRNA pairs thus increasing the probability of a functional interaction, which can be confirmed only through wet-lab experiments. Nonetheless, our study revealed a strong signal for the modulation of the hypoxic response pathway which is evident in Mtb infection but not in antigenic stimulation, thus indicating the hijacking of this machinery by Mtb specifically through cytoplasmically enriched lncRNAs. Finally, our prediction models are highly specific, but these can guide decisions in persons who may already be suspect for progression (confirmation) as opposed to a sensitive screen.

Conclusion

Our study concludes that a non-random enrichment of lncRNAs in the cytoplasm, specifically those associated with the hypoxic response pathways dictates the progression of Mtb from latent to active state. Our analyses indicate the mechanisms of such regulation and its potential to predict progression using machine learning models.

Conflict of Interest

Authors declare no conflict of interest.

Acknowledgements

This work was partly supported by the Wellcome Trust/DBT India Alliance Fellowship IA/CPHE/14/1/501504 awarded to Tavpritesh Sethi and the Tavpritesh Sethi also acknowledges support from the DBT Project BT/PR/34245/AI/133/9/2019 and the Center for Artificial Intelligence at IIIT-Delhi. Tavpritesh Sethi also acknowledges the valuable inputs provided by Dr. Mitali Mukerji (CSIR-IGIB) and Dr. Gaurav Ahuja (Computational Biology Department, IIIT-Delhi).

