

FUNDAÇÃO OSWALDO CRUZ  
INSTITUTO AGGEU MAGALHÃES  
DOUTORADO EM BIOCÊNCIAS E BIOTECNOLOGIA EM SAÚDE

Raul Emídio de Lima

**Fatores genéticos e epigenéticos associados ao surgimento do carcinoma  
hepatocelular (HCC)**

Recife

2022

Raul Emídio de Lima

**Fatores genéticos e epigenéticos associados ao surgimento do carcinoma  
hepatocelular (HCC)**

Tese apresentada ao Doutorado em  
Biociências e Biotecnologia em Saúde do  
Instituto Aggeu Magalhães, Fundação  
Oswaldo Cruz, para a obtenção do grau de  
Doutor em Ciências.

Orientador:

Dr<sup>o</sup> Luydson R. Silva Vasconcelos

Coorientadoras:

Dr<sup>a</sup> Clarice N. Lins de Moraes

Dr<sup>a</sup> Patrícia M. M. Freire de Moura

Recife

2022

**Catálogo na fonte: Biblioteca Luciana Borges Abrantes dos Santos  
Instituto Aggeu Magalhães/Fundação Oswaldo Cruz**

---

- L732f      Lima, Raul Emídio de.  
Fatores genéticos e epigenéticos associados ao surgimento do carcinoma hepatocelular (HCC) / Raul Emídio de Lima. — Recife: [s.n.], 2022.  
134 p. il., graf., tab.; 30 cm.
- Tese (doutorado) – Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2022.  
Orientador: Luydson R. Silva Vasconcelos; coorientadoras: Clarice N. Lins de Moraes, Patricia M. M. Freire de Moura.  
Programa de Pós-graduação em Biociências e Biotecnologia em Saúde.
1. Biomarcadores tumorais - genética. 2. Carcinoma Hepatocelular - genética. 3. Polimorfismo de nucleotídeo único. 4. MicroRNAs. 5. Cirrose hepática - genética. 6. Predisposição Genética para Doença. 7. Fatores de Risco. I. Vasconcelos, Luydson R. Silva. II. Moraes, Clarice Neuenschwander Lins de. III. Moura, Patricia M. M. Freire de. IV. Título.

---

CDU 595.771

Raul Emídio de Lima

**Fatores genéticos e epigenéticos associados ao surgimento do carcinoma  
hepatocelular (HCC)**

Tese apresentada ao Doutorado em  
Biociências e Biotecnologia em Saúde do  
Instituto Aggeu Magalhães, Fundação  
Oswaldo Cruz, para a obtenção do grau de  
Doutor em Ciências.

Aprovado em: 16 de junho de 2021

Banca Examinadora



Dr. Luydson Richardson Silva Vasconcelos

Instituto Aggeu Magalhães/ Fundação Oswaldo Cruz



Dr<sup>a</sup>. Norma Lucena Cavalcanti Licínio da Silva

Instituto Aggeu Magalhães/ Fundação Oswaldo Cruz

Dr. Marcelo Henrique Santos Paiva

Instituto Aggeu Magalhães/Fundação Oswaldo Cruz

Dr. Pablo Rafael Silveira Oliveira

Universidade Federal da Bahia

Dr. Rodrigo Feliciano do Carmo

Universidade Federal do Vale do São Francisco

A minha avó, Dinalva Emídio (*in memoriam*)

Aos meus pais, Jacqueline e Ailton

Aos meus amores, Janayna e Gabriela

A todos os portadores de Câncer de Fígado

## AGRADECIMENTOS

Inicialmente preciso agradecer à Vida, pois o fato de escolher Ciências Biológicas como curso de graduação e como profissão foi somente um pretexto para entender essa energia fascinante chamada de Vida. Para alguns ela se manifesta de maneira teológica, para outros de maneira filosófica, mas na verdade a vida é energia e não cabe ser colocada em argumentos. Eu, enquanto admirador da Vida tento observar, analisar e na verdade conclui que a Vida só é bela e rara devido a certeza de sua finitude.

Finitude esta que durante muitos momentos deste trabalho foi-me colocada à mostra de maneira explícita. Inclusive em momentos pessoais, por isso, eu preciso agradecer a tudo o que minha avó Dinalva Emídio (*in memoriam*) deixou de aprendizado para todas as pessoas que com ela conviveram. Agradeço também aos meus pais Ailton Alves e Jacqueline Emídio, por terem se sacrificado por mim, por acolherem minhas escolhas mesmo às vezes sem entendê-las. Meu pai, é o maior exemplo de determinação e altruísmo com quem tenho o prazer de conviver. E minha mãe, mulher forte como essa não existe, meu porto seguro, minha fortaleza e o equilíbrio da minha vida. Eu tenho uma sorte danada e poder dizer que meus pais são também meus grandes amigos.

E eles ainda me presentearam com uma irmã única, inigualável, parceira e amiga, Janayna, a quem também devo agradecimentos por sempre se fazer presente para dividir comigo as agonias da vida. Também agradeço muito a minha amada, Gabriela, a pessoa que escolhi para compartilhar a vida. Talvez o curso em Ciências Biológicas tenha sido pretexto do destino para encontrá-la. Crescemos juntos, erramos juntos, acertamos juntos, mas sempre tivemos a certeza de que é assim que devemos seguir, juntos, pois sabemos que nos completamos.

O fardo pesado da pós-graduação no Brasil, foi um pouco mais leve para mim por conta da companhia diária de pessoas especiais, Waltinho, Mercinha, Amanda(s), Rhaíssa, Gilberto, Elis e Cynthia, meus companheiros do departamento de Parasitologia. Agradeço por poder conviver com vocês, pois em nossos momentos de cafés e conversas vocês me ensinam sobre amizade e sobre ciência.

Aos de sempre e para sempre, meus amigos de graduação que levo para vida! Ana, Felipe, Wendell, Taynã, Alécia, Silas, Manuela, Heytor e Vanessa, agradeço por

saber que sempre posso contar com vocês nos momentos de aperreio e nos momentos de comemoração, pude testemunhar o quanto cada um cresceu com o tempo. Fico extremamente feliz com as conquistas de vocês, assim como, sei que comemoram as minhas junto comigo. Obrigado pelos nossos encontros dentro e fora do laboratório.

Falando em encontros, o doutorado me deu de presente pessoas maravilhosas, por determinação do destino fui para no departamento de virologia e lá conheci Lígia e Aninha, que juntamente comigo e Heytor formamos um quarteto coeso, alinhado, com uma sintonia tão forte que parece que sempre nos conhecemos. Vocês modificaram minha visão profissional e minha visão de vida, superamos dificuldades e também escapamos para os cafés unidos também. A convivência com vocês fez essa trajetória única e bem mais leve.

Devo agradecimentos também ao meu trio de orientadores Luydson, Patrícia e Clarice. Inicialmente quero deixar registrado meu agradecimento a minha primeira orientadora Patrícia Moura que foi a pessoa responsável por me introduzir no mundo acadêmico, foi quem me mostrou como é o universo da pesquisa. Clarice, mais um presente colocado em meu caminho, nossos caminhos acabaram de cruzando por força do destino e até hoje sou grato por esse encontro, com Clarice aprendi que a pesquisa acadêmica pode ser mais leve dependendo de como olhamos para ela e agradeço por todas as oportunidades que me foram ofertadas.

Luydson, meu orientador, acredito que nesse processo ambos crescemos, afinal eu sempre vou poder dizer que fui o seu primeiro aluno de doutorado oficial, esse título ninguém tira de mim. Agradeço por toda a relação de confiança e por todas as oportunidades que me foram depositadas. Por tornar esse projeto viável e por me permitir crescer profissionalmente desenvolvendo ele.

Preciso agradecer a todos os pacientes com Hepatite C e com Câncer de fígado por aceitarem contribuir e possibilitar as análises e o conhecimento gerado com a execução deste projeto. E agradeço aos doadores de órgãos, pessoas que mesmo em momentos difíceis autorizaram e contribuíram com a ciência desenvolvida no nosso país, a qual vem sendo sumariamente desacreditada e desmontada.

Agradeço a todos, funcionários, professores e colegas de grupos de pesquisa que de modo direto ou indireto contribuíram com esta jornada e peço desculpa se o

espaço ou a mente cansada me fizeram esquecer alguém. A todos, de coração,  
Obrigado!!



Na calada da noite  
Os estudantes fazem  
O futuro amanhecer  
Quem aprendeu a ler e escrever  
Sabe bem que analfabeto  
Jamais voltará a ser

Mesmo que o destino  
Reserve um presidente adoecido  
E sem amor  
A juventude sonha sem pudor  
Flor da idade, muito hormônio  
Não se curva ao opressor

Pode apostar  
A rebeldia do aluno é santa  
Não senta na apatia da injustiça  
Agita e inferniza e a rua avança  
Escola não tem medo de polícia

Pode apostar  
Balbúrdia de aluno é o que educa  
Ensina o governante que caduca  
Retroceder não é uma opção

Respeito é pra quem dá educação  
Respeito é pra quem dá educação  
Respeito é pra quem dá educação

(FERRO, 2021)

## RESUMO

LIMA, Raul Emídio de. **Fatores genéticos e epigenéticos associados ao surgimento do carcinoma hepatocelular (CHC)**. 2021. Tese (Doutorado em Biociências e Biotecnologia em Saúde) - Instituto Aggeu Magalhães, Fundação Oswaldo Cruz, Recife, 2021.

Em contraste com a maioria das outras neoplasias malignas, o carcinoma hepatocelular (CHC), responsável por aproximadamente 90% dos cânceres primários do fígado, surge quase exclusivamente no contexto de inflamação crônica. Independentemente da etiologia, uma sequência típica de necroinflamação crônica, regeneração hepática compensatória, indução de fibrose hepática e cirrose subsequente muitas vezes precede a hepatocarcinogênese. Tais eventos são fortemente influenciados por fatores de modulação genética e epigenética, por exemplo, presença de polimorfismos e expressão diferencial de microRNAs, respectivamente, deste modo, esta pesquisa teve como objetivo identificar o perfil de expressão de microRNA's, assim como, criar um classificador baseado em inteligência artificial a partir de polimorfismos associados ao surgimento do Carcinoma Hepatocelular. Para isso amostras de sangue, tecidos tumorais e não tumorais do fígado foram selecionados para análise de polimorfismos genéticos e sequenciamento de nova geração, e posterior, validação por RT-qPCR. Comparamos expressões de microRNA entre áreas hepáticas de CHC e não-neoplásicas. Encontramos que os polimorfismos rs1840680 e rs2305619 no gene *PTX3*, o -221 no *MBL2*, rs1800629 no *TNF- $\alpha$*  e rs2333227 no gene *MPO* estão associados ao risco de aparecimento de CHC. Além disso, foram construídos algoritmos computacionais de predição do surgimento do CHC, utilizando aprendizado de máquina com acurácia média de 70%. Os microRNAs hsa-miR-26a/b, hsa-let-7a/c e hsa-miR-125a-5p foram encontrados com níveis de expressão maior em CHC, tais microRNAs atuam desregulando vias de controle de ciclo celular, possivelmente potencializando o aparecimento do câncer primário de células hepáticas. Deste modo, propõe-se que genes *PTX3*, *MBL2*, *TNF $\alpha$*  e *MPO* podem desempenhar funções significativas na evolução do processo fibrótico em pacientes acometidos por esta hepatopatia. No entanto, mais amostras devem ser investigadas para confirmar e compreender seu papel durante o desenvolvimento do carcinoma hepatocelular.

Palavras-chave: biomarcadores tumorais; microRNA; polimorfismo de nucleotídeo único; cirrose hepática e neoplasias hepáticas.

## ABSTRACT

LIMA, Raul Emídio de. **Genetic and epigenetic factors associated with the appearance of hepatocellular carcinoma (CHC)**. 2021. Thesis (PhD in Biosciences and Biotechnology in Health) - Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, 2021.

In contrast to most other malignancies, hepatocellular carcinoma (CHC), responsible for approximately 90% of primary liver cancers, appears almost exclusively in the context of chronic inflammation. Regardless of the etiology, a typical sequence of chronic necroinflammation, compensatory liver regeneration, induction of liver fibrosis, and subsequent cirrhosis often precede hepatocarcinogenesis. Such events are strongly influenced by factors of genetic and epigenetic modulation, for example, the presence of polymorphisms and differential expression of microRNAs, respectively. Thus, this research aimed to identify the expression profile of microRNA's, as well as to create a classifier based on artificial intelligence from polymorphisms associated with the appearance of Hepatocellular Carcinoma. For this, blood, tumor, and non-tumor tissues of the liver were selected for analysis of genetic polymorphisms and sequencing of new generation, and later, validation by RT-qPCR. We compared microRNA expressions between liver areas of CHC and non-neoplastic. We found that the rs1840680 and rs2305619 polymorphisms in the PTX3 gene, the -221 in MBL2, rs1800629 in TNF- $\alpha$ , and rs2333227 in the MPO gene are associated with the risk of developing CHC. In addition, computational algorithms were built to predict the onset of CHC, using the machine with an average accuracy of 70%. The microRNAs hsa-miR-26a / b, hsa-let-7a / c, and hsa-miR-125a-5p were found to have higher levels of expression in CHC, such microRNAs act by deregulating cell cycle control pathways, possibly potentiating or causing cancer primary liver cell. Thus, it is proposed that the PTX3, MBL2, TNF $\alpha$ , and MPO genes can play important roles in the evolution of the fibrotic process in patients affected by this liver disease. However, more should be investigated to confirm and understand their role during the development of hepatocellular carcinoma.

**KEYWORDS:** tumor biomarkers; microRNA, single nucleotide polymorphism, liver cirrosis and liver neoplasms

## LISTA DE ILUSTRAÇÕES

Figura 1 -	Taxas de incidência e prevalência de câncer hepático.....	28
Figura 2 -	Número estimado de casos, de câncer hepático, incidentes de 2018 a 2040, para ambos os sexos e todas as idades.....	29
Figura 3 -	Evolução da fibrose hepática à cirrose por diferentes fatores etiológicos.....	31
Figura 4 -	Organização Molecular do Vírus da Hepatite B (VHB).....	32
Figura 5 -	Organização genética e processamento de poliproteína do vírus da hepatite C (VHC).....	35
Figura 6 -	História Natural da DHGNA/EHNA associada ao CHC.....	37
Figura 7 -	Representação esquemática da progressão do processo da fibrose e promoção da tumorigênese hepática.....	41
Figura 8 -	Interação entre a predisposição genética, fatores ambientais e ocorrência do CHC.....	44
Figura 9 -	Mecanismos epigenéticos de regulação hereditária da expressão gênica.....	47
Quadro 1	Definição das variáveis independentes verificadas no estudo.	56
Figura 10	Arquitetura do sistema.....	60
Figura 11	Diagrama do protocolo experimental.....	63
Figura 12	Fluxograma do classificador genômico SVM-ANN para CHC.	64
Figura 13	Análise do desequilíbrio de ligação, em variantes proximais supostamente funcionais.....	73
Figura 14	Diagrama da matriz de confusão.....	79
Figura 15	Boxplots de contagens de leitura bruta (esquerda) e normalizada (direita).....	86
Figura 16	Distribuição das amostras baseada nos dados normalizados.	87
Figura 17	Análise de variação dos dados realizada via PCA (Principal Component Analysis).....	87
Figura 18	MA-plot (s) e Volcano plot (s) em cada tipo de estratégia de análise.....	88
Figura 19	Diagrama de Venn para microRNAs expressos diferencialmente.....	90

Figura 20	Gráfico evidenciando os miRNAs diferencialmente expressos considerando todas as combinações de análises.....	91
Figura 21	Mapa de calor evidenciando o miRNA expresso diferencial Top10.....	92
Figura 22	Contagem de <i>reads</i> no tecido do fígado por RNA-Seq.....	93
Figura 23	Heatmap evidenciando o miRNA expresso diferencial Top10 por RT-qPCR.....	96
Figura 24	Expressão relativa de microRNAs validados por RT-qPCR.....	95
Figura 25	Correlação entre o número de reads e o Cycling Treshold.....	97
Figura 26	Rede de interação mRNA-miRNA de miRNAs diferenciais validados por RT-qPCR.....	98
Figura 27	Análise de enriquecimento funcional de mRNAs previstos regulados pelos 5 miRNAs validados em RT-qPCR.....	99

## LISTA DE TABELAS

Tabela 1 -	Exemplo de codificação Recessiva e One-Hot para um dado SNP.....	59
Tabela 2 -	Diferentes configurações do conjunto de dados classificadas neste trabalho.....	63
Tabela 3 -	Características clínicas e bioquímicas relacionadas aos grupos com diferentes níveis de fibrose hepática.....	70
Tabela 4 -	Frequência alélica e genotípica dos SNPs estudados, de acordo com a severidade da fibrose hepática.....	72
Tabela 5 -	Modelos Genéticos utilizados para análise de associação ao risco de progressão para doença hepática grave.....	74
Tabela 6 -	Especificação da proporcionalidade de classes presentes do conjunto de dados.....	77
Tabela 7 -	Melhores configurações encontradas, demonstrando os parâmetros de Acurácia, F1, Precisão, Recall e AUC/ROC.....	78
Tabela 8 -	Resultados dos classificadores com o conjunto de dados sem aplicação de seleção de características.....	80
Tabela 9 -	Resultados dos classificadores com a aplicação de algoritmo genético para seleção de características no conjunto (F0,F1,F2,F3 X CHC).....	81
Tabela 10 -	Resultados dos classificadores com a aplicação de algoritmo genético para seleção de características no conjunto (F4 X CHC).....	81
Tabela 11 -	Resultados dos melhores classificadores para cada uma das configurações com a aplicação de algoritmo genético para seleção de características.....	82
Tabela 12 -	Porcentagens de incidência das características em cada conjunto de divisão dos dados nos diferentes classificadores.....	83
Tabela 13 -	Características clínicas e bioquímicas relacionadas às amostras utilizadas na fase de descoberta (experimento RNA-Seq).....	85

Tabela 14 -	MicroRNAs diferencialmente expressos em ambos os tecidos, apenas no CHC e apenas no tecido do tumor adjacente.....	89
Tabela 15 -	MicroRNAs diferencialmente expressos em cada material biológico, comparando a taxa de variação para cada técnica.....	97
Tabela 16 -	Correlação entre os microRNAs validados e variáveis bioquímicas.....	97

## LISTA DE ABREVIATURAS E SIGLAS

AASLD	- Associação Americana para o Estudo de Doenças Hepáticas
AFB1	- Anflatoxina B1
AFP	- Alfafetoproteína
ALD	- Álcool Desidrogenase
ALT	- Alalina Aminotransferase
AMR	- Ad Mixed American
AST	- Aspartato Aminotransferase
BCLC	- <i>Barcelona clinic liver cancer</i>
BT	- Bilirrubina Total
cDNA	- Ácido desoxirribonucleico complementar
CYP2E1	- Cirocromo P450 2E1
DAA	- Agentes antivirais de ação direta
DNA	- Ácido desoxirribonucleico
DT	- Decision Tree Learnig
ECM	- Matriz Extracelular
EDTA	- Ácido etilenodiamino tetra-acético
EHE	- Esquistossomose Hepato-espênica
EHW	- Equilíbrio de Hardy-Weinberg
FA	- Fosfatase Alcalina
GO	- Gene Ontology
HAS	- Hipertensão Arterial Sistêmica
VHB	- Vírus da Hepatite B
HCA	- Adenoma Hepatocelular
CHC	- Carcinoma Hepatocelular
VHC	- Vírus da Hepatite C
HDV	- Vírus da Hepatite D ou Delta
HSC	- Células estreladas hepáticas
IC	- Intervalo de Confiança



IFNL3	- Interferon Lambda 3
IL-10	- Interleucina 10
IL-6	- Interleucina 6
IMC	- Índice de Massa Corporal
IRES	- Sítio interno de entrada do ribossomo
KEGG	- Kyoto Encyclopedia of Genes and Genomes
KNN	- K-Nearest Neighbor
lincRNA	- RNA longo intergênico
lncRNA	- RNA longo não codificante
MAPK	- Via MAP Quinase
MBL2	- Lectina Ligante de Manose 2
miRNA	- microRNA
MPO	- Mieloperoxidase
mRNA	- Ácido ribonucleico mensageiro
Mt rRNA	- RNA ribossomal mitocondrial
Mt tRNA	- RNA transportador mitocondrial
NAFLD	- Doença Hepática Gordurosa Não Alcoólica
NASH	- Esteatohepatite Não-Alcoólica
ncRNA	- RNA não codificante
OR	- Razão de Chance
ORF	- Matriz de leitura aberta
PCR	- Reação em cadeia da polimerase
PEG-IFN $\alpha$	- Interferon-alfa peguilado
PTX3	- Pentraxina 3
RBV	- Ribavirina
RM	- Ressonância Magnética
RNA	- Ácido ribonucleico
RT-PCR	- Reação em cadeia da polimerase por transcriptase reversa
scaRNA	- RNA pequenos de corpos-específicos de Cajal
SNP	- Polimorfismo de base única

snRNA	- RNA pequeno nuclear
SOD2	- Superóxido Desmutase
SUS	- Sistema único de saúde
SVM	- Support Vector Machine
SVM-RBF	- Support Vector Machine-Radial Basis Function
SVR	- Resposta virológica sustentada
TACE	- Quimioembolização transarterial
TC	- Tomografia Computadorizada
TGF $\alpha$	- Fator de transformação do crescimento alfa
US	- Ultrassonografia
UTR	- Região não traduzida
VPM	- Volume Plaquetário Médio
$\gamma$ GT	- Gama-glutamil Transpeptidase
$\chi^2$	- Qui-quadrado

## SUMÁRIO

1	INTRODUÇÃO.....	24
2	REFERENCIAL TEÓRICO.....	27
2.1	Epidemiologia Do Carcinoma Hepatocelular.....	27
2.2	Etiologia Do Carcinoma Hepatocelular (CHC).....	30
2.2.1	<i>Hepatite B</i> .....	32
2.2.2	<i>Hepatite C</i> .....	33
2.2.3	<i>Ingestão de Álcool</i> .....	36
2.2.4	<i>Doença hepática gordurosa não alcoólica e Esteatohepatite não alcoólica</i> .....	37
2.3	Patogênese Do Carcinoma Hepatocelular.....	38
2.3.1	<i>Aspectos genéticos do carcinoma hepatocelular</i> .....	42
2.3.2	<i>Aspectos epigenéticos do carcinoma hepatocelular</i> .....	46
2.3.3	<i>microRNA's e Hepatocarcinoma Celular</i> .....	47
3	JUSTIFICATIVA.....	50
4	OBJETIVO GERAL.....	52
4.1	Objetivos Específicos.....	52
5	METODOLOGIA.....	53
5.1	Desenho de estudo.....	53
5.2	População Alvo.....	53
5.2.1	<i>Grupos de Estudos</i> .....	53
5.2.1.1	Para a primeira etapa do estudo.....	53
5.2.1.2	Para a segunda etapa do estudo.....	54
5.3	Tipo De Amostragem E Definição Do Tamanho Da Amostra.....	55
5.4	Definição e categorização das variáveis.....	55
5.5	Métodos de coleta.....	57
5.6	Extração de DNA.....	57
5.7	Detecção das variantes polimórficas.....	57
5.8	Construção do algoritmo genético para seleção de características por meio de aprendizado de máquina.....	58
5.8.1	<i>Codificação</i> .....	58

<b>5.8.2</b>	<b><i>Subgrupos e Seleção de Características.....</i></b>	<b>58</b>
<b>5.8.3</b>	<b><i>Classificadores.....</i></b>	<b>58</b>
<b>5.8.4</b>	<b><i>Métricas.....</i></b>	<b>59</b>
5.8.4.1	Acurácia.....	60
5.8.4.2	Precisão.....	60
5.8.4.3	Recall.....	61
5.8.4.4	F1.....	61
5.8.4.5	AUC/ROC.....	61
<b>5.9</b>	<b>Protocolo Experimental para construção de algoritmo genético.....</b>	<b>62</b>
<b>5.10</b>	<b>Obtenção das amostras para Sequenciamento de RNA.....</b>	<b>65</b>
<b>5.11</b>	<b>Extração de microRNA.....</b>	<b>65</b>
<b>5.12</b>	<b>Sequenciamento de Small RNAs.....</b>	<b>65</b>
<b>5.13</b>	<b>Validação dos miRNAs por RT-qPCR.....</b>	<b>66</b>
<b>5.14</b>	<b>Análise da rede de interação mRNA-miRNA.....</b>	<b>66</b>
<b>5.15</b>	<b>Plano de Tabulação e análise estatística dos dados.....</b>	<b>67</b>
<b>5.16</b>	<b>Considerações Éticas.....</b>	<b>67</b>
<b>6</b>	<b>RESULTADOS.....</b>	<b>69</b>
<b>6.1</b>	<b>Análise dos Polimorfismos.....</b>	<b>69</b>
<b>6.2</b>	<b>Construção da Rede Neural Artificial.....</b>	<b>76</b>
<b>6.2.1</b>	<b><i>Conjunto de dados.....</i></b>	<b>76</b>
<b>6.2.2</b>	<b><i>Melhores configurações.....</i></b>	<b>77</b>
<b>6.2.3</b>	<b><i>Configuração do conjunto de dados completo para classificação dos modelos.....</i></b>	<b>79</b>
<b>6.2.4</b>	<b><i>Conjunto de dados F0,F1,F2,F3 X CHC.....</i></b>	<b>80</b>
<b>6.2.5</b>	<b><i>Conjunto de dados F4 X CHC.....</i></b>	<b>81</b>
<b>6.2.6</b>	<b><i>Melhores modelos em relação à métrica de AUC\ROC.....</i></b>	<b>82</b>
<b>6.3</b>	<b>RNA-Seq e validação dos dados.....</b>	<b>84</b>
<b>6.3.1</b>	<b><i>Caracterização da amostra.....</i></b>	<b>84</b>
<b>6.3.2</b>	<b><i>Análise diferencial de miRNAs em CHC, tecido adjacente e controles.....</i></b>	<b>86</b>
<b>6.4</b>	<b>Validação dos miRNAs candidatos por RT-qPCR.....</b>	<b>94</b>
<b>6.5</b>	<b>Determinação funcional via análise de enriquecimento.....</b>	<b>98</b>

<b>7</b>	<b>DISCUSSÃO.....</b>	<b>100</b>
<b>8</b>	<b>CONCLUSÃO.....</b>	<b>109</b>
	<b>REFERÊNCIAS.....</b>	<b>110</b>
	<b>ANEXO A - TERMO DE APROVAÇÃO DO COMITÊ DE ÉTICA DA UNIVERSIDADE DE PERNAMBUCO .....</b>	<b>123</b>
	<b>ANEXO B - TERMO DE APROVAÇÃO DO COMITÊ DE ÉTICA DO INSTITUTO AGGEU MAGALHÃES.....</b>	<b>126</b>
	<b>APÊNDICE A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO DA UNIVERSIDADE DE PERNAMBUCO.....</b>	<b>130</b>
	<b>APÊNDICE B - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO DO INSTITUTO AGGEU MAGALHÃES.....</b>	<b>131</b>
	<b>APÊNDICE C - QUESTIONÁRIO AVALIATIVO ELABORADO PARA O ESTUDO.....</b>	<b>133</b>

## 1 INTRODUÇÃO

O fígado tem funções metabólicas e de depuração vitais que envolvem a absorção de nutrientes, produtos residuais e agentes patogênicos do sangue. Além disso, as suas funções imunorreguladoras singulares mediadas pela expressão local de receptores coinibitórios e mediadores imunossupressores que ajudam a prevenir danos inadvertidos ao órgão, fazem do fígado uma víscera de importância única para a manutenção da vida humana (PROTZER; MAINI; KNOLLE, 2012). O câncer hepático corresponde a um importante conjunto de doenças que podem acometer o fígado, gerando um problema relevante no âmbito da saúde coletiva, causando consideráveis morbidade e mortalidade em escala global (HALLIDAY; KLENERMAN; BARNES, 2011; KIM; BOSE; RAY, 2014). Em contraste com a maioria das outras neoplasias malignas, o carcinoma hepatocelular (CHC), que é responsável por aproximadamente 90% dos casos de câncer de fígado, surge quase exclusivamente no contexto de inflamação crônica (RINGERLHAN, 2019).

Considerado o principal tipo de câncer primário de fígado, o CHC, possui uma alta mortalidade e independentemente da etiologia ocorre uma sequência típica de necroinflamação crônica, regeneração hepática compensatória, indução de fibrose hepática e cirrose subsequente que frequentemente precedem a hepatocarcinogênese (LAFARO; DEMIRJIAN; PAWLINK, 2015). É responsável globalmente por aproximadamente 800.000 mortes ao ano (GLOBAL BURDEN OF DISEASE CANCER, 2017) e contribui fortemente para a crescente taxa de mortalidade associada a neoplasias malignas e corresponde a principal causa de morte relacionada ao câncer em muitas partes do mundo, sendo estimada como a quinta causa mais comum de morte relacionada ao câncer em todo o mundo é o quinto câncer mais comumente diagnosticado em todo o mundo (LAFARO; DEMIRJIAN; PAWLINK, 2015).

Somente na China, país mundialmente mais populoso, são registrados metade dos casos de câncer hepático de todo mundo (LAFARO; DEMIRJIAN; PAWLINK, 2015). No Brasil, os casos de CHC são predominantes nos estados do Espírito Santo e Bahia, porém, são escassos estudos que representem a real incidência de CHC nacionalmente, se fazendo assim, necessários maiores estudos relacionados a esse tema (GOMES *et al.*, 2013).

Dentre as principais causas para o surgimento do carcinoma hepatocelular, destacam-se as hepatites virais B e C, que são responsáveis pela maioria dos casos de câncer hepático relatados em todo o mundo. Além das hepatites virais e da cirrose hepática, outros fatores também contribuem para surgimento do CHC, tais como o consumo excessivo de álcool, a doença hepática gordurosa não alcoólica (DHGNA), em especial sua forma mais grave, conhecida como, Esteatohepatite não alcoólica (EHNA) (YANG *et al.*, 2019).

