Predição de microRNAs hepáticos como biomarcadores do desenvolvimento e avanço da DHGNA

Thomaz Guadagnini Ramalheira

July 31, 2022

Abstract

Com o aumento da prevalência de obesidade e doenças hepáticas ao redor do mundo, comumente relacionadas ao desequilíbrio dietético e sedentarismo, diversos estudos surgem com o intuito de compreender mecanismos ligados a esses distúrbios para que haja um entendimento mais amplo de fatores que podem ou não ter influência em seu desenvolvimento e progressão. Entre os mecanismos, um muito estudado nos dias atuais é a modulação epigenética que, envolve desde alterações químicas no DNA e histonas, até a transcrição de pequenos RNAs não codificantes chamados microRNAs, que participam da modulação da expressão gênica. Os microRNAs atuam principalmente da inibição da tradução de diversos genes, entre eles genes envolvidos com o metabolismo de lipídios, oncogênese, resistência à insulina, ou seja, diversos genes relacionados à obesidade e ao desenvolvimento e progressão de diversas comorbidades associadas. Dentre essas comorbidades, destaca- se a doença hepática gordurosa não alcoólica, que pode progredir para estágios mais avançados, levando ao desenvolvimento de cirrose e carcinoma hepatocelular. Com a necessidade de um diagnóstico precoce, principalmente em estágios iniciais e intermediários da DHGNA, possibilitando alguma intervenção na progressão da doença, o presente estudo surge com o desafio de encontrar e confirmar microRNAs isolados ou grupos de microRNAs hepáticos, se seus respectivos alvos, como biomarcadores da progressão da DHGNA, auxiliando em um diagnóstico precoce mais preciso. A partir da tecnologia de microarray, que nos permite monitorar a expressão de milhares de microRNAs de uma só vez, foram extraídos os dados referentes ao perfil de expressão de múltiplas amostras de conjuntos diferentes, encontrados em repositório específico para base de dados de expressão

gênica, sendo todos relacionados com estágio inicial da doença hepática gordurosa não alcoólica, a esteatohepatite não alcoólica. Os dados extraídos dos microarrays foram tratados e analisados estatisticamente, gerando uma lista de microRNAs e genes que, dentre milhares deles, podem ser considerados diferencialmente expressos entre as amostras analisadas. A partir dessa lista, os resultados foram contrastados entre as amostras e organizados para compreender a distribuição do perfil de expressão dos microRNA e genes e, através de ferramentas de bioinformática, utilizados para busca de genes- alvos e análise de enriquecimento funcional para que sejam compreendidas em quais vias de sinalização e processos biológicos estão envolvidos. A partir da análise de perfil de expressão, foi possível encontrar vinte microRNAs diferencialmente expressos que foram eleitos para análise de alvos e futuras análises. Pela análise de enriquecimento funcional foi possível observar que diversos genes-alvos parecem estar envolvidos em vias relacionadas ao desenvolvimento e progressão da DHGNA, dentre elas a via de "síntese de novo" de esfingolipídios e ceramidas. A próxima etapa envolve a confirmação em laboratório com tecido hepático humano.

Abstract

As the prevalence of obesity and hepatic diseases increases around the world, which are usually related to the dietary imbalance and a sedentary lifestyle, many studies emerge aiming to comprehend mechanisms connected to these disorders, so there is a wider and better understanding of factors that could or could not have an impact on its development and progression. Among these mechanisms, one of the most studied is the epigenetic modification which involves not only chemical alterations in the DNA and histones, but also the transcription of non-coding small RNAs known as microRNAs, which participate in the modification of gene expression. The microRNAs act mainly as an inhibitor of many genes translation such as genes involved with the lipid metabolism, oncogenesis and insulin resistance, meaning that they are deeply related to the obesity and also to the development and progression of many comorbidities. Between these comorbidities, in this work, we will highlight the non-alcoholic fatty liver disease (NAFLD) which can progress to more advanced stages of the disease, leading to the development of cirrhosis and hepatocellular carcinoma. The need of an early diagnosis is blatant, especially for initial and intermediate phases of the disease enabling an intervention in the progression, and for this reason the current study aims in finding and establishing isolated or groups of hepatic microRNAs, and also its own targets, as biomarkers of the progression of the NAFLD assisting in a more precise early diagnosis. Starting with the microarray

