



**UNIVERSIDADE FEDERAL DO MARANHÃO  
CENTRO DE CIÊNCIAS BIOLÓGICAS E SAÚDE  
CURSO DE FARMÁCIA**

**JOÃO PEDRO COSTA RAMALHO**

**ESTRUTURAÇÃO DE UMA BASE DE DADOS DE EXPRESSÃO GÊNICA PARA  
ANÁLISE *IN SILICO* DE GENES ASSOCIADOS À TUBERCULOSE**

São Luís – MA

2022

JOÃO PEDRO COSTA RAMALHO

**ESTRUTURAÇÃO DE UMA BASE DE DADOS DE EXPRESSÃO GÊNICA PARA  
ANÁLISE *IN SILICO* DE GENES ASSOCIADOS À TUBERCULOSE**

Monografia apresentada como requisito  
para conclusão do Curso de Farmácia da  
Universidade Federal do Maranhão.

Orientador: Prof. Dr. Paulo Vítor Soeiro  
Pereira

São Luís - MA

2022

Ficha gerada por meio do SIGAA/Biblioteca com dados fornecidos pelo(a) autor(a).  
Diretoria Integrada de Bibliotecas/UFMA

Costa Ramalho, João Pedro.

ESTRUTURAÇÃO DE UMA BASE DE DADOS DE EXPRESSÃO GÊNICA  
PARA ANÁLISE IN SILICO DE GENES ASSOCIADOS À TUBERCULOSE /  
João Pedro Costa Ramalho. - 2022.

53 f.

Orientador(a): Paulo Vitor Soeiro Pereira.  
Monografia (Graduação) - Curso de Farmácia,  
Universidade Federal do Maranhão, São Luís, 2022.

1. Bioninformática. 2. Microarray. 3. Resposta  
imune. 4. RNAseq. I. Soeiro Pereira, Paulo Vitor. II.  
Título.

JOÃO PEDRO COSTA RAMALHO

**ESTRUTURAÇÃO DE UMA BASE DE DADOS DE EXPRESSÃO GÊNICA PARA  
ANÁLISE *IN SILICO* DE GENES ASSOCIADOS À TUBERCULOSE**

Defesa de Monografia como requisito para  
conclusão do Curso de Farmácia da  
Universidade Federal do Maranhão.

Orientador: Prof. Dr. Paulo Vitor Soeiro  
Pereira

Data da defesa: 27 / 07 / 2022

BANCA EXAMINADORA

---

Profa. Dra. Angela Falcai  
Universidade CEUMA

---

Profa. Dra. Rosimary de Jesus Gomes Turri  
Universidade Federal do Maranhão – UFMA

---

Prof. Dr. Paulo Vitor Soeiro Pereira  
Universidade Federal do Maranhão - UFMA

*Dedico este trabalho a Deus e minha família,  
que estiveram sempre atuando com o  
objetivo de me tornar uma pessoa valorosa e  
com princípios.*

## **AGRADECIMENTOS**

Me faltam palavras para agradecer a todos que me auxiliaram nessa imensa caminhada rumo a finalização deste trabalho e deste curso. Primeiramente, quero agradecer a Deus, meu eterno protetor e guia, aquele que está sempre olhando por mim e me mostrando o caminho correto a seguir. Gostaria de agradecer também aos meus pais, Lia Costa Ramalho e José Ribamar Moraes Ramalho, meu irmão Jefferson Felipe Costa Ramalho, estes estiveram sempre ao meu lado e, mesmo nos meus momentos mais difíceis, mantiveram-se provendo apoio e carinho a mim

Não poderia deixar de dizer um grande obrigado ao meu orientador, Dr. Paulo Vitor Soeiro, um grande professor e exímio pesquisador que me acolheu em seu laboratório, estando sempre disposto a ensinar e dar suporte aos seus alunos. Aos meus amigos, Giuliana da Silva Miranda, Anna Gabrielle de Sousa Ferreira, Ronald Antônio Coelho de Azevedo, Maria Rita Enes Torres, entre diversos outros que fizeram parte da minha vida acadêmica e sempre estarão nas minhas felizes lembranças.

*Aquilo que se tem como presente é, inúmeras vezes, puro passado. E o passado, como ente sepultado, está quase sempre se recriando nas elaborações do presente e na imaginação do futuro.*

*José Flávio Sombra Saraiva*

## RESUMO

Entender e combater a tuberculose tem sido, por décadas, o objetivo de grande parcela da comunidade científica, tendo em vista seu potencial patogênico e o grande número de óbitos causados pelo *Mycobacterium tuberculosis*. Eventualmente, as ciências biológicas ganharam um grande aliado no processo de entendimento de doenças e microrganismos: a transcriptômica. Esse segmento da biologia molecular permite um estudo mais aprofundado e detalhado do comportamento transcricional de seres vivos, compilando seus dados experimentais em extensas bases de dados. Nesse cenário, o objetivo do presente trabalho foi estruturar uma base de dados com informações referentes ao perfil de transcrição gênica de macrófagos no contexto da tuberculose. Além disso, foi delimitado o estudo do padrão de expressão gênica da via IL-18/SIGIRR na tuberculose, tendo em vista seu potencial como via importante no combate ao *M. tuberculosis*. O processo de estruturação da base de dados consistiu, inicialmente, na elaboração de vias de sinalização no STRING e inBio Discover que forneceram, respectivamente, 54 e 26 genes. Após serem colocados em um diagrama de Venn, os 19 genes presentes simultaneamente nas duas bases de dados foram selecionados para direcionarem a análise de DEGs, que resultou na elaboração de tabelas com padrão de regulação e *heatmaps*. Em relação aos genes da via IL-18/SIGIRR, os genes IL18, IRAK2 e IRF7 estavam regulados positivamente (expressão aumentada), indicando um padrão de resposta reativo ao *M. tuberculosis*, bem como está documentado na literatura. O gene SIGIRR apresentou um padrão de expressão inconclusivo e faz-se necessário analisar outros datasets para o estabelecimento de seu padrão transcricional. Sendo assim, o presente estudo demonstrou todo o processo de tratamento e normalização de dados de expressão gênica, provando ser possível trabalhar com informações de transcriptômica em contextos mais objetivos de estudo, visando uma via de sinalização ou proteína específica. Ainda, este trabalho sinaliza a possibilidade de pesquisas em condições de orçamento modesto e sem grandes instalações laboratoriais, além de incentivar a reutilização eficiente de grandes volumes de dados compilados, que seriam apenas armazenados e agora podem ser empregados em análises de grande valor científico.

**Palavras-chave:** bioinformática; microarray; RNAseq; resposta imune.

## ABSTRACT

Understanding and combating tuberculosis has been, for decades, the objective of a large part of the scientific community, given its pathogenic potential and the large number of deaths caused by *Mycobacterium tuberculosis*. Eventually, the biological sciences gained a great ally in the process of understanding diseases and microorganisms: transcriptomics. This segment of molecular biology allows for a more in-depth and detailed study of the transcriptional behavior of living beings, compiling their experimental data in extensive databases. In this scenario, the objective of the present work was to structure a database with information regarding the gene transcription profile of macrophages in the context of tuberculosis. In addition, the study of the gene expression pattern of the IL-18/SIGIRR pathway in tuberculosis was delimited, in view of its potential as an important pathway in the fight against *M. tuberculosis*. The database structuring process consisted, initially, in the elaboration of signaling pathways in STRING and inBio Discover, which provided, respectively, 54 and 26 genes. After being placed in a Venn diagram, the 19 genes present simultaneously in the two databases were selected to direct the analysis of DEGs, which resulted in the elaboration of tables with regulation pattern and heatmaps. Regarding the genes of the IL-18/SIGIRR pathway, the IL18, IRAK2 and IRF7 genes were upregulated (increased expression), indicating a pattern of reactive response to *M. tuberculosis*, just like is documented in the literature. The SIGIRR gene showed an inconclusive expression pattern and it is necessary to analyze other datasets to establish its transcriptional pattern. Therefore, the present study demonstrated the entire process of treatment and normalization of gene expression data, proving that it is possible to work with transcriptomic information in more objective study contexts, targeting a specific signaling pathway or protein. Still, this work signals the possibility of research under modest budget conditions and without large laboratory facilities, in addition to encouraging the efficient reuse of large volumes of compiled data, which would only be stored and can now be used in analyzes of great scientific value.

**Key words:** bioinformatics; microarray; RNAseq; immune response.

## Lista de Ilustrações

### FIGURAS

Figura 1 - Perfil de disseminação e desenvolvimento da tuberculose .....	17
Figura 2 - Resposta imune a <i>M. tuberculosis</i> .....	19
Figura 3 - Desenho experimental para avaliação de expressão gênica por análise de mRNA, utilizando Microarray e RNA-seq. .....	22
Figura 4 - Diagrama de interação proteína-proteína STRING na via de sinalização IL-18 e SIGIRR .....	32
Figura 5 - Diagrama de interação de genes/proteínas inBio Discover envolvidos(as) na via de sinalização IL-18 e SIGIRR .....	33
Figura 6 - Diagrama de Venn contendo os 19 genes que constam simultaneamente no STRING e no InBio Discover .....	34
Figura 7 - Heatmap demonstrativo do perfil de expressão dos genes selecionados em macrófagos infectados ou não com Mtb .....	36
Figura 8 - Heatmap demonstrativo do perfil de expressão dos genes selecionados em macrófagos infectados ou não com Mtb .....	40
Figura 9 - Heatmap demonstrativo do perfil de expressão dos genes selecionados em macrófagos infectados ou não com Mtb .....	42

## QUADROS

Quadro 1 - Caracterização dos 7 datasets encontrados no NCBI GEO DataSets .....	29
--	----

## Lista de tabelas

Tabela 1 -	Análise de expressão diferencial dos genes selecionados a partir do GSE52819.....	35
Tabela 2-	Análise de expressão diferencial dos genes selecionados a partir do GSE139871.....	38
Tabela 3 -	Análise de expressão diferencial dos genes selecionados a partir do GSE148731.....	41

## Lista de abreviaturas e siglas

BCG	Bacilo de Calmette-Guérin
CD4+	Linfócito T auxiliar
cDNA	DNA complementar
Células NK	Natural Killer
DEG	Gene diferencialmente expresso
DNA	Ácido Desoxirribonucleico
IFN- $\gamma$	Interferon gama
HIV	Vírus da Imunodeficiência Humana
IL-1	Interleucina-1
IL-12	Interleucina-12
IL-18	Interleucina-18
IL18R	Receptor de interleucina-18
Macrófagos M1	Macrófago com padrão de resposta M1
Macrófagos M2	Macrófago com padrão de resposta M2
<i>M. tuberculosis</i>	<i>Mycobacterium tuberculosis</i>
Mo	Monócitos
mRNA	RNA mensageiro
OMS	Organização Mundial da Saúde
PBMC	Células mononucleares de sangue periférico
RNA	Ácido ribonucléico
RNAseq	RNA sequencial
scRNA-seq	RNA sequencial de célula individual
TB	Tuberculose
TLR 2	<i>Toll-like</i> receptor do tipo 2
TLR 7	<i>Toll-like</i> receptor do tipo 7
TNF- $\alpha$	Fator de necrose tumoral alfa

## SUMÁRIO

<b>1. INTRODUÇÃO.....</b>	<b>14</b>
<b>2. REFERENCIAL TEÓRICO.....</b>	<b>16</b>
<b>2.1 Tuberculose.....</b>	<b>16</b>
<b>2.2 Resposta imune no combate à Tuberculose.....</b>	<b>18</b>
<b>2.3 Fatores genéticos envolvidos no estabelecimento da Tuberculose....</b>	<b>20</b>
<b>2.4 Análise da Expressão Gênica.....</b>	<b>21</b>
<b>2.5 Meta análise de dados de expressão gênica.....</b>	<b>23</b>
<b>3.0 OBJETIVOS.....</b>	<b>25</b>
<b>3.1 Objetivos gerais.....</b>	<b>25</b>
<b>3.2 Objetivos específicos.....</b>	<b>25</b>
<b>4.0 METODOLOGIA.....</b>	<b>26</b>
<b>4.1 Obtenção de dados.....</b>	<b>26</b>
<b>4.2 Análise RNASeq.....</b>	<b>26</b>
<b>4.3 Análise de expressão diferencial de dados em massa de RNA-seq e microarray.....</b>	<b>27</b>
<b>4.4 Análise das interações de genes envolvidos na resposta imune na Tuberculose.....</b>	<b>27</b>
<b>4.5 Enriquecimento de termos GO.....</b>	<b>28</b>
<b>5.0 RESULTADOS.....</b>	<b>29</b>
<b>5.1 DataSets selecionados para análise.....</b>	<b>29</b>
<b>5.2 Mineração e refinamento dos dados brutos.....</b>	<b>34</b>
<b>6.0 DISCUSSÃO.....</b>	<b>44</b>
<b>7.0 CONSIDERAÇÕES FINAIS.....</b>	<b>49</b>
<b>REFERÊNCIAS.....</b>	<b>50</b>

## 1. INTRODUÇÃO

Não é uma tarefa fácil encontrar uma enfermidade que assola a humanidade há tanto tempo e com tanto impacto quanto a tuberculose. Podendo atingir virtualmente qualquer tecido do corpo humano, mas com tropismo pelos pulmões, a tuberculose é uma doença infectocontagiosa que já ceifou a vida de milhares de pessoas por todo o planeta, o que lhe rendeu a alcunha de “peste branca”. Os agentes etiológicos da tuberculose estão agrupados no grupo de micobactérias do complexo *Mycobacterium tuberculosis*, com destaque para *Mycobacterium tuberculosis* e, mesmo que em menor escala, *Mycobacterium bovis* (MAYER-BARBER; BARBER, 2015).