Proveniente dos hepatócitos, consideradas as principais células do parênquima hepático, o CHC surge a partir de uma doença hepática crônica, pois ao final este processo normalmente desencadeia-se em cirrose hepática, que por sua vez é considerada uma doença pré-maligna. Modulando a hepatocarcinogênese existem variações genéticas e epigenéticas, tais como mutações, metilação da região promotora e desregulações nos RNAs não codificantes (ncRNAs), como os microRNAs (miRNAs) (GOMES *et al.*, 2013; WONG; TSANG; OI-LIN, 2018). A soma desses fatores pode tornar o paciente ainda mais suscetível ao surgimento da doença.

Nossa compreensão acerca da patogênese molecular do carcinoma hepatocelular melhorou significativamente na última década. Certamente, os dados cumulativos advindos de análises de *high-throughput* realizadas em um montante considerável de amostras forneceram um panorama das alterações genéticas e epigenéticas ocorridas no CHC. Sabe-se que a estratégia terapêutica utilizada no manejo clínico deve ser avaliada de acordo com o estado da doença e da compreensão molecular do CHC dependente de etiologia (RINGERLAN, 2019). Apesar das melhorias recentes nas intervenções terapêuticas, o CHC ainda está associado a um mau prognóstico em pacientes com doença em estágio avançado. Estudos anteriores relataram a influência benéfica da vigilância regular para CHC em populações de alto risco para detectar CHC em um estágio inicial (KUDO *et al.*, 2016).

O montante de dados associados a compreensão clínica, genética e epigenética envolvidas com o CHC devem ser avaliados de maneira curada para determinação da terapêutica (YANG *et al.*, 2019). A complexidade da hepatocarcinogênese ainda permanece um desafio para os cientistas devido a sua heterogeneidade, deste modo, ao idealizar a medicina de precisão, como ferramenta utilizada para elevar o sucesso na evolução clínica faz-se necessário desvendar

possíveis padrões biológicos existentes no surgimento e progressão do CHC dependente de etiologia (SATO, 2019; XU *et al.*, 2019).

O aprendizado de máquina é um campo multidisciplinar que combina ciência da computação e matemática e é focado na implementação de algoritmos de computador capazes de maximizar a precisão preditiva de dados estáticos ou dinâmicos usando modelos analíticos ou probabilísticos. Combinar dados clínicos e/ou genéticos usando esta ferramenta analítica pode permitir o desenvolvimento de um novo modelo para a previsão de CHC (SATO, 2019).

Dentro desta ótica, o presente estudo se propõe a investigar a influência entre diferentes modificações genéticas e de regulação epigenética associadas ao surgimento do CHC, empregando uma estratégia combinatória entre análise de RNA-Seq e biologia computacional afim de investigar novos ncRNA, em especial miRNA, e seus potenciais mecanismos de ação em CHC. Assim como, desenvolver uma estrutura de aprendizado de máquina para estabelecer o modelo mais apropriado, baseado em dados genéticos, e aplicar este modelo aos dados existentes de pacientes com CHC para desenvolver um modelo apropriado para predição de CHC. Culminando na tradução desse conhecimento para o desenvolvimento de novas metas e biomarcadores que possam impactar a tomada de decisão médica durante o manejo clínico do paciente com CHC, e, finalmente, potencializando seus resultados.



## 2 REFERENCIAL TEÓRICO

O câncer de fígado pode ocorrer em dois tipos: primário (com início no próprio órgão) e secundário ou metastático (com origem em outro órgão e, com a evolução da doença, atinge também o fígado). Dentre os tumores iniciados no fígado, o mais comum é o carcinoma hepatocelular, que apresenta característica agressiva. Nesta seção do trabalho é feita uma breve apresentação dos principais assuntos pertinentes à pesquisa

### 2.1 Epidemiologia Do Carcinoma Hepatocelular

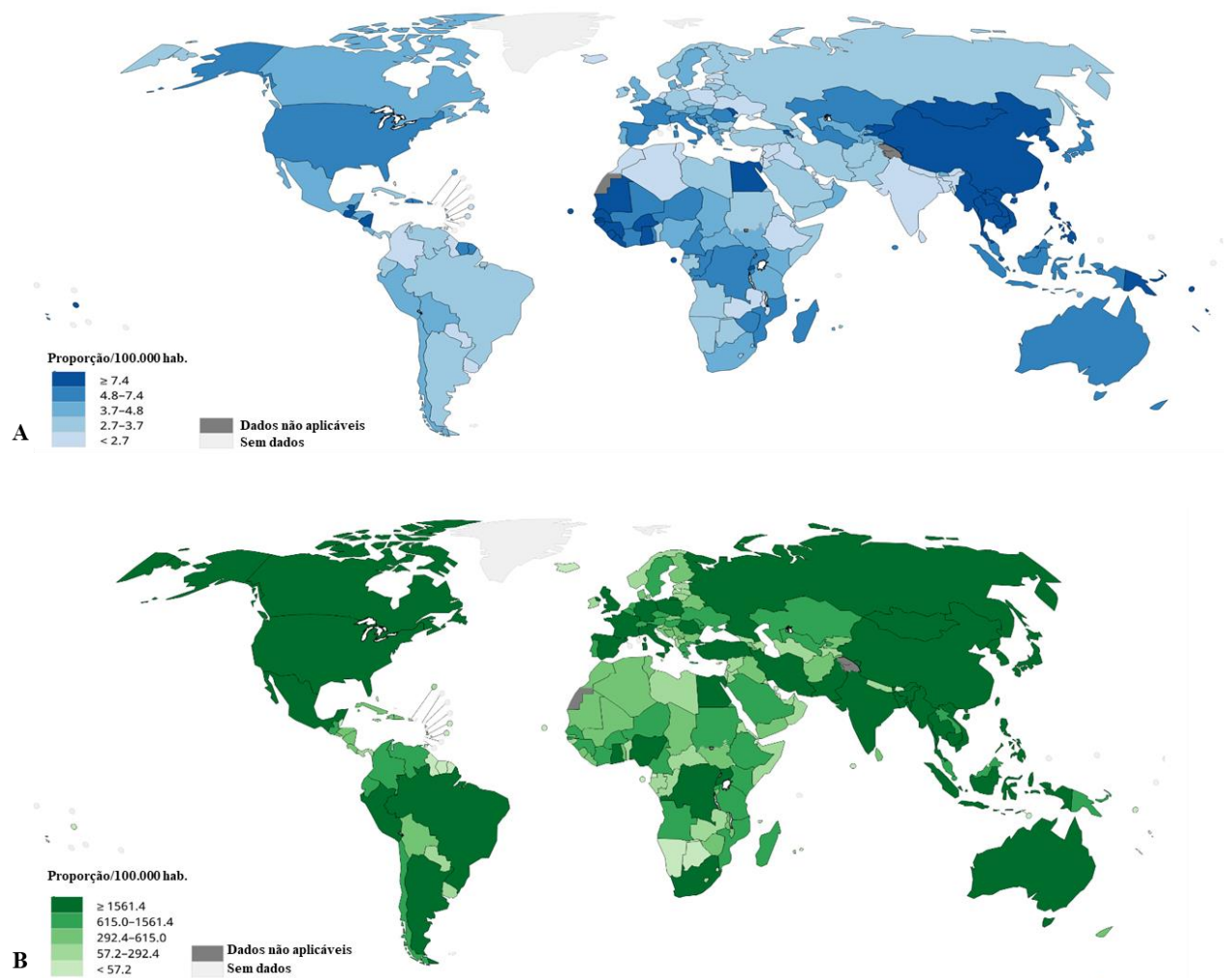
O Carcinoma hepatocelular (CHC) é o principal tipo de câncer primário de fígado, correspondendo de 85% a 90% dos casos, sendo considerado o quinto câncer mais comum no mundo, sendo o quinto tipo de câncer mais comum em homens e o nono em mulheres, e o terceiro quando se trata de mortalidade por neoplasias malignas (LAFARO; DEMIRJIAN; PAWLIK, 2015). Encontramos substanciais variações globais na incidência e mortalidade de CHC que existem devido a diferenças como: o nível de exposição a fatores de risco ambientais e infecciosos, disponibilidade de recursos de saúde e capacidade de detectar CHC em estágio inicial e fornecer tratamento potencialmente curativo (YANG *et al.*, 2019).

Estima-se que quase 85% dos casos de CHC ocorram em países de poucos ou médios recursos, particularmente na Ásia Oriental e na África Subsaariana (GLOBAL BURDEN OF DISEASE CANCER, 2019). O país com maior incidência de câncer hepático no mundo é a China, responsável por mais da metade dos casos de CHC existentes, seguido por países também asiáticos como Mongólia, Japão e Coreia e países da África subsaariana como Moçambique (FERLAY; SHUN, 2013; KULIK; EL-SERAG, 2019).

Na América do Sul, do Norte e na Europa, a incidência de CHC ainda é considerada baixa quando comparada aos países asiáticos e africanos, no entanto o número de novos casos vem aumentando com o passar do tempo (GBDC, 2019). No Brasil, os estados com maiores casos de CHC relatados são Espírito Santo e Bahia, porém estudos relacionados a esse tema ainda são escassos (ALMEIDA-CARVALHO *et al.*, 2017). Nacionalmente a taxa de incidência padronizada por idade é estimada em 4,5 por 100.000 habitantes para o total de casos, de câncer hepático, considerando indivíduos de 0 a 85 anos e ambos os sexos (Figura 1A). O número estimado da taxa

de prevalência é de 5,9 em 100.000 em todo o Brasil (Figura 1B), a taxa de mortalidade estimada nacional está em 4,3/100.000 habitantes, considerando ambos os sexos (WHO, 2020).

**Figura 1** - Taxas de incidência e prevalência de câncer hepático.



Dados obtidos do Global Cancer Observatory ([www.gco.iarc.fr](http://www.gco.iarc.fr)).

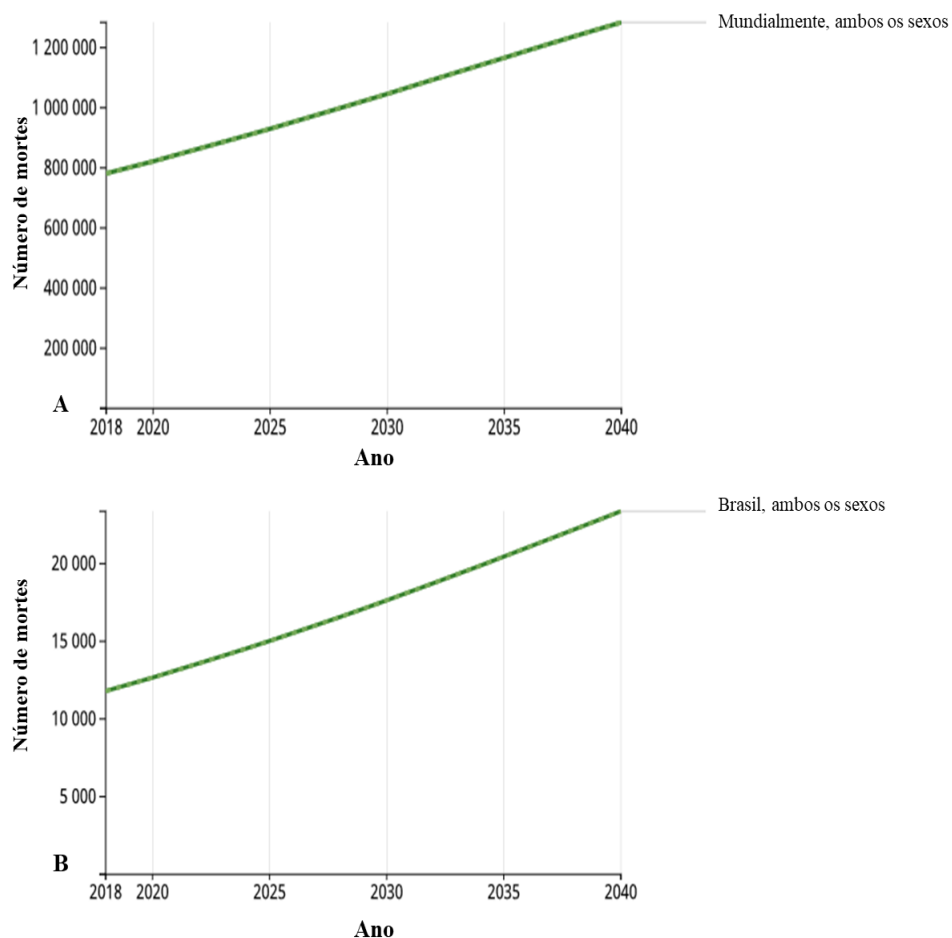
(A) Evidência a taxa de incidência padronizada por idade, considerando indivíduos de 0 a 74 anos e ambos os sexos. Para o Brasil está taxa encontra-se em 3,7/100.000 habitantes. (B) Número estimado de casos prevalentes, de câncer hepático, entre 2014 e 2018 para ambos os sexos e todas as idades.

Dados de inquérito nacional brasileiro indicam predomínio da doença em homens maiores de 50 anos com cirrose e infecção pelo VHC. No entanto, esses achados diferem entre as diferentes regiões do país, principalmente entre Norte/Nordeste e Sul/Sudeste (CARRILHO *et al.*, 2010). Além disso, há uma escassez de dados dessas regiões quanto ao diagnóstico, tipo de população de estudo (em

serviços privados ou públicos) e tratamento (ALMEIDA-CARVALHO *et al.*, 2017). É noticiada a existência de quase duas vezes maior incidência de CHC entre homens afrodescendentes em comparação aos homens caucasianos, uma proporção similar é observada entre indivíduos do gênero feminino (PARANAGUÁ-VEZOZZO *et al.*, 2014).

No Brasil, em 2019, foram registrados um total de 10.901 óbitos por câncer hepático, sendo 6.317 homens e 4.584 mulheres. (INCA, 2019). A mortalidade relacionada ao câncer hepático segue em taxa crescente juntamente com a prevalência de casos, é previsto que o número de mortes continue a aumentar vertiginosamente até 2040 tanto em âmbito nacional, quanto mundial (Figura 2) (WHO, 2020). A taxa de sobrevivência de 5 anos para estes pacientes melhorou em 60% de 1975 a 2005 (NJEI, *et al*, 2015).

**Figura 2** - Número estimado de casos, de câncer hepático, incidentes de 2018 a 2040, para ambos os sexos e todas as idades.



Dados obtidos do Global Cancer Observatory ([www.gco.iarc.fr](http://www.gco.iarc.fr)).

(A) Gráfico evidenciando o número de mortes por ano a nível mundial. O gráfico (B) mostra o mesmo dado, no entanto apenas para o Brasil.

Quando descoberto nos estados iniciais, o CHC tem maior potencial curativo, porém, quando diagnosticado tardiamente, ou quando não é realizada nenhuma intervenção sobre o tumor, o paciente pode evoluir para óbito em aproximadamente 10 meses (KUMAR; ABBAS; ASTER, 2014). Os equipamentos de imagem são fundamentais para o diagnóstico de CHC e raros nos hospitais públicos brasileiros. Além disso, o transplante de fígado está disponível apenas nas grandes cidades brasileiras. Deste modo, mais estudos sobre o CHC são necessários para a elaboração de estratégias diagnósticas e terapêuticas para população brasileira (ALMEIDA-CARVALHO *et al.*, 2017).

O CHC é uma doença heterogênea que geralmente se desenvolve associada a um evento de cirrose hepática, que por sua vez, pode ser o estágio final de várias doenças hepáticas, por exemplo, a infecção pelo vírus da Hepatite B (VHB), com ou sem a presença do vírus da hepatite Delta (VHD), exposição a Aflatoxina B1 (AFB1), as quais são as etiologias mais frequentes na Ásia e na África, ao passo que a infecção pelo vírus da hepatite C (VHC), abuso crônico de álcool e síndrome metabólica estão frequentemente relacionadas ao CHC nos países ocidentais (FORNER; LOVET; BRUIX, 2012; KULIK; EL-SERAG, 2019). Em casos raros, o CHC desenvolve-se no fígado normal, com alguns destes tumores resultando potencialmente de transformação maligna de adenomas hepatocelulares (HCAs) (EL-SERAG, *et. Al.*, 2012).

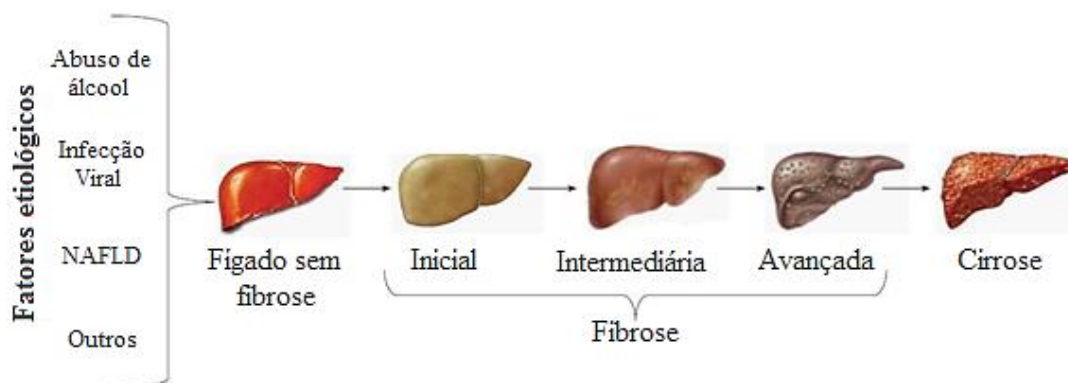
## **2.2 Etiologia Do Carcinoma Hepatocelular (CHC)**

O CHC possui uma amplitude de fatores de risco etiológicos, com alguns que mostram ter uma forte associação com o desenvolvimento de câncer hepático (KULIK; EL-SERAG, 2019). Os principais fatores de riscos incluem infecções por vírus hepatotrópicos, além destes, o consumo excessivo de álcool, a Doença hepática gordurosa não alcoólica (DHGNA), bem como, sua forma mais grave chamada de Esteatohepatite não alcoólica (EHNA), a exposição a Aflatoxinas, tabagismo e Diabetes *Mellitus* podem ser considerados fatores de predisposição ao Câncer hepático (KUDO *et al.*, 2016; RINGERLHAN, 2019). Cerca de 80% a 90% dos casos

de CHC ocorrem no contexto de uma cirrose subjacente e de um microambiente inflamatório (RINGERLAN, 2019; ZHANG; FRIEDMAN, 2012).

Além disso, há um efeito incremental da presença de mais de um fator de risco responsável pelo CHC, pois a presença de coinfeções VHB/VHC e VHB/VHD aumenta o risco de CHC em duas a seis vezes, assim como, da mesma forma, o abuso de álcool aumenta ainda mais esse risco (EL-SERAG *et al.*, 2011). No entanto, o estágio final para a maioria das doenças hepáticas crônicas é a cirrose do fígado, independentemente do fator causal da hepatopatia (Figura 3). O espectro limitado de biomarcadores fiéis que demonstrem a evolução da cirrose hepática até sua manifestação mais grave, o CHC, dificulta um manejo clínico personalizado (LUO *et al.*, 2020; XU *et al.*, 2019).

**Figura 3** - Evolução da fibrose hepática à cirrose por diferentes fatores etiológicos.



Adaptado de Ramos-Lopez *et al.*, 2015.

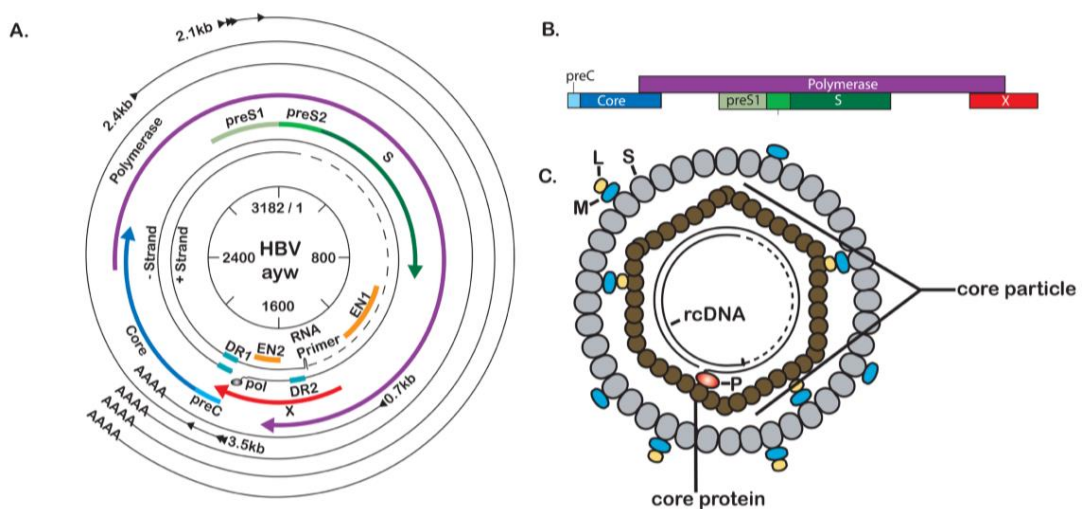
NAFLD – Non Alcoholic Fatty Liver Disease/Doença gordurosa não alcoólica.

A cirrose hepática pode ser definida como a substituição do parênquima hepático por nódulos regenerativos e tecido fibroso, resultante de agressões ao fígado causadas por diferentes etiologias de doenças hepáticas crônicas (REIS *et al.*, 2018). Uma das principais consequências da cirrose é a hipertensão portal, e é a partir desse quadro que surgem as maiores complicações, tais como o surgimento do CHC. Sabe-se que em 80 a 90% dos casos o paciente com câncer de fígado apresenta cirrose hepática, e a maioria dos fatores de risco podem levar a formação e progressão da mesma (FATTOVICH; STROFOLINI; ZAGNI, 2004; GOMES, *et al.*, 2013).

### 2.2.1 Hepatite B

O vírus da Hepatite B (VHB) pertence à família *Hepadnaviridae*, possui um pequeno genoma de DNA circular (3.2 kb), parcialmente em cadeia dupla que codifica quatro quadros de leitura aberta (*Open Reading Frame* - ORFs) sobrepostas (Figura 4) (LAMONTAGNE; BAGGA; BOUCHARD, 2016). As formas de transmissão do vírus podem ser tanto por via parenteral, como o compartilhamento de agulhas, tatuagens, *piercings* e procedimentos cirúrgicos, quanto por via sexual, através de relações desprotegidas e de forma vertical. Atualmente, a vacinação é a forma principal de prevenção e combate a hepatite B, além de reduzir o risco de Carcinoma hepatocelular (GOMES *et al.*, 2013). No Brasil, a vacina para prevenção da Hepatite B faz parte do calendário de vacinação de crianças e é disponibilizada nos postos de saúde pelo governo brasileiro.

**Figura 4** - Organização Molecular do Vírus da Hepatite B (VHB).



Adaptado de Lamontagne, Bagga e Bouchard, 2016.

(A) Representação em escala do genoma do VHB. O círculo interno mostra a posição genômica em relação ao sítio EcoRI na posição 1. O genoma parcialmente de cadeia dupla é representado com o iniciador de RNA ligado e a proteína polimerase. Os quadros de leitura abertos (ORFs) são indicados pelas linhas coloridas mais espessas. Os círculos negros mais externos representam os transcritos virais com o sítio de poliadenilação compartilhado; (B) representação esquemática da natureza sobreposta das ORF do VHB; (C) o *virion* VHB maduro consiste em duas partes principais: um nucleocapsídeo (ou partícula central) consistindo de um genoma de DNA de fita dupla parcialmente ligado à polimerase (P), um capsídeo formado por dímeros de proteína do núcleo, e um envelope viral consistindo principalmente de S-HBsAg (S), com uma quantidade intermediária de M-HBsAg (M) e níveis mais baixos de L-HBsAg (L).

O VHB possui uma grande variabilidade genética, apresentando 8 genótipos virais diferentes, do A ao H, com ampla distribuição geográfica (BRASIL, 2018). Na África Subsaariana, onde predominam casos de CHC relacionados com infecção por hepatite B, o genótipo predominante é o A. O genótipo C é o mais associado a uma lesão hepática grave e com maiores riscos do desenvolvimento de CHC, sendo predominantemente presente em países como China, Coreia e Japão (GOMES *et al.*, 2013). Já no Brasil, a distribuição é considerada variada e podem ser encontrados os genótipos A, D e F (ARAÚJO *et al.*, 2004; BRASIL, 2018; MELLO *et al.*, 2007).

O VHB se diferencia dos outros fatores de risco pelo fato de tornar o paciente suscetível ao CHC com ou sem a presença da cirrose, tal processo pode ocorrer de diferentes formas. O papel da integração do material genético do vírus da hepatite B ao da célula hospedeira no desenvolvimento do CHC ainda não é totalmente compreendido. Além disso, o fenômeno de CHCs induzidos por VHB que ocorrem no contexto de um fígado não cirrótico em adultos jovens infectados cronicamente suporta fortemente um papel específico para integração do VHB na carcinogênese hepática (PROTZER; MAINI; KNOLLE, 2012). Quando ocorre a integração do genoma viral, isto produz uma gama de alterações genéticas no genoma do hospedeiro, criando uma instabilidade genômica generalizada. Tais alterações presumivelmente levam à seleção de clones de células hepáticas que têm uma vantagem de crescimento e expansão. Dentro deste contexto, clones neoplásicos invasivos podem então surgir, os quais têm um conjunto de alterações que resultam em proliferação celular desregulada e apoptose suprimida (BONILLA; ROBERTS, 2005; FERBER; MONTROYA; YU, 2003).

### **2.2.2 Hepatite C**

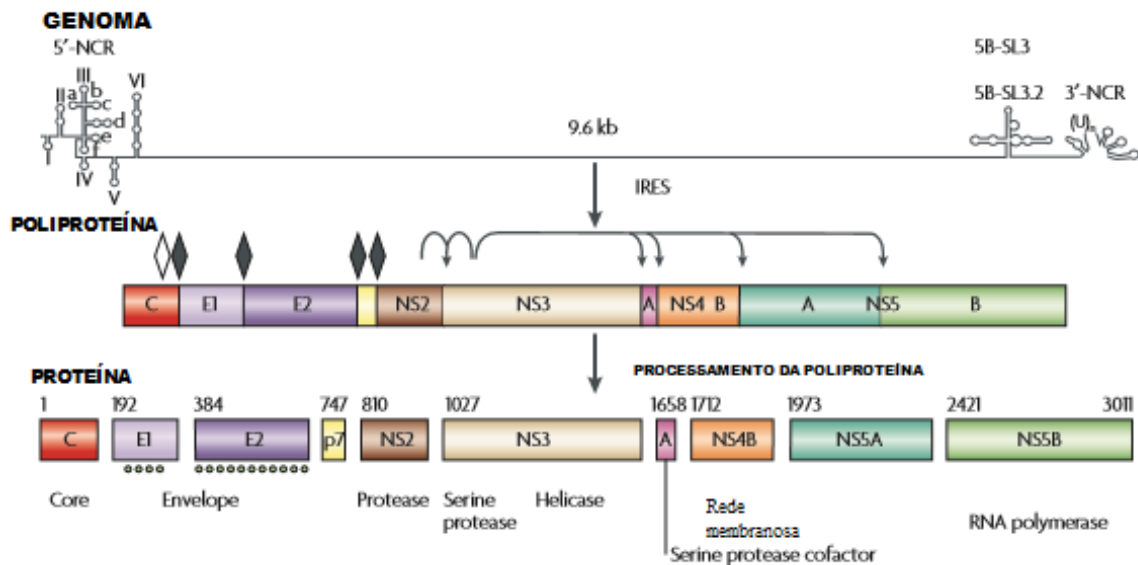
Outra hepatopatia viral relevante é causada pelo vírus da hepatite C (VHC), o qual é um vírus de RNA pertencente ao gênero *Hepacivirus*, da família *Flaviviridae*, com genoma de RNA em fita simples de polaridade positiva, medindo 9,6 kb de comprimento (Figura 5). Contém um único quadro de codificação de leitura aberta (ORF), para uma poliproteína, o qual é flanqueado por regiões 5' e 3' não-traduzidas (UTR). As formas de transmissão do vírus podem ocorrer por via parenteral, pelo compartilhamento de seringas como no uso de drogas injetáveis e por vias sexual e vertical, apesar destas serem consideradas pouco frequentes. A hepatite C acomete

hoje mais de 200 milhões de pessoas mundialmente e a maioria dos pacientes, em torno de 85%, evoluem para a forma crônica, atualmente é considerada uma das principais causas de cirrose e de transplantes de fígado no ocidente (BRASIL, 2018).

Existem sete genótipos do vírus, que vão do 1 ao 7, capazes de se dividir em subtipos, conferindo ao VHC uma grande variabilidade genética, assim como o VHB. O genótipo 1 é o mais prevalente na América do Norte e do Sul e na Ásia, enquanto o 4 domina o Egito e África central (MANNNS *et al.*, 2017). A eficácia no tratamento da hepatite C varia de acordo com o genótipo, o que dificultou por muito tempo uma boa resposta virológica sustentada (SVR), porém, com o avanço de novos medicamentos, que evoluiu desde o tratamento com Interferon até os agentes antivirais de ação direta (DAAs), hoje já são obtidos bons resultados. Assim como acontece com o VHB, o tratamento contra a hepatite C também reduz as chances de o paciente desenvolver CHC (MORGAN; BAACK; SMITH, 2013).

Apesar do vigente Plano de Erradicação da Hepatite C criado pelo Governo Federal Brasileiro, que visa eliminar o surgimento de novos casos da doença até 2030 e da distribuição dos DAAs para o tratamento via Sistema Único de Saúde, sabe-se que o surgimento do CHC é gradual e silencioso, de modo que, os indivíduos que atualmente apresentam fibrose avançada e cirrose apresentam grande risco do aparecimento do câncer hepático, tornando-se assim uma população de risco (BRASIL, 2018).



**Figura 5** - Organização genética e processamento de poliproteína do vírus da hepatite C (VHC).

Adaptado de Moradpour & Penin, 2007.