technology, which allows us to reveal the expression of thousands of microRNAs at once, data referred to the expression profiling of multiple samples was extracted from many datasets found in a repository that contains a database for gene expression related to the early stage of the NAFLD, which is the non-alcoholic steatohepatitis (NASH). All the data extracted from the microarrays were treated and used in statistical analysis, generating a list of microRNAs and genes that could be considered differentially expressed between thousands of them and all the samples used. Using the differentially expressed microRNAs (DEMs) and differentially expressed genes (DEGs) list the results were organized and used to comprehend the distribution of the expression profile of these microRNAs and genes, and by using bioinformatics tools target genes were found, and functional enrichment was done to understand signaling pathways and biological process involved with the DEMs/DEGs. From the expression profile analysis twenty microRNAs were found to be differentially expressed comparing the datasets and selected to be used in further analysis. Using the functional enrichment analysis, it was possible to find that many target genes seemed to be involved in pathways related to the development and progression of the NAFLD, e.g. pathway of de novo synthesis of sphingolipids and ceramides. The next step for the current work involves the confirmation of the potential biomarkers in wet-lab analysis using human hepatic tissue.

List of Figures

List of Tables

Lista de abreviaturas e siglas

- **miRNA:** microRNA
- **DHGNA:** Doença Hepática Gordurosa Não Alcoólica
- **OMS:** Organização Mundial da Saúde
- **mRNA:** RNA mensageiro
- **EHNA:** Esteatohepatite não alcoólica
- **ANOVA:** Análise de variância

- **EROS:** Espécies reativas de oxigênio
- **GEO:** *"do inglês"* Gene Expression Omnibus
- **NCBI:** *"do inglês"* National Center for Biotechnology Information
- **NGS:** *"do inglês"* Next Generation Sequencing
- **DEM:** *"do inglês"* Differentially Expressed microRNAs
- **DEG:** *"do inglês"* Differentially Expressed Genes
- **limma:** *"do inglês"* linear models for microarray data
- **logFC:** *"do inglês"* log Fold Change
- **KEGG:** *"do inglês"* Kyoto Encyclopedia of Genes and Genomes
- **GO:** *"do inglês"* Gene Ontology
- **MCODE:** *"do inglês"* Molecular Complex Detection

Contents

1 Introdução

Estimativas atuais demonstram claramente um grande aumento na prevalência de obesidade e sobrepeso ao redor do mundo. Dados divulgados pela Organização Mundial da Saúde representam uma preocupação latente nos dias de hoje e reforçam o grave problema de saúde pública relacionado à obesidade e

sobrepeso enfrentado por todos os países no mundo. Os números da OMS de 2016 revelam que houve um aumento aproximado de 300% no número de indivíduos obesos entre os anos de 1975 e 2016 ao redor do mundo. A soma de pessoas acima de 18 anos com sobrepeso, até o ano de 2016, era superior a 1.9 bilhão, sendo que destes, 650 milhões apresentavam obesidade (OMS, 2016).

O número de jovens e crianças também demonstra um cenário alarmante, com mais de 340 milhões de jovens e adolescentes entre 5 e 19 anos de idade que apresentavam sobrepeso ou obesidade em 2016. Em crianças abaixo de 5 anos de idade, os que apresentaram sobrepeso ou obesidade ultrapassaram 38 milhões em 2019 (OMS, publicado em 2016; OMS, publicado em 2019).