De acordo com a Organização Mundial de Saúde (OMS), o Brasil está entre os 30 países com maior número de pessoas acometidas pela tuberculose, mesmo com todos os programas de conscientização e políticas públicas mundiais e nacionais desenvolvidos nas últimas décadas. Só no ano de 2019, estima-se que 1 milhão e 300 mil pessoas tenham morrido em decorrência da infecção por *M. tuberculosis* em todo o mundo, ainda que os tuberculostáticos sejam disponibilizados em larga escala e, muitas vezes, gratuitamente. Em relação ao estado do Maranhão, uma região historicamente afetada por infecções causadas por *M. tuberculosis*, o Ministério da Saúde estima que houve 2.194 novos casos da doença no ano de 2021, com uma frequência calculada de 30,7 casos por 100 mil habitantes, acompanhado de 150 óbitos no ano de 2020 (OMS, 2021; BRASIL, 2022).

Assim como em outras doenças infecto contagiosas, o estabelecimento e a progressão da tuberculose dependem em grande escala da atuação do sistema imune frente a *M. tuberculosis*. Basicamente, a grande maioria da população nem sequer desenvolve a tuberculose ativa, mesmo entrando em contato com o patógeno, tendo em vista a atuação eficiente das células de defesa do infectado (CHURCHYARD, 2017). Com isso, apenas 5% a 10% das pessoas imunocompetentes irão realmente desenvolver a condição clínica patológica. Outra possibilidade de desfecho clínico é o desenvolvimento de uma forma latente da doença, onde há a formação de um granuloma constituído de várias micobactérias circundadas por diversos tipos

celulares, como macrófagos, linfócitos T e B, entre outras células (MÖLLER, 2018; PAI, 2016).

O que intriga a comunidade científica é a relação existente entre o componente genético e o desenvolvimento da tuberculose. Nesse aspecto, tem ganhado destaque a biologia molecular focada na avaliação imunogenética de pacientes com doenças infecciosas. Uma das estratégias de análise utilizada têm sido o estudo dos transcritos de mRNA produzidos por células de humanos infectados com *M. tuberculosis*, onde é crucial a utilização de métodos analíticos mais complexos que permitam detectar o nível de expressão de cada gene em estudo, como faz a análise de RNAseq e o Microarray. Muitos destes trabalhos acabam armazenando seus resultados em bases de dados que são disponibilizadas para o público em geral, criando assim um acervo genético útil para novas pesquisas que busquem dar outro enfoque aos dados desses trabalhos (KAN 2017; WANG, 2009).

Portanto, o presente trabalho foi desenvolvido com o intuito de estruturar uma base de dados objetiva de expressão gênica a partir de informações de transcriptoma humano no contexto da infecção por *M. tuberculosis*, utilizando para isso conjuntos de dados já disponibilizados em plataformas internacionais. Seu papel é servir de referencial para estudos subsequentes que visem analisar o comportamento de células mononucleares de sangue periférico (do inglês PBMC), como monócitos, no contexto da tuberculose (LATORRACA, 2019; MAYER-BARBER, 2015).

## 2. REFERENCIAL TEÓRICO

### 2.1 Tuberculose

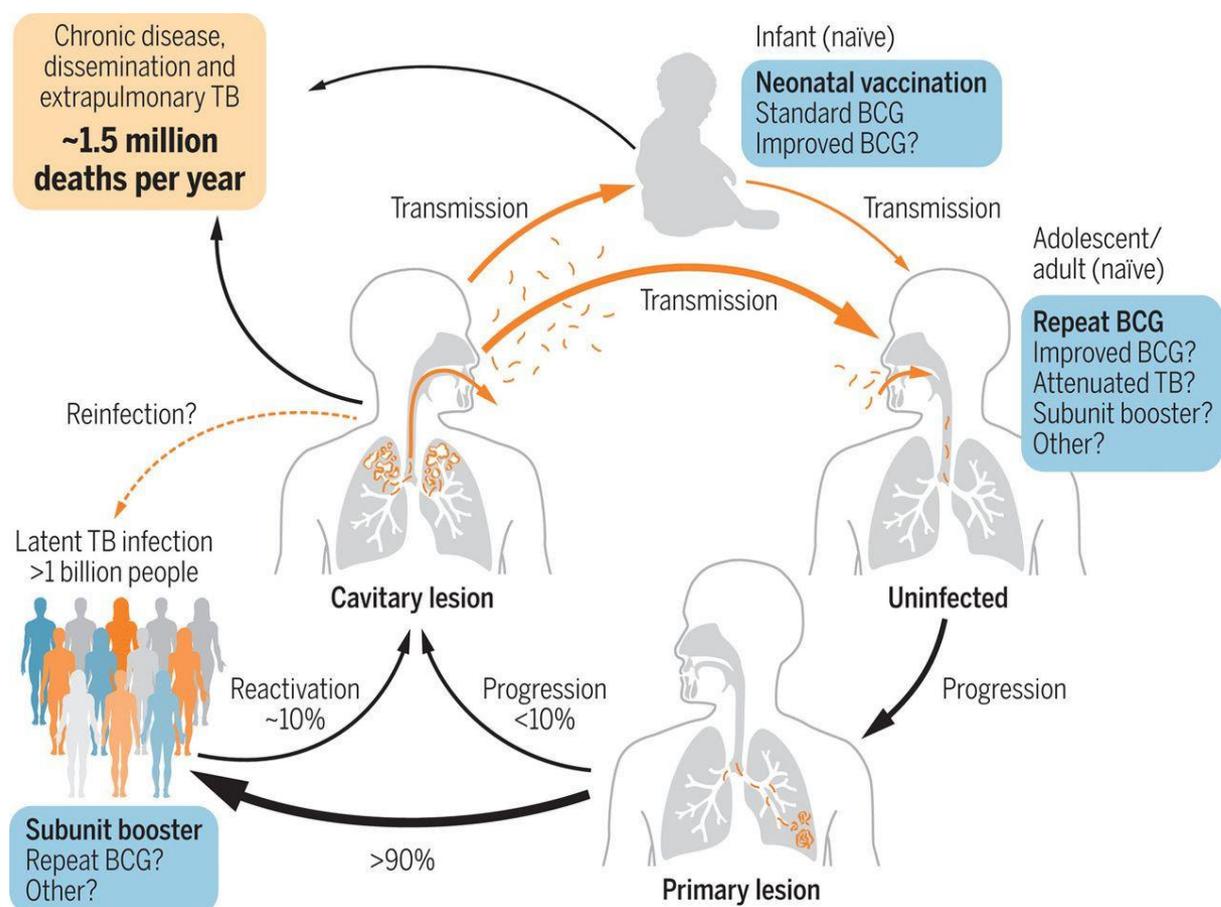
A tuberculose é uma doença que pode apresentar diversas manifestações e desfechos clínicos, onde tudo depende de uma ampla gama de fatores, como: local do corpo onde o bacilo se aloja, que pode ser, teoricamente, qualquer órgão do corpo; grau de infecção e tempo de exposição à micobactéria; estado do sistema imunológico do hospedeiro, onde a homeostase do corpo se faz fundamental; presença de comorbidades associadas, com maior atenção para a coinfeção HIV-Tuberculose; características genéticas do indivíduo, como a presença de erros inatos no sistema imune; entre outros fatores que têm sido descobertos e estudados com afinco pela comunidade científica. Geralmente, a forma pulmonar da doença é a de maior relevância clínica, onde os sintomas variam entre febre vespertina que não costuma ultrapassar os 38,5°C, sudorese noturna, tosse seca ou produtiva por mais de três semanas, emagrecimento e fraqueza muscular (MÖLLER *et al*, 2018; BRASIL, 2019).

Apesar dos inúmeros avanços obtidos pela ciência no combate e na prevenção à tuberculose, a mesma foi responsável pela morte de 1,5 milhões de pessoas no mundo todo só no ano de 2018 (WHO, 2019). No Brasil, a situação não é muito diferente, pois a OMS afirma que o mesmo figura na lista dos 30 países com maior associação da presença de *Mycobacterium tuberculosis* com infecção por HIV (WHO, 2019). Atualmente, o avanço no aumento no número de casos fatais de tuberculose tem sido visto como uma consequência da infecção concomitante com o vírus da imunodeficiência humana (HIV, sigla em inglês). Além disso, o crescimento desordenado da malha urbana em grandes cidades, de países em desenvolvimento, também tem sido um dos focos de disseminação de *Mycobacterium tuberculosis*, aliado com péssimas condições sanitárias e à subnutrição, principalmente nos chamados bolsões de pobreza, onde a crescente desigualdade social tem feito dessas populações carentes alvos fáceis para o bacilo (BRASIL, 2019).

A transmissão do bacilo causador da tuberculose ocorre, principalmente, por meio da inalação de aerossóis contendo *M. tuberculosis* em sua forma ativa, onde o sítio primário de infecção, geralmente, são os pulmões do indivíduo. A partir de então, o sistema imunológico determinará o rumo que a infecção irá tomar, dando origem a

quatro possíveis desfechos (CHURCHYARD, 2017). Na primeira possibilidade, temos a ação efetiva e equilibrada do sistema imune, onde todos os bacilos infecciosos são eliminados. Uma segunda, e mais frequente, opção é o desenvolvimento de uma forma latente de tuberculose (ILT), onde há a formação de um granuloma constituído, majoritariamente, por linfócitos T ativados e macrófagos, isolando uma área com várias micobactérias e células infectadas. Essa forma de contingenciamento bacteriano, na maioria das vezes, é bem eficiente e consegue suprimir a reativação da tuberculose. Entretanto, 5% a 10% dos indivíduos imunocompetentes podem desenvolver a forma ativa da doença em algum momento de sua vida. A terceira possibilidade de desfecho é a apresentação de uma forma subclínica da doença, com sintomas que variam de leves a moderados. A quarta, e última, possibilidade de apresentação clínica é o desenvolvimento da forma ativa da tuberculose, sem a formação prévia de granulomas, onde o indivíduo tem uma rápida evolução da doença (Figura 1) (MÖLLER *et al*, 2018; PAI *et al*, 2016).

**Figura 1. Perfil de disseminação e desenvolvimento da tuberculose.**



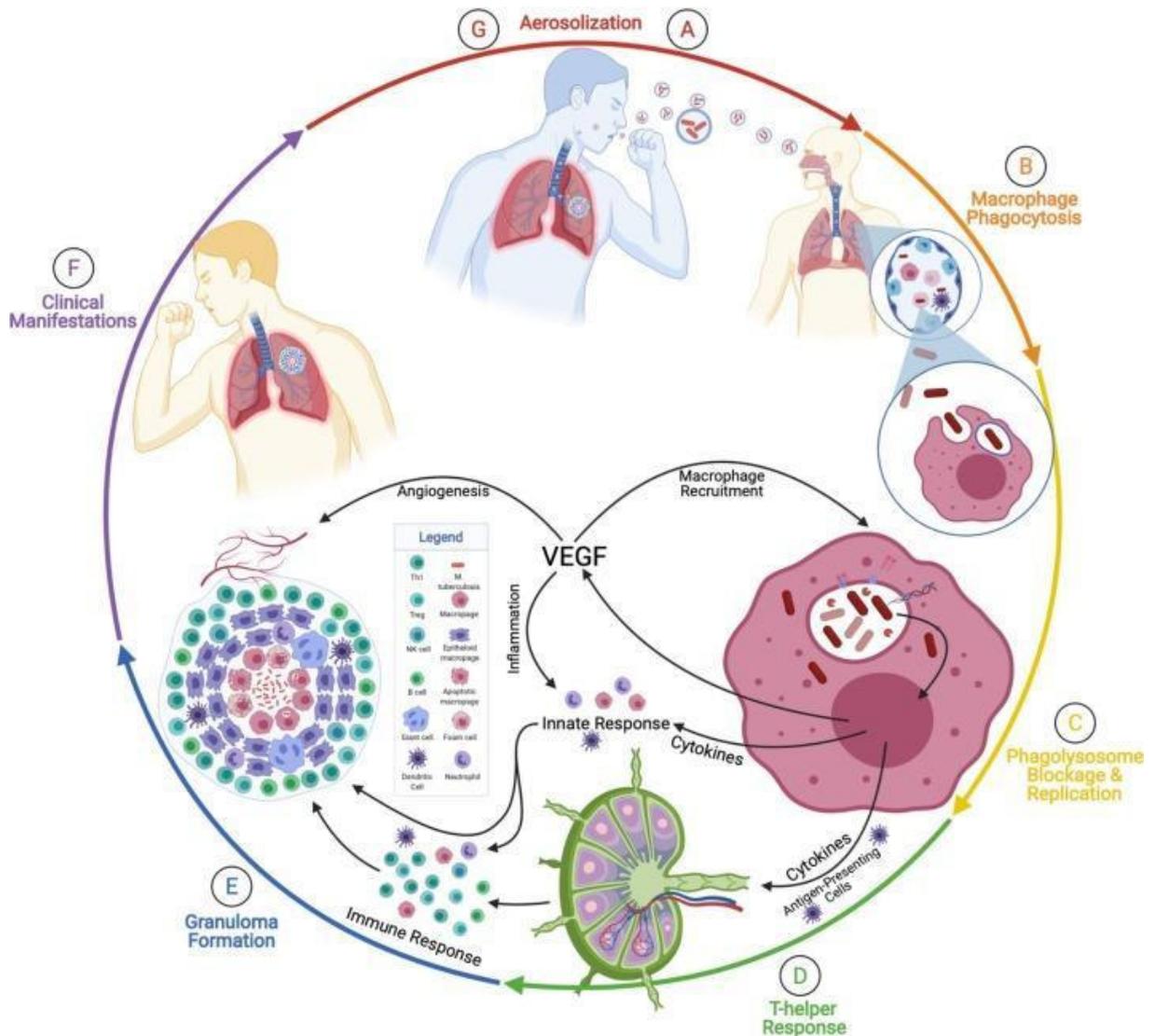
A transmissão por via aérea de *M. tuberculosis* é muito eficiente, mas a maioria das pessoas não desenvolve uma forma agressiva da doença. A implementação de uma estratégia de imunização, com a vacina BCG, garante proteção à maioria dos indivíduos imunocompetentes. A utilização de vacinas BCG melhoradas pode reduzir ainda mais o número de casos de tuberculose. No entanto, a forma da doença na qual o bacilo encontra-se latente, mas viável, abrange cerca de 1 bilhão de indivíduos no mundo. Essa variação no estabelecimento ou não da doença grave continua sendo objetivo de estudos no mundo inteiro, visando métodos mais eficientes de imunização, diagnóstico, definição de prognósticos e tratamentos.