O genoma de RNA de cadeia positiva de 9,6 kb está representado esquematicamente no topo. São mostradas estruturas secundárias de RNA simplificadas nas regiões não codificadoras 5' e 3' (NCR) e o gene nuclear. A tradução mediada pelo sítio de entrada ribossomal interno (IRES) origina um precursor de poliproteína que é processado nas proteínas estruturais e não estruturais maduras. Os números de aminoácidos são apresentados acima de cada proteína. Os losangos pretos denotam locais de clivagem do precursor de poliproteína de VHC pela peptidase de sinal do retículo endoplasmático. O losango branco indica um processamento C-terminal adicional da proteína do núcleo pelo peptídeo-sinal. As setas indicam clivagens pelas proteases NS2-3 e NS3-4A do VHC. Os pontos em E1 e E2 indicam a glicosilação das proteínas de envelope. Nota-se que o processamento de poliproteínas, ilustrado aqui como um passo separado para simplicidade, ocorre co- e pós-traducional.

O grau de agressividade em relação ao CHC é maior em pacientes infectados por VHC do que por VHB, isso se deve ao fato da maioria dos pacientes com hepatite C crônica desenvolverem cirrose (GOMES *et al.*, 2013). Os mecanismos pelos quais o VHC promove a hepatocarcinogênese ainda precisam ser elucidados. Se o VHC provoca câncer de fígado indiretamente, através de inflamação crônica, fibrose e regeneração hepática contínua, ou diretamente, através da expressão de proteínas virais promotoras de tumor, de maneira análoga a outros vírus oncogênicos, como papilomavírus humano e Vírus Epstein–Barr, permanece desconhecido (HAROUAKA; ENGLE; WOLLENBERG, 2016). A hipótese mais aceita é que as proteínas virais do VHC contribuem para o desenvolvimento do câncer hepático. Tais proteínas atuam no evento de apoptose, transdução de sinais, proliferação celular e resposta imunológica, esse último através da regulação positiva de proteínas como interleucina

1, interleucina 2 e o fator de necrose tumoral, que são moléculas pró-inflamatórias que ajudam na formação câncer (LAFARO; DEMIRJIAN; PAWLIK, 2015; GOMES, *et al.*, 2013; ARAÚJO, 2004).

### **2.2.3 Ingestão de Álcool**

O álcool é metabolizado e eliminado primariamente por ação do fígado por meio de uma série de reações metabólicas oxidativas (THIELE; KLASSEN; TUMA, 2008; TUMA; CASEY, 2013). As três principais etapas são: (1) oxidação reversível do etanol em acetaldeído, o qual é tóxico; (2) metabolismo irreversível do acetaldeído tóxico ao acetato; e (3) decomposição do acetato em água e dióxido de carbono. O metabolismo oxidativo do álcool é efetuado por enzimas chave, incluindo a álcool desidrogenase (ALD), o Citocromo P450 2E1 (CYP2E1) e a catalase. A ALD é a principal enzima oxidante; tem alta afinidade pelo álcool e quebra o etanol no citoplasma. A CYP2E1 é utilizada por uma via distinta que é induzida pelo consumo crônico de álcool e resulta na formação de acetaldeído nos peroxissomas (FREEMAN; TUMA; THIELE, *et. Al.*, 2005).

Modelos experimentais em animais forneceram evidências convincentes de que o álcool tem potencial mutagênico e que o acetaldeído deve ser considerado carcinogênico. O papel do álcool/acetaldeído como carcinogênico é particularmente relevante para a patogênese (SETSHEDI; WANDS; LA MONDE, 2005). O processo de formação do CHC associado a bebidas alcoólicas podem ocorrer de duas formas diferentes: primeiro, o metabolismo oxidativo do álcool gera excesso da forma reduzida da Nicotinamida-Adenina-Dinucleotídeo (NADH), desequilibrando a proporção de NADH para NAD<sup>+</sup> nas células hepáticas, promovendo assim a inibição da oxidação de ácidos graxos e o aumento de gordura. A segunda forma é a geração de ROS (espécies reativas de oxigênio) nos hepatócitos, aumentando também os níveis de NADH no fígado (SETSHEDI; WANDS; LA MONDE, 2005).

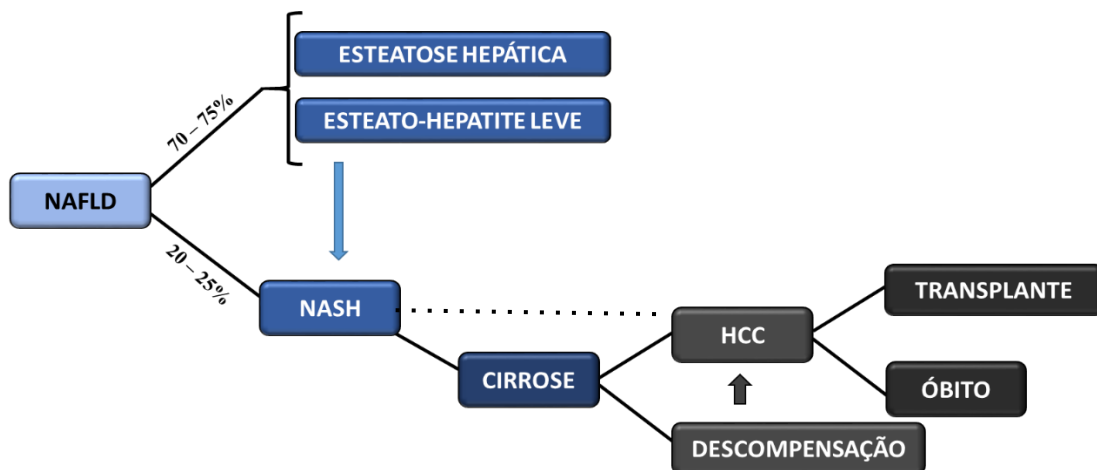
Assim como o VHC e os demais fatores de risco, o consumo excessivo de álcool pode resultar em cirrose hepática. Estudos mostraram que consumir mais de 80 g/dia de bebida alcoólica num período de 10 anos aumenta em 5 vezes a probabilidade de desenvolver CHC (ZHU; JIA; MISRA, 2012; DONATO; TAGGER; GELATTI, 2002). Além disso, cessar a ingestão de álcool reduz de 6 a 7% por ano o risco de carcinoma hepatocelular, e que é necessário um período de 23 anos após parar de beber para

que o risco se iguale aos das pessoas que não consomem (GREWAL; VISWANATHEN; 2012).

#### 2.2.4 Doença hepática gordurosa não alcoólica e Esteatohepatite não alcoólica

A Esteato-hepatite não alcoólica (EHNA) é a forma mais grave da doença hepática gordurosa não alcoólica (DHGNA), ambas são caracterizadas pela presença de dano significativo ao parênquima hepático somado a ausência significativa de consumo de álcool e outras doenças hepáticas, tais como as hepatites virais (GOMES *et al.*, 2013). A DHGNA é resultante de uma síndrome metabólica, nela, estão inclusas hipertensão, resistência à insulina, obesidade e dislipidemia, que é o excesso de colesterol e triglicérides no sangue. Sabe-se que cerca de 20-25% dos indivíduos com DHGNA evoluem para a EHNA, além disso, indivíduos com esteatose hepática ou esteatohepatite leve também apresentam potencial de evolução para EHNA (Figura 6) (YOUNOSSI *et al.*, 2012).

**Figura 6** - História Natural da DHGNA/EHNA associada ao CHC.



Fonte: o autor.

A progressão do processo de hepatocarcinogênese associada a Doença Hepática Gordurosa Não Alcoólica (DHGNA ou NAFLD). A linha tracejada indica a possível evolução da Esteatohepatite Não Alcoólica (EHNA ou NASH) diretamente para o carcinoma hepatocelular, não passando pela cirrose.

A nova epidemia da doença hepática crônica está associada à DHGNA, paralelamente com o aumento da obesidade ao redor do mundo. A prevalência global de DHGNA atualmente é estimada em 24% variando de acordo com região e etnia,

sendo altamente prevalente em todos os continentes. Entretanto, as taxas mais altas relatadas são da América do Sul (31%) e do Oriente Médio (32%), seguido pela Ásia (27%), Estados Unidos (24%) e Europa (23%), seguida da África (14%) (YOUNOSSI *et al.*, 2018). No Brasil, estudos epidemiológicos têm revelado que a DHGNA é um problema de saúde pública. Parise e Cotrim (2016), utilizando a ultrassonografia (US) como método diagnóstico, encontraram em torno de 20% de esteatose hepática na população geral brasileira, sendo 2-3% EHNA. Estudos mostram que o número de transplante hepático por EHNA tem aumentado consideravelmente nos últimos anos, ficando em equivalência com a Hepatite C, na maioria dos países, e já ultrapassando em países ocidentais (GADDIKERI; McNEELEY; WANG, 2014; PARISE E COTRIM 2016).

Diversos estudos têm relacionado à EHNA com o surgimento do CHC e a cirrose, o processo de hepatocarcinogênese classicamente perpassa por meio da cirrose, através de alterações inflamatórias e da formação de novos vasos que se dá devido ao desenvolvimento da doença, no entanto estudos indicam a evolução direta de NASH para CHC sem presença de cirrose (ANSTEE *et al.*, 2019; NEGRO, 2020). Essas alterações inflamatórias estão ligadas ao fator de necrose tumoral (TNF- $\alpha$ ) e a interleucina 6 (IL-6) que são promotoras essenciais do câncer (GOMES, *et. Al.*, 2013; SETSHEDI; WANDS; LA MONDE, 2010).

### **2.3 Patogênese Do Carcinoma Hepatocelular**

O câncer primário de fígado compreende um grupo heterogêneo de tumores malignos, que não incluem metástases para o fígado de outros locais. A carcinogênese hepática é um processo multifatorial, e os fatores predisponentes para os vários subtipos de câncer hepático diferem a depender da etiologia do tumor (LUO *et al.*, 2020; RINGELHAN *et al.*, 2018). No macroambiente tumoral, os hepatócitos sofrem transformação maligna através de mecanismos que impedem a destruição do tumor, evitam a apoptose e promovem proliferação e neovascularização tumoral. A cirrose induz alterações carcinogênicas e é encontrada em 90% dos pacientes com diagnóstico de CHC. Nos restantes 10% dos pacientes, mecanismos não-cirróticos da carcinogênese são responsáveis por doenças malignas (PELLICORO *et al.*, 2014; ZHANG; FRIEDMAN, 2012).

Estudos demonstram que o acompanhamento de pacientes com fatores de risco para o desenvolvimento do CHC reduz a mortalidade de indivíduos com cirrose, portanto, o rastreamento é fundamental para evitar o surgimento do CHC e geralmente é realizada através de ultrassonografia (ALMEIDA-CARVALHO *et al.*, 2017; GADDIKERI; McNEELEY; WANG, 2014), reforçando que marcadores biológicos poderiam intervir e auxiliar no rastreamento e análise prognóstica. Para o diagnóstico do carcinoma hepatocelular, são realizadas tomografia computadorizada (TC) e ressonância magnética (RM) ambas com contraste endovenoso (CLARK *et al.*, 2015). Pode ser, ainda, realizada uma biópsia hepática, porém a mesma deve ser evitada devido a complicações que possam surgir, como o extravasamento de sangue para a cavidade peritoneal (CLARK; MCNEELEY; MAKI, 2014).

O CHC provém das células do parênquima hepático, conhecidas como hepatócitos, e, na maioria dos casos, é derivado de uma doença hepática crônica, que são consideradas fatores de risco para o surgimento desse tipo de câncer, normalmente passando por um evento de cirrose hepática (PROTZER; MAINI; KNOLLE, 2012). Esses fatores de risco podem desencadear mutações gênicas, levando a multiplicação desordenada que resulta no processo de hepatocarcinogênese (GOMES *et al.* 2013).

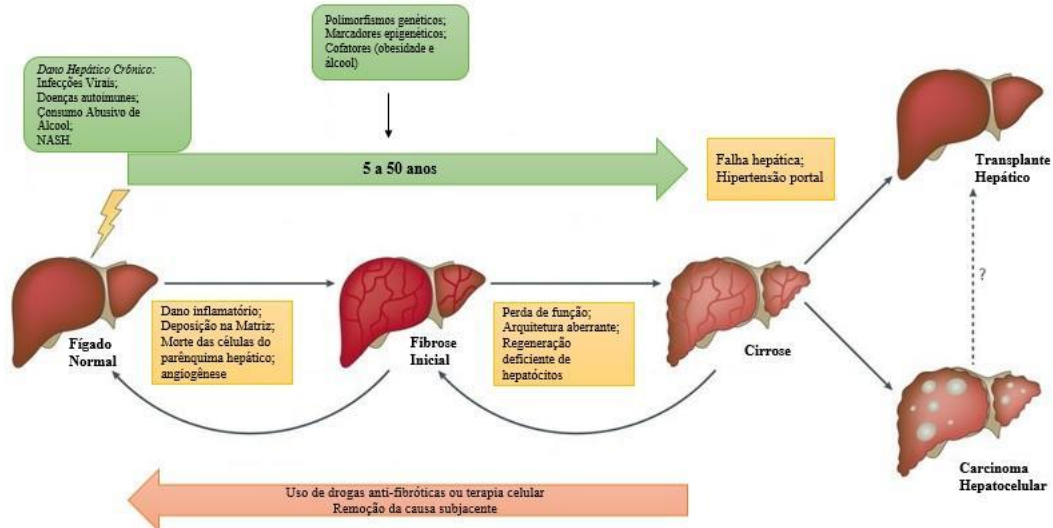
O modelo atualmente aceito de hepatocarcinogênese é um processo de várias etapas desde o início do tumor até a malignidade estabelecida. A evidência para a progressão do CHC é que os hepatócitos normais são transformados em lesões pré-neoplásicas, que ocorrem sob a forma de focos e nódulos displásicos (IOANNOU, 2021; FATTOVICH; STROFOLINI; ZAGNI, 2004). Com a inflamação crônica em curso, essas lesões precoces progridem para displasia de baixo e alto grau, ambas com o potencial de progredir para o CHC (FATTOVICH; STROFOLINI; ZAGNI, 2004; HERNANDEZ-GEA; TOFANIN; FRIEDMAN, 2013).

O mecanismo subjacente da progressão sequencial ainda não é completamente entendido, mas é pensado ser devido a desdiferenciação progressiva do hepatócito devido à diminuição da expressão de genes tecido específico e à alteração de inúmeras vias de sinalização, levando a proliferação celular desregulada e resistência apoptose (LOVET, *et al.*, 2016). Pacientes com CHC que apresentam padrões expressão gênica de marcadores de proliferação celular e genes antiapoptóticos significativamente maiores, correspondem aos pacientes com pior

prognóstico, reforçando à sua significância na patogênese do CHC. Ainda não existem biomarcadores eficazes para o diagnóstico precoce e prognóstico do CHC. Atualmente, apenas cerca de 30% a 40% dos pacientes com tumores hepáticos podem receber tratamento efetivo no momento certo. É extremamente necessário descobrir novos biomarcadores para diagnóstico preciso, prognóstico e eficiência do tratamento do CHC (LOU *et al.*, 2020).

O fígado possui uma singular e considerável capacidade de reparação após um evento de dano agudo. Hepatócitos já diferenciados são capazes de iniciar novamente o ciclo celular e servir como sua principal fonte de reposição (RINGELHAN *et al.*, 2018; SAI; VILLANUEVA; FRIEDMAN *et al.* 2017). No entanto, quando ocorre uma situação de necroinflamação crônica, a morte celular constante, a regeneração compensatória e a ativação de células não-parenquimatosas, juntamente com uma resposta imune alterada, promovem a fibrose hepática e a tumorigênese (Figura 7). Assim como, alteração em sinais bioquímicos de sobrevivência e proliferação durante a necroinflamação, estresse celular, modificações epigenéticas, alterações mitocondriais e senescência promovem a tumorigênese, sendo assim, a necroinflamação também induz estresse de replicação celular associado à proliferação, dano ao DNA e instabilidade genética, que é detectável antes que ocorram alterações neoplásicas (SAI; VILLANUEVA; FRIEDMAN *et al.*, 2017).

**Figura 7** - Representação esquemática da progressão do processo da fibrose e promoção da tumorigênese hepática.



Adaptado de Pellicoro, et. al., 2014.

A fibrose hepática é a resposta de cicatrização do fígado a muitas causas de lesão crônica, das quais a infecção viral, o álcool e a esteatohepatite não alcoólica (NASH) são os mais comuns. Independentemente da causa subjacente, a lesão iterativa causa dano inflamatório, deposição de matriz, morte celular parenquimatosa e angiogênese levando à fibrose progressiva. A matriz cicatricial normalmente se acumula muito lentamente, mas uma vez que a cirrose é estabelecida, o potencial para reverter esse processo é diminuído e as complicações se desenvolvem. Polimorfismos genéticos, marcas epigenéticas e co-fatores (como obesidade e álcool) podem modular o risco de progressão da fibrose. Se a causa da fibrose for eliminada, pode ocorrer a resolução de fibrose hepática precoce. Na cirrose, embora a resolução não seja possível, a regressão (isto é, a melhora, mas não a reversão) da fibrose melhora os resultados clínicos. Estão surgindo terapias anti-fibróticas que podem retardar, parar ou reverter a progressão da fibrose. Atualmente, o transplante de fígado é o único tratamento disponível para insuficiência hepática ou em alguns casos de câncer primário do fígado.

Os fatores genéticos e epigenéticos formam a base molecular do surgimento do CHC, sabe-se que a patogênese das etiologias do CHC, é atualmente considerada umas das principais áreas de estudo dentro da Hepatologia, deste modo tal conhecimento fornece uma compreensão adicional do *background* mutacional e do panorama molecular do CHC. Além disso, técnicas como o sequenciamento de nova geração aplicadas ao estudo de tumores sólidos, tais como, o CHC traz novas informações sobre a influência genética no surgimento do tumor.

A ocorrência de CHC resulta de uma interação complexa entre fatores genéticos e não genéticos do hospedeiro, exposição a agentes carcinogênicos ambientais e vírus potencialmente oncogênicos, e desenvolvimento de doença

hepática crônica subjacente, que, em sua fase final é chamada de cirrose, torna-se um certo campo procarcinogênico. Embora a cirrose seja a base onde a maioria dos CHCs se desenvolvem, o surgimento em fígado não cirrótico indica-nos reavaliar os fatores de risco e mecanismos que levam ao desenvolvimento de CHC sem danos hepáticos (BONILLA; ROBERTS, 2005; PROTZER; MAINI; KNOLLE, 2012)

O requisito para a carcinogênese é um meio permissivo onde genes e vias de sinalização que regulam o destino de todas as células, ou seja, diferenciação, proliferação e morte, são alterados. Nesse contexto, mutações de oncogenes e/ou genes supressores tumorais no CHC tornam-se determinantes relevantes como objetos de pesquisa (XU *et al.*, 2019).

### **2.3.1 Aspectos genéticos do carcinoma hepatocelular**

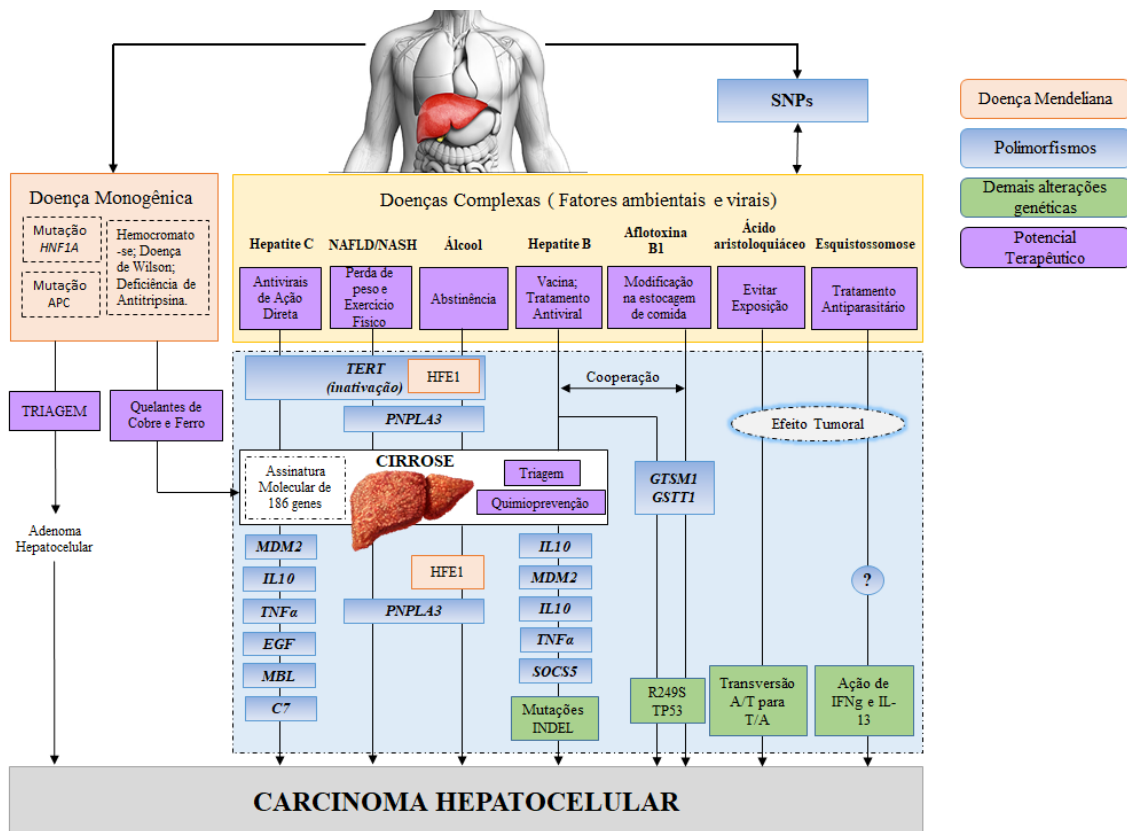
O fígado é um alvo muito raro de síndromes clássicas de predisposição ao câncer. As síndromes monogênicas mais frequentes que predispoem ao câncer de mama, ovário ou colorretal geralmente não estão associadas ao desenvolvimento de tumores hepáticos (TOMASO, 2010). Excepcionalmente, o CHC pode se desenvolver em pacientes com mutações na linhagem germinativa dos genes *HNF1A* e *APC* (LI; GERBER; KORUDA, 2012; WILLSON; GODWIN; WIGGINS, 2013). No entanto, predisposições para o CHC são observadas em várias doenças metabólicas genéticas, principalmente através do desenvolvimento de cirrose. Estas doenças monogênicas raras incluem sobrecarga de ferro (hemocromatose, gene *HFE1*) ou cobre (doença de Wilson, gene *ATP7B*), tirosinemia tipo 1 (gene *FAH*), porfiria aguda intermitente (gene *HMBS*) ou cutânea tarda (gene *UROD*) e deficiência de antitripsina (Gene *SERPINA1*) (ZUCMAN-ROSSI *et al.*, 2015).

Ao se levar em consideração a análise de predisposição ao CHC por um ângulo poligênico, sabe-se que vários polimorfismos de nucleotídeo único (SNPs) foram identificados como associados ao risco de CHC. Entre eles, muitos polimorfismos alteram as vias biológicas de carcinogênese, incluindo inflamação via *TNFA*, *IL1B* ou *TGFB* (TARHUNI; RUFAT; SUTTON, 2014; WANG; KATO; HOSHIDA, 2014); estresse oxidativo através de *SOD2*, *LGALS3* e *MPO* (DO CARMO *et al.*, 2014, SOUSA *et al.*, 2016); metabolismo do ferro com uma cooperação entre a ingestão de álcool e a variante C282Y de *HFE1* (LAURET; RODRÍGUEZ; GONZÁLEZ, *et al.* 2002; NAHON; SUTTON; RUFAT, 2008). Reconhecimento e atuação do sistema imune



inato por meio de *PTX3*, *MBL* e *C7*(CARMO, 2016; DE LIMA, 2018; HALLA, 2010). Reparo de DNA (*MTHFR* ou *XRCC3*) (LONG; MA; QU; *et al*, 2008); ciclo celular com um papel importante de *MDM2* (DHAREL, KATO; MUROYAMA, 2006) e *TP53*(YOON; CHANG; AHN, 2008); ou fator de crescimento com *EGF* (DAYYEH; YANG; FUCHS, 2011).

Polimorfismos possuem a capacidade de modular o risco para o CHC em diferentes etapas da doença, incluindo predisposição a fatores de risco, à gravidade da doença hepática crônica e sua evolução para cirrose, ou para a transformação maligna e progressão tumoral (Figura 8) (NAHON; SUTTON; RUFAT, 2008). Eles podem ser usados para estratificar os pacientes em políticas de vigilância personalizadas e atuar como alvos candidatos para estratégias de quimioprevenção (YANG *et al.*, 2019). A influência exercida pela maioria dos polimorfismos associados ao desenvolvimento do CHC na doença hepática crônica está relacionada a fatores de risco específicos, ou seja, etiologia tumoral dependente. Esta observação destaca as relações próximas entre a natureza da exposição, o *background* genético e o mecanismo de transformação/proliferação dos hepatócitos. Ambos, os fatores genéticos e epigenéticos formam a base molecular do surgimento do CHC (LIU *et al.*, 2020).

**Figura 8** - Interação entre a predisposição genética, fatores ambientais e ocorrência do CHC.

Adaptado de Zucman-Rossi *et al*, 2015.

A imagem mostra o padrão de herança mendeliana (Laranja), polimorfismos envolvidos na ocorrência do CHC (azul), outras alterações genéticas, tais como mutações e variações de expressão (verde) e terapias potenciais utilizadas para reduzir a incidência do CHC (Roxo).

O CHC provoca impactos devastadores na sociedade, causando milhares de mortes anualmente. A Organização Mundial da Saúde (OMS) em 2016 colocou os erros de diagnóstico como um problema de alta prioridade. Para auxiliar profissionais médicos no diagnóstico do câncer de fígado, sistemas de suporte à decisão (SSD) podem ser empregados para aumentar as chances de um diagnóstico correto, já que um diagnóstico da doença no seu estágio inicial é crucial para a sobrevivência do paciente (SATO, 2019)

Nos dias atuais, modelos de aprendizagem de máquina se tornaram ferramentas centrais em estudos que envolvem a biologia computacional e a medicina informática (MAGOULAS; PRENTZA, 2001). Com a popularização desses modelos é possível identificar padrões em complexas bases de dados que envolvem características genômicas, por exemplo, Cruz e Wishart (2006) construíram um survey contendo um apanhado de técnicas e modelos amplamente utilizados em estudos envolvendo o diagnóstico e a prevenção de câncer. No estudo de Davi e

colaboradores (2019) avaliou-se o uso de modelos como Support Vector Machine (SVM) e Artificial Neural Networks (ANN) na predição de casos de dengue severa. Além disso, Chen e Ishwaran (2012) estudaram o modelo de Random Forests (RF) nas construções de aplicações para a predição, classificação e seleção de características a partir de base de dados contendo o genoma humano.

No campo da bioinformática o uso de modelos de aprendizagem de máquina tem sido amplamente utilizado no que diz respeito ao reconhecimento de padrões de doenças genéticas. Estes modelos promovem discernimentos acerca das relações entre genótipo e fenótipo que até então eram difíceis de serem concebidas apenas pelo uso da biologia e estatística. Tais modelos preditivos estão sendo cada vez mais aplicados nos estudos das implicações de variantes genéticas e ambientais, ou outros fatores não genéticos, possuem no surgimento de certas anomalias biológicas (OTKLEY, 2014).

Algoritmos genéticos são uma família de modelos computacionais inspirados na teoria da evolução biológica (WHITLEY, 1994). Uma das aplicações de algoritmos genéticos é a seleção de características, para problemas de classificação. A classificação de padrões e problemas de descoberta de conhecimento requerem um subconjunto de características para representar os padrões a serem classificados. Algoritmos genéticos oferecem uma opção atrativa para encontrar uma solução quase ótima para tais problemas de otimização (YANG; HONAVAR, 1998).

Estudos recentes utilizaram técnicas de aprendizagem de máquina aplicadas ao problema do câncer de fígado. Sato em 2018 aplicou os classificadores: SVM, Logistic Regression, Support Vector Machine (SVM), Gradient Boosting (GB) Random Forest (RF), Artificial Neural Network e Deep Learning à classificação entre indivíduos quanto à presença CHC. Ksia e colaboradores em 2019 utilizou um 2level genetic optimizer para otimizar os parâmetros dos classificadores e realizar seleção de características com as entradas. Alguns dos classificadores usados foram: SVM, MLP, K-nearest neighbors (KNN), classificador de Naive Bayes, Logistic Regression e Random Forest.

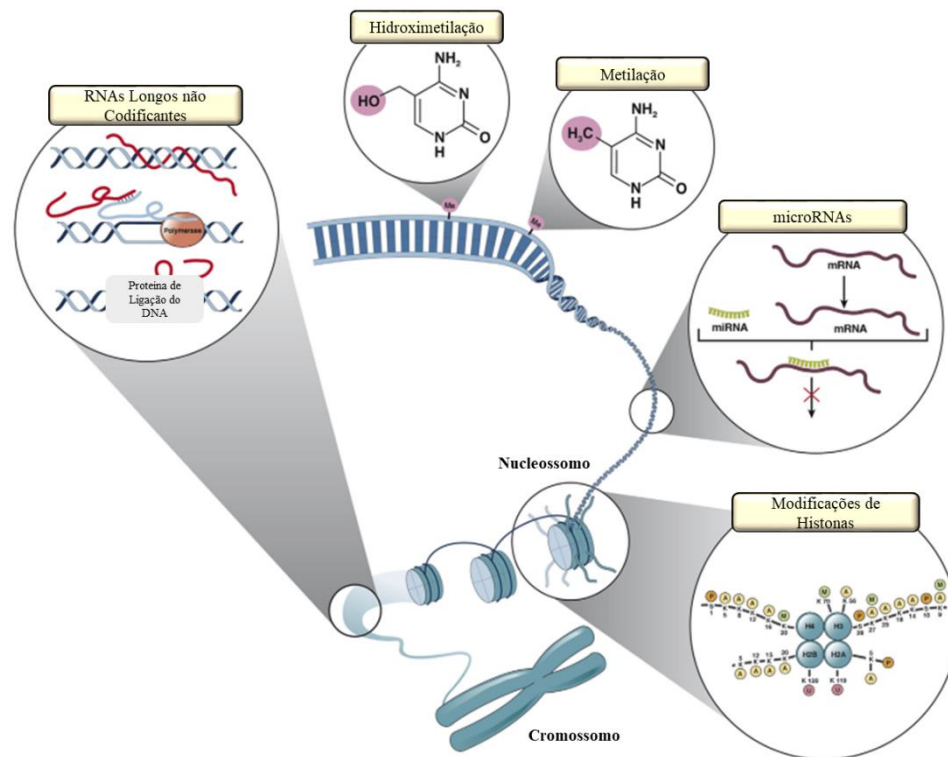
Todavia, aplicações genômicas que envolvem modelos preditivos estão, na maioria das vezes, associadas à bases de dados de alta dimensão. Um conjunto de dados com essa característica apresenta uma série de desafios no processo de aprendizagem, haja vista o grande espaço de busca, dificultando a extração de

padrões altamente relevantes. Neste caso, a seleção das características mais expressivas não só elimina informações redundantes e irrelevantes, como também auxilia os algoritmos de classificação a se tornarem mais genéricos no domínio de interesse (PES, 2019). Contudo, os métodos de seleção de características podem ser classificados de acordo com a sua abordagem. Existem aqueles em que os seus critérios de seleção se baseiam na correlação que cada característica possui com a variável dependente (esta correlação é oriunda de cálculos estatísticos ou entrópicos). E também, existem aqueles que realizam buscas por subgrupos do conjunto de dados à procura dos que promovem a melhor performance num dado classificador (DENG *et al.*, 2018).