Diversos estudos vêm sendo realizados na tentativa de compreender os mecanismos envolvidos com a gênese da obesidade em todo o mundo. Sabe-se que hábitos de vida pouco saudáveis estão relacionados com o surgimento da obesidade e são capazes de agravar esse quadro, com destaque para o sedentarismo associado à ingestão de uma dieta desequilibrada, principalmente com predominância de gorduras saturadas (Gakidou et al., 2014). A obesidade resulta de períodos crônicos de balanço energético positivo, caracterizados por ingestão calórica maior que gasto energético, levando a um aumento na síntese de triacilglicerol, com hipertrofia e hiperplasia de adipócitos (Drolet et al., 2008). O aumento excessivo da gordura visceral está associado ao desenvolvimento de comorbidades, como doença arterial coronariana, hipertensão, diabetes melito tipo 2, dislipidemias, além de acúmulo ectópico de gordura no tecido hepático, denominado como doença hepática gordurosa não-alcoólica (DHGNA) (Meldrum et al., 2017). A DHGNA tem sido considerada, há alguns anos, como a manifestação hepática da síndrome metabólica (Angulo, 2007).

A síndrome metabólica, atualmente, é considerada um estado patofisiológico complexo sendo originada normalmente de um desbalanço na razão ingestão-gasto de calorias, além de diversos outros fatores associados como diferenças genéticas e epigenéticas de cada indivíduo, um estilo de vida sedentário, ou seja, baixa predominância de atividade física, e fatores relacionados à nutrição, como qualidade e composição dos alimentos (Sakalyen, 2018). Um grupo de fatores de risco como obesidade, resistência à insulina, dislipidemia e hipertensão que, em conjunto, resultam em um risco aumentado de diabetes mellitus tipo 2 e doenças cardiovasculares, estão diretamente ligados à síndrome metabólica de maneira geral (O'Neill e O'Driscoll, 2015).

A DHGNA é caracterizada pelo acúmulo de lipídios no fígado não decorrente do consumo de álcool e pode progredir para estágios ainda mais avançados da doença. A esteatohepatite não-alcoólica (EHNA) se dá através do desenvolvimento de um quadro inflamatório associado ao acúmulo de

gorduras no fígado e acomete aproximadamente 40% dos indivíduos ao longo de, em média, 6 anos (McPherson et al., 2015; Woo Baidal e Lavine, 2016). Com o agravamento do quadro, há o risco do desenvolvimento de fibrose, o que leva à cirrose hepática não-alcoólica e, em estágios mais avançados da doença, pode-se progredir para o carcinoma hepatocelular (Zhen He et al., 2016).

Dados recentes indicam que aproximadamente 25% da população mundial é acometida pela DHGNA, sendo que na América do Sul o valor chega próximo de 30% da população (Younossi, 2019).

A patogênese da EHNA, que é um dos estágios da progressão da DHGNA, tem sido explicada pela hipótese dos múltiplos hits, iniciando com a predominância do acúmulo de triglicérides em excesso nos hepatócitos juntamente com um aumento da lipotoxicidade caracterizado pelo aumento de ácidos graxos livres, colesterol livre e diversos outros metabólitos. Posteriormente há a caracterização do estresse oxidativo com a geração de espécies reativas de oxigênios (ROS) e ativação de mecanismos associados com o estresse do retículo endoplasmático. Com isso, alterações na flora intestinal acabam levando a uma produção e absorção aumentada de ácidos graxos, elevando os níveis de moléculas que contribuem para a ativação de vias inflamatórias e liberação de citocinas pró inflamatórias (Buzzetti et al., 2016).

A partir de todo o cenário da doença pelo mundo, e sua estreita relação com sobrepeso e obesidade, fica evidente a necessidade de maior atenção à DHGNA, visando diagnóstico e terapêutica eficiente. Sendo assim, diversas ferramentas e estratégias metodológicas foram utilizadas para caracterizar e diagnosticar a DHGNA ou seus diversos estágios.