References

1. Marchese, F., Raimondi, I. and Huarte, M. (2017). The multidimensional mechanisms of long noncoding RNA function. Genome Biology, 18(1).
2. Huang, S., Huang, Z., Luo, Q. and Qing, C. (2018). The Expression of lncRNA NEAT1 in Human Tuberculosis and Its Anti-tuberculosis Effect. BioMed Research International, 2018, pp.1-8.
3. Yang, X., Yang, J., Wang, J., Wen, Q., Wang, H., He, J., Hu, S., He, W., Du, X., Liu, S. and Ma, L. (2016). Microarray analysis of long noncoding RNA and mRNA expression profiles in human macrophages infected with Mycobacterium tuberculosis. Scientific Reports, 6(1).
4. Chen, Z., Wei, L., Shi, L., Li, M., Jiang, T., Chen, J., Liu, C., Yang, S., Tu, H., Hu, Y., Gan, L., Mao, L., Wang, C. and Li, J. (2017). Screening and identification of lncRNAs as potential biomarkers for pulmonary tuberculosis. Scientific Reports, 7(1).
5. Li, M., Cui, J., Niu, W., Huang, J., Feng, T., Sun, B. and Yao, H. (2019). Long non-coding PCED1B-AS1 regulates macrophage apoptosis and autophagy by sponging miR-155 inactive tuberculosis. Biochemical and Biophysical Research Communications, 509(3), pp.803-809.
6. Wang, Y., Zhong, H., Xie, X., Chen, C., Huang, D., Shen, L., Zhang, H., Chen, Z. and Zeng, G. (2015). Long noncoding RNA derived from CD244 signalling epigenetically controls CD8+ T-cell immune responses in tuberculosis infection. Proceedings of the National Academy of Sciences, 112(29), pp.E3883-E3892.
7. Pieters, J. (2008). Mycobacterium tuberculosis and the Macrophage: Maintaining a Balance. Cell Host & Microbe, 3(6), pp.399-407.
8. Jasenosky, L., Scriba, T., Hanekom, W. and Goldfeld, A. (2015). T cells and adaptive immunity to Mycobacterium tuberculosis in humans. Immunological Reviews, 264(1), pp.74-87.
9. Pathak, S., Basu, S., Basu, K., Banerjee, A., Pathak, S., Bhattacharyya, A., Kaisho, T., Kundu, M. and Basu, J. (2007). Direct extracellular interaction between the early secreted antigen ESAT-6 of Mycobacterium tuberculosis and TLR2 inhibits TLR signalling in macrophages. Nature Immunology, 8(6), pp.610-618.
10. Karbalaei Zadeh Babaki, M., Soleimanpour, S. and Rezaee, S. (2017). Antigen 85 complex as a powerful Mycobacterium tuberculosis immunogene: Biology, immune-pathogenicity, applications in diagnosis, and vaccine design. Microbial Pathogenesis, 112, pp.20-29.
11. Rustad, T., Harrell, M., Liao, R. and Sherman, D. (2008). The Enduring Hypoxic Response of Mycobacterium tuberculosis. PLoS ONE, 3(1), p.e1502.
12. Belton, M., Brilha, S., Manavaki, R., Mauri, F., Nijran, K., Hong, Y., Patel, N., Dembek, M., Tezera, L., Green, J., Moores, R., Aigbirhio, F., Al-Nahhas, A., Fryer, T., Elkington, P. and Friedland, J. (2016). Hypoxia and tissue destruction in pulmonary TB. Thorax, 71(12), pp.1145-1153.
13. Lin, P. and Flynn, J. (2015). CD8 T cells and Mycobacterium tuberculosis infection. Seminars in Immunopathology, 37(3), pp.239-249.
14. Yang, Y., Ju, J., Deng, M., Wang, J., Liu, H., Xiong, L. and Zhang, J. (2017). Hypoxia-Inducible Factor 1α Promotes Endogenous Adaptive Response in Rat Model of Chronic Cerebral Hypoperfusion. International Journal of Molecular Sciences, 18(1), p.3.
15. Scriba, T., Penn-Nicholson, A., Shankar, S., Hraha, T., Thompson, E., Sterling, D., Nemes, E., Darboe, F., Suliman, S., Amon, L., Mahomed, H., Erasmus, M., Whatney, W., Johnson, J., Boom, W., Hatherill, M., Valvo, J., De Groote, M., Ochsner, U., Aderem, A., Hanekom, W. and Zak, D. (2017). Sequential inflammatory processes define human progression from M. tuberculosis infection to tuberculosis disease. *PLOS Pathogens*, 13(11), p.e1006687.
16. Davis, S. and Meltzer, P. (2007). GEOquery: a bridge between the Gene Expression Omnibus (GEO) and BioConductor. Bioinformatics, 23(14), pp.1846-1847.
17. Reimers, M. and Carey, V. (2006). [8] Bioconductor: An Open Source Framework for Bioinformatics and Computational Biology. Methods in Enzymology, pp.119-134.
18. Clough, E. and Barrett, T. (2016). The Gene Expression Omnibus Database. Methods in Molecular Biology, pp.93-110.
19. Kim, D., Langmead, B. and Salzberg, S. (2015). HISAT: a fast-spliced aligner with low memory requirements. *Nature Methods*, 12(4), pp.357-360.
20. Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G., Abecasis, G. and Durbin, R. (2009). The Sequence Alignment/Map format and SAMtools. *Bioinformatics*, 25(16), pp.2078-2079.
21. Sun, Z., Nair, A., Chen, X., Prodduturi, N., Wang, J. and Kocher, J. (2017). UClncR: Ultrafast and comprehensive long non-coding RNA detection from RNA-seq. *Scientific Reports*, 7(1).