### **2.3.2 Aspectos epigenéticos do carcinoma hepatocelular**

O termo "epigenética" refere-se a todas as alterações estáveis de características fenotípicas que não são codificadas diretamente por conta da sequência de DNA (MORAN-SALVADOR; MANN, 2017). Mecanismos epigenéticos podem ser vistos como uma interface entre o genoma, o fator de risco, estilo de vida e influência ambiental. Diversos padrões de modificação epigenética podem afetar os tipos e frequências de alterações genéticas nas regiões de cromatina adjacentes (LIU; JIN; ZHOU, 2015). A análise integrativa do sequenciamento do genoma humano somado ao sequenciamento do epigenoma identificaram como o panorama epigenético influencia no acúmulo de alterações genéticas durante a hepatocarcinogênese (BERGER, 2007).

O epigenoma é regido por, ao menos 3 tipos principais de sistemas de regulação, são eles: metilação do DNA, modificações de histonas e silenciamento gênico mediado por RNAs não codificantes (ncRNA) (Figura 9) (MORAN-SALVADOR; MANN, 2017). Esses mecanismos epigenéticos separados, porém interativos e que se sobrepõem, são atualmente considerados fundamentais para iniciar e sustentar modificações no DNA e na cromatina que conduzem ao fenótipo celular, modulando o controle da transcrição gênica por intermédio de fatores de transcrição específicos.

**Figura 9** - Mecanismos epigenéticos de regulação hereditária da expressão gênica.

Adaptado de Moran-Salvador e Mann, 2017.

Existem vários mecanismos epigenéticos altamente interdependentes que são importantes no controle da expressão gênica, por exemplo, metilação do DNA (e hidroximetilação), modificações pós-traducionais da histona e vias baseadas em RNAs não codificantes (ncRNA), incluindo os tipos pequenos e longos de ncRNA.

Estudos prévios e emergentes começam a incorporar ncRNAs ao nosso entendimento molecular acerca da patogênese e progressão do CHC. Neste aspecto, os miRNAs ganharam destaque, estes são moléculas que possuem ~22 nucleotídeos de comprimento, sendo as menores moléculas de RNAs eucarióticas funcionais conhecidas (LIU; JIN; ZHOU, 2015). Foram descobertos em *Caenorhabditis elegans*, e são capazes de reprimir muitos genes. O miRNA liga-se por complementariedade à região 3' não traduzida (3'-UTR) do RNA mensageiro alvo (mRNA) através da sua *seed sequence*, que corresponde entre as posições 2 a 8 da molécula de miRNA, para mediar a readenilação do mRNA ou o bloqueio total do processo de tradução (BERGER, 2007).

### 2.3.3 *microRNA's e Hepatocarcinoma Celular*

Tais moléculas são conservadas entre as espécies, existem aproximadamente 500 genes de miRNA, e mais de 3000 destas moléculas, foram identificadas e

encontrados para serem componentes importantes do complexo funcional de vias que controlam cruciais processos celulares, como proliferação, diferenciação e apoptose (RUPAIMOOLE; SLACK, 2017) . Cada tipo de miRNA é estável e pode diminuir a regulação de centenas de genes de cada vez, eles podem controlar grandes eventos de transcrição que determinam os recursos celulares fundamentais (VERDUCE *et al.*, 2019). Tal diversidade em funções permite que miRNAs possam ser usados como ferramentas de diagnóstico para a detecção precoce do câncer, avaliação de risco e prognóstico, e como novos alvos terapêuticos (ZHOU *et al.*, 2020).

Os miRNAs associados ao desenvolvimento de CHC têm sido investigados como biomarcadores para fins de diagnóstico. Alguns desses miRNAs demonstraram prever com precisão o mau prognóstico no CHC (NI, *et al.*, 2018). Por exemplo, estudos indicaram que miR-200a e miR-200b, estão desregulados durante o desenvolvimento de CHC e fibrose hepática (ZHANG *et al.*, 2018). Os níveis aumentados de miR-21 sérico têm sido usados para distinguir casos de CHC de hepatite crônica e controles saudáveis de maneira mais eficiente quando é feita a mesma análise utilizando a alfafetoproteína como um biomarcador em CHC (CAO *et al.*, 2018). Os miR-15b e miR-130b, encontrados no soro dos pacientes, são potenciais marcadores de miRNA adicionais que são significativamente regulados positivamente no CHC (WANG *et al.*, 2019).

Neste contexto sabe-se que os miRNAs contribuem para os vários estágios patológicos da doença hepática e participam do controle da transdiferenciação de células estreladas hepáticas (HSC), fundamentais para a progressão da fibrogênese hepática, tornando-os potenciais biomarcadores e ferramentas terapêuticas (HERCEG; PALIWAL, 2011). O número de publicações que descrevem o papel dos miRNAs na fibrose hepática cresceram exponencialmente na última década, criando, portanto, uma lista de miRNAs profibróticos (miR-21, -221/222, -181b ou -150) e antifibróticos (miR-29b, -101, -122 ou -214-3p) (HAN *et al.*, 2018; ZHOU *et al.*, 2020).

Wilson e colaboradores em 2014 compilaram a função de miRNAs circulantes, em uma metanálise que incluiu 14 estudos envolvendo 1848 casos com CHC e 1187 controles e concluiu que os painéis de miRNAs podem ser biomarcadores para o CHC com AUC=0,99 (sensibilidade de 96% e especificidade de 96%). Muitos autores recomendam prestar atenção ao papel e à função dos miRNAs no diagnóstico, prognóstico e terapia da doença (SZABO; CSAK, 2016; WILSON *et al.*, 2017;).

Devido à estabilidade dos pequenos RNAs é possível a realização de análises de *high-throughput*. No entanto, este tipo de abordagem enfrenta, alguns desafios à frente, tais como o debate sobre o papel da heterogeneidade do tumor e sua avaliação em estágios avançados tem alimentado certo ceticismo acerca do futuro dos tratamentos. Entretanto, por fim, este trabalho visa investigar e compreender, de forma abrangente o mecanismo patogênico e o valor clínico dos SNPs, da expressão diferencial de genes alvos e dos miRNAs associados ao CHC por diversas etiologias, tornando-os biomarcadores de surgimento do CHC, assim como, propiciar uma visão geral das mudanças genéticas e epigenéticas envolvidas no desenvolvimento e progressão do CHC, e conseqüentemente, discutir o papel dos marcadores para estratificar os pacientes com base em seu prognóstico e resposta as terapias, culminando em uma terapêutica eficaz e assim, em melhores condições clínicas para os indivíduos acometidos pelo câncer de fígado.

### 3 JUSTIFICATIVA

O Carcinoma hepatocelular (CHC) é um dos mais comuns e mais letais tipos de câncer no mundo. A sobrevida geral após o diagnóstico é frequentemente inferior a 6 meses, pois o CHC apresenta-se frequentemente em estágio avançado, quando as possibilidades terapêuticas são limitadas, deste modo, poucos pacientes se beneficiam de um tratamento potencialmente curativo.

Os tratamentos para CHC são considerados de alto custo e, principalmente, quando o diagnóstico acontece em estágio muito avançado onde as condições de um pior prognóstico estão mais presentes. As estratégias terapêuticas variam de acordo com o estágio da doença e da quantidade de recursos disponíveis, podendo variar entre: cirurgias de ressecção ou ablação da víscera, quimioembolização ou radioembolização transarteriais, uso de Sorafenib ou Lenvatinib. No entanto em casos mais graves o transplante do fígado passa a ser a única opção terapêutica, porém, grande parte dos pacientes não consegue o transplante devido ao baixo número de doadores. Dados da Associação Brasileira de Transplantes de Órgãos mostram que no ano de 2019 no Brasil foram realizados 2.245 transplantes de fígado (Associação Brasileira de Transplantes de Órgãos, 2019). E o custo médio total do transplante hepático em hospitais universitários públicos brasileiros varia de US\$ 11.384,30 a US\$ 54.698,34, com média de US\$ 20.605,01 e é um procedimento de alta complexidade realizado em poucas instituições do Brasil (PORTELA *et al.*, 2010).

Seguindo os avanços em tecnologias moleculares e genômica, biomarcadores de câncer, como microRNAs, mutações de DNA foi demonstrado que as regiões diferencialmente expressas têm potencial para vigilância e podem ser utilizadas para detecção precoce de CHC. No entanto, esses candidatos a biomarcadores ainda não foram usados na prática clínica. Sendo assim, encontrar marcadores biológicos eficientes, específicos e baratos correspondem a uma estratégia para potencializar a terapêutica administrada ao paciente, possuindo minimamente duas vantagens: (1) Aumentar a eficiência da estratégia de terapia e (2) propiciar uma racionalização das políticas de saúde objetivando uso consciente de recursos financeiros públicos. E ainda melhorar os planos de rastreio para diagnóstico de CHC com maior brevidade.



E ainda a fusão da pesquisa biológica com a ciência de dados gera abordagens únicas as vantagens do aprendizado de máquina incluem flexibilidade e escalabilidade em comparação com métodos bioestatísticos tradicionais, o que o torna implementável para muitas tarefas, como estratificação de risco, diagnóstico e classificação e previsões de sobrevivência. Outra vantagem dos algoritmos de aprendizado de máquina é a capacidade de analisar diversos tipos de dados e incorporá-los em previsões de risco de doença, diagnóstico, prognóstico e tratamentos apropriados, permitindo a identificação de pacientes propensos a progredir mais brevemente para o câncer hepático.

E assim, propiciando um evento de medicina personalizada melhorando a terapêutica, de acordo com a etiologia do CHC e, conseqüentemente, ocasionando um decréscimo nos pacientes onde o transplante hepático passa a ser considerada a única opção terapêutica. O rastreio prévio de populações de risco para desenvolvimento do câncer hepático, mostra-se como um importante e eficaz ferramenta bloquear o surgimento de novos casos, utilizando-se de uma terapêutica eficiente, gerando assim, investimento eficiente e economia dos gastos públicos direcionados a este setor. Bem como uma melhora no manejo clínico dos pacientes, tornando-o mais adequado ocasionando uma melhoria na qualidade de vida dos pacientes.

## 4 OBJETIVO GERAL

Identificar o perfil de expressão de microRNA's candidatos criar um classificador baseado em inteligência artificial a partir de polimorfismos associados ao surgimento do Carcinoma Hepatocelular, atendidos em centros de referência do estado de Pernambuco, no período de 2017 a 2020.

### 4.1 Objetivos Específicos:

- Determinar a frequência alélica e genotípica de SNPs nos genes *PTX3*, *MBL2*, *IL-10*, *TNF $\alpha$* , *MPO*, *SOD2* e *IFNL3* em genes da resposta imunológica em pacientes com CHC;
- Desenvolver um algoritmo de inteligência artificial com base em polimorfismos genéticos para a identificação de padrões de variação genética associados ao desenvolvimento de doença hepática crônica.
- Determinar a expressão de microRNA's candidatos em tecidos hepáticos (Tumoral e não Tumoral) de pacientes com carcinoma hepatocelular (CHC);
- Correlacionar o nível de expressão de miRNA's candidatos com a quantificação de marcadores bioquímicos para dano hepático;

## **5 METODOLOGIA**

Esta pesquisa pode ser classificada como um estudo transversal analítico do tipo caso- controle com comparação de grupos.

### **5.1 Desenho de estudo**

O presente trabalho trata-se de um estudo analítico transversal com grupos de comparação, tem três vertentes principais, na primeira verificou-se associação entre polimorfismos genéticos do tipo SNP (*single nucleotide polymorphism*) e a gravidade da doença hepática. Na segunda etapa, com base nos SNP's avaliados, foi construído um algoritmo de inteligência artificial para reconhecimento de padrões genéticos. Em seguida, para a concretização da última etapa, a mesma, foi dividida em duas fases: (1) Fase de descoberta, onde foi realizada a análise de expressão diferencial de microRNAs por RNA-Seq e (2) fase de validação, onde os resultados da fase de descoberta foram verificados por RT-qPCR.

### **5.2 População Alvo**

A população deste estudo é constituída por pacientes acima de 18 anos, com câncer hepático, tipo Carcinoma Hepatocelular (CHC) atendidos no Serviço de Hepatologia do Instituto do Fígado e Transplantes de Pernambuco, centro de referência para tratamento de hepatopatas na cidade do Recife. Este serviço tem uma média de 900 (novecentos) atendimentos/mês, na sua maioria, pacientes portadores de doença hepática. Cabe destaque que há ambulatório de seguimento de pacientes portadores de nódulos e transplantes hepáticos (03 vezes por semana). Este estudo será realizado em um período de 2017 a 2020.

#### **5.2.1 Grupos de Estudos**

5.2.1.1 Para a primeira etapa do estudo:

**Grupo 01: Pacientes com fibrose leve (F0, F1, F2)**

Critérios de inclusão: pacientes diagnosticados com Hepatite C, maiores de 18 anos, e acompanhados no ambulatório do Serviço de Gastroenterologia do Instituto do Fígado de Pernambuco. Com pelo menos um exame de biopsia hepática para estadiamento da fibrose.

Critérios de exclusão: pacientes que apresentarem outras infecções crônicas associadas.

**Grupo 02: Pacientes com fibrose grave (F3 e F4)**

Critérios de inclusão: pacientes diagnosticados com Hepatite C, maiores de 18 anos, e acompanhados no ambulatório do Serviço de Gastroenterologia do Instituto do Fígado de Pernambuco. Com pelo menos um exame de biopsia hepática para estadiamento da fibrose.

Critérios de exclusão: pacientes que apresentarem outras infecções crônicas associadas.

**Grupo 03: Pacientes com Carcinoma Hepatocelular**

Critérios de inclusão: pacientes diagnosticados com Hepatite C, maiores de 18 anos, com diagnóstico de CHC e acompanhados no ambulatório do Serviço de Gastrohepatologia do Instituto do Fígado de Pernambuco.

Critérios de exclusão: pacientes que apresentarem outras infecções crônicas associadas.

5.2.1.2 Para a segunda etapa do estudo:

***Para a fase de descoberta e para a fase de validação*****Grupo 01: Pacientes com Carcinoma Hepatocelular**

Critérios de inclusão: pacientes, maiores de 18 anos, com diagnóstico de CHC e acompanhados no ambulatório do Serviço de Gastrohepatologia do Instituto do Fígado de Pernambuco, com indicação para transplante hepático.

Critérios de exclusão: pacientes que apresentarem outras infecções crônicas associadas.

**Grupo 02: Grupo controle**

Critérios de inclusão: Tecido hepático saudável oriundo do implante utilizado em eventos de transplante hepático.

Critérios de exclusão: pacientes que apresentarem testes positivo para *VHC*, *VHB*, HIV, HTLV e Doença de Chagas, Sífilis e Citomegalovírus.

**5.3 Tipo De Amostragem E Definição Do Tamanho Da Amostra**

Os pacientes incluídos no estudo foram obtidos como amostra de conveniência e foi formada por pacientes maiores de 18 anos com carcinoma hepatocelular, atendidos no serviço de referência citado acima. As amostras foram coletadas mensalmente de todos os pacientes que concordaram em participar da pesquisa, durante o período de março de 2017 até outubro de 2020. O estudo do polimorfismo dos genes, no que se refere a pacientes brasileiros com CHC são escassos e devido a este fato, para definição do tamanho da amostra não temos uma prevalência da frequência dos polimorfismos para calcular a amostragem, por este motivo, o nosso estudo é considerado de caráter exploratório.

**5.4 Definição e categorização das variáveis**

**Variável dependente:** Neste estudo foi considerada como variável dependente a doença hepática crônica sendo seu desfecho mais grave do Carcinoma Hepatocelular definida pelo grau de invasão celular e acúmulo de matriz extracelular, no parênquima hepático, segundo os padrões definidos pelo *The International Consensus Group for Hepatocellular Neoplasia* (THE INTERNATIONAL CONSENSUS GROUP FOR HEPATOCELLULAR NEOPLASIA, 2009). Assim como, os critérios definidos pelo BCLC (*Barcelona Clinic Liver Cancer*) para diagnóstico de CHC (REIG *et al.*, 2022)

## Variáveis Independentes

**Quadro 1.** Definição das variáveis independentes verificadas no estudo.

Nome da variável	Definição	Categorização
Idade	Anos completos de vida do paciente, no momento do diagnóstico do CHC	Em 2 classes 1-- $\geq 40$ ; 2-- $< 40$
Sexo	Categorizado como masculino e feminino	1--masculino 2--feminino
Polimorfismos SNPs	Alelo 1, Alelo 2	1--0; 2--1
Expressão dos microRNAs	Quantificação relativa da expressão dos miRNAs quando comparados os grupos de estudo (RQ-- Relative quantification)	Variável continua indicada pela sigla RQ
Uso de bebida alcoólica (etilismo)	Classificação de usuários de bebidas alcoólicas classificados de acordo com a quantidade de drinks ingeridos por dia, semana ou mês.	Homens: 1-- $> 40g/dia$ -- Bebedor leve 2-- $< 40g/dia$ -- Bebedor crônico  Mulheres: 1-- $> 20g/dia$ -- Bebedor leve 2-- $< 20g/dia$ -- Bebedor crônico
Diabetes mellitus	glicose alterada em jejum quando os valores forem $\geq 126mg/dl$ . SBEM (Sociedade Brasileira de Endocrinologia e Metabologia)	1--sim 2--não
Uso contraceptivos oral		1--sim; 2--não
Tabagismo	História de tabagismo progresso ou atual.	1--nunca fumou; 2--ex--fumante; 3--fumou até um ano antes da pesquisa; 4--fumante (fuma atualmente)

**A metodologia descrita abaixo foi planejada para contemplar os objetivos específicos:**

- Determinar a frequência alélica e genotípica de SNPs *PTX3*, *MBL2*, *IL-10*, *TNF $\alpha$* , *MPO*, *SOD2* e *IFNL3* em genes da resposta imunológica em pacientes com CHC.

### **5.5 Métodos de coleta**

O paciente foi convidado a participar do estudo, foi informado sobre os objetivos da pesquisa e posteriormente assinou o termo de consentimento livre e esclarecido (Apêndices A) confirmando sua aceitação em participar da pesquisa. Após a assinatura do termo, os dados clínicos e demográficos de cada paciente foram coletados em uma ficha previamente elaborada para este estudo (Apêndice C). O sangue foi coletado por punção venosa a vácuo, em tubo contendo anticoagulante (EDTA) para extração do DNA.

### **5.6 Extração de DNA**

O DNA das amostras foi extraído a partir de sangue total coletado em tubo com EDTA com kit *QIAamp Mini Spin Columns* (QIAGEN) seguindo as instruções do fabricante, e após a extração o material foi identificado e armazenado em refrigerador -20° até o momento da identificação dos polimorfismos. A quantificação do DNA foi feita por NanoDrop 2000 espectrofotômetro (*Thermo Scientific, Waltham, MA*).

### **5.7 Detecção das variantes polimórficas**

Foi utilizada a metodologia de PCR em tempo real, através do sistema TAQMAN®, para detecção dos polimorfismos (SNPs), que consiste de sondas marcadas com fluorocromos desenhadas especificamente para ser complementar aos alelos em estudo. O sistema funciona com um par de sondas específicas para a variação quando presente ou ausente. As sondas têm em sua estrutura molecular um *Quencher* que inibe o sinal emitido pelo fluorocromo ou *repórter*, assim que ocorre o pareamento e hidrólise da sonda o composto fluorescente é liberado proporcionando liberação de energia que é captado pela máquina e analisado em um gráfico, dessa

forma a detecção do polimorfismo é observada. Este tipo de sonda possui uma tecnologia chamada de MGB (*Minor Groove Binder*) a qual tem a capacidade de se ligar na cavidade menor do DNA aumentando a especificidade e diminuindo o *background* da reação.

Todas as sondas e primers utilizados nesse estudo foram obtidas através de ensaios previamente validados através da busca no site da *Life Technologies/Thermo Fisher Scientific* que disponibiliza o banco de dados com os SNPs dos genes que serão utilizados. Para realização desta técnica foi utilizado o PCR em tempo real QuantStudio 5 (*Applied Biosystems, CA, USA*) com o seguinte protocolo: Desnaturação Inicial de 95°C por 10 minutos, seguido de 40 ciclos de desnaturação a 95°C por 15 segundos e extensão/anelamento a 60°C por 1 minuto e uma extensão final de 60°C por 1 minuto.

**A metodologia descrita abaixo foi planejada para contemplar o objetivo específico:**

- Desenvolver um algoritmo de inteligência artificial com base em polimorfismos genéticos para a identificação padrões de variação genética associados ao desenvolvimento de doença hepática crônica.

## **5.8 Construção do algoritmo genético para seleção de características por meio de aprendizado de máquina**

### **5.8.1 Codificação**

Os experimentos deste foram construídos a partir da biblioteca *sklearn* escrita em Python. Entretanto, os modelos que ela dispõe aceitam somente entradas numéricas. Como a base de dados é 100% categórica, foi preciso converter cada uma das entradas para o formato numérico. Para tanto utilizamos a codificação One-Hot, nele cada ocorrência de uma categoria de um dado SNP se transforma numa nova coluna de zeros e uns. Utilizamos também a *recessive-encoding*, abordagem específica para o cenário de dados genômicos, cada genótipo é codificado em duas novas colunas binárias, uma para cada alelo possível. Uma coluna é definida como 0 se o alelo correspondente não estiver presente, e 1 se estiver presente pelo menos uma vez. A tabela 1 ilustra um



exemplo.

**Tabela 1** - Exemplo de codificação Recessiva e One-Hot para um dado SNP.

SNP	Recessive-encoding		One-Hot		
	A	B	AA	AB	BB
AA	1	0	1	0	0
AB	1	1	0	1	0
BB	0	1	0	0	1

### 5.8.2 Subgrupos e Seleção de Características

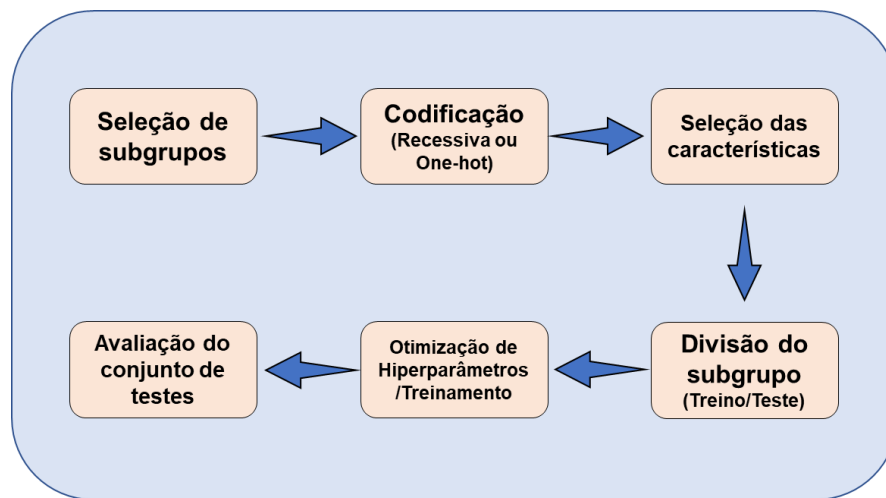
Tendo em vista a frequência dos pacientes com CHC dentro da base de dados, foi visto um cenário de configuração CHC e não-CHC bastante desbalanceado. Neste caso, apenas 14.1% da base de dados são pacientes CHC positivos. Considerando-se detectar padrões até mais diversos, combinações balanceadas do banco de dados foram geradas tendo em vista encontrar indícios sobre como o câncer está relacionado a tipos mais brandos de fibrose. Os algoritmos SVM-RFE e Chi2 são as ferramentas indicadas quando se trata de selecionar características de dados categóricos. Portanto, ambos foram escolhidos como responsáveis por remover as informações irrelevantes do conjunto de entradas.

### 5.8.3 Classificadores

Para trabalhar com grupos que possuem poucas ocorrências, torna-se necessário a introdução de algumas abordagens que amplifique as capacidades de classificação destes modelos. Para este fim, optou-se pela técnica de otimização de hiperparâmetros, aliada à validação cruzada, que encontra a melhor combinação de parâmetros em um espaço de busca pré-definido. Ao encontrar tais parâmetros,

os modelos são reavaliados em um conjunto de testes que representa um total de 15 ou 20% do subgrupo de dados em questão (a proporção das ocorrências das classes é mantida nos conjuntos de treinamento e teste). Para tal, utiliza-se uma série de métricas para enriquecer a etapa de avaliação de resultados deste trabalho. Acurácia, Precisão, Recall e F1 são métricas encontradas na biblioteca sklearn. A Figura 10 ilustra como a arquitetura foi organizada.

**Figura 10** - Arquitetura do sistema.



Fonte: Adaptado de Júnior, et al, 2019.

Ilustração demonstrando as etapas necessárias para codificação, organização, categorização e avaliação do conjunto de testes e treino.

#### 5.8.4 Métricas

As métricas que utilizamos para avaliar os experimentos deste trabalho foram: Acurácia, Precisão, Recall, F1 e AUC/ROC.

##### 5.8.4.1 Acurácia

Basicamente, corresponde ao percentual das previsões que foram corretamente classificadas. Contudo, em casos em que há disparidade entre classes, um alto valor não necessariamente significa o melhor dos modelos. Esta métrica deve ser levada em conta em cenários onde a base de dados é balanceada. Ela é definida pela seguinte equação.

$$\frac{\text{Número de previsões corretas}}{\text{Número total de previsões}}$$

#### 5.8.4.2 Precisão

Existem quatro valores associados às previsões de um modelo. A taxa de verdadeiros positivos refere-se ao percentual das previsões positivas que foram corretamente classificadas. Além disso, a taxa de falsos positivos diz respeito a fração de previsões positivas que foram erroneamente classificadas. Já taxas de verdadeiros e falsos negativos, referem-se as previsões negativas que foram corretamente e erroneamente classificadas, respectivamente. Em outras palavras, esta métrica diz o quanto das previsões positivas são de fato corretas. A métrica de Precisão é definida pela seguinte equação.

$$\frac{\textit{Verdadeiros positivos}}{\textit{Verdadeiros positivos} + \textit{falsos positivos}}$$

#### 5.8.4.3 Recall

A métrica Recall é definida pela seguinte equação.

$$\frac{\textit{Verdadeiros positivos}}{\textit{Verdadeiros positivos} + \textit{falsos negativos}}$$

Ela representa a fração dos reais positivos que foram identificados corretamente, sendo um bom indicativo da performance de modelos onde há um alto custo associado a taxa de falsos negativos.

#### 5.8.4.4 F1

F1 é uma média harmônica entre Precisão e Recall, indicando o quanto estas duas métricas estão equilibradas.

$$\frac{2 * \textit{Precisão} * \textit{Recall}}{\textit{Precisão} + \textit{Recall}}$$

Valores altos de Acurácia não são uma regra para modelos precisos. Então F1 serve como uma ferramenta para averiguar o real significado dessa métrica. Se

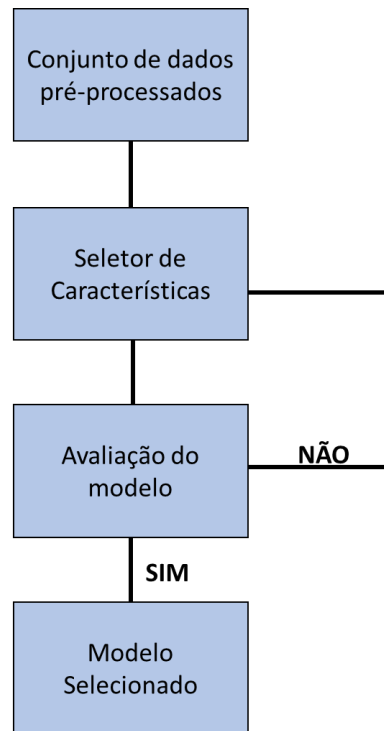
um modelo apresenta um bom F1, então conclui-se haver baixas taxas de falsos positivos e falsos negativos.

#### 5.8.4.5 AUC/ROC

A sigla AUC, do inglês *Area Under Curve*, representa a área abaixo de uma curva num espaço 2D. Já ROC, do inglês *Receive Operation Characteristics*, representa uma curva de probabilidades entre a taxa de verdadeiros positivos versus a taxa de falsos positivos conforme um limiar de discriminação varia. Juntando essas duas definições obtém-se a área abaixo da curva ROC. A literatura indica que modelos com esta métrica abaixo de 0.75 não são classificadores úteis. Já aqueles com valores maiores que 0.97 tem um alto índice de acerto. Portanto, esta métrica revela, num contexto geral, o quão promissores os modelos preditivos podem ser.