Fonte: adaptado de Porcelli e Jacobs Jr, 2019.

## 2.2 Resposta imune no combate à Tuberculose

O correto entendimento da interação patógeno-hospedeiro é crucial para a elucidação dos mecanismos imunológicos envolvidos no controle de infecções. Entender, detalhadamente, porque algumas pessoas nunca desenvolvem uma forma ativa de tuberculose, e por que outras acabam manifestando formas graves e extrapulmonares, é um ponto que ainda precisa ser esclarecido. No contexto imunológico, a comunicação entre imunidade inata e imunidade adaptativa é um ponto chave na resposta a *M. tuberculosis*, podendo definir os rumos que a infecção irá tomar. O primeiro contato entre o sistema imune e o bacilo da tuberculose ocorre nos pulmões, onde a micobactéria é reconhecida, principalmente, por macrófagos alveolares. Após a formação do sítio inflamatório primário, no parênquima pulmonar, as células apresentadoras de antígenos migram em direção aos linfonodos regionais, onde os linfócitos T farão o reconhecimento dos antígenos, e a resposta imune adaptativa tomará forma. A partir de então, pode ocorrer a formação do granuloma, que se trata de um microambiente constituído de várias células que suprimem e policiam a replicação bacteriana, por um período de tempo indeterminado (Figura 2) (PAI *et al*, 2016).

Figura 2. Resposta imune a *M. tuberculosis*.



Fonte: MAISON, 2022.

A resposta imune a micobactérias tem sido muito estudada e documentada na literatura internacional. Um dos pontos elucidados é o padrão de resposta predominante no combate a infecções por *M. Tuberculosis*, onde se destaca a resposta Th1 (T helper 1). Tal padrão é observado, principalmente, em infecções por microrganismos intracelulares, onde as principais células envolvidas no processo são os linfócitos TCD4+, os macrófagos e as células dendríticas. As citocinas clássicas são a IL-12, o IFN- $\gamma$  e o TNF- $\alpha$ , ou seja, mediadores inflamatórios que facilitam a eliminação de patógenos que residem dentro de células do hospedeiro (MAYER-BARBER; BARBER, 2015).

No contexto das micobactérias, há quatro tipos celulares que se destacam na resposta imune: macrófagos, células dendríticas, linfócitos T e células Natural Killer (NK). Após entrar em contato com o patógeno, seja por fagocitose da micobactéria ou de algum antígeno específico, os macrófagos e células dendríticas passam a produzir uma citocina chamada de IL-12, que tem o papel de induzir a produção de IFN- $\gamma$  em linfócitos T e células NK. Conseqüentemente, esse IFN- $\gamma$  secretado servirá com um estímulo para os macrófagos e células dendríticas adjacentes secretarem ainda mais IL-12, bem como aumentarem a produção de espécies reativas de oxigênio e nitrogênio, e a expressão de MHC de classe 2, criando-se assim um mecanismo de feedback positivo entre imunidade inata e adaptativa (RAMIREZ-ALEJO; SANTOS-ARGUMEDO, 2014).

### **2.3 Fatores genéticos envolvidos no estabelecimento da Tuberculose**

Durante a replicação celular e produção de proteínas, todos os processos que envolvem a transcrição e tradução do código genético são controlados por mecanismo “antifalhas” que, na maior parte das vezes, conseguem impedir a ocorrência de mutações no DNA. Entretanto, há momentos em que esses mecanismos são insuficientes para a reparação de danos sofridos no material genético. Além disso, muitos indivíduos nascem com falhas gênicas previamente definidas pela hereditariedade, onde a carga genética de seus ancestrais é fator de peso na determinação do seu genótipo. Como resultado dessas alterações genéticas, alguns indivíduos acabam nascendo com uma predisposição a desenvolverem determinadas doenças. Conseqüentemente, entender toda a dinâmica imunológica envolvida nessa dicotomia existente entre genes e doença, abre um leque de oportunidades que resulta na criação de novas terapêuticas, principalmente aquelas focadas nos pacientes imunodeficientes (MENCK; MENEGHINI, 2015).

Em se tratando da tuberculose, há uma miríade de genes, já relatados na literatura, que interferem na correta atuação do sistema imune na resposta ao *M. tuberculosis*. A extensa lista que elenca esses genes faz menção a receptores *Toll-like* como o 2 (TLR2) e o 4 (TLR4), interferon gama (IFN- $\gamma$ ), receptor de vitamina D, interleucina 12-R $\beta$ 1, entre vários outros que ainda estão em processo de investigação e prometem ajudar a esclarecer a patogênese e progressão da doença, bem como seus vários possíveis desfechos clínicos (ABEL *et al*, 2018).

## 2.4 Análise da Expressão Gênica

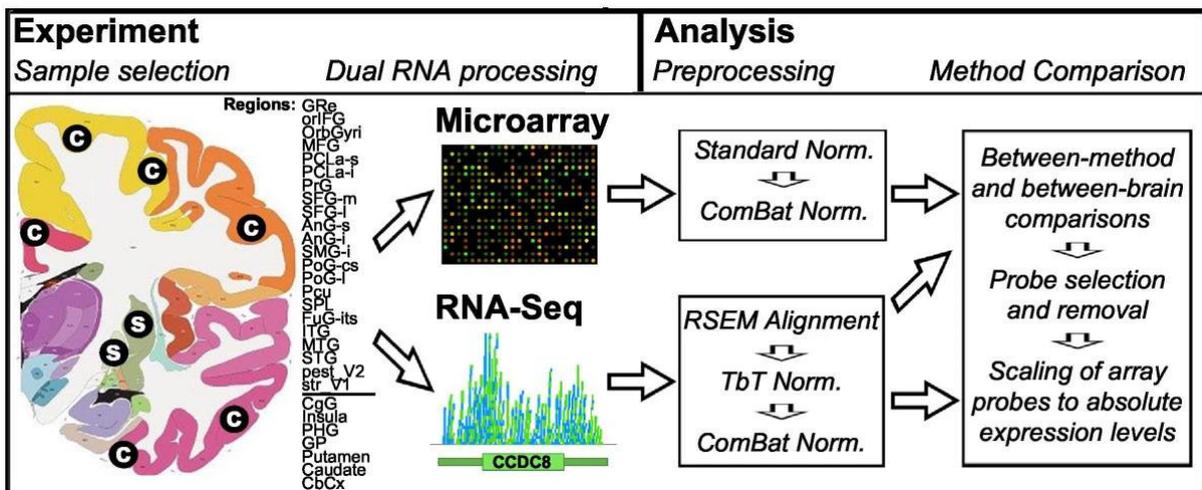
A missão de estudar sistemas biológicos transformou-se completamente na segunda metade do século XX. A descoberta da molécula do DNA e a criação de métodos capazes de analisá-la em escala microscópica acabou criando um novo ramo da ciência: a biologia molecular. Os mistérios envolvidos na transmissão de características hereditárias, as diferenças mais sutis entre espécies semelhantes, o porquê de muitas bactérias conseguirem resistir a antimicrobianos de alta eficiência, são só algumas das perguntas resolvidas pelo advento da biologia molecular. Com mais estudos e a criação e o aprimoramento das técnicas e metodologias analíticas, o novo objetivo passou a ser catalogar todos os genes presentes no ser humano e nas espécies de interesse científico, através do sequenciamento de seus genomas (LEANDRO, 2015).

Atualmente, apenas o estudo do genoma do ser humano tornou-se insuficiente para entender a dinâmica de funcionamento do nosso genótipo. O estudo da informação genética humana frente a ataques de microrganismos invasores tornou-se uma nova vertente da biologia molecular, no momento em que os cientistas perceberam que em situações de estresse biológico alguns genes tendem a ser expressos enquanto outros se mantêm silenciados, mesmo quando se espera que sigam uma sistemática de funcionamento que mantenha o *status quo* daquilo presente na literatura científica. Tais questionamentos fez surgir o transcriptoma, que se baseia no estudo de transcritos de RNA, em suas mais variadas formas, quantificando-os e classificando-os de forma a entender como o organismo funciona frente a uma situação pré-estabelecida. Com isso, surge também a bioinformática, responsável por organizar e estudar essa grande quantidade de dados complexos oriundos da biologia molecular, preparando-os para serem estudados e esquematizados em bases de dados, de forma a resolver hipóteses e perguntas de interesse técnico-científico (WANG; GERSTEIN; SNYDER, 2009).

Dentre as técnicas mais utilizadas para análise de expressão gênica destacam-se o *RNA-seq* e o *Microarray* (Figura 3). A técnica de microarranjo, ou *microarray*, consiste na utilização de pequenos chips para a análise genética, sendo a metodologia mais antiga entre as duas. Inovou no sentido de permitir o estudo simultâneo de vários genes, apesar de não conseguir avaliar duas espécies ao mesmo tempo e ser um

método de baixa reprodutibilidade. A metodologia do *RNA-seq*, por sua vez, consegue analisar várias sequências por amostra, não se limitando a apenas uma espécie por vez, além de ser capaz de identificar transcritos com baixo nível de expressão. Todos esses benefícios encarecem a técnica, já que ela se limita a pequenas quantidades de amostras. Geralmente, nas duas técnicas necessita-se transformar RNA em cDNA, com o intuito de deixar a sequência que abrigará o gene de interesse mais estável. Acompanhando o ritmo do progresso, a imunologia buscou incorporar tais inovações ao seu campo de estudo, utilizando o transcriptoma como ferramenta complementar no estudo do comportamento de linhagens celulares no contexto de infecções, bem como o perfil de resposta desempenhando nessas situações de estresse celular (MALONE *et al*, 2006; MOITINHO JUNIOR, 2019).

**Figura 3. Desenho experimental para avaliação de expressão gênica por análise de mRNA, utilizando Microarray e RNA-seq.**



Exemplo de estratégia de avaliação de amostras de RNA mensageiro usando microarray e RNA-Seq em diferentes amostras biológicas (e.g. tecidos ou células isoladas). Os níveis de expressão gênica são calculados por métodos comparativos. Geralmente, os resultados do microarray são avaliados, filtrados e aprimorados por meio da técnica de RNA-Seq. Fonte: Miller *et al.*, 2014.

## 2.5 Meta análise de dados de expressão gênica

As metanálises são realizadas quando um autor necessita compilar uma grande quantidade de estudos ou informações em um só trabalho, criando sua própria base de dados e submetendo-a a uma análise que objetiva extrair o máximo de informações possíveis dentro de uma nova roupagem experimental. No campo da transcriptômica, estudos com metanálise são utilizados com certa frequência, tendo em vista a imensa

quantidade de transcritos de RNA que precisam ser analisadas, fazendo uma busca sistematizada em conjuntos de dados dispersos e com temáticas muitas vezes dissonantes, mas que mantêm em seu cerne um ponto em comum. Dentro desse cenário metodológico é comum que se busque identificar, quando se trata de estudos na área da saúde, genes com padrão de transcrição alterados ou diferenciados, o que fez com que se cunhasse o termo “genes diferencialmente expressos”, ou DEGs (sigla em inglês). A utilidade dos DEGs varia muito do enfoque que o autor busca dar à alteração da expressão gênica, entretanto, a maior parte sempre utiliza essas informações para entender o comportamento celular frente a um certo estímulo, já que a produção proteica de um organismo costuma variar à medida que é exposto a diferentes condições ambientais (FILHO *et al*, 2014; ZHAO; ERWIN; XUE, 2018).

Cada trabalho determina de forma particular o que caracteriza um gene como diferencialmente expresso, contudo o meio científico tem predileção por utilizar os valores de *fold change* como parâmetro de referência no momento de classificar os DEGs. De maneira sucinta, o *fold change* expressa o quanto determinado gene é transcrito em uma rodada de análises experimentais, adotando um valor que pode variar muito de trabalho para trabalho, mas geralmente os valores perpassam de 1 a 2. A escala utilizada na determinação do valor numérico é um pouco diferente do comumente adotado na matemática, elevando-se o valor de *fold change* em uma unidade quando o resultado da contagem de transcritos contabilizados for equivalente ao dobro do valor inicial. Por exemplo: o valor basal de transcrição é 30, logo uma população celular que exprime um valor de 60 terá um *fold change* de 1. (JUNG; FRIEDE; BEISSBARTH, 2015; ZHAO; ERWIN; XUE, 2018).