O presente trabalho foi desenvolvido com o intuito de correlacionar alguns microRNAs como possíveis biomarcadores do desenvolvimento da DHGNA ou de sua progressão, utilizando de métodos de bioinformática a partir do perfil de expressão dos microRNAs. A alteração do perfil de expressão de pequenos RNAs não traduzidos, como é o caso dos microRNAs (miRNAs) vêm sendo amplamente estudados por se tratar de um fator importante na modulação da expressão gênica. Os miRNAs são pequenas moléculas de RNA não codificantes que agem como reguladores pós-transcricionais da expressão gênica, ligando-se à região 3' não traduzida de RNAs mensageiros-alvo (mRNA), revelando um papel fundamental dos miRNAs em diversas vias regulatórias já que, ao se ligar no mRNA, realiza a clivagem ou a desestabilização do mRNA, impedindo a sua tradução em proteína (Kim, 2005).

A bioinformática e a estatística auxiliam de maneira muito positiva na análise e estudo de dados submetidos pela comunidade científica pois, de forma geral, com o agrupamento de diversos dados se torna possível um olhar amplo,

com número amostral grande suficiente e muito maior quando comparado com estudos individuais, principalmente se tratando de amostras de humanos. Com base nas informações, o presente estudo tem como foco a utilização de diversas ferramentas de bioinformática e análise estatística para que seja possível predizer microRNAs como biomarcadores para a progressão da doença hepática gordurosa não-alcoólica utilizando de dados de microarrays obtidos de amostras de fígado de pacientes em dois estágios iniciais, seja EHNA ou DHGNA, comparando com amostras controle obtidas em cada agrupamento de dados utilizado. A utilização de microRNAs como biomarcadores pode ser uma ferramenta útil não apenas no diagnóstico precoce da DHGNA, nos estágios iniciais do acúmulo de gordura no fígado, mas, principalmente pode predizer maiores riscos de progressão da DHGNA para estágios mais avançados e que trariam maiores consequências para o paciente, auxiliando na caracterização da progressão da doença a partir de padrões de valores de expressão dessas moléculas.

2 Materiais e métodos

2.1 Seleção e filtragem dos dados

Para o presente estudo fez-se necessário analisar os perfis de expressão dos miRNAs e mRNAs em tecido hepático de humanos utilizando conjuntos de microarrays de tecido hepático de humanos. A tecnologia de microarray é uma poderosa ferramenta de alta capacidade que monitora a expressão de milhares de microRNAs de uma só vez entre dezenas de amostras processadas em paralelo em um único experimento (Liu et al., 2008).

Os dados foram coletados da plataforma Gene Expression Omnibus (GEO, criada pelo NCBI - Centro Nacional de Informação de Biotecnologia) que é um repositório público internacional que armazena e distribui gratuitamente dados genômicos funcionais de alta capacidade como microarrays e NGS (Next-generation sequencing) submetidos pela comunidade (Barrett et al., 2012).

A seleção dos microarrays com dados de microRNAs que foram utilizados para as análises foi realizada na plataforma de submissão de dados citada anteriormente, a partir de filtragem com as palavras-chave “nash”, “nafld”, “microRNA”, “miRNA”, “miR” e “liver”, além das filtragens oferecidas pela própria plataforma: Non-coding RNA profiling by array, microarrays de RNAs não codificantes, espécie homo sapiens.

Para a seleção dos microarrays com dados do perfil de expressão de RNAs mensageiros foram utilizadas as palavras “nash”, “nafld” e “liver” com

filtragem do tipo de experimento “Expression profiling by array”, na espécie homo sapiens.

2.2 Análise de genes e microRNAs diferencialmente expressos

Após a seleção dos microarrays que atendiam aos requisitos previamente estipulados, foi utilizada a plataforma GEO2R, diretamente integrada à GEO, para realização das análises estatísticas que compõe a extração de DEMs (Differentially Expressed microRNAs) ou DEGs (Differentially Expressed Genes).

O GEO2R é uma ferramenta web que permite que usuários comparam dois ou mais grupos de amostras de uma GEO Series para identificar microRNAs, mRNAs, entre outros, que são diferencialmente expressos nas condições experimentais [https://www.ncbi.nlm.nih.gov/geo/info/geo2r.html].