### 5.9 Protocolo Experimental para construção de algoritmo genético

Como alguns classificadores não dão suporte a características categóricas, é necessário codificar o conjunto de dados em uma representação numérica. Um *Label Encoder* é utilizado para transformar as SNPs em características numéricas. Os valores de um dado SNP passam de valores: AA, AB, BB para valores numéricos: 0.333, 0.667, 1.0. Foi escolhida a classificação binária a partir do valor da Fibrose separando as classes em dois grupos, em 3 diferentes configurações do conjunto de dados. O objetivo do classificador é prever se um indivíduo do conjunto possui câncer de fígado ou não. As diferentes configurações podem ser vistas na Tabela 2 A Figura 11 apresenta o digrama geral do protocolo experimental.

**Figura 11** - Diagrama do protocolo experimental.

Fonte: Adaptado de Leal, *et. al.*, 2019  
 Protocolo utilizado para a seleção do modelo.

**Tabela 2** - Diferentes configurações do conjunto de dados classificadas neste trabalho.

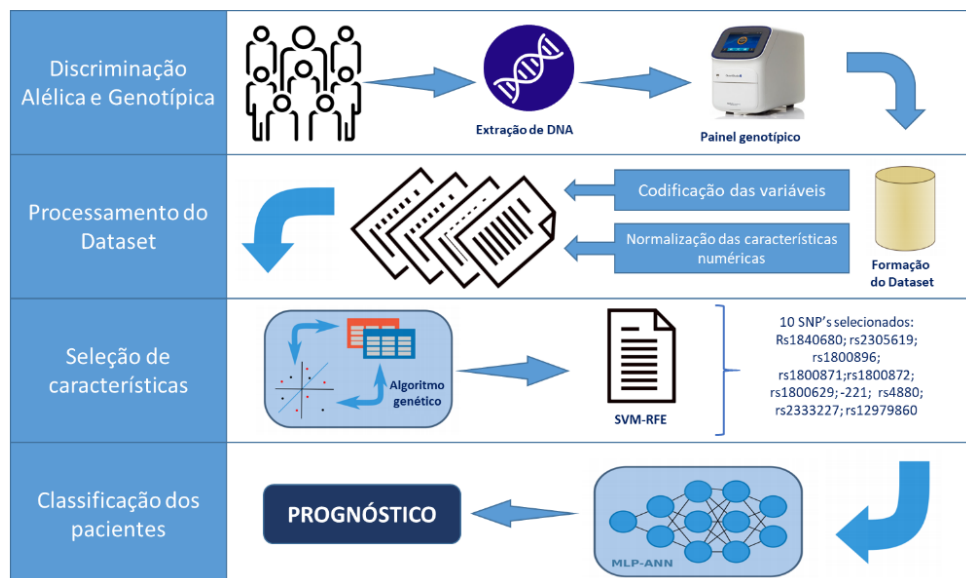
Classe 0	Classe 1	Ocorrências Classe 1
F0, F1, F2, F3, F4	CHC	14.1%
F0, F1, F2, F3	CHC	19.6%
F4	CHC	50.6%

Antes de classificar um modelo, aplicamos a técnica de KFold para criação dos pares de conjuntos de treinamento e teste, são criados k pares treinamento/teste de modo que estes conjuntos não possuem elementos em comum com a finalidade de oferecer uma forma de avaliar o desempenho de um classificador justamente. Na maioria dos modelos foi aplicado um 5Fold, com exceção do One Class SVM onde foi feita a aplicação de um Leave One out. Para cada par treinamento/teste do conjunto de Folds:

1. É aplicada a seleção de características ao conjunto de dados.
2. O conjunto de treinamento é balanceado através de Oversampling, que consiste em repetir de modo aleatório linhas da classe minoritária até que ambas as classes tenham a mesma proporção.
3. O classificador é treinado com o conjunto de treinamento balanceado. No entanto, caso o classificador seja o One Class SVM, o treinamento ocorre somente com as linhas da classe 1 do conjunto de treinamento.
4. O classificador é avaliado com o conjunto de teste, as métricas são computadas.

Após essa etapa são calculadas as médias e desvios padrão das métricas para o conjunto de folds. Nossa abordagem para o prognóstico do câncer hepático usando técnicas de Machine Learning pode ser dividida em quatro estágios: Aquisição de dados, pré-processamento de dados, seleção de recursos e classificação do paciente. Todo o processo é ilustrado na Figura 12.

**Figura 12** - Fluxograma do classificador genômico SVM-ANN para CHC.



Adaptado de Davi, C. *et al.* 2019

1-A aquisição dos dados foi realizada por meio da genotipagem em PCR em tempo real de todos os pacientes com fibrose e CHC, em seguida, armazenada em um banco de dados. 2- O pré-processamento de dados foi realizado para codificar e normalizar os dados em um formato adequado para a etapa de aprendizado de máquina. 3- A seleção de recursos foi realizada para encontrar o melhor conjunto de combinação entre os SNP's. 4- Um classificador de ANN (Rede Neural Artificial) foi treinado para o prognóstico de CHC grave com base nas características recentemente selecionadas.

**A metodologia descrita abaixo foi planejada para contemplar os objetivos específicos:**

- Determinar a expressão de microRNA's candidatos em tecidos hepáticos (Tumoral e não Tumoral) de pacientes com carcinoma hepatocelular (CHC);
- Correlacionar o nível de expressão de miRNA's candidatos com a quantificação de marcadores bioquímicos para dano hepático.

### **5.10 Obtenção das amostras para Sequenciamento de RNA**

O paciente foi convidado a participar do estudo, foi informado sobre os objetivos da pesquisa e posteriormente assinou o termo de consentimento livre e esclarecido (Apêndice 02) confirmando sua aceitação em participar da pesquisa. A coleta do material biológico (tecido hepático) foi realizada, por equipe especializada, durante o momento do transplante hepático. São 2 pontos de coleta por paciente, o primeiro correspondente ao CHC e secundariamente em uma área do parênquima hepático não acometida pela neoplasia. Os tecidos hepáticos saudáveis foram coletados de implantes que foram utilizados em transplantes hepáticos.

### **5.11 Extração de microRNA**

Para a extração de miRNA Total foi utilizado o KIT *RNAeasy Spin colum* (QIAGEN) a partir de tecido, a extração foi realizada, segundo orientações do fabricante. A concentração de miRNA e a integridade do RNA serão verificadas por Qubit 3.0 (*Applied Biosystems*) e Bioanalyzer 2100 (*Agilent Technologies*), respectivamente. O RNA extraído foi imediatamente armazenado em refrigerador -80° até o momento da análise de expressão.

### **5.12 Sequenciamento de Small RNAs**

Para o estudo de expressão diferencial dos microRNAs foi realizado um estudo de comparação relativa via RNA-Seq, considerando 2 amostras por paciente, referentes ao tecido do nódulo de CHC e área adjacente não acometida pela neoplasia. A determinação do padrão de expressão foi feita por sequenciamento de

nova geração, pela técnica de RNA-Seq, através do sequenciador de nova geração MiSeq (Illumina, USA), onde as bibliotecas foram construídas com o Nextflex Small RNA-Seq kit v3 (Bioo Scientific, PerkinElmer, USA). Após a remoção dos adaptadores dos dados brutos os “reads limpos” são comparados com o banco de dados conhecido miRBase 19.0 para identificar miRNAs conservados. Para análise de expressão diferencial foi utilizado o t-test com filtro de >1.5 de fold change, será considerado significativo  $p < 0.05$ . Os Achados mais significativos serão confirmados por qRT-PCR. Para determinar as similaridades e diferenças entre os padrões de expressão para as diferentes etiologias será construído o Diagrama de Venn.

### **5.13 Validação dos miRNAs por RT-qPCR**

O miRNA extraído foi imediatamente armazenado em refrigerador  $-80^{\circ}\text{C}$  até o momento da análise de expressão. A concentração do miRNA foi determinada em Qubit 3.0 (*Applied Biosystems, CA, USA*). Para obtenção do cDNA foram utilizados primers específicos para cada miRNA estudado de acordo com kit microRNA reverse transcription (*Applied Biosystems, CA, USA*). Para a quantificação dos miRNAs utilizamos a tecnologia TAQMAN MicroRNA Assays (*Applied Biosystems, CA, USA*) que possui sondas específicas e sensíveis para detecção de forma quantitativa dos miRNA estudados. Para normalização foi utilizado a estratégia de genes endógenos (RNU48) e cálculo da expressão relativa através do método  $\Delta\Delta\text{Ct}$  pelo software expression Suite v1.1 (*Life Technologies, CA, USA*) e DataAssist v3.0.1 (*Applied Biosystems, CA, USA*).

### **5.14 Análise da rede de interação mRNA-miRNA**

As interações entre microRNAs e RNAs mensageiros foi predita utilizando o miRWalk 2.0, o qual compreende 12 algoritmos (targetScan, RNAhybrid, RNA22, PITA, Pictar2, miRWalk, Microt4, miRNAMap, miRDB, mirbridge, miRanda e miRMap). Os miRNAs diferencialmente expressos foram determinados nos pacotes edgeR e DeSeq2 do BioConductor, utilizamos o filtro de  $|\log_2(\text{foldchange})| > 1$  e p valor ajustado  $< 0,05$ . A rede de interação proteína-proteína (PPI) foi estabelecida pelo STRING (v10.5) e visualizada pelo Cytoscape 3.6.1. As análises de enriquecimento



foram realizadas utilizando o Gene Ontology (GO) e o KEGG e foram conduzidas no clusterProfiler, um pacote do R para classificação funcional e enriquecimento de agrupamentos gênicos utilizando distribuição hipergeométrica.

Para verificar a rede de interação miRNAs-mRNA foi utilizado o software miRNet e para a análise da assinatura funcional de acordo com as manifestações clínicas utilizaremos o miR2Disease e o PhenomiR. As redes funcionais de regulação miRNA-mRNA serão construídas usando o cytoHubba v2.1 baseado na análise de enriquecimento do Cytoscape v3.6.1.

### **5.15 Plano de Tabulação e análise estatística dos dados**

As análises foram realizadas no programa *GraphPad Prism v.7.0*. Realizamos a comparação dos indivíduos positivos com CHC, oriundos de diferentes etiologias. A prevalência de diferentes genótipos foi comparada pelo teste Qui-quadrado. As frequências alélicas foram estimadas pelo método da contagem gênica e softwares específicos como o Arlequin e Haploview e pLINK que foram utilizados para montagem dos mapas de desequilíbrio de ligação e testes estatísticos. O teste Qui-quadrado foi utilizado para verificar se a distribuição genotípica está de acordo com a hipótese do equilíbrio de Hardy-Weinberg. Para as análises de correlação foram utilizados o teste de Correlação de Spermann ou de Pearson, a depender da distribuição normal. A existência de associações entre variáveis categóricas foi avaliada pelos testes Qui-quadrado de Pearson e exato de Fisher quando apropriado para as variáveis quantitativas foram utilizados o teste-t ou Mann-Whitney obedecendo a regra de distribuição normal. As diferenças foram consideradas significativas para valores de  $p < 0,05$ . A magnitude destas associações foi estimada como Odds ratio (OR), utilizando intervalos de confiança de 95%. Regressão logística foi utilizada para o ajuste do OR de acordo com os possíveis fatores de confusão.

### **5.16 Considerações Éticas**

O projeto tem aprovação do Comitê de Ética do HUOC-UPE sob o protocolo CAAE: 02755212.2.0000.5207 (Anexo A), assim como tem aprovação do Comitê de Ética em Pesquisa do Instituto Aggeu Magalhães (CAAE: 34976619.4.0000.5190)

(Anexo B). O estudo será informado para o paciente atendido, segundo a rotina do serviço de Hepatologia, o mesmo, estando disposto a participar da pesquisa assinará um termo de consentimento livre e esclarecido. Após a assinatura do termo, os dados de cada paciente serão coletados em um questionário previamente elaborado para este estudo.

O questionário conterá dados pessoais como nome do paciente, prontuário, gênero, idade, procedência, fatores de risco, exame físico e exames laboratoriais. Após a obtenção dos dados de cada paciente, será iniciada a etapa de obtenção do material biológico para o uso específico da pesquisa. Todos os dados obtidos serão mantidos em sigilo e utilizados exclusivamente para os fins da pesquisa científica.

## 6 RESULTADOS

Por conta das diferentes técnicas utilizadas para a construção do trabalho esta seção está subdividida em tópicos, que serão apresentados na seguinte ordem: Iniciaremos apresentando os resultados voltados para os ensaios de frequências dos SNP's, os quais foram utilizados para a construção do algoritmo genético de classificação. Em seguida, versaremos sobre os principais achados referentes a análise de transcriptoma, por fim, culminando na confirmação dos resultados do RNA-Seq por RT-qPCR. E concatenando estas informações apresentaremos os resultados de anotação funcional.

Os dados referentes aos SNPs utilizados como objetos de pesquisa neste trabalho foram selecionados com base em dados publicados na literatura e que apresentassem significado biológico para a análises dentro da ótica de progressão do processo fibrótico e, concomitantemente, possível relevância/influência para o estabelecimento da doença hepática crônica. Os dados de discriminação alélica foram obtidos de base de dados própria pré-existente. No entanto, a estratégia analítica, considerando a utilização destes dados para a construção de algoritmo computacional é inédita. A qual além desta tese gerou como produto dois trabalhos de conclusão de curso: (1) Júnior, N *et al.*, 2019 e (2) Leal, B *et al.*, 2019.

### 6.1 Análise dos Polimorfismos

Dentre os 297 pacientes infectados por *VHC*, 144 foram classificados como fibrose leve, correspondente aos níveis F0, F1 e F2, neste trabalho. 111 com fibrose septal ou avançada e cirrose, e, por fim, 42 pacientes compunham o grupo CHC. A caracterização da população amostral encontra-se na tabela 01. Pacientes com CHC foram significativamente mais idosos em comparação com os demais grupos (média de idade  $53.03 \pm 11.13$ ,  $57.86 \pm 9.37$  e  $61.29 \pm 7.98$ , respectivamente de acordo com os níveis de acometimento hepático;  $p < 0.0001$ ). Foi visto também, no grupo CHC, a prevalência do gênero masculino 66.6% ( $p < 0.0061$ ).

Em relação as características bioquímicas, para as variáveis bilirrubina total, AST (aspartato aminotransferase),  $\gamma$ GT (Gama-Glutamiltranspeptidase), FA (Fosfatase Alcalina) e AFP (alfafetoproteína) todas foram significativamente maiores no grupo de pacientes em estado mais grave de doença hepática ( $p < 0.0001$ ). Como

imaginado os índices bioquímicos de gravidade de doença hepática, APRI e FIB-4, foram maiores em pacientes com fibrose avançada e CHC ( $p < 0.0001$ ). Entretanto, a taxa plaquetária foi maior em pacientes ainda em estágio inicial da doença. Não foram encontradas diferenças significativas dentre a análise da carga viral, do genótipo viral, ALT, presença de *Diabetes Mellitus* e Hipertensão Arterial. (Tabela 3).

**Tabela 3** - Características clínicas e bioquímicas relacionadas aos grupos com diferentes níveis de fibrose hepática.

Variáveis	Fibrose Leve (n=144)	Fibrose Avançada (n=111)	CHC (n=42)	p-Value
<b>Clínicas</b>				
Idade (anos)	53.03±11.13	57.86±9.37	61.29±7.98	<0.0001
Sexo masculino (%)	62 (43,05)	32 (28,82)	28 (66,66)	0.0061
IMC (kg/m <sup>2</sup> )	25.5 (19.1-38.1)	26.9 (18.1-40.4)	25.3 (17.3-37.1)	0,2967
<i>Diabetes Mellitus</i>	26 (18,05)	24 (21,62)	9 (21,42)	0,1784
Hipertensão Arterial	53 (36,80)	51 (45,94)	10 (23,80)	0,1582
APRI	0.59 (017-4.11)	1.24 (0.12-5.89)	2.65 (0.35-10.23)	<0.0001
FIB-4	1.65 (0.37-8.6)	3.42 (0.40-11.64)	6.62 (1.27-18.59)	<0.0001
<b>Bioquímicas</b>				
Bt [mg/dL] (min-max)	0.68 (0.20-3.35)	0.90 (0.35-3.30)	1.10 (0.50-7.22)	<0.0001
AST (U/L) (min-max)	45.5 (8.0-254.0)	68.0 (12.0-438.0)	94.0 (16.0-235.0)	<0.0001
ALT (U/L) (min-max)	69.0 (14.0-561.0)	74.0 (7.0-267.0)	72.5 (18.0-491.0)	0,2435
γGT (U/L) (min-max)	56.0 (12.0-484.0)	93.5 (14.0-823.0)	172.0 (41.0-525)	<0.0001
FA (U/L) (min-max)	70.0 (13.0-410.0)	91.0 (23.0-332.0)	138.5 (51.0-310.0)	<0.0001
Contagem de Plaquetas (min-max)	196,000 (53,000-365,000)	141,500 (58,000-370,000)	100,500 (36,000-324,000)	<0.0001
AFP (ng/mL) (min-max)	2.81 (0.50-82.0)	6.5 (0.92-155.0)	25.32 (1.97-9,352.0)	<0.0001

Carga Viral	612,660 (539.0- 28,900,000)	583,500 (277.0- 5,531,380)	22,928 (4,400- 1,130,000)	0,2435
VHC genótipo 1 (%)	101(70,13)	74 (66,66)	20 (47,61)	0.91

Fonte: O autor.

Valores expressos em Média  $\pm$  Desvio Padrão, Mediana (variações) ou número absoluto (%) Abreviações: Bt – Bilirrubina total; AST aspartato aminotransferase; ALT – Alanina aminotransferase;  $\gamma$ GT – Gama glutamil transpeptidase; FA – Fosfatase alcalina; AFP: alfafetoproteína; IMC: índice de Massa Corpórea.  $p$  - Análise entre todos os grupos – ANOVA ou Kruskal-Wallis.

A Tabela 4 sintetiza as frequências alélicas e genotípicas das variantes estudadas nos genes *PTX3*, *MBL2*, *IL-10*, *TNF $\alpha$* , *MPO*, *SOD2* e *IFNL3* de acordo com o nível de evolução da fibrose e a ocorrência de CHC. A distribuição de todos os polimorfismos analisados encontra-se de acordo com a hipótese do equilíbrio de Hardy-Weinberg. A Figura 13 mostra todas as variantes em desequilíbrio de ligação acima de  $r^2=0,8$  em população AMR (*Ad Mixed American*). Foi comprovado que as variantes rs1800871/rs1800872 do gene *IL-10* estão em total desequilíbrio, assim como as variantes rs2305619 e rs1840680 ambas do gene *PTX3* encontram-se em  $r^2=0.895$ .

**Tabela 4** - Frequência alélica e genotípica dos SNPs estudados, de acordo com a severidade da fibrose hepática.

<b>Polimorfismos</b>	<b>Fibrose Leve (144/288)</b>	<b>Fibrose Avançada (111/222)</b>	<b>CHC (42/84)</b>
<b><i>rs1840680 PTX3</i></b>			
G	181 (62,84)	127 (57,20)	44 (52,38)
A	107 (37,16)	95 (42,80)	40 (47,62)
GG	58 (40,28)	33 (29,74)	12 (28,59)
GA	65 (45,14)	61 (54,95)	20 (47,61)
AA	21 (15,58)	17 (15,31)	10 (23,80)
<b><i>rs2305619 PTX3</i></b>			
G	154 (54,47)	113 (50,90)	37 (44,04)
A	134 (46,53)	109 (49,10)	47 (55,96)
GG	43 (29,87)	23 (20,73)	9 (21,44)
GA	68 (47,22)	67 (60,36)	19 (45,23)
AA	33 (22,91)	21 (18,91)	14 (33,33)
<b><i>-221 MBL2</i></b>			
Y	234 (81,25)	189 (85,13)	65 (77,38)
X	54 (18,75)	33 (14,87)	19 (22,62)
YY	96 (66,66)	79 (71,17)	26 (61,90)
YX	42 (29,16)	31 (27,92)	13 (30,95)
XX	6 (4,18)	1 (0,91)	3 (7,15)
<b><i>rs1800896 IL-10</i></b>			
G	101 (35,06)	76 (34,23)	27 (32,14)
A	187 (64,94)	146 (65,77)	57 (67,87)
GG	18 (12,51)	14 (12,62)	4 (9,54)
GA	65 (45,13)	48 (43,24)	19 (45,23)
AA	61 (42,36)	49 (44,14)	19 (45,23)
<b><i>rs1800871/rs1800872 IL-10</i></b>			
C	199 (69,09)	147 (66,21)	51 (60,71)
T/A	89 (30,91)	75 (33,79)	33 (39,28)
CC	71 (49,30)	47 (42,34)	15 (35,71)
CT/CA	57 (39,58)	53 (47,74)	21 (50,00)
TT/AA	16 (11,12)	11 (9,92)	6 (14,29)

**rs1800629 TNF $\alpha$** 

G	243 (84,37)	197 (88,73)	77 (91,66)
A	45 (15,63)	25 (11,27)	7 (8,34)
GG	102 (70,83)	87 (78,37)	35 (83,33)
GA	39 (27,08)	23 (20,72)	7 (16,64)
AA	3 (2,09)	1 (0,91)	0 (0)

**rs4880 SOD2**

G	145 (50,34)	118 (53,15)	46 (54,76)
A	143 (49,66)	104 (46,85)	38 (45,24)
GG	40 (27,77)	31 (27,92)	12 (28,57)
GA	65 (45,13)	56 (50,45)	22 (52,38)
AA	39 (27,10)	24 (21,63)	8 (19,05)

**rs2333227 MPO**

G	208 (72,22)	180 (81,08)	64 (76,19)
A	80 (27,78)	42 (18,92)	20 (23,80)
GG	74 (51,38)	73 (65,76)	25 (59,52)
GA	60 (41,66)	34 (30,63)	14 (33,33)
AA	10 (6,96)	4 (3,61)	3 (7,15)

**rs12979860 IFNL3**

C	151 (52,43)	109 (49,10)	44 (52,38)
T	137 (47,57)	113 (50,90)	40 (47,61)
CC	34 (23,61)	23 (20,72)	13 (30,95)
CT	83 (57,63)	63 (56,72)	18 (42,85)
TT	27 (18,76)	25 (22,53)	11 (26,20)

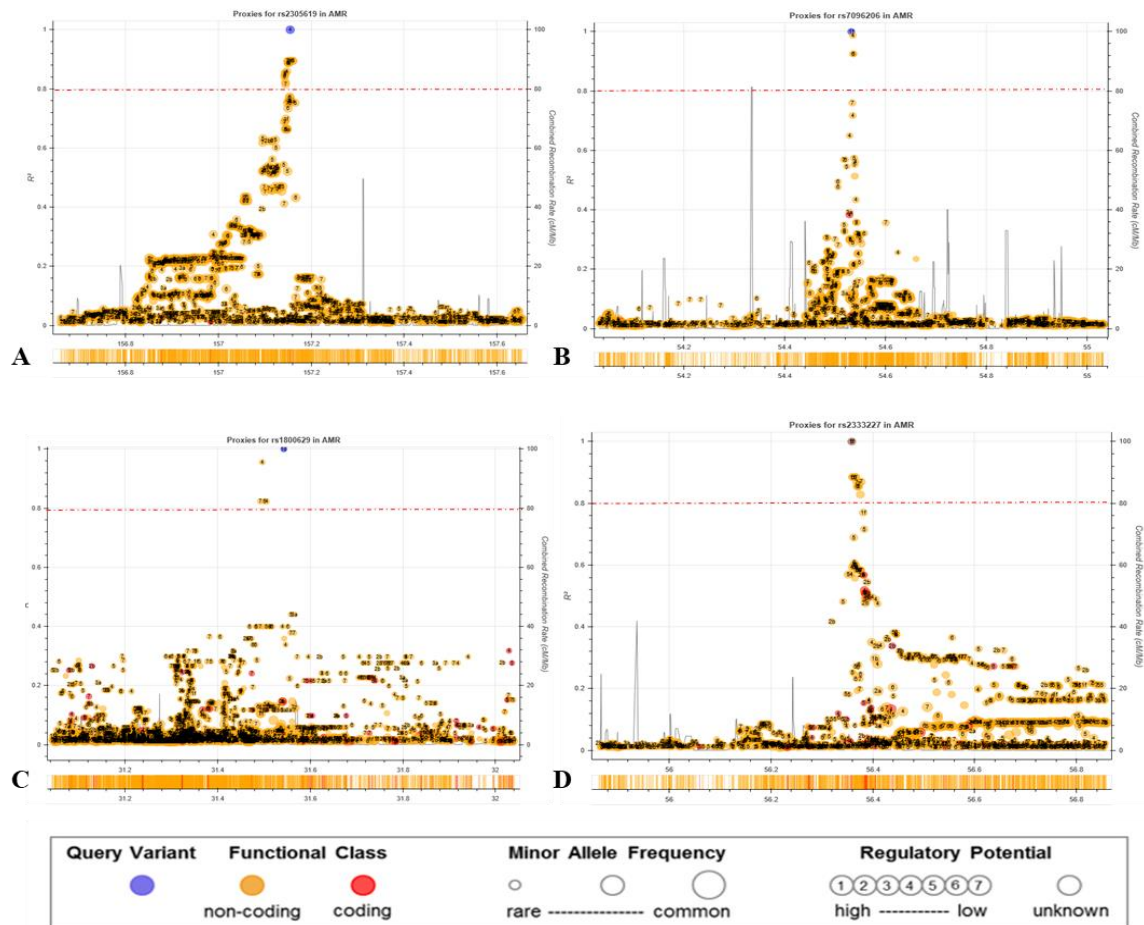
---

Fonte: O Autor

Dados apresentados em valor absoluto e (porcentagem).

Os resultados significativos para todas as combinações de grupos, assim como, os modelos genéticos referentes às análises comparativas de risco estão sumarizadas na tabela 5.

**Figura 13** - Análise do desequilíbrio de ligação, em variantes proximais supostamente funcionais.



Fonte: O autor

Imagem retratando os demais pontos de variação genética em desequilíbrio com os SNPs alvos, seguindo os parâmetros Desequilíbrio de ligação (LD) acima de  $R^2 = 0,8$  e taxa de recombinação acima de 80 Cm / Mb, representados pela linha vermelha tracejada. (A) Imagem representativa para o rs2305619, gene *PTX3*. (B) representa a análise feita para a região -221 (rs7096206) do gene *MBL2*, em (C) para o rs1800629 pertencente ao gene *TNF $\alpha$*  e (D) retrata o resultado para o rs2333227, gene *MPO*. A análise foi realizada no pacote LDproxy da ferramenta LDlink v3.3.0.

Para variantes polimórficas analisadas para o gene *PTX3* foi observada associação do rs1840680 em modelo alélico e dominante com  $p=0.0441$   $OR=1.33$   $95\% IC 0.96-1.85$  e  $p=0.032$   $OR=1.61$   $95\% 1.004-2.64$ , respectivamente quando comparado o grupo de fibrose leve vs. fibrose avançada somado ao grupo de carcinoma hepatocelular. Mas também foram encontrados valores significativos para o rs2305619 em modelos overdominant e GG vs GA ( $p=0.0251$   $OR=1.70$   $95\% IC 1.01-2.80$  e  $p=0.0337$   $OR=1.84$   $95\% IC 1,003-3.38$ , respectivamente) em combinação de fibrose leve vs avançada. Assim como, evidenciando o CHC como desfecho positivo em contrabalanço com estado avançado de fibrose hepática com os valores  $p=0.0466$



OR=2.14 95% IC 0.96-4.76 e  $p=0.0375$  OR=2.35 95% IC 1.002 - 5.48 para os modelos recessivo e GA vs. AA, na devida ordem.

**Tabela 5** - Modelos Genéticos utilizados para análise de associação ao risco de progressão para doença hepática grave.