Independentemente do enfoque da metanálise, esse tipo de trabalho está se tornando cada vez mais frequente, seja pelo seu caráter de pesquisa holística, seja pelos seus mais variados usos e formas de adaptar metodologias. Essa tendência no meio científico, principalmente na área da saúde, permite tornar o estudo e investigação de doenças um processo mais dinâmico e completo, possibilitando aos autores que empreguem os mais diversos métodos de busca e reunião de informações, com vistas a responder perguntas instigantes e complexa (HAIDICH, 2010; AKOBENG, 2005).

### 3. OBJETIVOS

#### 3.1 Geral

Realizar análise *in silico* de transcriptomas para construção de banco de dados de expressão gênica associada à resposta imune no contexto da tuberculose.

#### 3.2 Específicos

- Identificar os estudos com dados de *RNAseq* e *microarray* disponíveis publicamente para análise;
- Realizar a mineração dos dados selecionados com relação direta com vias importantes de respostas imunológicas na tuberculose;
- Construção de banco de dados com as informações refinadas para uso por estudos futuros.

## 4. MATERIAL E MÉTODOS

### 4.1 Obtenção de dados

Foi realizada uma análise integrativa pesquisando o banco de dados NCBI GEO DataSets e o banco de dados ArrayExpress para identificar informações de expressão gênica publicamente disponíveis no contexto da infecção por *M. tuberculosis*, envolvendo a resposta de macrófagos. São abrangidos estudos publicados entre março de 2012 e março de 2022. Como critérios de inclusão, foram utilizados:

- 1) dados de expressão gênica para leucócitos, em especial os macrófagos de infecção *in vitro* com *M. tuberculosis*;
- 2) estudos em humanos compostos por paciente com tuberculose ativa ou latente, com mais de duas amostras por grupo, com análise de expressão em macrófagos;
- 3) estudos com células humanas de linhagem envolvendo a infecção *in vitro* com cepas de *M. tuberculosis*;
- 4) presença de grupos de controle para comparação.

Todas as plataformas de análise de expressão gênica serão consideradas, e apenas os estudos que forneceram dados de transcriptoma serão incluídos para as avaliações integrativas. Os critérios de exclusão foram: estudos com amostras não humanas; estudos com tratamento antes da análise genética molecular; e estudos de revisão. Os estudos de *RNAseq* e *MicroArray* serão incluídos em nossa análise integrativa.

### 4.2 Análise RNASeq

Foi obtido o objeto Seurat contendo dados de scRNA-seq de VRIES et al. (2020), que será depositado no banco de dados eQTLGen Consortium database (<https://eqtlgen.org/candida.html>). Seguindo o *pipeline* Seurat padrão ([https://satijalab.org/seurat/articles/pbmc3k\\_tutorial.html](https://satijalab.org/seurat/articles/pbmc3k_tutorial.html)) conforme descrito por Stuart et al (2019), para realizar análise de expressão diferencial e visualização de dados (UMAP, dotplot e mapa de calor), quando necessário.

### 4.3 Análise de expressão diferencial de dados em massa de RNA-seq e microarray

Para caracterizar a assinatura imunológica dos perfis transcricionais globais na infecção por *M. tuberculosis*, as contagens de leitura de cada estudo de RNA-seq foram transformadas em contagem em log2 por milhão. A ferramenta web Network Analyst 3.0 (<https://www.networkanalyst.ca/>) foi usada para realizar a análise de expressão diferencial, aplicando o *pipeline* DESeq2, conforme ZHOU et al (2019). Os estudos de *microarray* foram analisados por meio do aplicativo da web GEO2R (BARRET et al, 2012; <https://www.ncbi.nlm.nih.gov/geo/geo2r/>), usando o pipeline limma-voom (LAW et al, 2014). Foram usados os pontos de corte estatísticos de *log2-fold-change* > 1 (regulado para cima) ou <-1 (regulado para baixo) e o valor de p ajustado <0,05 para selecionar genes regulados para cima e para baixo entre os grupos com infecção por *M. tuberculosis* e grupos normais.

A partir dessa análise, as informações foram selecionadas para compor um banco de dados específico para análises imunogenéticas específicas em estudos posteriores.

### 4.4 Análise das interações de genes envolvidos na resposta imune na Tuberculose

Para elencar o que seria analisado nos datasets, 4 genes/proteínas relacionados à via IL18/SIGIRR e a resposta imune à tuberculose (IL18, TLR4, IFNG, SIGIRR) foram submetidos às plataformas de interação entre proteínas STRING (<https://string-db.org/>) e InBio Discover (<https://inbio-discover.com/>). A partir dessas aplicações foram definidas redes de interações entre as proteínas IL18, TLR4, IFNG e SIGIRR, e outras relevantes para a análise. Foram considerados válidos os genes com *interaction score* a partir de 0.4 e *networks* de no máximo 54 genes. Ainda, a remoção de termos redundantes foi feita atribuindo um valor de similaridade permitido de 0,5 aos parâmetros do STRING e InBio Discover. A partir dos dois diagramas de interação criados por essas bases de dados, os genes elencados foram colocados em um diagrama de Venn que destacou os 19 genes em comum nos dois conjuntos. Essa última lista serviu de direcionamento para a análise de expressão diferencial gênica nos datasets.

#### **4.5 Enriquecimento de termos GO (*Gene ontology*)**

A análise de enriquecimento GO, feita para demonstrar quais termos de ontologia gênica que estavam estatisticamente enriquecidos no conjunto de genes diferencialmente expressos. A análise, bem como a visualização dos perfis de expressão, foi realizada com o programa R/R Studio, seguindo os parâmetros padrões do pacote *Heatmap3*. Os Heatmaps que mostram a expressão dos DEGs foram gerados também usando o programa R/R Studio.

## 5. RESULTADOS

### 5.1 DataSets selecionados para análise

Após as buscas feitas no NCBI GEO DataSets, foram encontrados 7 datasets que estavam dentro dos critérios de inclusão estabelecidos. Os datasets envolvem estudos com análise de transcritos de RNA de populações celulares inseridos em contextos específicos da tuberculose, obtidas de indivíduos humanos e sem comorbidades (e.g. HIV, hepatite C), ou em tratamento medicamentoso para tuberculose, entre outras condições excludentes (Quadro1).

**Quadro 1. Caracterização dos 7 datasets encontrados no NCBI GEO DataSets.**

<b>Datasets</b>	<b>Descrição do Estudo</b>	<b>Tipo de Análise</b>	<b>Amostra</b>	<b>Resultados pertinentes</b>	<b>Número de genes</b>
<b>GSE 52819</b>	Análise de macrófagos infectados com Mtb e tratados com Vitamina D	<i>Microarray</i>	Monócitos THP-1 infectados com Mtb cepa H37Rv e tratados com Vitamina D.	A vitamina D induziu a produção de IL- $\beta$ através da sinalização parácrina de macrófagos	22119 genes
<b>GSE 139825</b>	Análise do RNA total de macrófagos alveolares de pacientes com TB em comparação a controles (posteriorment e infectados), e análise de macrófagos esplênicos infectados <i>in vitro</i> pelo Mtb e grupo controle	<i>Microarray</i>	Macrófagos infectados com Mtb das cepas H37Rv e UT205	Pacientes com tuberculose apresentaram maior produção de genes relacionados a sinalização por IFN e do sensor intracelular AIM2	47324 genes
	GSE 143627 - Análise da			Os macrófagos	

<b>GSE 143627</b>	expressão diferencial de genes de macrófagos expostos a cepas de Mtb ativadas e inativadas.	<i>Microarray</i>	Macrófagos infectados com cepas de Mtb vivos e inativados pelo calor	responderam de maneira similar quando em contato com o Mtb vivo e o inativado pelo calor	60675 genes
<b>GSE 54992</b>	Análise da expressão de componentes C1q (A,B e C) por PBMCs de pacientes com TB.	<i>Microarray</i>	PBMCs de pessoas portadoras de tuberculose	Os níveis de C1q estavam aumentados nos macrófagos de pacientes com tuberculose	54675 genes
<b>GSE 139871</b>	Análise do mRNA de monócitos de pacientes com TB comparados com monócitos de pacientes saudáveis. Também foram comparados os mRNA de monócitos infectados com cepas de Mtb (UT127 e UT205).	<i>Microarray</i>	Monócitos saudáveis infectados com cepas de Mtb UT127 e UT205	Foram observadas diferenças significativas de expressão gênica entre os macrófagos saudáveis e os infectados, tendo sido identificados 48 DEGs.	47323 genes
<b>GSE 20050</b>	Análise da expressão gênica de macrófagos	<i>Microarray</i>	Macrófagos obtidos de pacientes com granulomas	O metabolismo lipídico pode ser influenciado pelo Mtb. A via do estresse do retículo endoplasmático pode induzir apoptose em macrófagos de granulomas. O gene ITK pode atuar no	61359 genes

	presentes em granulomas.			combate ao Mtb em casos com formação de granulomas.	
<b>GSE 148731</b>	Análise do impacto do Mtb no metabolismo de macrófagos M1 e M2 oriundos de indivíduos sadios e, posteriormente, infectados com Mtb H37Rv.	<i>RNAseq</i>	Macrófagos diferenciados com fenótipos M1 e M2 infectados com cepas de Mtb H37Rv	A infecção por Mtb pode influenciar no metabolismo dos macrófagos, aumentando a síntese de NAD <sup>+</sup> e glutatona, na glicólise, e na via da creatina fosfato.	60448 genes

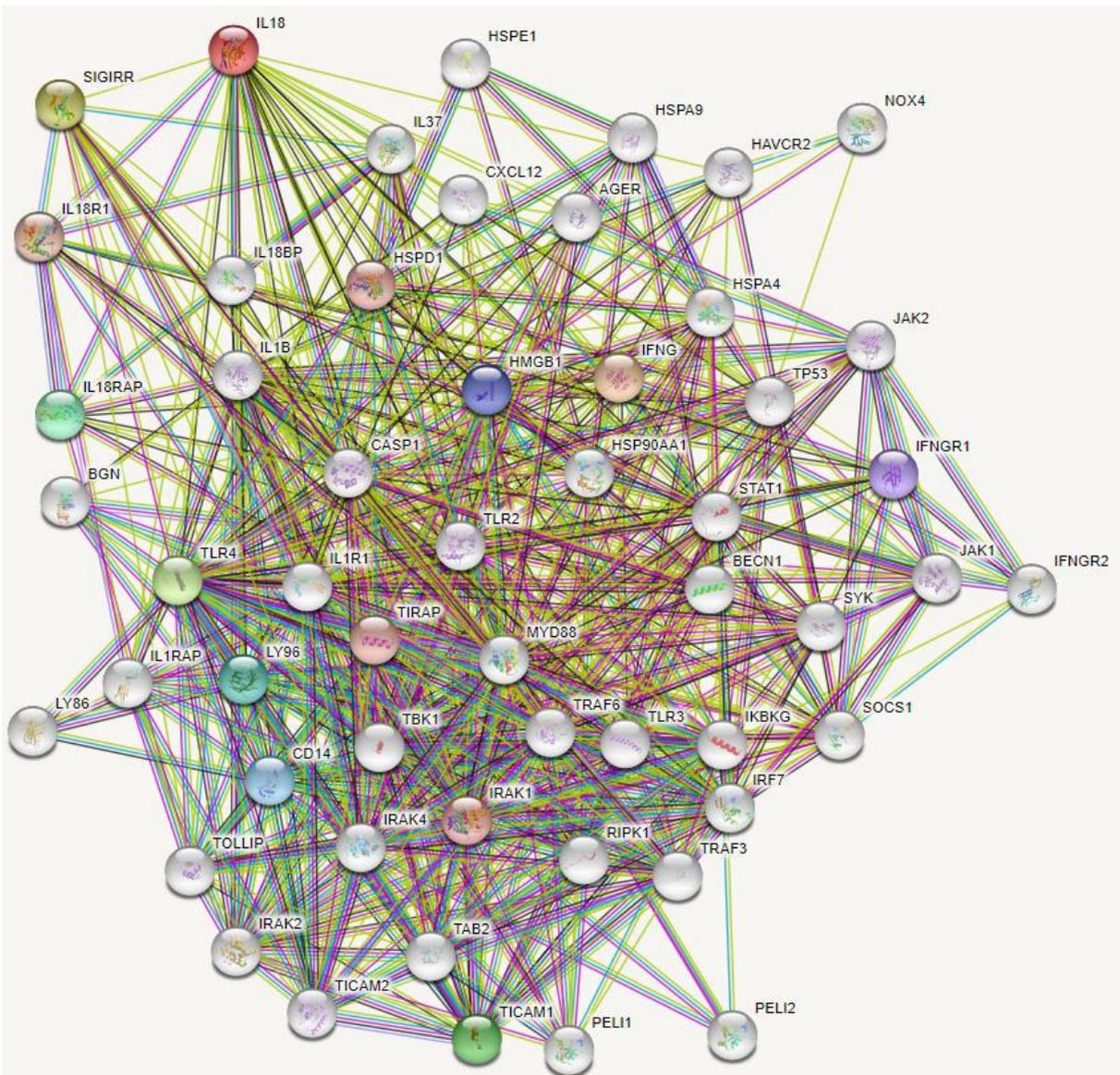
Fonte: Autoria própria (2022)

Para as etapas seguintes do estudo, três dos sete DataSets foram selecionados para composição da base de dados para uma análise mais detalhada, sendo eles GSE52819, GSE139871 e GSE148731.