22. Frankish, A., Diekhans, M., Ferreira, A., Johnson, R., Jungreis, I., Loveland, J., Mudge, J., Sisu, C., Wright, J., Armstrong, J., Barnes, I., Berry, A., Bignell, A., Carbonell Sala, S., Chrast, J., Cunningham, F., Di Domenico, T., Donaldson, S., Fiddes, I., García Girón, C., Gonzalez, J., Grego, T., Hardy, M., Hourlier, T., Hunt, T., Izuogu, O., Lagarde, J., Martin, F., Martínez, L., Mohanan, S., Muir, P., Navarro, F., Parker, A., Pei, B., Pozo, F., Ruffier, M., Schmitt, B., Stapleton, E., Suner, M., Sycheva, I., Uszczynska-Ratajczak, B., Xu, J., Yates, A., Zerbino, D., Zhang, Y., Aken, B., Choudhary, J., Gerstein, M., Guigó, R., Hubbard, T., Kellis, M., Paten, B., Reymond, A., Tress, M. and Flicek, P. (2018). GENCODE reference annotation for the human and mouse genomes. Nucleic Acids Research, 47(D1), pp.D766-D773.
23. Anders, S., Pyl, P. and Huber, W. (2014). HTSeq--a Python framework to work with high-throughput sequencing data. Bioinformatics, 31(2), pp.166-169.
24. Pruitt, K., Tatusova, T. and Maglott, D. (2007). NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic Acids Research, 35(Database), pp.D61-D65.
25. Smedley, D., Haider, S., Ballester, B., Holland, R., London, D., Thorisson, G. and Kasprzyk, A. (2009). BioMart – biological queries made easy. BMC Genomics, 10(1), p.22.
26. Kim, Y. and Cribbie, R. (2017). ANOVA and the variance homogeneity assumption: Exploring a better gatekeeper. British Journal of Mathematical and Statistical Psychology, 71(1), pp.1-12.
27. Li, Q., Yu, X., Chaudhary, R., Slebos, R., Chung, C. and Wang, X. (2019). lncDIFF: a novel quasi-likelihood method for differential expression analysis of non-coding RNA. BMC Genomics, 20(1).
28. Robinson, M., McCarthy, D. and Smyth, G. (2009). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), pp.139-140.
29. Ge, S., Jung, D. and Yao, R. (2019). ShinyGO: a graphical gene-set enrichment tool for animals and plants. Bioinformatics.
30. Gene Ontology Consortium: going forward. (2014). Nucleic Acids Research, 43(D1), pp.D1049-D1056.
31. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2016). KEGG: new perspectives on genomes, pathways, diseases and drugs. Nucleic Acids Research, 45(D1), pp.D353-D361.
32. Anand, L. and Rodriguez Lopez, C. (2019). chromoMap: An R package for Interactive Visualization and Annotation of Chromosomes.
33. Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S., Tan, G., Baranasic, D., Arenillas, D., Sandelin, A., Vandepoele, K., Lenhard, B., Ballester, B., Wasserman, W., Parcy, F. and Mathelier, A. (2017). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Research, 46(D1), pp.D260-D266.
34. Gearing, L., Cumming, H., Chapman, R., Finkel, A., Woodhouse, I., Luu, K., Gould, J., Forster, S. and Hertzog, P. (2019). CiiiDER: A tool for predicting and analysing transcription factor binding sites. PLOS ONE, 14(9), p.e0215495.
35. Mas-Ponte, D., Carlevaro-Fita, J., Palumbo, E., Hermoso Pulido, T., Guigo, R. and Johnson, R. (2017). LncATLAS database for subcellular localization of long noncoding RNAs. RNA, 23(7), pp.1080-1087.
36. Kong, L., Zhang, Y., Ye, Z., Liu, X., Zhao, S., Wei, L. and Gao, G. (2007). CPC: assess the protein-coding potential of transcripts using sequence features and support vector machine. Nucleic Acids Research, 35(suppl_2), pp.W345-W349.
37. Zhao, Y., Sun, L., Wang, R., Hu, J. and Cui, J. (2018). The effects of mitochondria-associated long noncoding RNAs in cancer mitochondria: New players in an old arena. Critical Reviews in Oncology/Hematology, 131, pp.76-82.
38. Ncbi.nlm.nih.gov. (2020). GRCh38.p12 - Genome - Assembly - NCBI. [online] Available at: https://www.ncbi.nlm.nih.gov/assembly/GCF_000001405.38/ [Accessed 21 Feb. 2020].
39. Klec, C., Prinz, F. and Pichler, M. (2018). Involvement of the long noncoding RNA NEAT1 in carcinogenesis. Molecular Oncology, 13(1), pp.46-60.
40. Zhang, F., Wu, L., Qian, J., Qu, B., Xia, S., La, T., Wu, Y., Ma, J., Zeng, J., Guo, Q., Cui, Y., Yang, W., Huang, J., Zhu, W., Yao, Y., Shen, N. and Tang, Y. (2016). Identification of the long noncoding RNA NEAT1 as a novel inflammatory regulator acting through the MAPK pathway in human lupus. Journal of Autoimmunity, 75, pp.96-104.