Modelos Genéticos	Fibrose Leve VS Fibrose Avançada	Fibrose Avançada VS CHC	Fibrose Leve VS CHC	Fibrose Leve VS Fibrose Avançada + CHC
<b>rs1840680</b> <b>PTX3</b>				
Alélico	0.2302 / 1.26 (0.65-3.07)	0.5288 / 1.21 (0.73-2.01)	<u>0.0548 / 1.54</u> (0.94-2.51)	<b>0.0441 / 1.33</b> (0.96-1.85)
Dominante	0.8934 / 0.89 (0.43-1.34)	0.9533 / 1,05 (0.48-2.31)	0.2313 / 1.68 (0.79-3.56)	<b>0.032 / 1.61</b> (1.004-2.64)
GG vs GA	<u>0.0501 / 1.64</u> (0.95-2.81)	0.9756 / 0.90 (0.39-2.07)	0.4364 / 1.48 (0.66-3.30)	<u>0.0502 / 1.33</u> (0.96-1.85)
<b>rs2305619</b> <b>PTX3</b>				
Dominante	0.1315 / 1.62 (0.91-2.91)	0.8992 / 0.95 (0.40-2.28)	0.3809 / 1.56 (0.68-3.54)	<u>0.0505 / 1.60</u> (0.94-2.73)
Recessivo	0.5351 / 0.78 (0.42-1.44)	<b>0.0466 / 2.14</b> (0.96-4.76)	0.2439 / 1.68 (0.79-3.56)	0.8967 / 0.99 (0.94-1.71)
Overdominant	<b>0.0251 / 1.70</b> (1.01-2.80)	0.4938 / 1.32 (0.69-2.52)	0.9593 / 1.08 (0.54-2.15)	0.1518 / 1.69 (0.97-2.96)
GG vs GA	<b>0.0337 / 1.84</b> (1,003-3.38)	0.6589 / 0.72 (0.28-1.82)	0.6700 / 1.33 (0.55-3.21)	0.0416 / 1.69 (0.97-2.96)
GA vs AA	0.2396 / 0.64 (0.33-1.22)	<b>0.0375 / 2.35</b> (1.002 - 5.48)	0.5412 / 1.40 (0.62-3.19)	0.3757 / 0.94 (0.94-1.43)
<b>-221 MBL2</b>				
YY vs XX	0.2207 / 0.20 (0.02-1.71)	<b>0.0489 / 9.11</b> (0.90-1.48)	0.6727 / 1.84 (0.43-7.88)	0.6659 / 0.60 (0.16-2.22)
YY vs YX	<b>0.0354 / 0.54</b> (0.32-0.93)	0.1613 / 0.78 (0.35-1.72)	0.8812 / 1.14 (0.53-2.43)	0.9695 / 0.95 (0.57-1.58)
<b>rs1800629</b> <b>TNF<math>\alpha</math></b>				
Alélico	0.1970 / 1.45 (0.86-2.46)	0.5908 / 0.71 (0.29-1.72)	0.0646 / 2.03 (0.91-4.88)	<b>0.0399 / 1.58</b> (0.97-2.57)
Dominante	0.2225 / 1.49 (0.83-2.65)	0.6490 / 1.37 (0.54-3.49)	0.1558 / 2.05 (0.84-5.0)	<b>0.0498 / 1.62</b> (0.95-2.76)
<b>rs2333227 MPO</b>				

Alélico	<b>0.0263 / 1.64</b> <b>(1.07-2.51)</b>	0.4292 / 1.33 (0.73-2.45)	0.5605 / 0.81 (0.46-1.42)	<b>0.0403 / 1.51</b> <b>(1.03-2.21)</b>
Dominante	<b>0.0295 / 1.81</b> <b>(1.09-3.02)</b>	0.5966 / 0.87 (0.36-1.58)	0.4508 / 1.39 (0.69-2.79)	<b>0.0364 / 1.68</b> <b>(1.05-2.68)</b>
Overdominant	<b>0.0464 / 1.61</b> <b>(0.95-2.72)</b>	0.8994 / 0.88 (0.41-1.88)	0.4285 / 0.70 (0.34-1.44)	<b>0.0424 / 1.56</b> <b>(0.97-2.51)</b>
GG vs GA	<b>0.0272 / 1.74</b> <b>(1.02-2.95)</b>	0.7871 / 1.20 (0.55-2.59)	0.4222 / 0.69 (0.33-1.44)	<b>0.0273 / 1.65</b> <b>(1.01-2.68)</b>
GG vs GA vs AA	<b>0.0305</b>	0.5797	0.6146	<b>0.0422</b>

Fonte: O Autor

Dados apresentados em valor de  $p$  / Odds Ratio (Intervalo de confiança 95%) os valores considerados significativos encontram-se destacados, e valores considerados próximos a significância estão sublinhados. Para a análise de contingenciamento foi utilizado o teste de Qui-Quadrado com correção de Yates, assim como, o teste Exato de Fisher quando necessário. (\*) Teste unicaudal.

Para a variação presente na região promotora -221 do gene *MBL2* foram encontrados os valores significativos comparando o modelo YY vs YX, na combinação de estágios de fibrose leve vs avançada, revelando a única associação onde o fator de exposição positivo (apresentar a variação genética) foi considerado fator protetivo com  $p=0.0354$  OR=0.54 95% IC 0.32-0.93. Ao passo que quando realizada a combinação de Fibrose avançada vs CHC foi encontrado o fator de risco considerado mais elevado, comparado com todos os demais deste estudo com  $p=0.0489$  OR=9.11 95% IC 0.90-1.48, em modelo comparativo para os dois genótipos homocigóticos.

Em relação a variante rs1800629 no gene *TNF $\alpha$*  foram encontrados resultados relevantes apenas quando feita a combinação do estado mais leve de fibrose contra os grupos de fibrose avançada somado ao grupo CHC, revelando  $p=0.0399$  OR=1.58 95% IC 0.97-2.57 para o modelo alélico e  $p=0.0498$  OR=1.62 95% IC 0.95-2.76 para o modelo dominante.

No gene *MPO* foi avaliada a variação rs2333227, e para a combinação Leve vs Avançada foram encontrados valores significativos para os modelos Alélico, Dominante, *Overdominant*, GG vs GA e GG vs GA vs AA com os valores de  $p=0.0263$  OR=1.64 95% IC 1.07-2.51,  $p=0.0295$  OR=1.81 95% IC 1.09-3.02,  $p=0.0464$  OR=1.61 95% IC 0.95-2.72,  $p=0.0272$  OR=1.74 95% IC 1.02-2.95 e  $p=0.0305$ . Estes mesmos modelos foram significativos para a combinação de fibrose leve e em contrapartida os pacientes com estado mais grave da doença hepática, com os valores  $p=0.0403$  OR=1.51 95% IC 1.03-2.21,  $p=0.0364$  OR=1.68 95% IC 1.05-2.68,  $p=0.0424$  OR=1.56 95% IC 0.97-2.51,  $p=0.0273$  OR=1.65 95% IC 1.01-2.68 e  $p=0.0422$ . Não foram

encontrados valores estatisticamente relevantes nas demais combinações, tampouco, para as variações rs1800896, rs1800871 e rs1800872 do gene *IL-10*, rs4880 do gene *SOD2* e rs12979860 do *IFNL3*.

## **6.2 Construção da Rede Neural Artificial**

A construção do modelo computacional se deu por meio de técnicas de aprendizado de máquina, visando o reconhecimento de padrões, no caso combinações de genótipos que se apresentam de modo frequente e repetitivo em grupo de doença hepática crônica e diferencial em grupo de indivíduos que não desenvolveram a forma mais grave da hepatopatia. Esta vertente do trabalho tem por objetivos: (1) avaliar o desempenho de diferentes classificadores aplicados ao banco de dados gerado a partir dos polimorfismos anteriormente avaliados frente ao aparecimento do CHC. (2) E, por fim, propor uma abordagem de algoritmo genético realizando seleção de características para melhorar a métrica AUROC dos modelos.

### **6.2.1 Conjunto de dados**

Este conjunto possui 297 registros, cada registro tem características genéticas de um paciente e o diagnóstico do grau de Fibrose correspondente (sendo considerado o CHC como o estado de maior gravidade). As características genéticas correspondem à 10 SNPs, advindos da análise anterior, de alguns genes associados à resposta imune e com papel decisivo em doenças hepáticas. As distribuições das classes no conjunto de dados e o significado de cada uma delas estão presentes na Tabela 06. Estes SNPs assumem valores sem ordem associada, e.g. AA, GG, GA, TT etc., o que torna a base de dados 100% categórica. Como grande parte dos modelos de aprendizagem de máquina foram construídos em função de entradas contínuas, este trabalho está inserido num cenário um tanto desafiador, oferecendo suporte a natureza inovadora em que se encontra a ideia da geração de um classificador baseado em inteligência artificial para reconhecimento de padrões biológicos.

**Tabela 6.** Especificação da proporcionalidade de classes presentes do conjunto de dados.

<b>Classe</b>	<b>Quantidade</b>	<b>Proporção</b>	<b>Nível de Fibrose (Estadiamento Clínico)</b>
F0	6	02.0%	Sem fibrose
F1	58	19.5%	Fibrose Leve
F2	80	26.9%	Fibrose Intermediária
F3	70	23.6%	Fibrose Avançada
F4	41	13.8%	Cirrose
CHC	42	14.1%	Câncer Hepático

Nota-se que a base de dados possui múltiplas classes e poucas ocorrências, como é o caso de F0, CHC e F4 as quais possuem frequência menor que 50 cada. Como o objetivo deste trabalho é discernir entre casos de câncer e não câncer, então alguns experimentos foram propostos tendo em vista utilizar apenas modelos de classificação binária, ao invés de alguma abordagem multiclasse. Segue-se o princípio de que qualquer entrada cuja classe seja CHC, é assumido o valor 1, caso contrário 0. Todavia, o conjunto de entradas apresenta-se bastante desbalanceado, com 42 ocorrências positivas e 256 não-positivas. Um total de 80 experimentos foram executados. A grande maioria gerou resultados com acurácia menor que 50%, sendo descartados deste trabalho. No fim, somente os experimentos com resultados mais promissores serão abordados aqui.

### **6.2.2 Melhores configurações**

Em cada subgrupo, o modelo em questão utiliza 85% das entradas como conjunto de treino e 15% como conjunto de testes. A proporção entre ocorrências positivas e negativas se mantém a mesma para cada conjunto. A Tabela 7 traz os resultados mais promissores e passíveis de serem explorados de forma mais elaborada no futuro. Ao utilizar tais classificadores encontramos os melhores resultados para os parâmetros analisados com a divisão nos subgrupos CHC vs. F4 (realizado com classificador Random Forest) e CHC + F4 vs. F0 + F1 (feito com o classificador Support Vector Machine), ambos classificadores já descritos para

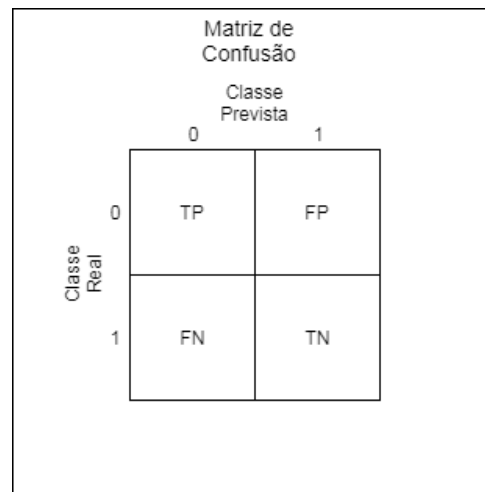
trabalhos com dados biológicos. Nota-se que de modo geral é possível por meio dos SNP's avaliados separar formas graves e leves da hepatopatia, assim como, individualizar cirróticos e pacientes com câncer.

Os resultados obtidos nos experimentos anteriores foram utilizados como base para a construção do algoritmo genético próprio e personalizado para a nossa base de dados. Tais experimentos apresentados aqui correspondem à análise dos classificadores com a aplicação de seleção de características com a configuração do conjunto de dados completa, isto é a configuração F0, F1, F2, F3, F4 vs CHC. Assim como os resultados dos classificadores com a aplicação de seleção de características com a configuração do conjunto de dados F0, F1, F2, F3 vs CHC e com os classificadores com a aplicação de seleção de características com a configuração do conjunto de dados F4 X CHC.

**Tabela 7** - Melhores configurações encontradas, demonstrando os parâmetros de Acurácia, F1, Precisão, Recall e AUC/ROC.

<b>Experimento</b>	<b>Acurácia</b>	<b>F1</b>	<b>Precisão</b>	<b>Recall</b>	<b>AUC\ROC</b>
<b>CHC vs F4</b>	<b>0.77</b>	<b>0.77</b>	<b>0.83</b>	<b>0.71</b>	<b>0.77</b>
CHC e F3 vs F2	0.56	0.65	0.56	0.77	0.55
CHC e F4 vs F3	0.65	0.69	0.64	0.75	0.65
<b>CHC e F4 vs F0 e F1</b>	<b>0.74</b>	<b>0.79</b>	<b>0.73</b>	<b>0.85</b>	<b>0.72</b>
CHC vs F0, F1, F2, F3, F4	0.64	0.38	0.25	0.83	0.72

Cada par classificador/subconjunto de características é avaliado a partir de 2 diferentes métricas: Acurácia e a AUC\ROC. A matriz confusão (Figura 14) também é apresentada e fornece um quadro geral do desempenho do modelo.

**Figura 14** - Diagrama da matriz de confusão

Fonte: Júnior, *et al.*, 2019

Diagrama da matriz de confusão. Ilustração explicativa sobre o conceito interpretativo e construção da matriz de confusão. A classe 0 corresponde aos indivíduos sem CHC e a classe 1 corresponde aos indivíduos com CHC.

### **6.2.3 Configuração do conjunto de dados completo para classificação dos modelos**

Para classificação e avaliação dos modelos foram testadas 3 estratégias de análise: (1) aplicação dos classificadores com o conjunto de dados sem aplicação de seleção de características, (2) classificadores com a aplicação de seleção de características e ainda (3) a execução do algoritmo de seleção de características aplicado ao classificador One Class SVM com Leave One Out.

A Tabela 08 apresenta os resultados da execução dos modelos com o conjunto de dados completos. Nota-se que na estratégia onde não foi feita a seleção de características e foi possível perceber que o modelo de Gradient Boosting apresentou os melhores resultados tanto para acurácia quanto para AUC\ROC. Ao passo que quando aplicada a execução dos modelos com a utilização de algoritmo genético para seleção de características, com foco na melhora da métrica AUC\ROC, pode-se perceber que o modelo de SVM apresentou melhor índice para acurácia, no entanto o modelo em MLP foi melhor para AUC\ROC.

Quanto a execução do modelo One Class SVM em uma configuração de conjunto treinamento/teste diferente dos demais experimentos, a Leave One Out, equivalente ao KFold com k igual à 297. O treinamento do modelo One Class SVM

é feito somente com a classe de indivíduos com CHC, que é um subconjunto com 42 indivíduos e o teste é feito por indivíduos das duas classes, onde foram encontradas as melhores métricas.

**Tabela 8** - Resultados dos classificadores com o conjunto de dados sem aplicação de seleção de características.

<b>Modelo</b>	<b>Acurácia</b>	<b>AUC\ROC</b>	<b>Matriz de Confusão</b>
Classificação sem seleção de características			
SVM	0,677 (0,061)	0,499 (0,057)	(192, 33) (63, 9)
Random Forest	0,610 (0,039)	0,478 (0,036)	(168, 29) (87, 13)
MLP	0,704 (0,007)	0,501 (0,038)	(201, 34) (54, 8)
Gradient Boosting	<b>0,744 (0,040)</b>	<b>0,504 (0,070)</b>	(215, 36) (40, 6)
Classificação com seleção de características			
SVM	<b>0,667 (0,045)</b>	0,605 (0,058)	(177, 21) (78, 21)
Random Forest	0,627 (0,114)	0,612 (0,135)	(165, 21) (90, 21)
MLP	0,643 (0,036)	<b>0,615 (0,085)</b>	(169, 20) (86, 22)
Gradient Boosting	0,647 (0,059)	0,609 (0,062)	(169, 19) (86, 23)
One Class Classification com seleção de características			
One Class SVM	<b>0,754</b>	<b>0,688</b>	(199, 17)
Leave One Out			(56, 25)

#### **6.2.4 Conjunto de dados F0,F1,F2,F3 X CHC**

Neste experimento, foi removida a classe F4 do conjunto de dados, objetivando analisar fenótipos mais extremos. A Tabela 09 apresenta os resultados da execução dos modelos com a utilização de algoritmo genético para seleção de características, com foco na melhora da métrica AUC\ROC.

**Tabela 9** - Resultados dos classificadores com a aplicação de algoritmo genético para seleção de características no conjunto (F0,F1,F2,F3 X CHC).

Modelo	Acurácia	AUC\ROC	Matriz de Confusão
SVM	0,629 (0,045)	0,577 (0,057)	(139, 20) (75, 22)
Random Forest	<b>0,660 (0,054)</b>	0,587 (0,036)	(148, 21) (66, 21)
MLP	0,609 (0,056)	<b>0,634 (0,092)</b>	(130, 16) (84, 26)
Gradient Boosting	0,617 (0,046)	0,592 (0,047)	(134, 18) (80, 24)

### 6.2.5 Conjunto de dados F4 X CHC

Neste experimento, foi mantida a classe F4 e removidas as classes F0, F1, F2 e F3 do conjunto de dados, visando tentar melhorar os índices encontrados anteriormente. A Tabela 10 apresenta os resultados da execução dos modelos com a utilização de algoritmo genético para seleção de características, com foco na melhora da métrica AUC\ROC. Onde os melhores resultados foram vistos com o modelo do tipo multilayer perceptron (MLP).

**Tabela 10** - Resultados dos classificadores com a aplicação de algoritmo genético para seleção de características no conjunto (F4 X CHC).

Modelo	Acurácia	AUC\ROC	Matriz de Confusão
SVM	0,638 (0,055)	0,657 (0,068)	(28, 17) (13, 25)
Random Forest	0,615 (0,054)	0,615 (0,062)	(31, 22) (10, 20)



MLP	<b>0,676 (0,106)</b>	<b>0,693 (0,100)</b>	(31, 17) (10, 25)
Gradient Boosting	0,615 (0,078)	0,664 (0,083)	(28, 19) (13, 23)

### 6.2.6 Melhores modelos em relação à métrica de AUC\ROC

O classificador MLP obteve os melhores resultados na configuração de treinamento 5Fold. No entanto, o melhor resultado geral foi o One Class SVM Leave One Out, alcançando métrica AUC\ROC igual à 0.688, dados evidenciados na tabela 11.

Uma hipótese inicial foi a de que indivíduos da classe F4, pacientes com diagnóstico de cirrose, possuísem maior proximidade aos indivíduos com CHC, o que tornaria a separação entre F4 e CHC mais difícil. No entanto, observando os resultados foi verificada uma piora na maioria dos resultados com a remoção da classe F4 e os resultados da classificação somente entre F4 e CHC obtiveram métricas mais altas.

**Tabela 11** - Resultados dos melhores classificadores para cada uma das configurações com a aplicação de algoritmo genético para seleção de características.

Modelo	Configuração do dataset	Configuração treinamento / teste	Acurácia	AUC\ROC	Matriz de Confusão
MLP	F0,F1,F2,F3,F4 X CHC	5Fold	0,643 (0,036)	0,615 (0,085)	(169,20) (86, 22)
One Class SVM	F0,F1,F2,F3,F4 X CHC	Leave One Out	0,754	0,688	(199, 17) (56, 25)
MLP	F0,F1,F2,F3 X CHC	5Fold	0,609 (0,056)	0,634 (0,092)	(130, 16) (84, 26)
MLP	F4 X CHC	5Fold	0,676 (0,106)	0,693 (0,100)	(31, 17)

(10, 25)

Na Tabela 12 apresentamos a porcentagem de presença de cada característica em cada um dos conjuntos de divisão dos dados. As características com maior presença no conjunto de características selecionadas variam de uma configuração de divisão de dados para outra. Para a configuração F0,F1,F2,F3,F4 X CHC destacam-se 4 características com pelo menos 50% de incidência, para a configuração F0,F1,F2,F3 X CHC esse número é de 3 e para a configuração F4 X CHC esse número é de 6. Apenas 3 características aparecem em todas as configurações, são elas: IL-10-1082, TNF-308 e SOD2.

**Tabela 12** - Porcentagens de incidência das características em cada conjunto de divisão dos dados nos diferentes classificadores.

Características	Configuração: F0, F1, F2, F3, F4 X CHC	Configuração: F0, F1, F2, F3 X CHC	Configuração: F4 X CHC
PTX3 rs1840680	00,00%	12,50%	75,00%
PTX3 rs2305619	00,00%	12,50%	00,00%
MBL -221	20,00%	00,00%	75,00%
IL-10 -1082	60,00%	12,50%	25,00%
IL-10 -819	80,00%	00,00%	50,00%
IL-10 -590	20,00%	00,00%	50,00%
TNF-308	60,00%	75,00%	25,00%
SOD2	20,00%	75,00%	75,00%
MPO C-463T	00,00%	12,50%	50,00%
IL-28b rs12979860	80,00%	75,00%	00,00%

### 6.3 RNA-Seq e validação dos dados

Nesta seção apresentamos os dados referentes a última etapa de experimentos, a mesma, foi dividida em duas fases: (1) Fase de descoberta, onde foi realizada a análise de expressão diferencial de microRNAs por RNA-Seq e (2) fase de validação, onde os resultados da fase de descoberta foram verificados por RT-qPCR.

#### 6.3.1 Caracterização da amostra

Em relação à coorte de descoberta, homens com idade média de  $60,88 \pm 5,02$  anos, formaram a maior parte do nosso grupo amostral (75%). Todos apresentam resultados negativos para Esquistossomose Hepatoesplênica e apenas 25% relatam baixo consumo de álcool ( $< 40\text{g}/\text{dia}$ ). Seis pacientes (75%) têm *Diabetes Mellitus*. Observa-se que os parâmetros bioquímicos glicose, bilirrubina total, albumina, gama-glutamil transferase, creatinina, tamanho do tumor, índice de massa corporal, seguiram a distribuição normal e apresentaram média de  $181,9 \pm 104,0$ ,  $2,77 \pm 0,92$ ,  $3,57 \pm 3,03$ ,  $109,0 \pm 51,18$ ,  $0,77 \pm 0,17$ ,  $2,26 \pm 0,24$  e  $23,27 \pm 3,05$ , respectivamente. Por outro lado, as variáveis Aspartato-Aminotransferase (AST), Alanina-Aminotransferase (ALT), contagem de plaquetas, Alfa-fetoproteína estavam fora da distribuição Gaussiana com mediana de 100,5 (59 - 2130), 66 (33 - 726), 121.000 (71.000 - 193.000) e 4,28 (1,30 - 1.590,0). Enquanto metade deles apresentava lesões uninodulares, porém, todos foram submetidos a tratamento cirúrgico por quimioembolização transarterial (TACE) e apenas uma amostra (12,5%) foi classificada como *Barcelona Clinic Liver Cancer* (BCLC) tipo A. Apresentamos todos os dados referentes clínicos e caracterização bioquímica de amostras utilizadas na fase de descoberta na Tabela 13.

Já no que se refere a coorte de validação, a mesma, foi composta por 40 amostras (20 CHC e 20 tecidos adjacentes pareados) e 10 tecidos hepáticos saudáveis. Homens com idade média de  $61,75 \pm 7,82$  anos formaram a maioria de nossa amostra (70%). A principal etiologia encontrada foi a infecção pelo *VHC* (55%), seguida por NASH (Esteatohepatite Não Alcoólica) com 20% das amostras. A infecção pelo *VHB* e a esquistossomose hepatoesplênica foram encontradas em 10%, além do

CHC para alto consumo de álcool (5%). O grupo controle foi equacionado por sexo para ambas fase de descoberta e validação, os tecidos saudáveis foram obtidos por meio de implantes hepáticos utilizados em transplantes, cujos pacientes sofreram morte encefálica por Acidente Vascular Encefálico Hemorrágico (AVEH) ou Traumatismo Craniano Encefálico (TCE) todos os constituintes desde grupo foram negativos para os testes Anti-VHC, Anti-HBs, Anti-HBc, Anti-HBS-Ag, Anti-HIV, Anti-HTLV I/II, Toxoplasmose, Citomegalovírus, VDRL e Chagas.

**Tabela 13** - Características clínicas e bioquímicas relacionadas às amostras utilizadas na fase de descoberta (experimento RNA-Seq).

	<b>Caso (8)</b>	<b>Controle (8)</b>	<b>p(Value)</b>
Gênero Masculino	6 (75%)	5 (62,5%)	>0,9999
Idade	60.8 ± 5.02	45.11 ± 15.73	<b>0,0073</b>
Consumo de Álcool	8 (100%)	7 (87.5%)	>0,9999
negativo		1 (12.5%)	
EHE	(+) 2 (25%) (-) 6 (75%)	(-) 8 (100%)	<b>0,0070</b>
DM	(-) 2 (25%) (+) 6 (75%)	(+) 2 (25%) (-) 6 (75%)	0,1319
HAS	(-) 4 (50%) (+) 4 (50%)	-	-
Glicose	140.50 (79 – 331)	-	-
Albumina	3.80 (2.3 – 4.2)	-	-
Bilirrubina Total	1.97 (0.57 – 5.24)	-	-
Fosfatase Alcalina	90.00 (64 – 180)	-	-
GGT	109.00 (46 – 208)	37 (14 – 116)	<b>0,0258</b>
INR	1.55±0.54	-	-
Creatinina	0.77 (0.6 – 1.02)	-	-
AST	100.5 (59 – 2122.00)	49 (27 – 217)	<b>0.0354</b>
ALT	66 (33 – 726)	50 (22 – 100)	0.1036
Contagem de	121.10 <sup>6</sup> (71.10 <sup>6</sup> -	-	-
Plaquetas	159.10 <sup>6</sup> )		
AFP	4.28 (1.3 – 1590)	-	-
BCLC	A - 7 (87,5%) B - 1 (12,5%)	-	-
Número de Nódulos	2.12±1.72	-	-
Tamanho médio do	2.26±0.24	-	-
Tumor			
IMC	23.26±3.04	-	-
Tipo de Tratamento	TACE – 8 (100%)	-	-

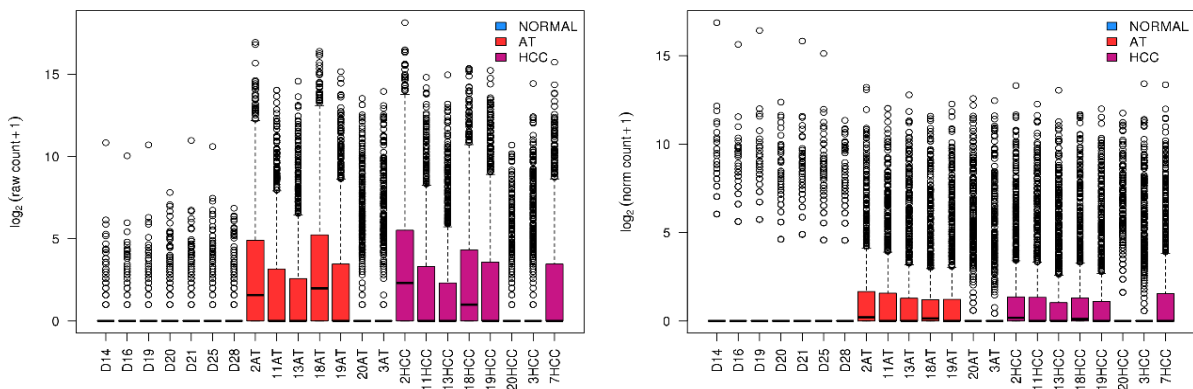
Fonte: O Autor

Valores expressos como Média ± Desvio Padrão, Mediana (alterações) ou número absoluto (%)  
 Abreviações: Bt - Bilirrubina total; AST aspartato aminotransferase; ALT - Alanina aminotransferase; gama de  $\gamma$ GT - glutamil transpeptidase; FA - fosfatase alcalina; AFP: alfa-fetoproteína; IMC: Índice de Massa Corporal. BCLC - *Barcelona Clinic Liver Cancer*, TACE - quimioembolização transarterial. Testes utilizados: Mann-Whitney, Teste T-Student e Qui-Quadrado

### 6.3.2 Análise diferencial de miRNAs em CHC, tecido adjacente e controles

Depois de carregar os dados, verificamos o dataset criado, o mesmo contém valores de contagem brutas que representam o número de *reads* mapeadas para os alvos. Para este projeto, existem 2883 *features* na tabela de dados de contagem, para melhor análise normalizamos os dados visando corrigir vieses técnicos sistemáticos nos dados, a fim de tornar as contagens de *reads* comparáveis entre as amostras (Figura 15). Nota-se que a principal variabilidade no experimento é advinda de diferenças biológicas entre as amostras. A análise de agrupamento mostra que os tecidos saudáveis usados como grupo de controle são semelhantes entre si, evidenciado pela técnica de VST (Variance Stabilizing Transformayion) (Figura 16), outra forma de visualizar a variabilidade do experimento foi realizado uma PCA (Principal Component Analysis), conforme mostrado na figura 17.

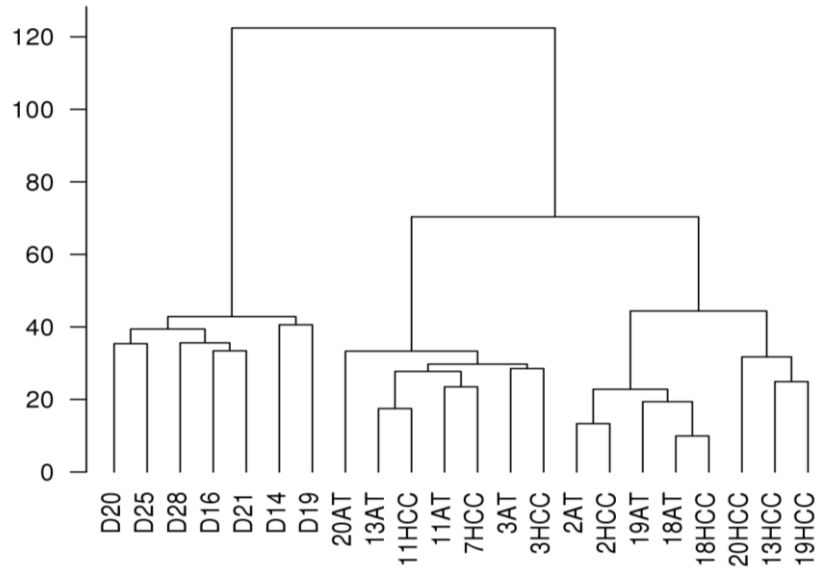
**Figura 15** - Boxplots de contagens de leitura bruta (esquerda) e normalizada (direita).



Fonte: O autor:

Os boxplots são frequentemente usados como uma medida qualitativa da qualidade do processo de normalização, pois mostram como as distribuições são afetadas globalmente durante esse processo. Percebe-se que a normalização estabiliza as distribuições entre as amostras. Na imagem são exibidos os gráficos referentes aos dados brutos (à esquerda) e normalizados (à direita), respectivamente. Os dados são exibidos de acordo com o material biológico: Tecido normal (D), tecido adjacente (AT) ou carcinoma hepatocelular (CHC) e o número correspondente da amostra.