## 5.2 Identificação *in silico* de genes relevantes para a resposta imunológica na tuberculose

Dentre as inúmeras vias de sinalização imunológica que podem ser avaliadas com os dados selecionados, optou-se por analisar a via da citocina IL-18, já reconhecidamente importante para a resposta imunológica na tuberculose. Foi agregada à via dessa citocina a proteína SIGIRR, tendo em vista seu possível envolvimento na patogênese da tuberculose (HASSUNA *et al*, 2022; RIVA *et al*, 2012). Para realizar uma análise mais minuciosa da via da IL-18 e a possível interação direta ou indireta com SIGIRR, foi realizada uma busca em duas bases de dados que criam redes de interação entre proteínas, STRING e InBio Discover. A análise via STRING revelou que um cluster de proteínas estava envolvido em resposta imune no contexto da IL-18. 54 genes/proteínas foram identificados, indicando a participação de SIGIRR na via de expressão do receptor de IL-18, IL18R. Ainda, foram identificados agrupamentos de genes/proteínas relacionados à via dos receptores do tipo Toll (TLR), relevantes para o reconhecimento do *M. tuberculosis* (Figura 4).

**Figura 4. Diagrama de interação proteína-proteína STRING na via de sinalização IL-18 e SIGIRR**



A análise de enriquecimento de GO de interações proteína-proteína entre genes/proteínas de interesse para o estudo.

No InBio Discover, apenas 26 genes/proteínas foram indicados pela base, seguindo os parâmetros de anotação e confiabilidade adotados (Figura 5). Nesta análise foi observada a formação de um cluster de interação envolvendo IL-18 e SIGIRR, no qual esta última mostra-se como proteína de sinalização entre o reconhecimento via TLR e a via intrínseca de IL-18.



**Figura 6. Diagrama de Venn contendo os 19 genes que constam simultaneamente no STRING e no InBio Discover.**

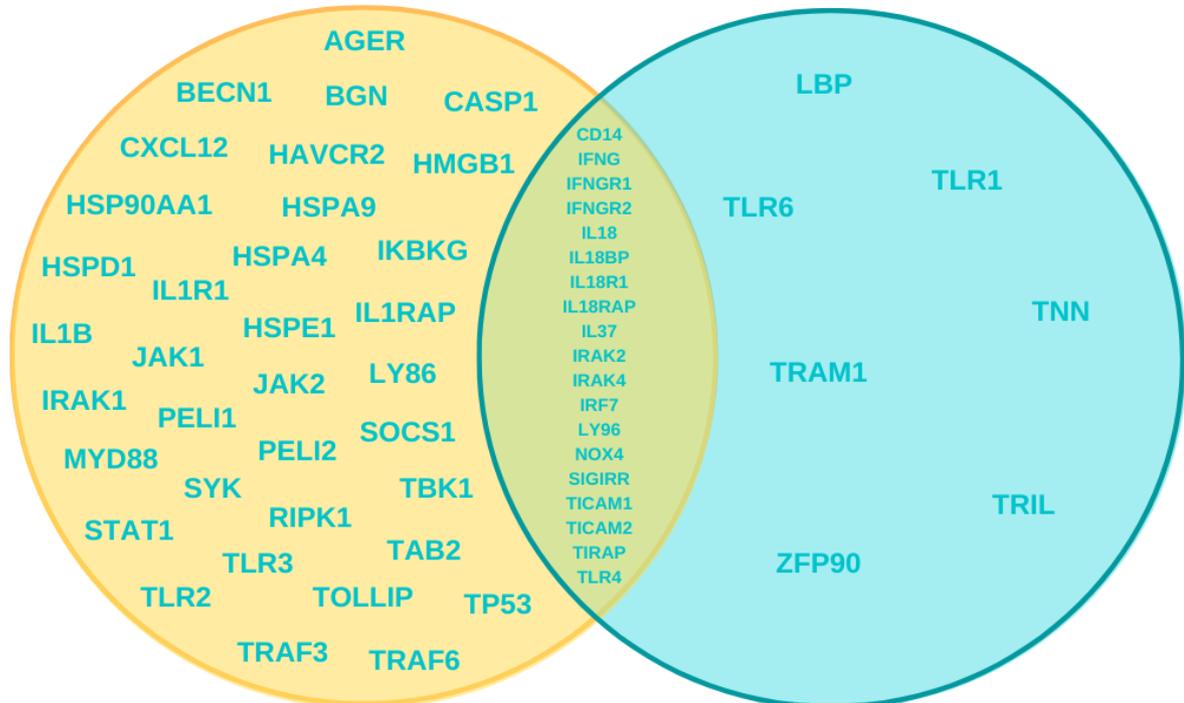


Diagrama de Venn abordando os genes presentes nas vias de sinalização obtidas através das bases de dados. O grupo em amarelo abrange os genes listados pelo STRING, enquanto o grupo azul engloba os genes do InBio Discover. A intersecção dos dois conjuntos elenca apenas os genes presentes simultaneamente em ambos os grupos.

### 5.3 Mineração e refinamento dos dados brutos

A partir da análise anterior e definição dos 19 genes-alvo para composição da base dados, as informações sobre estes genes foram extraídas dos DataSets GSE52819, GSE139871, GSE148731. O primeiro dataset a ser analisado foi o GSE52819, que tratou de estudar o padrão de transcrição gênica de macrófagos alveolares infectados com *M. tuberculosis* frente a estímulo com vitamina D. Os grupos de análise foram divididos em 4: não infectados, não infectados tratados com vitamina D, infectados, infectados tratados com vitamina D. Todos os testes ocorreram em triplicata e o *fold change* foi calculado utilizando o grupo não infectado com referência para os outros grupos. Dos 19 genes analisados, 7 apresentaram alteração na expressão entre o grupo não infectado e o infectado (CD14, IFNGR1, IL18R1, IL18RAP, IRF7, TICAM2, TLR4) (Tabela 1).

**Tabela 1. Análise de expressão diferencial dos genes selecionados a partir do GSE52819.**

<b>Macrófagos e DEGs</b>	<b>Não infectado + D</b>	<b>Infectado</b>	<b>Infectado + D</b>
<b>CD14</b>	Upregulated 15,584	Sem alteração 0,678	Upregulated 5,372
<b>IFNG</b>	Sem alteração 1,038	Sem alteração 0,973	Sem alteração 0,975
<b>IFNGR1</b>	Sem alteração 0,690	Upregulated 1,173	Sem alteração 0,889
<b>IFNGR2</b>	Upregulated 1,256	Upregulated 2,148	Upregulated 2,146
<b>IL18</b>	Sem alteração 0,650	Sem alteração 0,392	Sem alteração 0,441
<b>IL18BP</b>	Sem alteração 0,758	Sem alteração 0,813	Sem alteração 0,669
<b>IL18R1</b>	Sem alteração 0,858	Upregulated 2,296	Upregulated 1,929
<b>IL18RAP</b>	Upregulated 1,041	Upregulated 1,332	Upregulated 1,109
<b>IL37</b>	Upregulated 1,174	Upregulated 1,280	Upregulated 1,053
<b>IRAK2</b>	Upregulated 1,499	Upregulated 7,349	Upregulated 8,796
<b>IRAK4</b>	Sem alteração 0,684	Sem alteração 0,980	Sem alteração 0,950
<b>IRF7</b>	Sem alteração 0,948	Upregulated 2,403	Upregulated 1,642
<b>LY96</b>	Sem alteração 0,936	Upregulated 1,204	Upregulated 1,196
<b>NOX4</b>	Upregulated 1,131	Upregulated 1,088	Sem alteração 0,970
<b>SIGIRR</b>	Sem alteração 0,982	Sem alteração 0,797	Sem alteração 0,657
<b>TICAM1</b>	Upregulated 1,345	Upregulated 1,704	Upregulated 1,517
<b>TICAM2</b>	Sem alteração	Upregulated	Upregulated

	0,967	1,322	1,328
<b>TIRAP</b>	Sem alteração 0,873	Sem alteração 0,583	Sem alteração 0,583
<b>TLR4</b>	Sem alteração 0,776	Upregulated 1,356	Sem alteração 0,788

Fonte: Autoria própria (2022)

No contexto do presente trabalho, os dados relacionados a estimulação com vitamina D não ganham tanta relevância, embora sejam considerados, para todos os efeitos, os níveis de *fold change* e o grau de regulação da expressão dos genes presentes no dataset analisado. O grupo infectado recebeu maior destaque quando em comparação aos outros grupos deste dataset, tendo em vista sua melhor adequabilidade para fazer comparações com os dados dos outros datasets (Figura 7).

**Figura 7. Heatmap demonstrativo do perfil de expressão dos genes selecionados em macrófagos infectados ou não com Mtb.**

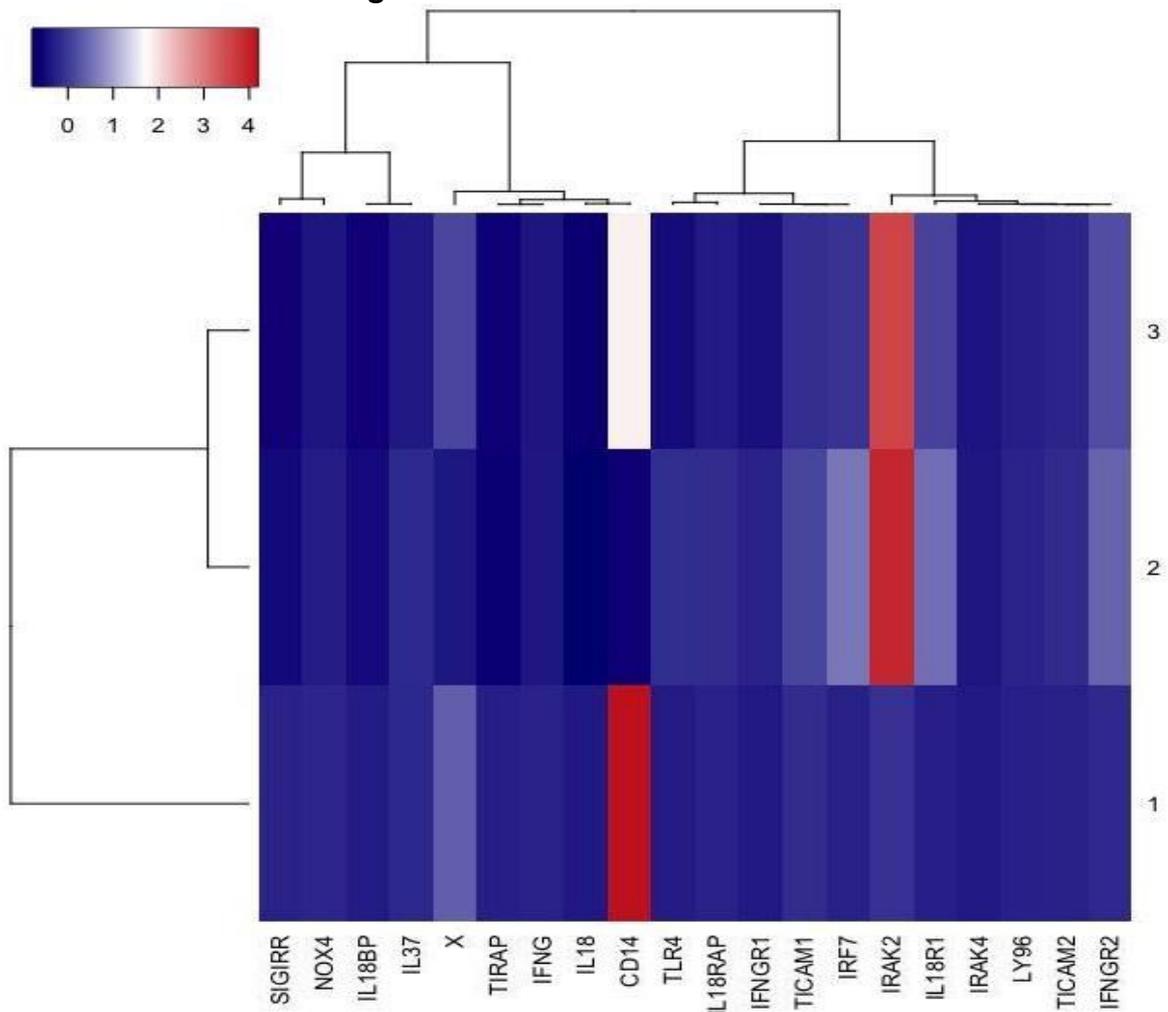


Gráfico de mapa de calor dos 19 genes selecionados com expressão diferencial, no contexto da infecção de macrófagos com *M. tuberculosis*. O padrão de expressão gênica mostra dois agrupamentos não supervisionados diferentes indicando um padrão de expressão divergente entre os diferentes grupos de tratamento: 1 - não infectado; 2 - infectado; 3 - infectado + vitamina D. A expressão alta é indicada em vermelho e a expressão baixa é indicada em azul.