Supplementary Figures

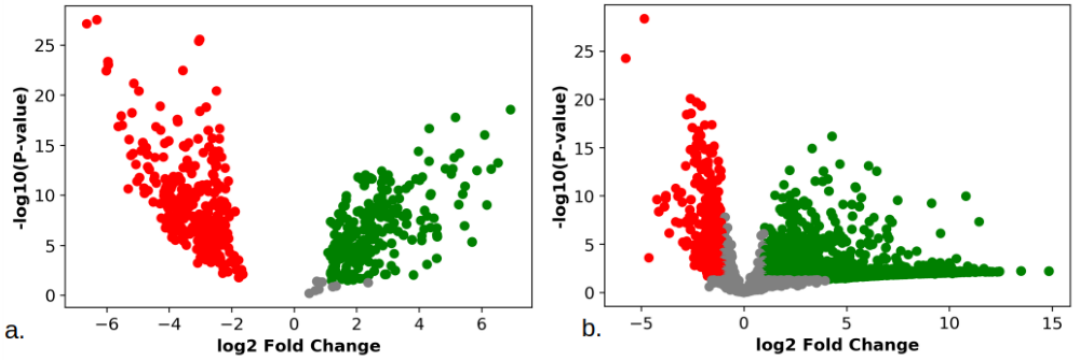

Figure 1 Volcano plot of expression type i.e. overexpressed/ underexpressed in progressor samples. Over-expressed genes are depicted in green (log2 (Fold Change)>1), Under-expressed genes are depicted in red(log2(Fold Change)<-1); (a) Volcano plot for differentially expressed lncRNAs; (b) Volcano Plot for differentially expressed mRNAs.

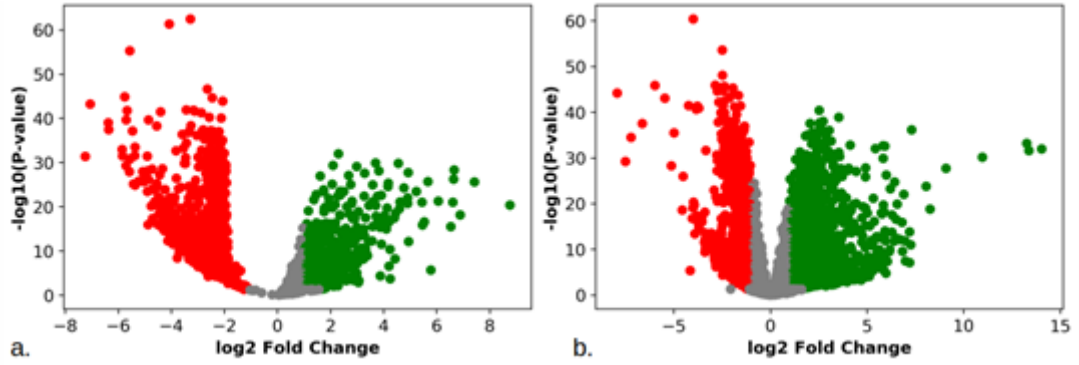

Figure 2 Volcano plot of expression type i.e. overexpressed/ underexpressed in non-progressor samples. Over-expressed genes are depicted in green (log2 (Fold Change)>1), Under-expressed genes are depicted in red(log2(Fold Change)<-1); (a) Volcano plot for differentially expressed lncRNAs; (b) Volcano Plot for differentially expressed mRNAs.