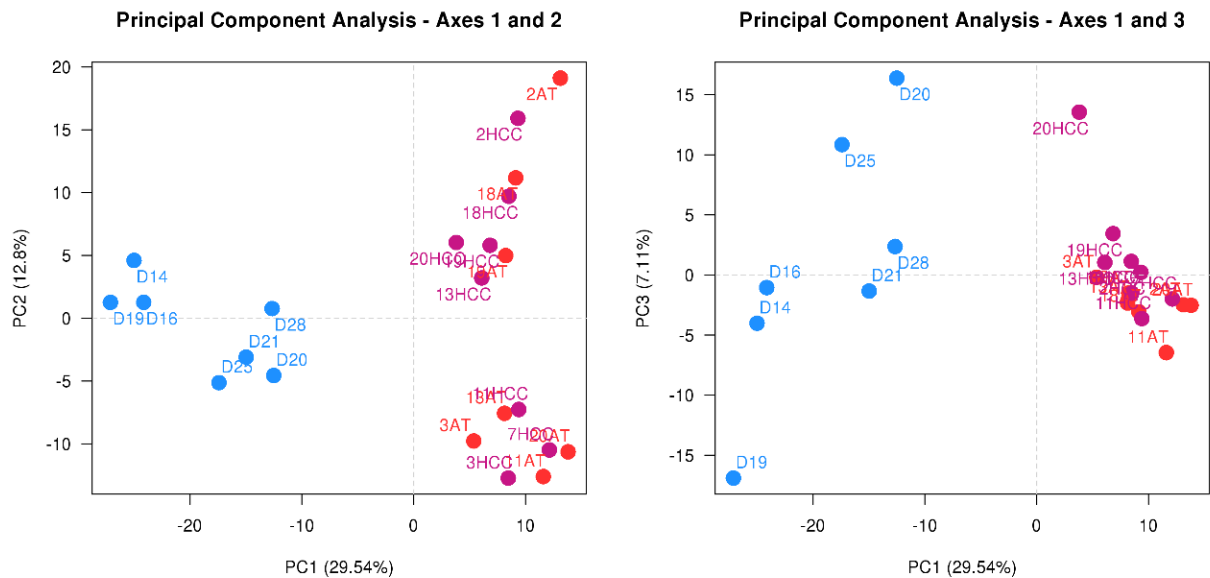
**Figura 16** - Distribuição das amostras baseada nos dados normalizados.



Fonte: O autor

A Figura 16 mostra o dendrograma obtido a partir de dados transformados com VST (*Variance Stabilizing Transformation*). Uma distância euclidiana é calculada entre as amostras, e o dendrograma é construído com base no critério de Ward. Pode-se perceber um agrupamento das réplicas e uma separação das condições biológicas. Os dados são exibidos de acordo com o material biológico: Tecido normal (D), tecido adjacente (AT) ou carcinoma hepatocelular (CHC) e o número correspondente da amostra.

**Figura 17** - Análise de variação dos dados realizada via PCA (Principal Component Analysis).

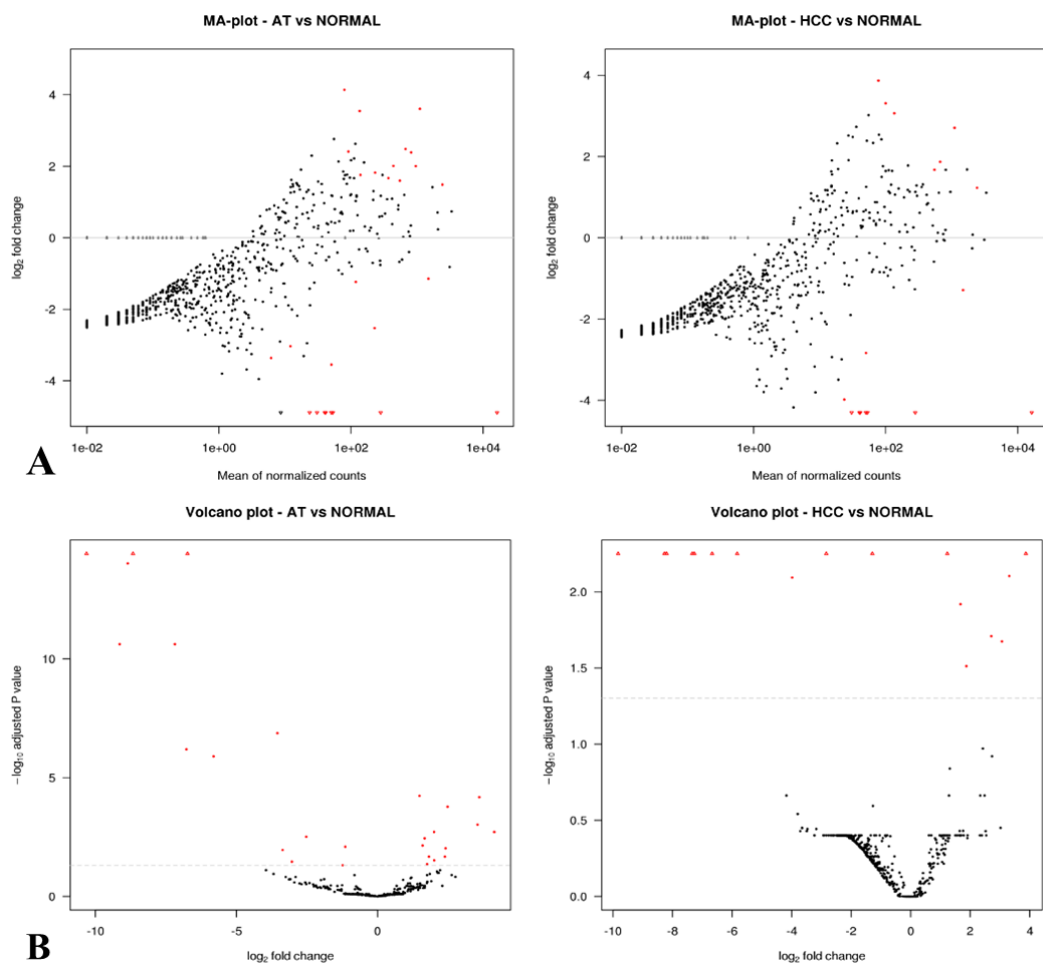


Fonte: O autor

Nesta figura, percebe-se que o primeiro componente principal (PC1) separa as amostras das diferentes condições biológicas, o que significa que a variabilidade biológica corresponde a principal fonte de variância dos dados.

Depois de normalizar os dados, encontramos 29 miRNAs expressos de forma diferencial. Entre esses 17 miRNAs são expressos diferenciais quando analisamos tecido CHC versus tecido hepático saudável, enquanto 27 miRNAs são diferenciais entre tecido adjacente e tecido hepático saudável e 15 miRNAs são diferencialmente expressos em ambas as análises (Tabela 14 e Figura 18).

**Figura 18** - MA-plot (s) e Volcano plot (s) em cada tipo de estratégia de análise.



Fonte: O autor

(A) A ilustração representa o MA-Plot para as comparações feitas, onde as características expressas diferencialmente são destacadas em vermelho. Um MA-Plot representa a razão logarítmica da expressão diferencial como uma função da intensidade média para cada característica. Os triângulos correspondem a características com um  $\log_2(\text{FC})$  muito baixo/alto para serem exibidos no gráfico. (B) Esta figura mostra os Volcano-Plot para as comparações realizadas e as características expressas diferencialmente ainda estão destacadas em vermelho. O Volcano-Plot representa o log do valor P ajustado como uma função da razão log da expressão diferencial. Os dados são exibidos de acordo

com o material biológico: Tecido normal (D), tecido adjacente (AT) ou carcinoma hepatocelular (CHC) e o número correspondente da amostra.

Entre os 17 analisados para CHC, observamos 7 regulados positivamente e 10 regulados negativamente. Em relação aos 27 miRNAs diferencialmente expressos no tecido adjacente, 13 foram regulados positivamente e 14 regulados negativamente em comparação com tecidos saudáveis. Um diagrama de Venn representa os miRNA diferencialmente expressos (Figura 19 e Figura 20). Depois de remover as contagens de leitura abaixo de 10, encontramos os 10 principais microRNA usados para validar a expressão (hsa-miR-1291, hsa-miR-21-3p, hsa-miR-26a, hsa-miR-125a-5p, hsa-let-7a-5p, hsa-miR-193b, hsa-miR-26b, hsa-miR-27b, hsa-let-7c, hsa-miR-148b). O heatmap evidenciou uma diferenciação entre as amostras de tumor hepático e os controles que formaram 2 grupos distintos (Figura 21).

**Tabela 14** - MicroRNAs diferencialmente expressos em ambos os tecidos, apenas no CHC e apenas no tecido do tumor adjacente.

miRNA	CHC vs <i>Tecido Saudável</i>			Tecido Adjacente vs <i>Tecido Saudável</i>		
	FC	Log2FC	p-value adj	FC	Log2FC	p-value adj
<b>Ambos</b>						
hsa-miR-1291	0.001	-9,831	<b>9,26E-29</b>	0.001	-10.315	<b>4,33E-31</b>
hsa-miR-4449	0.006	-7,338	<b>1,13E-03</b>	0.002	-8,668	<b>2,04E-07</b>
hsa-miR-3609	0.01	-6,671	<b>2,57E-02</b>	0.009	-6,735	<b>9,68E-01</b>
hsa-miR-3651	0.003	-8,201	<b>3,26E+01</b>	0.002	-8,853	<b>9,68E-01</b>
hsa-miR-4521	0.003	-8,273	<b>2,17E+04</b>	0.002	-9,14	<b>2,42E+03</b>
hsa-miR-6516-5p	0.006	-7,269	<b>1,81E+03</b>	0.007	-7,184	<b>2,42E+03</b>
hsa-miR-1248	0.018	-5,83	<b>0.00010</b>	0.009	-6,774	<b>6,44E+05</b>
hsa-miR-21-3p	0.14	-2,836	<b>0.00016</b>	0.085	-3,549	<b>1,35E+07</b>
hsa-miR-10395-3p	0.063	-3,983	<b>0.00807</b>	0.018	-5,812	<b>1,28E+08</b>
hsa-miR-26a	2.345	1,23	<b>0.00563</b>	2.804	1,487	<b>5,84E+09</b>
hsa-let-7b-5p	6.531	2,707	<b>0.01958</b>	12.181	3,607	<b>6,68E+09</b>
hsa-miR-125a-5p	3.656	1,87	<b>0.03080</b>	5.592	2,483	<b>0.00016</b>
hsa-miR-150	8.365	3,064	<b>0.02119</b>	11.672	3,545	<b>0.00095</b>
hsa-miR-26b	3.191	1,674	<b>0.01207</b>	3.024	1,597	<b>0.00812</b>
hsa-miR-27b	0.41	-1,288	<b>0.00563</b>	0.452	-1,145	<b>0.00812</b>



**Apenas no CHC**

hsa-miR-224	14.608	3,869	<b>0.00563</b>	3.131	1.647	0.358021
hsa-miR-221	9.923	3,311	<b>0.00789</b>	4.498	2.169	0.101457

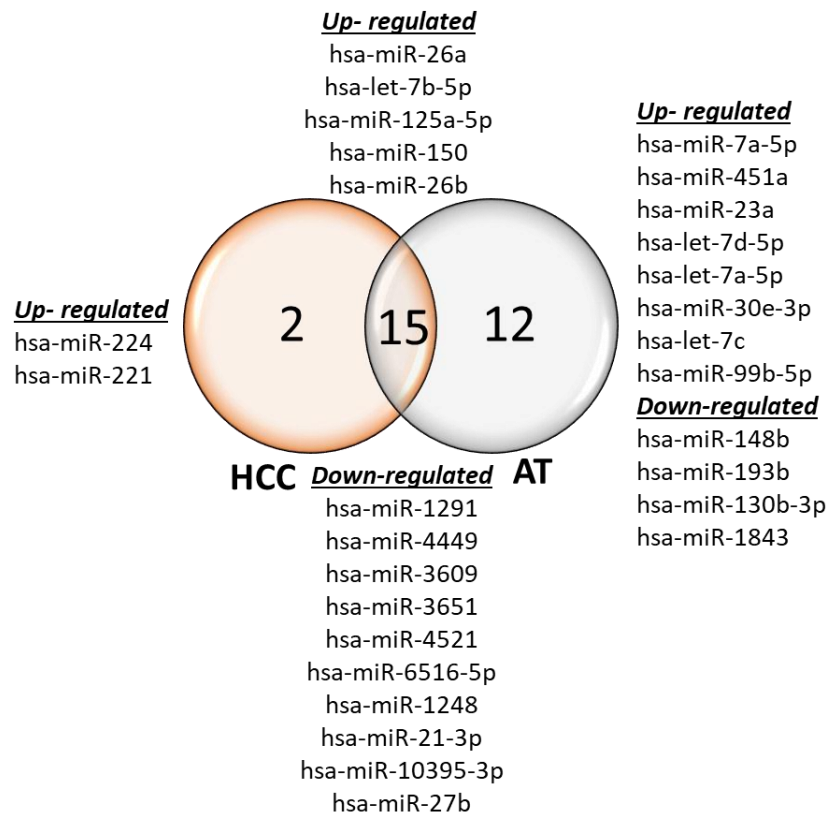
**Apenas no Tecido Adjacente**

hsa-let-7a-5p	2.187	1.129	0.397408	4.01	2,004	<b>0.001934</b>
hsa-miR-451a	5.806	2.537	0.397408	17.612	4,138	<b>0.001934</b>
hsa-miR-193b	0.486	-1.04	0.397408	0.174	-2,527	<b>0.003077</b>
hsa-miR-23a	2.487	1.314	0.144871	3.18	1,669	<b>0.003588</b>
hsa-let-7d-5p	3.389	1.761	0.371152	5.324	2,412	<b>0.009443</b>
hsa-miR-130b-3b	0.591	-0.759	0.642592	0.097	-3,363	<b>0.011136</b>
hsa-let-7a-5p	3.189	1.673	0.397408	5.234	2,388	<b>0.021231</b>
hsa-miR-30e-3p	2.194	1.133	0.397408	3.536	1,822	<b>0.021231</b>
hsa-let-7c	1.082	0.85	0.425871	4.028	2,01	<b>0.030325</b>
hsa-miR-1843	0.239	-2.067	0.397408	0.122	-3,034	<b>0.034515</b>
hsa-miR-99b-5p	2.088	1.062	0.397408	3.383	1,758	<b>0.043750</b>
hsa-miR-148b	0.579	-0.787	0.066016	0.425	-1.235	<b>0.048725</b>

Fonte: O autor

FC (Fold Change); Log2FC (Log2 Fold Change); p-value adj (p-value adjusted with False Discovery Rate Test)

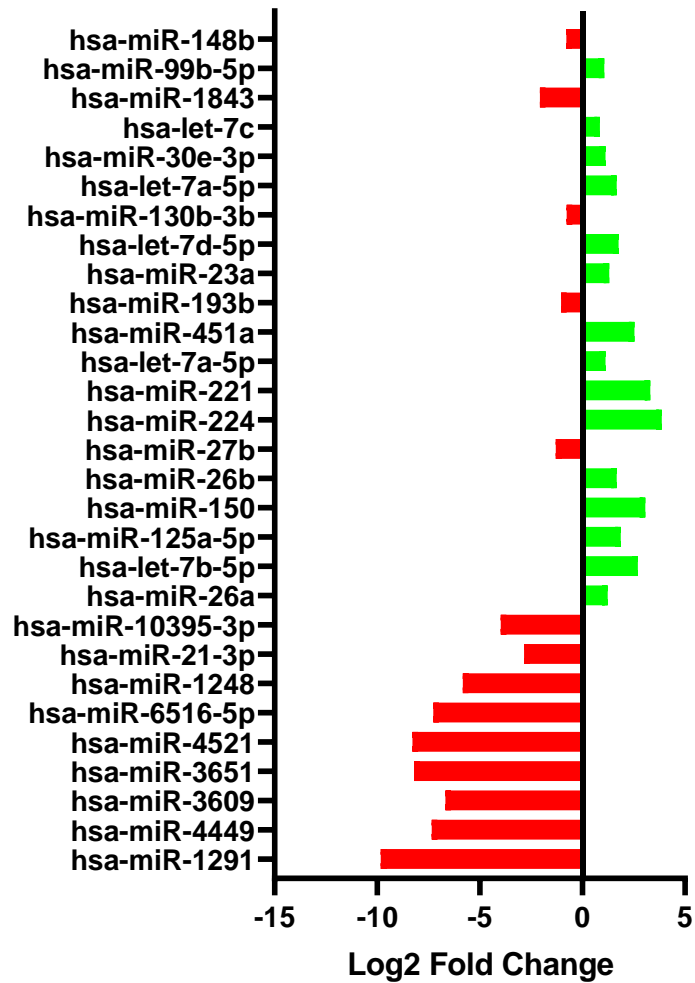
A figura 22 detalha a contagem absoluta de *reads* para cada microRNA em cada grupo, comparando o número médio de leituras, na maioria das análises feitas entre pares de grupos: tecido hepático saudável vs CHC e tecido hepático saudável vs tecido adjacente. No entanto, não observamos a mesma diferença na comparação entre o tecido tumoral e o tecido adjacente, novamente indicando a semelhança biológica entre o centro e a adjacência do tumor. Da mesma forma, foi observada significância quando os três grupos foram comparados, mostrando que há diferença entre as contagens da sequência absoluta.

**Figura 19** - Diagrama de Venn para microRNAs expressos diferencialmente.

Fonte: O autor

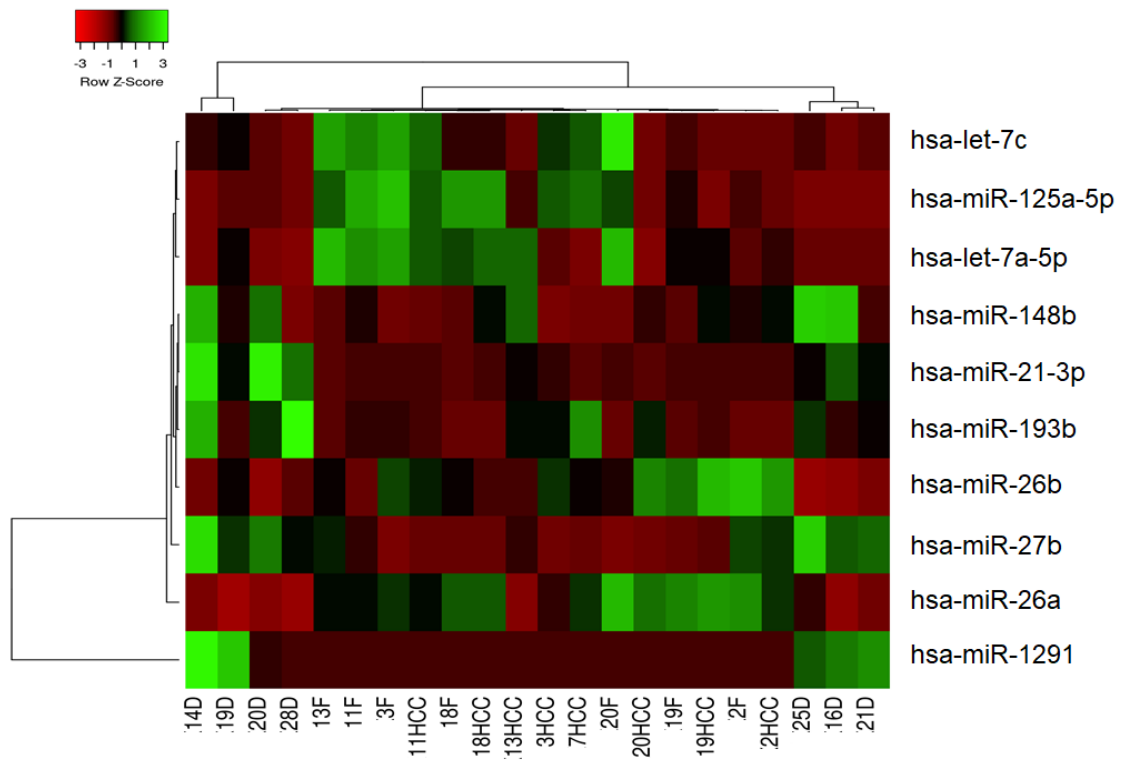
Observamos que dos 29 miRNAs expressos diferencialmente, apenas 2 apresentaram expressão diferencial apenas no carcinoma hepatocelular (CHC), enquanto 12 são exclusivos do tecido adjacente (AT) e 15 em ambas as condições. A imagem foi construída usando o pacote VennDiagram em linguagem R.

**Figura 20** - Gráfico evidenciando os miRNAs diferencialmente expressos considerando todas as combinações de análises.



Fonte: O autor

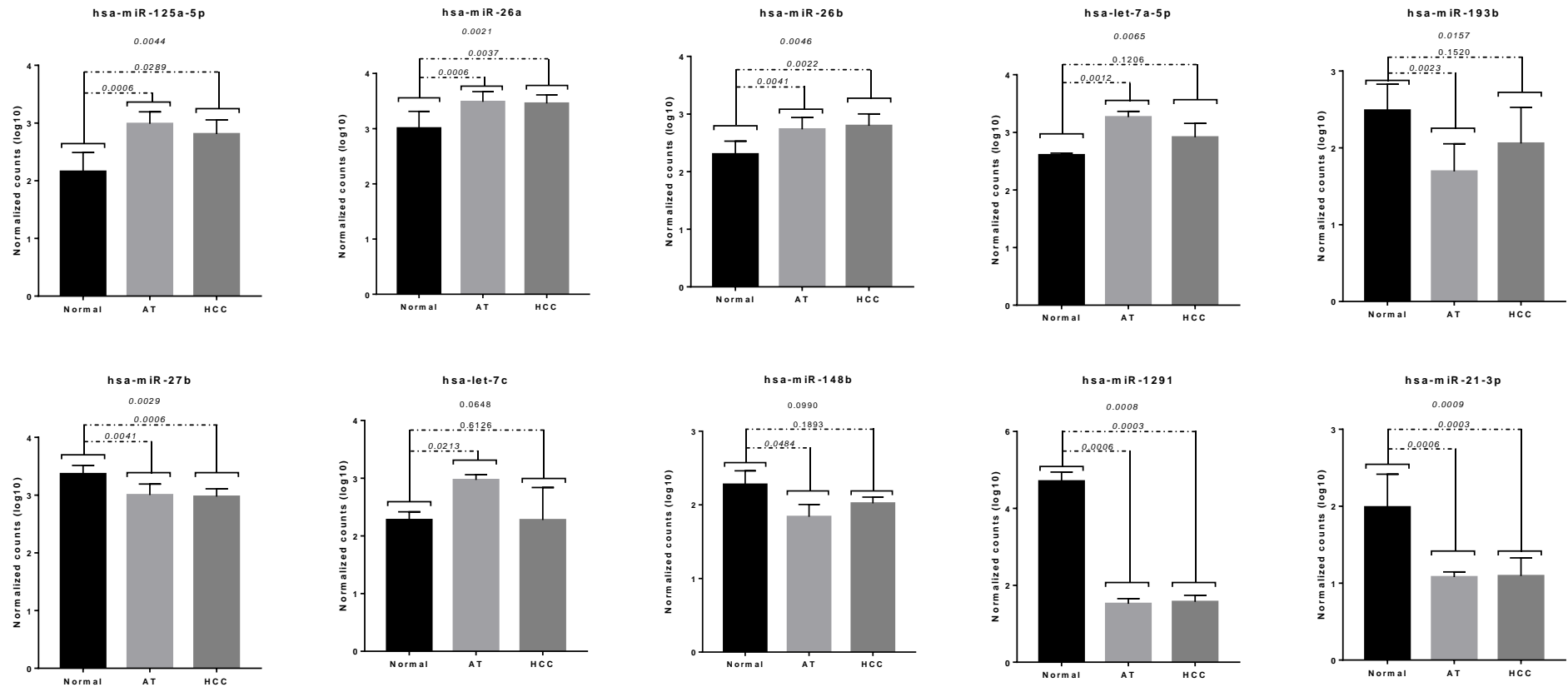
Imagem representativa dos miRNAs expressos de modo diferencial entre o CHC vs. Tecido Normal e Tecido Adjacente vs Tecido Normal. Evidenciando 14 microRNAs regulados negativamente e 15 regulados positivamente.

**Figura 21** - Mapa de calor evidenciando o miRNA expresso diferencial Top10.

Fonte: O autor

As amostras biológicas foram agrupadas de acordo com o padrão de expressão dos 10 miRNA's mais diferencialmente expressos após exclusão dos miRNA com contagem de reads menor que 10. O agrupamento foi feito por distribuição euclidiana usando StarTools. Os dados são exibidos de acordo com o material biológico: Tecido normal (D), tecido adjacente (AT) ou carcinoma hepatocelular (CHC) e o número correspondente da amostra.

Figura 22 - Contagem de *reads* no tecido do fígado por RNA-Seq.



Fonte: O autor

Na imagem evidenciamos a contagem normalizada absoluta de reads para tecido normal, tecido adjacente (AT) e carcinoma hepatocelular (CHC) visando representar o padrão de expressão de cada miRNA. A diferença entre os grupos foi calculada usando o teste de Kruskal-Wallis para 3 grupos e o teste de Mann-Whitney para 2 grupos.

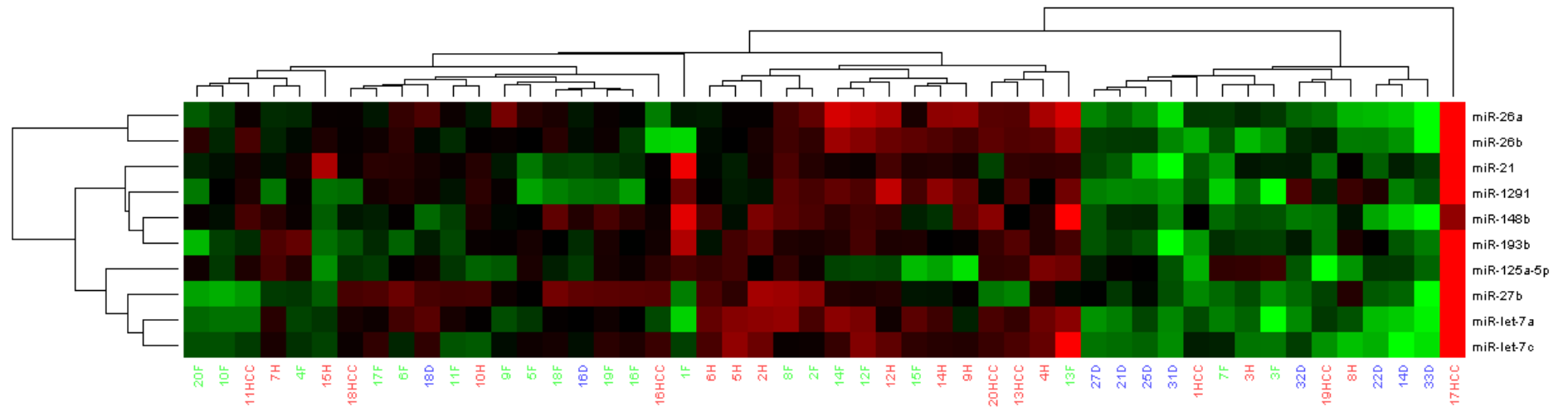
#### 6.4 Validação dos miRNAs candidatos por RT-qPCR

Usamos o Top10 dos microRNAs mais diferencialmente expressos, em ambas as condições (CHC vs tecido saudável e Tecido Adjacente vs Tecido saudável) em RT-qPCR. Pode-se perceber no heatmap que existem 3 subgrupos principais que compõem a amostra (Figura 23). Curiosamente, verificou-se que não existe um agrupamento completo de acordo com os grupos biológicos estudados, porém existem preponderâncias que permitem perceber a formação de pequenos grupos de amostras de CHC, tecidos adjacentes e tecidos saudáveis. A Figura 24 representa o nível de expressão relativa para o microRNA alvo estudado e os valores da taxa de variação são mostrados na tabela 15, notamos uma concordância entre o nível de expressão para a maioria dos alvos estudados em ambas as técnicas, porém alguns miRNAs, como hsa-miR-1291 e hsa-miR-27b não mostraram diferença estatisticamente significativa. Todas as outras análises mostraram resultados significativos.

Diante disso, investimos em 5 microRNAs que foram validados via RT-qPCR: miR-26a, miR-26b, miR-125a-5p, miR-let-7a-5p e miR-let-7c. Em nosso estudo foi possível encontrar também correlação entre tais microRNAs e as variáveis bioquímicas de relevância clínica Tempo de Protrombina (TP) e Aspartato aminotransferase (AST), dados evidenciados na tabela 16, para as demais variáveis não foi encontrado resultado significativo.

Para verificar a interseção entre as fases de descoberta e a fase da validação decidimos analisar a existência de correlação entre os valores de Ct (Cycling Threshold) obtidos na Rt-qPCR e a contagem bruta de reads obtidas no RNA-Seq, tais parâmetros foram analisados por demonstrar os resultados mais brutos para ambas as técnicas. Foram colocados na análise os dados referentes ao CHC, tecido adjacente e tecido saudável dos mesmos pacientes submetidos a ambas as análises. Foi encontrada significância estatística para os miRNAs validados miR-26a/b, miR-let-7a/c e miR-125a-5p (Figura 25). Os demais microRNAs foram testados, porém sem significância estatística.

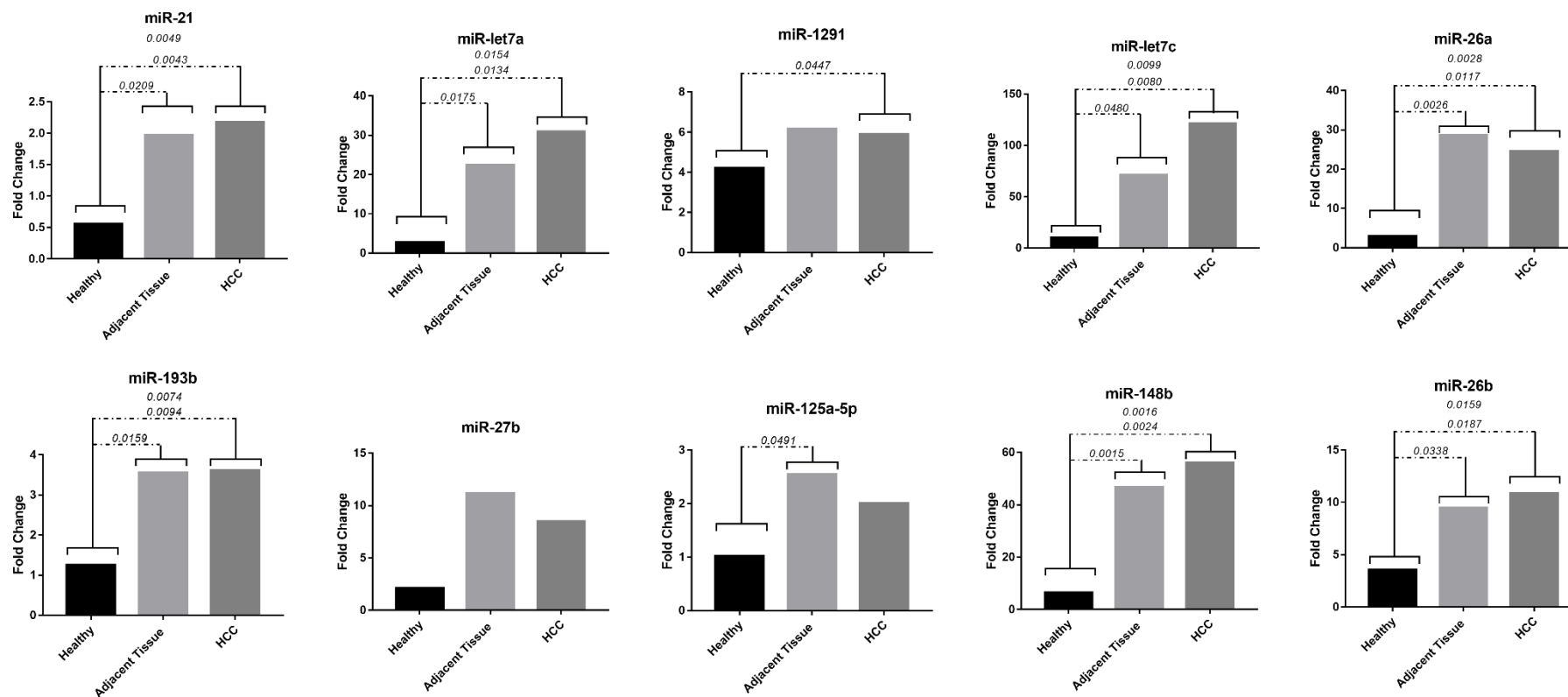
**Figura 23** - Heatmap evidenciando o miRNA expresso diferencial Top10 por RT-qPCR.