O dataset GSE139871 foi construído com base em um estudo do padrão de transcrição de macrófagos infectados com duas cepas do *M. tuberculosis*, UT127 e UT205, sendo a última mais virulenta que a primeira. O trabalho dividiu os grupos a serem analisados em: macrófagos controle oriundos de pessoas saudáveis, macrófagos infectados com a cepa UT127, macrófagos infectados com a cepa UT205, macrófagos oriundos de pessoas portadoras de tuberculose, macrófagos já infectados com reinfecção pela cepa UT127, macrófagos já infectados com reinfecção pela cepa UT205. A tabela dos dados disponibilizados pelo dataset continha 47323 genes, com os genes codificados pelo seu código ID Illumina. Para obter o *fold Change* deste estudo foi necessário primeiro obter e traduzir o código ID de cada um dos 19 genes selecionados, para só depois calcular manualmente o valor, tendo em vista que o trabalho disponibilizou apenas os *read counts* de cada gene em cada uma das 24 amostras. Com os valores de *fold Change* gerados observou-se a expressão diferencial em 13 genes no grupo com macrófagos não infectados e infectados com a cepa UT127 (CD14, IFNGR2, IL18, IL18BP, IL18R1, IL18RAP, IL37, IRAK2, IRF7, NOX4, SIGIRR, TICAM1, TIRAP), e 12 genes no grupo com macrófagos não infectados e infectados com a cepa UT205 (IFNG, IFNGR2, IL18, IL18R1, IL18RAP, IL37, IRAK2, IRF7, NOX4, SIGIRR, TICAM1, TIRAP). No grupo com macrófagos de pacientes com tuberculose sendo posteriormente infectados com a cepa UT127, foram encontrados 12 genes com expressão diferencial (IFNG, IFNGR1, IFNGR2, IL18, IL18R1, IL18RAP, IL37, IRAK2, IRF7, TICAM1, TICAM2, TIRAP), já os reinfecados com UT205 foram encontrados 12 genes (IFNG, IFNGR1, IFNGR2, IL18, IL18R1, IL37, IRAK2, IRF7, SIGIRR, TICAM1, TICAM2, TIRAP) (Tabela 2; Figura 8).

**Tabela 2. Análise de expressão diferencial dos genes selecionados a partir do GSE139871.**

Macrófagos e DEGs	FC Mo 127 saud.	FC Mo 205 saud.	FC Mo 127 TB	FC Mo 205 TB
-------------------	-----------------	-----------------	--------------	--------------

<b>CD14</b>	Upregulated 1,034006021	Sem alteração 0,890752799	Sem alteração 0,95995794	Sem alteração 0,992602289
<b>IFNG</b>	Sem alteração 0,99848761	Upregulated 1,018336509	Upregulated 1,016570583	Upregulated 1,054591938
<b>IFNGR1</b>	Sem alteração 0,965905809	Sem alteração 0,965905809	Upregulated 1,023063967	Upregulated 1,014837131
<b>IFNGR2</b>	Upregulated 1,043555315	Upregulated 1,046642621	Upregulated 1,048756322	Upregulated 1,04316893
<b>IL18</b>	Upregulated 1,048434793	Upregulated 1,038763344	Upregulated 1,049633033	Upregulated 1,045642472
<b>IL18BP</b>	Upregulated 1,000855136	Sem alteração 0,997800464	Sem alteração 0,998061921	Sem alteração 0,99960807
<b>IL18R1</b>	Upregulated 1,065053019	Upregulated 1,004303601	Upregulated 1,052657922	Upregulated 1,026668167
<b>IL18RAP</b>	Upregulated 1,007003859	Upregulated 1,005040641	Upregulated 1,000673799	Sem alteração 0,99730327
<b>IL37</b>	Upregulated 1,002208828	Upregulated 1,001688043	Upregulated 1,002417928	Upregulated 1,00471583
<b>IRAK2</b>	Upregulated 1,319119585	Upregulated 1,32604703	Upregulated 1,276984599	Upregulated 1,243662713
<b>IRAK4</b>	Sem alteração 0,970915273	Sem alteração 0,980852025	Sem alteração 0,995080093	Sem alteração 0,992539917
<b>IRF7</b>	Upregulated 1,022971133	Upregulated 1,093565328	Upregulated 1,066628949	Upregulated 1,11875447
<b>LY96</b>	Sem alteração 0,914621144	Sem alteração 0,939915315	Sem alteração 0,920831497	Sem alteração 0,981978729
<b>NOX4</b>	Upregulated 1,000215064	Upregulated 1,000445882	Sem alteração 0,99712269	Sem alteração 0,996168483
<b>SIGIRR</b>	Upregulated 1,003484989	Upregulated 1,003231224	Sem alteração 0,996039754	Upregulated 1,006825209
<b>TICAM1</b>	Upregulated 1,06381815	Upregulated 1,056628259	Upregulated 1,068748537	Upregulated 1,037420992
<b>TICAM2</b>	Sem alteração 0,968964667	Sem alteração 0,983961939	Upregulated 1,022408626	Upregulated 1,040092583

<b>TIRAP</b>	Upregulated 1,006695977	Upregulated 1,003279714	Upregulated 1,001366889	Upregulated 1,001124716
<b>TLR4</b>	Sem alteração 0,896565019	Sem alteração 0,913611226	Sem alteração 0,945190955	Sem alteração 0,961895619

Fonte: Autoria própria (2022)

**Figura 8. Heatmap demonstrativo do perfil de expressão dos genes selecionados em macrófagos infectados ou não com Mtb.**

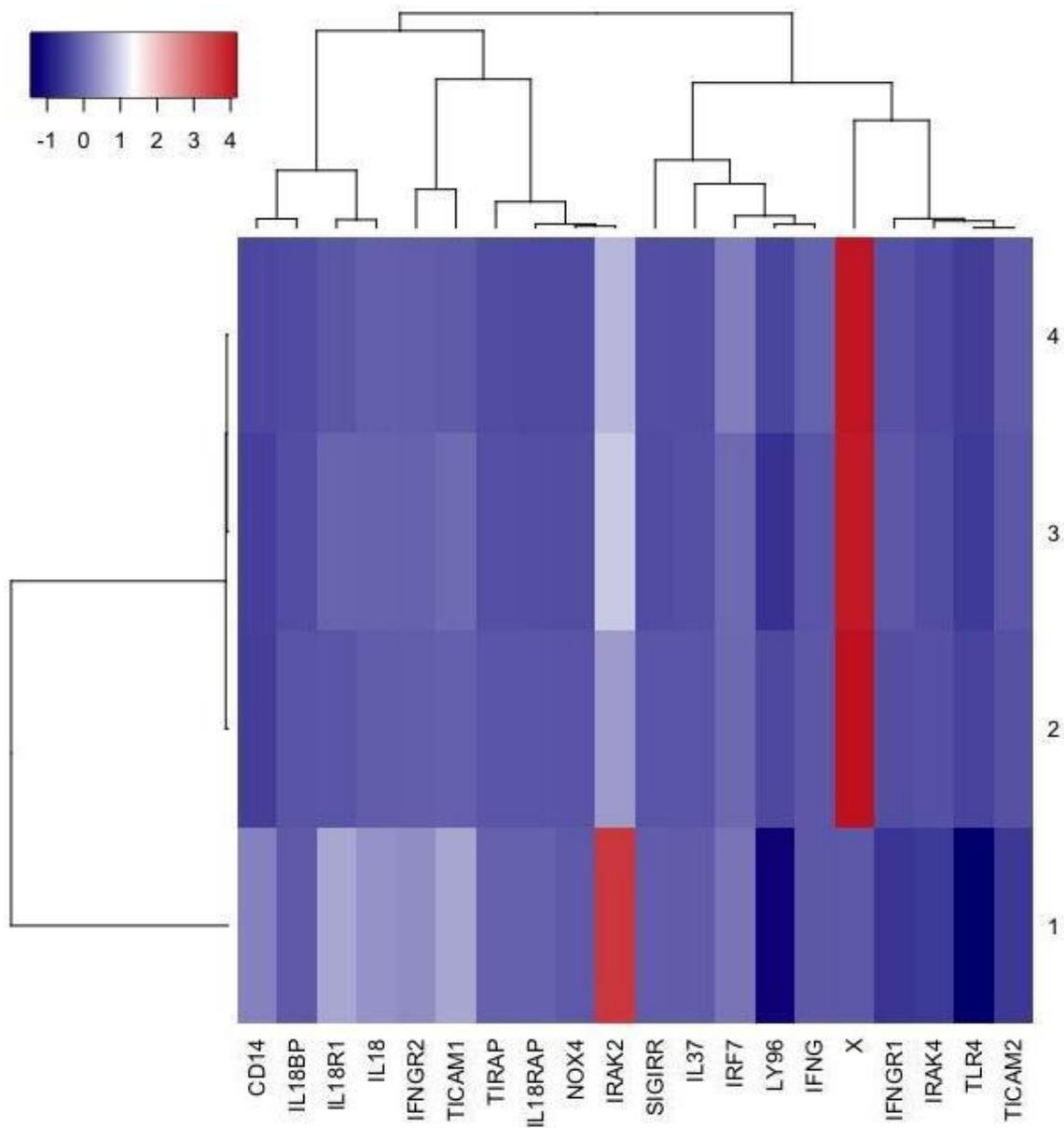


Gráfico de mapa de calor dos 19 genes selecionados com expressão diferencial, no contexto da infecção de monócitos com diferentes cepas de *M. tuberculosis* (127 e 205). O padrão de

expressão gênica mostra três agrupamentos não supervisionados, indicando um padrão de expressão divergente entre os diferentes grupos de tratamento: 1 - Monócitos padrão infectados com UT127; 2 - Monócitos padrão infectados com UT205; 3 - Monócitos de pacientes com TB infectados com UT127; 4 - Monócitos de pacientes com TB infectados com UT205. A expressão alta é indicada em vermelho e a expressão baixa é indicada em azul.

O conjunto de dados GSE148731 focou no estudo de diferentes fenótipos de macrófagos quando infectados pelo *M. tuberculosis*. Este dataset diferiu dos estudos anteriores pela polarização de macrófagos para os perfis M1 e M2, e de acordo com o tempo decorrido após os fagócitos em questão serem infectados pelo bacilo. Em ambos os grupos, as amostras foram coletadas em dois tempos, depois de 4 e 24 horas, permitindo observar o comportamento de transcrição precoce e tardio. Essa segregação possibilita entender um pouco mais sobre o papel dos macrófagos M1 e M2 na resposta a tuberculose, tendo em vista que, no geral, os padrões de resposta de cada subtipo costumam variar drasticamente e, até mesmo, terem efeitos que se antagonizam. Dessa forma, foi observado para o perfil M1 a alteração em todos os 19 genes avaliados em 4 horas de infecção e em 17 genes avaliados em 24 horas de infecção (Tabela 3; Figura 9). Para o perfil M2, 17 genes avaliados em 4 horas de infecção apresentaram expressão diferencial e em todos os 19 genes avaliados estavam com alterações em 24 horas de infecção (Tabela 3; Figura 9). Os macrófagos M1, com 4 horas de infecção, e os macrófagos M2, com 24 horas de infecção, apresentaram os maiores antagonismos de expressão gênica no contexto da infecção pelo bacilo.

**Tabela 3. Análise de expressão diferencial dos genes selecionados a partir do GSE148731.**

Macrófagos e DEGs	M1 4h	M1 24h	M2 4h	M2 24h
<b>CD14</b>	Downregulated -1,18558	Downregulated -3,21611	Downregulated -1,18724	Upregulated 3,028518
<b>IFNG</b>	Downregulated - 0, 2282	Sem alteração 0	Sem alteração 0	Upregulated 2,876364
<b>IFNGR1</b>	Downregulated -1,03091	Downregulated -1,5737	Downregulated -1,15268	Upregulated 1,035956
<b>IFNGR2</b>	Upregulated	Upregulated	Upregulated	Upregulated

	1,661891	1,456443	1,70802	1,210607
<b>IL18</b>	Upregulated 1,363564	Upregulated 1,207339	Upregulated 3,362602	Upregulated 1,29332
<b>IL18BP</b>	Upregulated 1,048306	Downregulated -1,16105	Downregulated -1,08318	Downregulated -1,84117
<b>IL18R1</b>	Upregulated 2, 436839	Upregulated 2,586802	Upregulated 9,365415	Upregulated 2,2903
<b>IL18RAP</b>	Upregulated 1,858253	Downregulated -1,22239	Upregulated 2,992157	Upregulated 4,356835
<b>IL37</b>	Upregulated 1,34329	Sem alteração 0,7861	Sem alteração 0	Downregulated -0,37152
<b>IRAK2</b>	Upregulated 4,65478	Upregulated 2,537495	Upregulated 10,91873	Upregulated 16,73689
<b>IRAK4</b>	Downregulated -1,01298	Upregulated 1,042094	Downregulated -1,49496	Downregulated -1,08324
<b>IRF7</b>	Upregulated 1,344878	Upregulated 16,00006	Upregulated 4,301516	Upregulated 8,68235
<b>LY96</b>	Downregulated -1,04279	Downregulated -1,10894	Upregulated 1,026252	Downregulated -1,17337
<b>NOX4</b>	Downregulated -1,29734	Upregulated 1,107313	Downregulated -1,30208	Downregulated -4,14935
<b>SIGIRR</b>	Downregulated -1,099	Downregulated -1,23697	Downregulated -1,1126	Downregulated -2,01283
<b>TICAM1</b>	Upregulated 1,15236	Upregulated 1,57452	Upregulated 2,524612	Upregulated 2,249624
<b>TICAM2</b>	Upregulated 1,110622	Upregulated 1,466251	Downregulated -1,34571	Upregulated 1,051014
<b>TIRAP</b>	Downregulated -1,06094	Upregulated 1,039004	Downregulated -1,76374	Upregulated 1,148763
<b>TLR4</b>	Downregulated -1,08348	Downregulated -1,17904	Downregulated -1,11162	Downregulated -1,91624

Fonte: Autoria própria (2022)

Figura 9. *Heatmap* demonstrativo do perfil de expressão dos genes selecionados em macrófagos infectados ou não com *Mtb*.