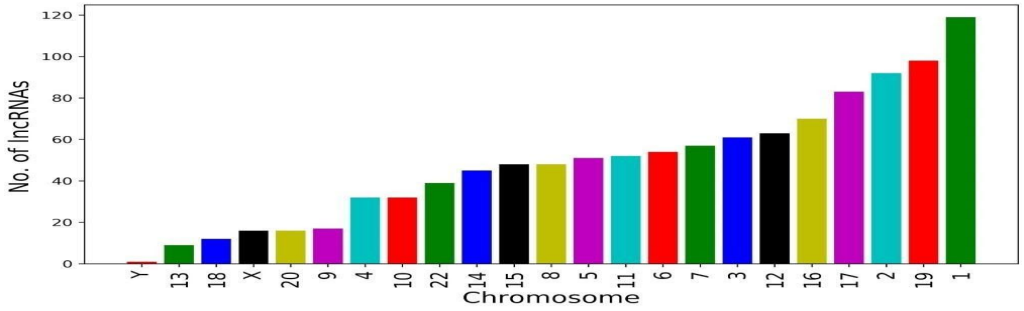

Figure 3 lncRNA distribution across the human genome for non-progressors; Chromosome 1 is the largest chromosome holds a higher number of lncRNA as expected

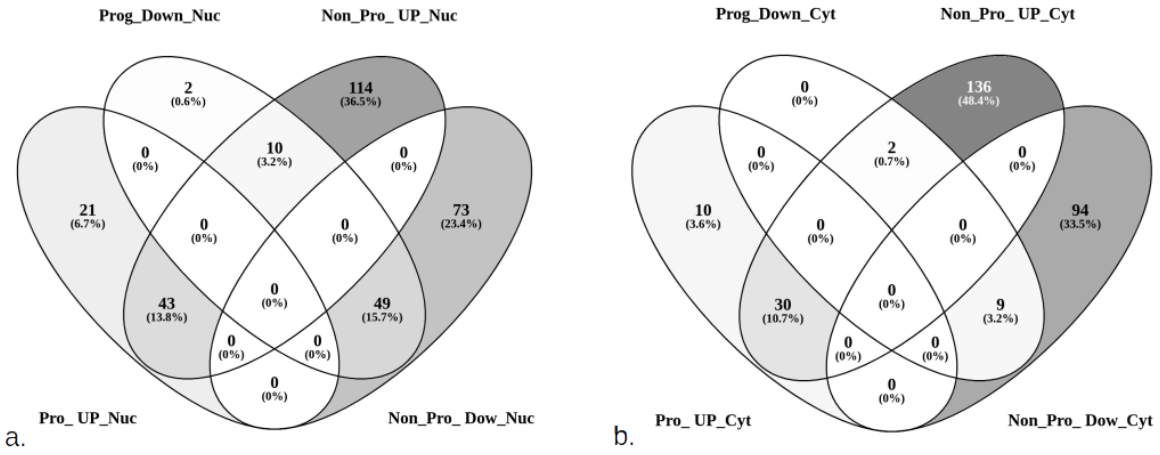

Figure 4 Comparative analysis of sub-cellular localization of the DE lncRNAs in progressor and non-progressor samples. The comparison shows the randomized expression of the DE lncRNAs in non-progressors (in cytosol and nucleus). The DE lncRNAs in progressive samples show more expression in the cytosol than the nucleus.

Supp. Table 1 (Gene Set Enrichment Analysis)

FDR	Enrichment Type	Enrichment Term	Count
0.006757268	KEGG	Thermogenesis	20
0.013159226	KEGG	Parkinson's Disease	14
0.015379000	KEGG	Systemic Lupus Erythematosus	13
0.000358735	GO: Molecular Function	Nucleic Acid Binding	190
0.000358735	GO: Molecular Function	Enzyme Binding	112
0.000840228	GO: Molecular Function	DNA Binding	123
0.000000001	GO: Biological Process	Organelle Organization	201
0.005500517	GO: Biological Process	Regulation of Gene Expression	198
0.005500517	GO: Biological Process	Phosphorylation	116
0.000000006	GO: Cellular Compartment	Nuclear Lumen	211
0.000000006	GO: Cellular Compartment	Nuclear Part	227
0.000000017	GO: Cellular Compartment	Nucleoplasm	184

Supp. Table 2 (Chromosomal Locations of the closely related lncRNA and mRNA)

mRNA Symbol	lncRNA Symbol	Chromosome	Distance
PNN	AL132639.3	14	298
HILPDA	AC010655.4	7	9
PBX2	AL662884.1	6	0
HILPDA	AC010655.2	7	-88
MNT	AC006435.1	17	-774