A análise foi realizada através de um pacote R, que é uma linguagem de programação com foco em análises estatísticas conhecido como limma (linear models for microarray analysis, ou modelos lineares para análises de microarrays) (Ritchie et al., 2015). Os DEMs/DEGs utilizados foram extraídos utilizando cutoff de p value ¡ 0.01 e logFC (log Fold Change) ¡ -1 — ¿ 1. A filtragem dos DEMs/DEGs que é feita para utilizá-los posteriormente se faz necessária para escolher, dentre todos os genes e microRNAs, os que são estatisticamente significantes, ou seja, que tem uma diferença considerada significante quando comparado entre os grupos utilizados tanto para o p value quanto para o logFC de cada DEM/DEG.

2.3 Representação das amostras dos datasets

Foram construídos heatmaps através dos valores de expressão brutos de cada um dos microarray selecionados. Os valores de expressão foram obtidos a partir da biblioteca Python GEOquery, que extrai os valores diretamente da plataforma GEO. Para construção dos heatmaps foram utilizadas duas bibliotecas de Python: SciPy e matplotlib.

Para a construção dos volcano plots foi utilizada a biblioteca Python bioinfokit, com a representação do valor de p no eixo Y do gráfico e o valor de logFC no eixo X, para cada microRNA.

2.4 Predição dos alvos dos microRNAs diferencialmente expressos

Para predição de genes potencialmente alvos dos miRNAs diferencialmente expressos nas análises anteriores, foi utilizada a plataforma mirDIP (mircroRNA Data Integration Portal), que realiza a busca dos genes alvos em diferentes bases de dados e agrupa os dados, gerando uma pontuação da provável interação do gene alvo com o miRNA. Foram considerados os genes com pontuação mínima classificada como "Very high" e foram selecionados genes-alvo com no mínimo 14 fontes confirmadas de 30 (algoritmos) utilizados pelo mirDIP.

Foi também realizada a análise reversa a partir dos DEGs extraídos dos 3 microarrays utilizados que continham perfis de expressão de genes para descoberta dos miRNAs que têm DEGs como alvo. A pontuação mínima dos microRNAs também foi classificada como "Very high" e selecionando os que tiveram sua interação confirmada por pelo menos 14 das 30 fontes que a plataforma utiliza.

2.5 Enriquecimento funcional

As análises de enriquecimento funcional utilizando genes alvos extraídos dos resultados da interação miRNA-mRNA foram executadas na plataforma DAVID (david.ncifcrf.gov/home.jsp), utilizando posteriormente os resultados de KEGG (Kyoto Encyclopedia of Genes and Genomes) para identificação de vias de sinalização enriquecidas e seus genes, e GO (Gene Ontology) para processos biológicos. Os resultados foram descritos em forma de tabela e de gráfico de barra construído com a biblioteca Python matplotlib, a partir dos valores de -log10 do valor de p.

Para a representação gráfica entre a relação dos genes que foram encontrados nas vias enriquecidas e os microRNAs, foi utilizado o Circos Plot através da ferramenta Circos Table Viewer [mkweb.bcgsc.ca/tableviewer/visualize/]. A pontuação de integração foi retirada de valores obtidos a partir da análise de interação miRNA-mRNA de mirDIP e as vias enriquecidas foram extraídas a partir dos resultados de análise por KEGG.

2.6 Interação proteína-proteína

Para a realização da análise de rede de interação proteína-proteína (PPI, Protein-protein interaction network) utiliza-se as interações entre os genes alvos dos microRNAs utilizados. As interações foram preditas utilizando o banco de dados STRING v11.0 (Jensen et al., 2009). A análise foi realizada

utilizando o software Cytoscape (Shannon et al., 2003) com um Confident interaction score ≥ 0.7. As redes de interação foram visualizadas também no software Cytoscape e as redes com maiores pontuações foram escolhidas utilizando o módulo MCODE (Molecular Complex Detection) com os critérios de: degree = 2, node score = 0.2, k-core = 2 (Bader e Hoguer, 2003).

3 Resultados e discussão

3.1 Filtragem e seleção dos microarrays