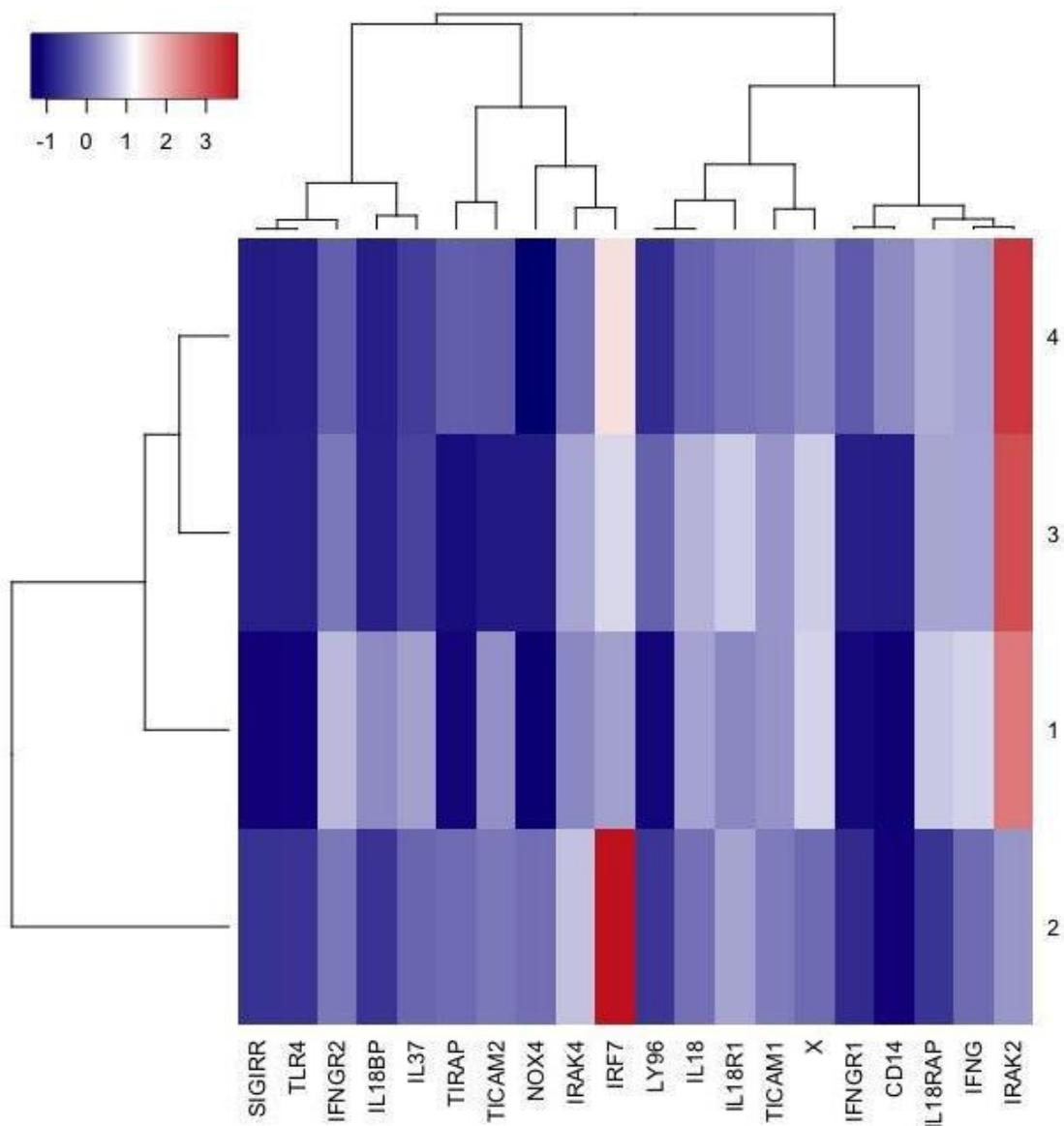


Gráfico de mapa de calor dos 19 genes selecionados com expressão diferencial, no contexto da infecção de macrófagos polarizados para os perfis M1 ou M2 com *M. tuberculosis*. O padrão de expressão gênica mostra quatro agrupamentos não supervisionados diferentes indicando um padrão de expressão divergente entre os diferentes grupos de tratamento: 1 - Macrófagos M1 após 4h; 2 - Macrófagos M1 após 24h; 3 - Macrófagos M2 após 4h; 4 - Macrófagos M2 após 24h. A expressão alta é indicada em vermelho e a expressão baixa é indicada em azul.

## 6. DISCUSSÃO

A transcriptômica tem se tornado na última década uma realidade para o estudo de diferentes situações biológicas, associadas ou não a patologias. Realizar a análise *in silico* de transcriptomas e a construção de bancos de dados de expressão gênica são determinantes para a prospecção de vias de sinalização nas mais diferentes doenças. No caso da tuberculose, uma das mais importantes doenças infectocontagiosas, a análise de genes associados à resposta imune é fundamental para definir e melhorar o diagnóstico à terapia da doença. No presente trabalho, selecionamos nas bases de dados públicas três DataSets com informações sobre a expressão gênica de proteínas importantes para resposta imune a *M. tuberculosis*. A partir desses DataSets foi realizada a mineração de dados para composição de uma base própria direcionada para a resposta de IL-18 na tuberculose. Foi incluída na análise a possível relação com SIGIRR, proteína com potencial participação nessa resposta. Os dados gerados permitem a definição de uma relação entre IL-18 e SIGIRR, além de elencar uma série de outras proteínas para estudos futuros de bioinformática, experimentais e/ou clínicos.

Antes mesmo de se iniciar a busca pelos datasets, foi elencada uma via de sinalização de interesse para que se pudesse delimitar o que seria analisado dentro dos datasets. O processo de seleção desses genes foi sistemático e girou em torno da via da citocina IL-18 com possível participação da proteína SIGIRR (IL 18/SIGIRR) e o processo de resposta ao bacilo *M. tuberculosis* (Mtb). Por serem proteínas reconhecidamente importantes para a eliminação Mtb e resolução dessa infecção, os genes IL18, SIGIRR, IFNG e TLR4 foram selecionados como parâmetros de entrada para análise nas plataformas STRING e o InBio Discover, resultando em duas redes de interações proteicas, ao redor das quatro selecionadas, com suas singularidades. A lista de genes/proteínas do STRING foi mais extensa e apresentou genes mais relacionados a IL18, SIGIRR, IFNG e, TLR4, enquanto a lista do InBio Discover foi mais sucinta e identificou os genes mais relacionados com os genes/proteínas de entrada. Com esses genes/proteínas foi elaborado um diagrama de Venn, salientando-se que tal ferramenta oferece uma melhor visualização do que é comum em dois conjuntos e o que é específico em cada um deles, ou seja, aquilo em que deve-se focar em uma análise, direcionando o processo de *datamining*.

Para a busca pelos dataset, foi crucial estabelecer corretamente as populações de estudo, o que diminuiu a possibilidade de erros de interpretação analítica e manteve o trabalho direcionado em seu objetivo principal, que é estudar macrófagos em casos de tuberculose ativa ou latente, ou até mesmo em situações laboratoriais controladas que simulem tais circunstâncias clínicas. Em relação ao datasets, dentre os 7 eleitos na base de dados GEO DataSets, os 3 que tiveram preferência no momento da escolha foram selecionados, principalmente, por abordarem a análise transcriptômica em um direcionamento mais próximo do ideal para o presente trabalho. A escolha dos datasets foi definida por parâmetros como a organização de dados, a possibilidade de análise de dados e grupos de forma independente do trabalho original, e, por último, a pertinência do assunto tratado no dataset em análise. Os critérios foram elaborados com o intuito de permitir uma seleção racional e direcionada dos dados.

Cada conjunto de dados possui dezenas de milhares de genes, todos analisados seguindo um mesmo protocolo e rigor metodológico, sendo possível considerar que diferentes genes em mesmas condições de análise estariam aptos a serem colocados em testes de hipóteses, assim como em testes de padrão de expressão pré e pós infecção. Essa grande quantidade de genes compilados, assim como em qualquer estudo que objetiva analisar determinada característica celular ou via de sinalização, passou por um processo de triagem, onde apenas genes envolvidos em determinados contextos foram selecionados. O processo de refinamento focou em coletar de cada dataset os 19 genes com maior *interaction score*, possibilitando ao trabalho uma área de trabalho delimitada e confiável.

A partir desse primeiro refinamento foi criada, para cada dataset, uma tabela informando o padrão de expressão de cada gene em cada tipo de experimento. Como cada estudo focou em uma análise distinta do contexto da tuberculose, o processo de montagem dessas tabelas teve de ser inicialmente padronizado e depois adaptado para cada dataset, já que é preciso extrair os dados de interesse sem deixar de lado a forma como cada trabalho obteve e categorizou essas informações.

A interleucina 18, base para nossa via de análise, é uma citocina da família da interleucina 1, com ação em diversas células, como linfócitos, macrófagos, mastócitos, basófilos, células epiteliais e nervosas, entre outras. Sua produção, por sua vez,

também é amplamente distribuída pelo corpo humano, sendo realizada por variadas células, como macrófagos, células dendríticas, células epiteliais, e nos mais variados tecidos biológicos. Sua principal característica, que também desperta interesse para esse trabalho, é a capacidade de intensificar a ação do IFN- $\gamma$  nas células mielóides, maximizando a atividade microbicida dessas células. Além disso, outros trabalhos já mostraram que a falta, ou produção irregular, de IL-18 acarreta em maior susceptibilidade a infecção por *M. tuberculosis*, seja através de estudos com camundongos *knockout* para IL-18 e IL-18R, seja por meio de estudos de polimorfismos no gene promotor de IL-18. (HASSUNA *et al*, 2022; YASUDA; NAKANISHI; TSUTSUI, 2019).

O gene IL18, codificante da IL-18, foi diferencialmente expresso nos datasets GSE139871 e GSE148731. Em ambos os casos ele estava regulado positivamente, ou seja, sua expressão estava aumentada quando comparado ao grupo controle. Esse achado corrobora com os achados de Wawrocki (2020), reiterando a atuação da IL-18 durante a infecção por *M. tuberculosis*. No caso do GSE52819, no qual os macrófagos foram infectados *in vitro* com Mtb, curiosamente, a expressão de IL-18 não sofreu alteração significativa em nenhuma das condições de análise. Este fato pode ser explicado por uma provável baixa sensibilidade do procedimento de análise do *microarray* empregado, já que é demonstrado por estudos anteriores que os níveis de IL-18 costumam ser elevados durante infecções por microrganismos que se alojam no meio intracelular (SCHNEIDER *et al*, 2010).

O *SIGIRR*, diferente do *IL18*, apresentou um padrão de expressão mais dissonante entre os datasets, já que no GSE52819 seus níveis não sofreram alteração e nos GSE139871 e GSE148731 estavam, respectivamente, regulados positivamente e negativamente. Esses resultados, que poderiam ser percebidos inicialmente como um fator de equívoco, podem ser reavaliados através de questionamentos analíticos em relação à geração dos dados brutos. No primeiro estudo, assim como aconteceu com a IL-18, a baixa sensibilidade do método pode ter atrapalhado a contagem dos transcritos. Nos dois trabalhos seguintes, o aumento e a baixa da transcrição provavelmente foram resultados das escolhas de grupos de análise, pois, enquanto o primeiro, GSE139871, operou separando monócitos por cepas do bacilo, o segundo

trabalho, GSE148731, optou por estudar os macrófagos em seus perfis de polarização, M1 e M2.

O SIGIRR, acrônimo para *single Ig IL-1- related receptor* (receptor órfão de imunoglobulina dependente de interleucina 1, em português), ou TIR8, ou ainda IL-1R8, faz parte de uma imensa família de receptores de interleucina 1 que desempenham os mais diversos papéis pelo corpo humano. O SIGIRR é expresso em uma infinidade de tecidos, a citar os pulmões, órgãos linfoides, fígado, trato digestivo e outros órgãos. Nos pulmões é expresso tanto em células epiteliais quanto na superfície dos leucócitos residentes daquela área. No âmbito da infecção pelo *M. tuberculosis*, camundongos deficientes em TIR/SIGIRR apresentaram significativo aumento da mortalidade em relação àqueles com expressão normal desse receptor (RIVA *et al*, 2012).

Dentre os demais genes, tiveram destaque no nível de expressão o IRAK2 e o IRF7. O IRAK2, ou quinase tipo 2 associada ao receptor 1 de interleucina 1 (sigla em inglês), é um fator de transcrição diretamente relacionado ao receptor de IL-1, presente também nos receptores *toll-like* (e.g. TLR 2 e 7). Esse fator de transcrição ganha destaque na indução de uma resposta celular pró-inflamatória e na produção de citocinas, auxiliando o sistema imune no combate a infecções. No contexto da tuberculose, e de encontro com o que foi encontrado neste trabalho, os níveis de transcrição estavam aumentados porque os receptores onde o IRAK2 está acoplado são o TLR 2 e TLR 7, ambos localizados em áreas estratégicas. O TLR2 é sensível à presença de lipopolissacarídeos e peptídeoglicanos, componentes característicos de bactérias, já o TLR7, por sua vez, é encontrado no compartimento lisossomal, exatamente onde os microrganismos invasores, incluindo o *M. tuberculosis* se encontram. Portanto, o aumento expressivo na expressão do IRAK2 pode ter uma relação com sua função biológica e a infecção pelo bacilo da tuberculose, conhecido por se alojar no ambiente intracelular dos macrófagos (WAN *et al*, 2011; MAYER-BARBER; BARBER, 2015).

O IRF7, ou fator regulador de interferon 7, é um fator de transcrição responsável por iniciar a produção de interferons do tipo I, foi outro gene com aumento no padrão de regulação, que, apesar de não ser o tipo de interferon padrão no combate ao *M. tuberculosis*, tem grande valor no combate ao bacilo. O bacilo de *M. tuberculosis*

busca o interior das células para se alojar e se evadir do sistema imunológico, sendo assim, a padrão de produção de citocinas costuma ter muitas similaridades nestes dois contextos, onde interferons do tipo I e do tipo II agem em conjunto na erradicação de patógenos que fazem das células seus hospedeiros (HOFFPAUIR *et al*, 2020; MACHADO *et al*, 2005).

Os genes diferencialmente expressos analisados provam a grande utilidade das bases de dados de transcriptômica, tendo em vista a possibilidade de testar hipóteses e expandir os conhecimentos sobre as vias de sinalização imunológica, isso tudo sem nem mesmo precisar realizar experimentos de bancada. Essa economia de tempo e material pode ser utilizada como elemento chave para o desenvolvimento de avaliações experimentais e clínicas, levando em conta não ser necessário grandes incentivos financeiros, nem extensas áreas físicas para implementação de um laboratório focado em análises *in silico*. Outro ponto que merece destaque é a praticidade de trabalhar com dados que, embora necessitem ser normalizados e ajustados, possibilitam uma reutilização constante de conjuntos de dados quando abordados em novos contextos e novas perspectivas e abre-se espaço para novas discussões e possíveis descobertas importantes no âmbito científico.

## 7. CONSIDERAÇÕES FINAIS

No presente trabalho foi possível a estruturação de uma base de dados de expressão de genes em macrófagos no contexto da resposta a tuberculose, sendo desenvolvidas matrizes de informações com grau de regulação gênica, além de *heatmaps*, e diagramas de interação de proteínas de vias de sinalização. Destaca-se que, além de proteínas já reconhecidamente envolvidas na resposta à tuberculose, as matrizes de dados construídas incluem proteínas com potencial envolvimento nessa patologia, que ainda não foram descritas em modelos experimentais ou clínicos.

Em relação aos genes diferencialmente expressos (DEGs), destacaram-se na análise dos datasets os genes *IL18*, *IRAK2* e *IRF7*, todos com expressão regulada positivamente na maioria dos contextos analisados, bem como o que a literatura tem documentado durante uma infecção por *M. tuberculosis*. Já o *SIGIRR*, outro gene analisado, apresentou um padrão de expressão irregular entre os estudos, o que dificultou o estabelecimento de uma conclusão direta sobre seu envolvimento na resposta, fazendo-se necessário analisar outros datasets em um contexto mais amplo para que se possa delinear seu comportamento.

A utilização de avançadas ferramentas de bioinformática, capazes de tratar uma vasta quantidade de dados, não demandou altos investimentos em infraestrutura e materiais, fazendo desse processo algo reprodutível em laboratórios de todos os portes. Dessa forma, foi estabelecida a base de dados para estudos futuros de bioinformática. Ainda, novas pesquisas clínicas experimentais com pacientes serão sustentadas pelas hipóteses geradas *in silico*. Toda essa informação agregada representa um ganho para as pesquisas em tuberculose em nosso estado e a aceleração do desenvolvimento do conhecimento.

## REFERÊNCIAS

- ABEL, Laurent *et al.* Genetics of human susceptibility to active and latent tuberculosis: present knowledge and future perspectives. **Lancet Infect Dis**, v. 18, n. 3, p. 64-75, mar. 2018. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8903186/>.
- AKOBENG, A. K. Understanding systematic reviews and meta-analysis. **Arch. Dis. Child**. v. 90, n. 8, p. 845–848, ago. 2005. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1720526/>.
- BARETT, Tanya *et al.* NCBI GEO: Archive for functional genomics datasets—update. **Nucleic Acids Res**. v. 41, n. 1, p. 991–995, jan. 2013. Disponível em: <https://academic.oup.com/nar/article/41/D1/D991/1067995?login=false>.
- BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. **Boletim Epidemiológico Tuberculose 2022**.
- BRASIL. Ministério da Saúde. Secretaria de Vigilância em Saúde. Departamento de Vigilância das Doenças Transmissíveis. **Manual de Recomendações para o Controle da Tuberculose no Brasil**. Brasília: Ministério da Saúde, 2019.
- CHURCHYARD, Gavin; *et al.* What We Know About Tuberculosis Transmission: An Overview. **The Journal of Infectious Diseases**, v. 216, n. 6, p. 629-635, nov. 2017. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5791742/>.
- FILHO, Dalson Britto Figueiredo *et al.* O que é, para que serve e como se faz uma metanálise? **Teoria e Pesquisa**, v. 23, n. 2, p. 205-228, jan. 2014. Disponível em: <https://www.teoriaepesquisa.ufscar.br/index.php/tp/article/view/401>.
- Haidich, A. B. Meta-analysis in medical research. **Hippokratia**, v. 14, n. 1, p. 29–37, dez. 2010. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3049418/>.
- HASSUNA, Noha A *et al.* Interleukin-18 and interferon- $\gamma$  single nucleotide polymorphisms in Egyptian patients with tuberculosis. **PLoS One**, v. 16, n. 1, jan. 2022. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7790531/>.
- HOFFPAUIR, Caitlyn T. *et al.* TRIM14 is a key regulator of the type I IFN response during *Mycobacterium tuberculosis* infection. **J Immunol**, v. 205, n. 1, p. 153-167, maio 2020. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7313415/>.
- JUNG, Klaus; FRIEDE, Tim; BEISSBARTH, Tim. Reporting FDR analogous confidence intervals for the log fold change of differentially expressed genes. **BMC Bioinformatics**, v. 12, n. 288, jul. 2015. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/21756370/>.
- KAN, Mengyuan; SHUMYATCHER, Mays; HIMES, Blanca E. Using omics approaches to understand pulmonary diseases. **Respir Res**, v. 18, n. 149, ago. 2017. Disponível em: <https://respiratory-research.biomedcentral.com/articles/10.1186/s12931-017-0631-9>.

LATORRACA, Carolina de Oliveira Cruz *et al.* Busca em bases de dados eletrônicas da área da saúde: por onde começar. **Diagn Tratamento**, v. 24, n. 2, p. 59-63, 2019. Disponível em: [https://docs.bvsalud.org/biblioref/2019/09/1015338/rdt\\_v24n2\\_59-63.pdf](https://docs.bvsalud.org/biblioref/2019/09/1015338/rdt_v24n2_59-63.pdf).

LAW, Charity W *et al.* voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. **Genome Biol.** v. 15, n. 29, fev. 2014. Disponível em: <https://genomebiology.biomedcentral.com/articles/10.1186/gb-2014-15-2-r29>.

LEANDRO M. MOREIRA. **Ciências genômicas: fundamentos e aplicações**. Ribeirão Preto: Sociedade Brasileira de Genética, 2015.

MACHADO, Paulo R. L. *et al.* Mecanismos de resposta imune às infecções. **Anais Brasileiros de Dermatologia**, v. 79, n. 6, p. 647-662, mar. 2005. Disponível em: <https://www.scielo.br/j/abd/a/3VZ9Fz6BH9HDGnPhkN3Ktbd/#>.

MAISON, David P. Tuberculosis Pathophysiology and anti-VEGF intervention. **Journal of Clinical Tuberculosis and Other Mycobacterial Diseases**, v. 27, n. 100300, maio 2022. Disponível em: <https://www.sciencedirect.com/science/article/pii/S2405579422000055>.

MALONE, Gaspar *et al.* Prospecção de genes em bibliotecas de cDNA. **R. Bras. Agrobiologia**, Pelotas, v. 12, n. 1, p. 07-13, jan./mar. 2006.

MAYER-BARBER, Katrin D.; BARBER, Daniel L. Innate and Adaptive Cellular Immune Responses to Mycobacterium tuberculosis Infection. **Cold Spring Harbor Perspective Medicine**, v.5, jul. 2015. Disponível em: <http://perspectivesinmedicine.cshlp.org/content/5/12/a018424.full.pdf+html>.

MENCK, Carlos Frederico Martins; MENEGHINI, Rogerio. Prêmio Nobel de Química 2015: Os mecanismos de reparo de DNA, **Quím. nov. esc.** São Paulo, v.37, n. 4, p. 264-269, nov. 2015. Disponível em: [http://qnesc.sbq.org.br/online/qnesc37\\_4/05-AQ116-15.pdf](http://qnesc.sbq.org.br/online/qnesc37_4/05-AQ116-15.pdf).

MILLER, Jeremy A. *et al.* Improving reliability and absolute quantification of human brain microarray data by filtering and scaling probes using RNA-Seq. **BMC Genomics**, v. 15, n. 1, fev. 2014. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/24564186/>.

MOITINHO JUNIOR, Valdomiro Silveira. **Perfil de transcriptoma imune revela um mecanismo de degradação de matriz extracelular na leishmaniose cutânea humana**. 2019. 70 f. Dissertação (Mestrado em Patologia) – Fundação Oswaldo Cruz, Instituto Gonçalo Moniz; Universidade Federal da Bahia, Salvador, 2019.

MÖLLER, Marlo; *et al.* Genetic Resistance to *Mycobacterium tuberculosis* Infection and Disease. **Frontiers in Immunology**, v. 9, n. 2219, set. 2018. Disponível em: <https://www.frontiersin.org/articles/10.3389/fimmu.2018.02219/full>.

OMS. Organização Mundial de Saúde. **Global Tuberculosis Report 2021**. Genebra: OMS, 2021.

PAI, Madhukar; *et al.* Tuberculosis. **Nature Reviews Disease Primers**, v. 2, out. 2016. Disponível em: <https://www.nature.com/articles/nrdp201676.pdf>.

PORCELLI, Steven A; JACOBS JR, William R. Exacting Edward Jenner's revenge: The quest for a new tuberculosis vaccine. **Science Translational Medicine**, v. 11, n. 490, maio 2019. Disponível em: <https://www.science.org/doi/10.1126/scitranslmed.aax4219>.

RAMIREZ-ALEJO, Noé; SANTOS-ARGUMEDO, Leopoldo. Innate defects of the IL-12/IFN- $\gamma$  axis in susceptibility to infections by *Mycobacteria* and *Salmonella*. **Journal of Interferon & Cytokine Research**, v. 34, n. 5, p. 307-317, maio 2014. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC4015507/pdf/jir.2013.0050.pdf>.

RIVA, Frederica *et al.* TIR8/SIGIRR is an interleukin-1 receptor/toll like receptor family member with regulatory functions in inflammation and immunity. **Front. Immunol**, v. 3, n. 322, out. 2012.

SCHNEIDER, Bianca E. *et al.* A role for IL-18 in protective immunity against *Mycobacterium tuberculosis*. **Eur J Immunol**, v. 40, n. 2, p. 396-405, fev. 2010. Disponível em: <https://pubmed.ncbi.nlm.nih.gov/19950174/>.

STUART, Tim *et al.* Comprehensive integration of single-cell data. **Cell**, v. 177, n. 7, p. 1888-1902, jun. 2019. Disponível em: [https://www.cell.com/cell/fulltext/S0092-8674\(19\)30559-8?\\_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867419305598%3Fshowall%3Dtrue](https://www.cell.com/cell/fulltext/S0092-8674(19)30559-8?_returnURL=https%3A%2F%2Flinkinghub.elsevier.com%2Fretrieve%2Fpii%2FS0092867419305598%3Fshowall%3Dtrue).

VRIES, Dylan H. de *et al.* Integrating GWAS with bulk and single-cell RNA-sequencing reveals a role for LY86 in the anti-*Candida* host response. **PLOS Pathog.** v. 16, n. 4, abr. 2020. Disponível em: <https://journals.plos.org/plospathogens/article?id=10.1371/journal.ppat.1008408> (alterar na citação em texto)

WAN, Youzhong *et al.* The dual functions of IRAK2 in TLR9-mediated interferon and proinflammatory cytokine production. **J Immunol**, v. 186, n. 5, p. 3006-3014, mar. 2011. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3163905/>.

WANG, Z.; GERSTEIN, M.; SNYDER, M. RNA-Seq: a revolutionary tool for transcriptomics. **Nat Rev Genet**, v. 10, p. 57-63, 2009.

WAWROCKI, Sebastian *et al.* Interleukin-18, Functional IL-18 Receptor and IL-18 Binding Protein Expression in Active and Latent Tuberculosis. **Pathogens**, v. 9, n. 6, jun. 2020. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7350328/>.

YASUDA, Koubun; NAKANISHI, Kenji; TSUTSUI, Hiroko. Interleukin-18 in health and disease. **Int J Mol Sci.** v. 20, n. 3, fev. 2019. Disponível em: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6387150/>.

ZHAO, Bi; ERWIN, Aqeela; XUE, Bin. How many differentially expressed genes: A perspective from the comparison of genotypic and phenotypic distances. **Genomics**, v. 110, n. 1, p. 67-73, jan. 2018. Disponível em: <https://www.sciencedirect.com/science/article/pii/S0888754317300745>.

ZHOU, Guangyan *et al.* NetworkAnalyst 30: A visual analytics platform for comprehensive gene expression profiling and meta-analysis. **Nucleic Acids Res.** v. 47, n. 1, p. 234–241, jul. 2019. Disponível em: <https://academic.oup.com/nar/article/47/W1/W234/5424072?login=false>.