Fonte: O autor

A análise de agrupamento de miRNAs regulados negativamente (verde) e regulados positivamente (vermelho) identificados em tecido de carcinoma hepatocelular, tecido hepático adjacente e tecido hepático saudável. Os dados são exibidos de acordo com o material biológico: Tecido normal (D), tecido adjacente (AT) ou carcinoma hepatocelular (CHC) e o número correspondente da amostra.

**Figura 24 -** Expressão relativa de microRNAs validados por RT-qPCR.



Fonte: O autor

Validação de dados utilizando usando a reação em cadeia da polimerase de transcrição reversa quantitativa (RT-PCR) com detecção via sistema Taqman para os principais miRNAs diferencialmente expressos na Fase de descoberta via RNA-Seq, desta forma, podemos ver perfis de miRNA diferenciados entre o carcinoma hepatocelular de tecidos tumorais adjacentes e tecido hepático saudável. A diferença entre os grupos foi calculada usando o teste de Kruskal-Wallis para 3 grupos e o teste de Mann-Whitney para 2 grupos.



**Tabela 15** - MicroRNAs diferencialmente expressos em cada material biológico, comparando a taxa de variação para cada técnica.

	CHC vs Controles				Tecido Adjacente vs Tecido Saudável			
	<u>FC (qPCR)</u>	<u>p value adj</u>	<u>FC (RNA-Seq)</u>	<u>p value adj</u>	<u>FC (qPCR)</u>	<u>p value adj</u>	<u>FC (RNA-Seq)</u>	<u>p value adj</u>
hsa-miR-21-3p	42.82	<b>0.0043</b>	0.14	<b>0.00016</b>	27.52	<b>0.0209</b>	0.085	<b>1,35E+07</b>
hsa-miR-1291	7.043	<b>0.0447</b>	0.001	<b>9,26E-29</b>	6.268	0.3171	0.001	<b>4,33E-31</b>
hsa-miR-let-7c	757.6	<b>0.0080</b>	1.082	<b>0.425871</b>	646.0	<b>0.0480</b>	4.028	<b>0.030325</b>
hsa-miR-26a	78.0	<b>0.0117</b>	2.345	<b>0.00563</b>	73.0	<b>0.0026</b>	2.084	<b>5,84E+09</b>
hsa-miR-193b	1581.0	<b>0.0094</b>	0.486	0.397408	255.9	<b>0.0159</b>	0.174	<b>0.003077</b>
hsa-miR-26b	52.89	<b>0.0187</b>	3.191	<b>0.0120</b>	25.56	<b>0.0338</b>	0.452	<b>0.00812</b>
hsa-miR-27b	51.17	0.1155	0.410	<b>0.00563</b>	95.53	0.4722	0.452	<b>0.00812</b>
hsa-miR-125a-5p	8.823	0.5014	3.656	<b>0.03080</b>	3.997	<b>0.0491</b>	5.592	<b>0.00016</b>
hsa-miR-148b	84.00	<b>0.0037</b>	0.579	0.066016	161.7	<b>0.0023</b>	0.425	<b>0.048725</b>
hsa-miR-let-7a	96.17	<b>0.0134</b>	2.187	0.397408	59.2	<b>0.0175</b>	0.085	<b>0.001934</b>

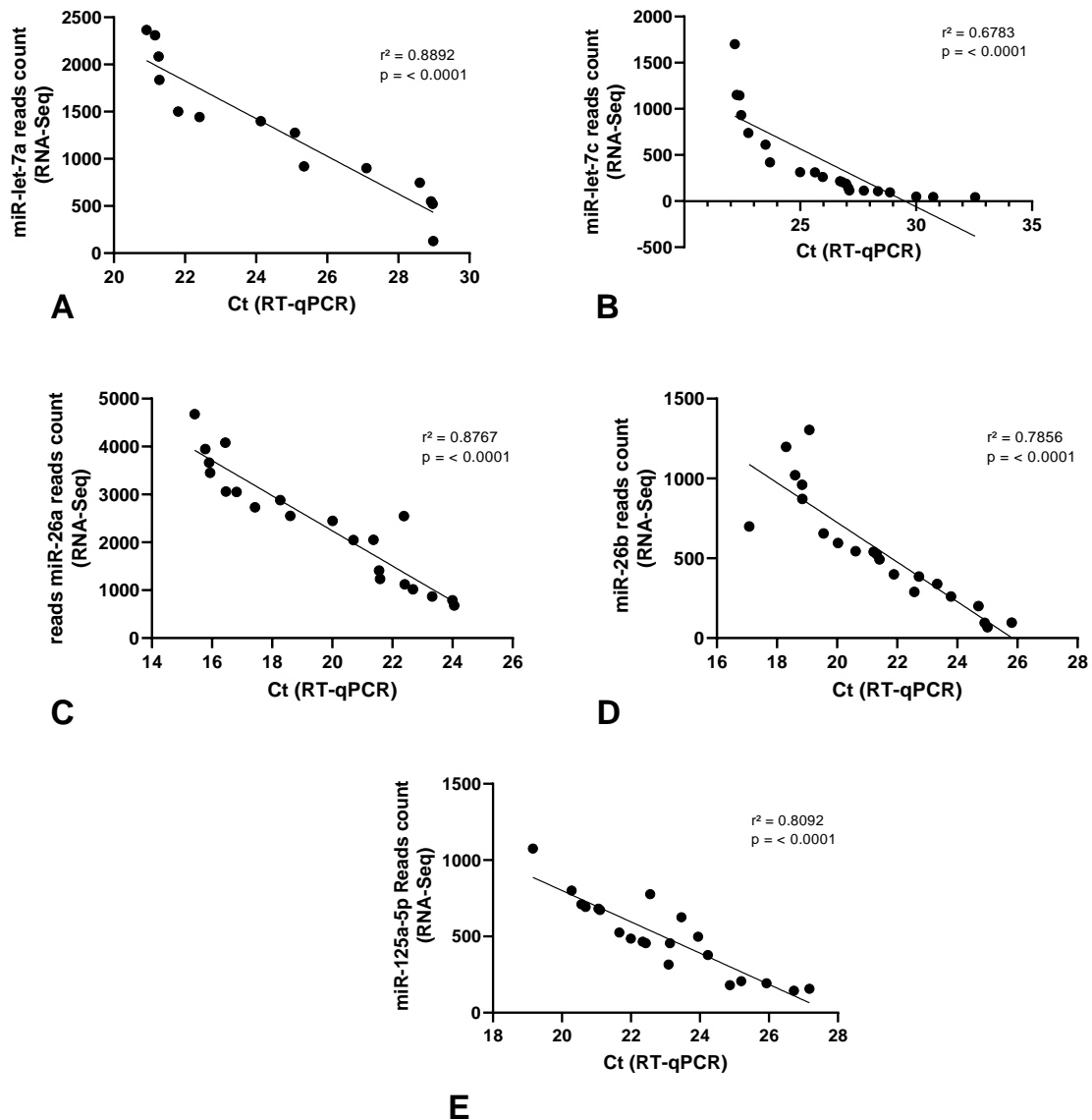
Fonte: O autor

Comparativo entre os valores de Fold Change (FC) advindo das técnicas utilizadas na fase de descoberta e na fase de validação.

**Tabela 16** - Correlação entre os microRNAs validados e variáveis bioquímicas.

	Tempo de protrombina		Aspartato amino-transferase	
	$r^2$	$p(\text{value})$	$r^2$	$p(\text{value})$
hsa-miR-26a	0.6768	<b>&lt;0.0001</b>	0.3381	<b>0.0144</b>
hsa-miR-26b	0.6950	<b>&lt;0.0001</b>	0.4507	<b>0.0032</b>
hsa-let-7a	0.5426	<b>0.0007</b>	0.2944	<b>0.0244</b>
hsa-let-7c	0.6674	<b>&lt;0.0001</b>	0.4469	<b>0.0034</b>
hsa-miR-125a-5p	0.5967	<b>0.0003</b>	0.3941	<b>0.0074</b>

Fonte: O autor

**Figura 25** - Correlação entre o número de reads e o Cycling Treshold.

Fonte: O autor

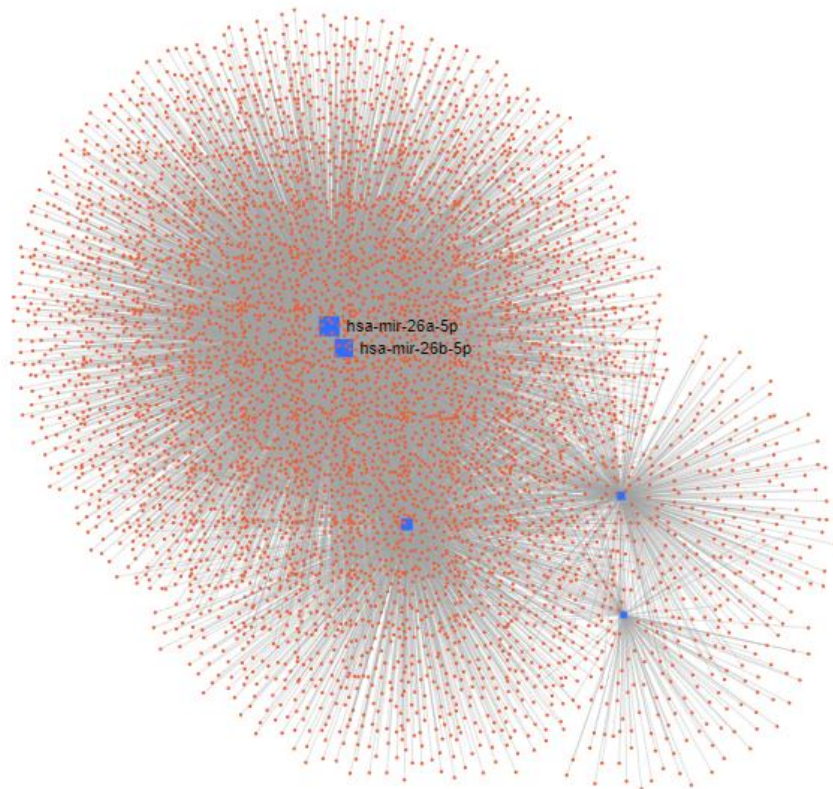
A imagem demonstra que existe uma forte correlação entre os dados obtidos no RNA-Seq, utilizado na fase de descoberta, e os valores de Ct (Cycling Treshold) obtidos na RT-qPCR, durante a fase de

validação. Pode-se perceber que de maneira geral existe uma correlação oposta entre as variáveis. Foi utilizado a Correlação de Pearson.

## 6.5 Determinação funcional via análise de enriquecimento

Essas análises foram feitas a fim de descobrir quais rotas biológicas estão sendo influenciadas pela expressão diferencial dos microRNAs observados anteriormente. Inicialmente, uma rede de interação miRNA-mRNA foi construída, por predição computacional de complementaridade, entre as sequências dos microRNAs e as regiões 3' e 5'-UTR de RNA's mensageiros, 7693 genes foram preditos regulados pelos 5 miRNAs diferenciais, validados por RT-qPCR (Figura 26). As principais vias biológicas influenciadas foram relacionadas ao evento do ciclo celular, bem como, as vias regulatórias da divisão celular (Figura 27).

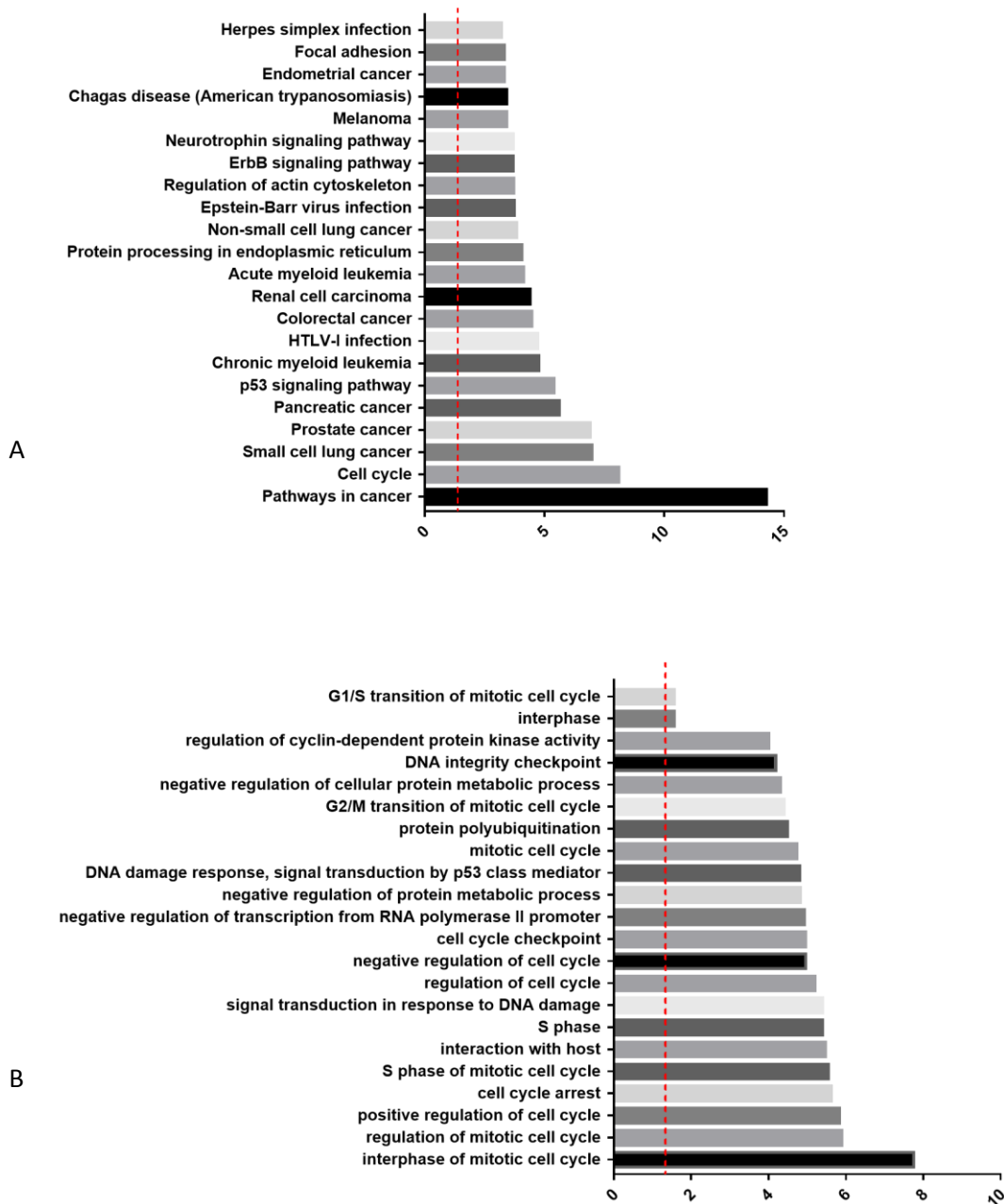
**Figura 26** - Rede de interação mRNA-miRNA de miRNAs diferenciais validados por RT-qPCR.



Fonte: O autor

Os miRNAs (quadrados azuis), que foram expressos diferencialmente e os mRNAs preditos para ocorrência da interação, são indicados com círculos laranja. A rede de interação é composta por mRNA 7963 alvos. A rede foi construída utilizando a ferramenta miRNet ([www.mirnet.ca](http://www.mirnet.ca)), sob parâmetros padrão de predição, cutoff = 1,0 e grau de proximidade não limitado.

**Figura 27** - Análise de enriquecimento funcional de mRNAs previstos regulados pelos 5 miRNAs validados em RT-qPCR.



Fonte: O autor

Os eventos mais significativos no processo biológico enriquecidos pela análise GO (A). As principais rotas são enriquecidas significativamente pela análise da rota KEGG (B). GO – Gene Ontology; KEGG - Kyoto Encyclopedia of Genes and Genomes. A linha tracejada vermelha indica o limite de significância.

## 7 DISCUSSÃO

No presente trabalho, combinamos estudos de associação de polimorfismos de base única, desenvolvimento de algoritmos computacionais de classificação e de miRNA e interação de vias miRNA-mRNA, com o objetivo de melhorar nossa compreensão acerca mecanismos moleculares envolvidos no processo de hepatocarcinogênese e ainda possibilitar a descoberta de biomarcadores associados a progressão da doença hepática e, conseqüentemente, o surgimento do carcinoma hepatocelular.

Os nossos achados em relação a caracterização clínica e laboratorial, correspondem ao que vem sendo reportado na literatura. Estudos de Sousa et. al., (2016) e Karagoz et. al., (2016) mostram que os níveis mais elevados de transaminases hepáticas são encontrados em pacientes mais crônicos, devido a um maior estabelecimento da doença o que ocasiona demasiadas lesões nas células do ambiente hepático, resultando em perda de função do órgão e exacerbação do processo fibrótico de reparo tecidual fisiológico.

Em relação a contagem da taxa plaquetária nossa pesquisa corrobora com o estudo de Karagoz et. al., (2016), o qual também observou redução deste componente hematológico em pacientes com cirrose e câncer de fígado. No entanto, no estudo de Gîrleanu, et. al., (2016), apesar de também relatarem trombocitopenia em pacientes crônicos, constatam que o volume plaquetário médio (VPM) encontra-se aumentado em pacientes com fibrose hepática avançada, afirmando que em estágios elevados de fibrose a diminuição da taxa de trombócitos é um evento multifatorial atribuído ao sequestro de plaquetas pelo baço, destruição plaquetária, supressão da medula óssea e diminuição de trombopoetina, que regula a maturação dos megacariócitos, por conseguinte, a produção plaquetária.

Deste modo, foi possível conceber uma diferenciação do quadro imunológico presente nos pacientes quando, os mesmos, apresentam fibrose leve/moderada, fibrose intensa e CHC. No caso, dos pacientes em estágio grave, sabe-se que estes apresentam um infiltrado linfocitário, células de Kupffer ativadas, células estreladas hepáticas ativadas e deposição de proteínas de matriz extracelular (PELLICORO et. al. 2014). Neste contexto, tal estado diferencial pode ser resultante de pontos de variação genética, que de algum modo podem estar associados ao agravamento da doença hepática.

Como o genótipo A/A PTX3 (tanto para o rs1840680 quanto para o rs2305619) conduz o CHC a sua ocorrência não é clara. Encontramos uma frequência maior desse genótipo no grupo CHC, assim como, valores de associação de risco para o modelo recessivo. Alguns estudos mostraram associação deste genótipo com níveis plasmáticos mais elevados em pacientes com CHC (CARMO, RF. 2016) e a literatura sugere que a proteína pentraxina 3, codificada pelo gene *PTX3* tem um importante papel como mediador inflamatório envolvendo a ocorrência de tumores (GÎRLEANU et. al., 2016; CHOI, LEE & PARK, 2015). Portanto, pacientes infectados pelo *VHC* com genótipo A/A podem ativar mediadores mais pró-inflamatórios do que aqueles com G/A ou Genótipos G/G e, conseqüentemente, aumentar o risco de desenvolver CHC.

Assim como para a pentraxina 3, a lectina ligadora de manose (MBL) é uma proteína de reconhecimento do sistema imune inato, atuando como receptor de reconhecimento de padrões, os quais estão associados aos patógenos. O polimorfismo -221 na região promotora do gene *MBL2* está associado aos níveis plasmáticos da proteína, onde o genótipo X/X está associado a um quantitativo menor de proteína circulante. Estudos anteriores encontraram também associação entre esta variação genética e o risco do agravamento da Hepatite C, por conta disso, acreditamos que a MBL emite um papel importante no estabelecimento da doença hepática, com um quantitativo menor da proteína resulta em uma menor ativação do sistema complemento, e por conseguinte, favorecendo a replicação viral e o estabelecimento da doença (HALLA et. al, 2010; STALLONE et al, 2014)

Sabe-se que o CHC compõe um grupo de tumores sólidos que necessitam que um microambiente inflamatório ideal para o seu aparecimento, deste modo moléculas que propiciem inflamação em modo exacerbado podem contribuir para o seu surgimento, como é o caso do fator de necrose tumoral- $\alpha$  (TNF- $\alpha$ ), que induz inflamação crônica e perfil de atuação de linfócitos T CD4+ de perfil Th1, conduzindo assim eventos inflamatórios com tendência a cronicidade EURICH et. al, 2011. Aroucha, et al, 2016 e Tharwat et al. 2018 também encontraram associação entre o mesmo ponto de variação genética e a progressão da fibrose hepática e, ainda, com o surgimento do carcinoma hepatocelular em população brasileira e egípcia, respectivamente.

O sistema imune mune-se de diversos fatores para combater uma infecção viral, além dos supracitados, é sabido que outro mecanismo eficaz para a eliminação

de patógenos é a formação da condição biológica conhecida como stress oxidativo, caracterizado por perturbações equilíbrio redox intracelular que podem provocar a produção de peróxidos e radicais livres que danificam todos os componentes celulares, incluindo proteínas, lipídios e o DNA. Uma proteína muito importante nesta condição chama-se mieloperoxidase (MPO), codificada por gene de mesmo nome. Em nosso estudo encontramos uma associação entre o genótipo G/G do rs2333227 com as formas avançadas da Hepatite C (incluindo o CHC), em modelo alélico e em outras combinações de modelos genotípicos.

Nahon *et al.*, 2012 verificou a mesma variante genética e encontrou associação do genótipo G/G com níveis mais elevados de MPO circulante, portanto, propomos que indivíduos G/G podem responder ao estímulo patogênico de modo exacerbado, na condição do stress oxidativo, gerando assim o desequilíbrio entre as espécies reativas de oxigênio e sua remoção por enzimas específicas gerando, assim, um ambiente propício para a ocorrência de eventos de mutação a nível de DNA, possibilitando o aparecimento do CHC.

Estudos que buscam a associação de variações pontuais são bastante aceitos e bem quistos pela comunidade científica, e bastante realizados quando visamos buscar um fator causal para um desfecho específico, assim deste modo possibilitando incriminar tal variação como biomarcador, no entanto, do ponto de vista biológico sabe-se que todas as moléculas estão sujeitas a uma ampla gama de interações, onde uma condição biológica passa a ser o resultado de tais interações. Buscando ampliar nossa análise, construímos um classificador, utilizando aprendizado de máquina, baseado exclusivamente em dados genéticos, no caso os 10 SNPs supracitados.

Técnicas de classificação já foram aplicadas ao problema de prever se um indivíduo possui uma doença. Austin *et al.*, 2013 aplicou Support Vector Machine (SVM), Random Forest (RF) e regressão logística para a predição de característica relacionada à insuficiência cardíaca. LG; AT 2013 aplicou Decision Tree (DT), SVM e Multi Layer Perceptron (MLP) na predição de reincidência de câncer de mama. Técnicas de otimização podem ser utilizadas em conjunto com técnicas de classificação, melhorando suas métricas de desempenho. Wang *et al.* 2005 aplicou as técnicas de cuckoo optimization algorithm e harmony search para construir um método de seleção de genes para o problema de classificação de câncer utilizando SVM.

Algoritmos genéticos são uma família de modelos computacionais inspirados na teoria da evolução biológica. Uma das aplicações de algoritmos genéticos é a seleção de características, para problemas de classificação. A classificação de padrões e problemas de descoberta de conhecimento requerem um subconjunto de características para representar os padrões a serem classificados.

Estudos recentes utilizaram técnicas de aprendizagem de máquina aplicadas ao problema do câncer de fígado. Sato, 2018 aplicou os classificadores: SVM, Logistic Regression, Support Vector Machine (SVM), Gradient Boosting (GB) Random Forest (RF), Artificial Neural Network e Deep Learning à classificação entre indivíduos quanto à presença CHC. Ksia, *et al.* 2019 utilizou um 2level genetic optimizer para otimizar os parâmetros dos classificadores e realizar seleção de características com as entradas. Alguns dos classificadores usados foram: SVM, MLP, K-nearest neighbors (KNN), classificador de Naive Bayes, Logistic Regression e Random Forest.

Compreender a influência da carga genética no CHC é de fundamental importância para descobrir de que forma alguns genes atuam na formação e na evolução tumoral, assim como para buscar novas formas de prevenção mais precisas e diretas.

Atualmente sabe-se que devido a sua estabilidade em fluidos corporais e sua variação decorrente de eventos biológicos definidos, os microRNAs vem sendo apontados como potenciais moléculas úteis para a refletir um estado clínico de indivíduos submetidos às mais diversas condições clínicas (YOUNOSSI *et al.*, 2018). Nas últimas décadas, com o desenvolvimento de métodos de sequenciamento de alta produtividade, bioquímica e biologia computacional, um grande número de miRNAs tiveram suas funções reveladas em vários tecidos e células. Um estudo recente revelou os importantes papéis dos miRNAs em uma miríade de doenças humanas, incluindo tumores malignos (YUAN; LU; VAN DEN BERG, 2018). Além disso, devido à alta especificidade celular e tecidual dos miRNAs, seus papéis em diferentes neoplasias não estão completamente elucidadas. No CHC, um número crescente de miRNAs, como miR-21 (ZHANG P. *et al.*, 2018), miR-182 (ZHANG; FRIEDMAN, 2012) miR-197-3p (ZHU; JIA; MISRA, 2012) e miR-139 (ZUCMAN-ROSSI *et al.*, 2015), tem exercido papéis importantes na regulação do processo fisiopatológico e na orientação



do diagnóstico clínico e tratamento. No entanto, ainda há muitos miRNAs que precisam ser desvendados.

A identificação e validação de novos biomarcadores, como assinaturas de miRNA, têm sido um foco de pesquisa na pesquisa do câncer. No presente estudo, identificamos uma assinatura de miRNA específica de tumor que consiste em cinco miRNAs (has-miR-26a, has-miR-125a-5p, has-let-7a-5p, has-miR-26b, has-let-7c) que foram significativamente associados à ocorrência e desenvolvimento em pacientes com CHC. Na fase de validação, avaliamos 20 CHC, 20 tecidos adjacentes e 10 fígados saudáveis como controle, a direção não era consistente com os dados da fase de descoberta. Na fase de descoberta, usamos apenas CHC causado por infecção por *VHC*. Em contrato, na fase de validação, avaliamos várias etiologias diferentes.

A falha na replicação dos resultados da fase de descoberta também pode ser em parte devido aos diferentes métodos de análise nas fases de descoberta e validação, como ocorreu com outros estudos. Na fase de descoberta usamos o sequenciamento de nova geração, então, o número de reads de miRNAs individuais foi usado para quantificação. No entanto, não há consenso sobre a aplicação de métodos de quantificação para resultados de sequenciamento de microRNA's com métodos desenvolvidos para lidar de forma adequada com a variação esperada proveniente de RNA-Seq experimental. Em estudos comparativos anteriores de técnicas de expressão diferencial de RNA-Seq, DESeq, o método de análise aplicado em nosso estudo, mostrou-se mais conservador em comparação com outros pacotes e é recomendado se o número de falso-positivos for uma grande preocupação. Além disso, os resultados da expressão do transcriptoma são geralmente estimativas justas para alvos que são expressos abundantemente em um determinado tecido. Em nosso estudo, observamos que, para transcritos raros, as estimativas de expressão fornecidas pelo sequenciamento de RNA não se correlacionam com qRT-PCR.

Em contrapartida, para a fase de validação onde utilizamos ensaios TaqMan individuais, foi necessário padronizar os dados para um controle, em nosso caso usamos RNU48 porque tem alta abundância e baixa variabilidade entre os tecidos normais. No entanto, não está claro se o U48 é um controle interno adequado para identificar com precisão miRNAs especificamente no tecido hepático, além disso, o uso de primers personalizados parece ser importante para reduzir a divergência entre

as técnicas uma vez, quando os primers são projetados a partir de junções que participam de splicing alternativo resulta em correlação não ideais. E é preferível o design primer ter como modelo não apenas a anotação de referência, porque é possível que os primers comerciais possam atingir variantes de emenda para causar tal discrepância.

Nossos resultados mostraram que ao comparar a área central do tumor com o tecido circundante, uma variação biológica não é perceptível nem na fase de descoberta ou validação, ambas as amostras de casos foram agrupadas no mesmo cluster. Tal semelhança entre esses pontos no fígado pode ser explicada pela necessidade de criar um microambiente inflamatório para aparecimento e estabelecimento do Carcinoma Hepatocelular, por isso o microambiente hepático deve ser totalmente afetado durante a ocorrência de tumor hepático (RINGELHAN *et al.*, 2018). Além disso, o fato de a hepatocarcinogênese alterar a maioria das vias imunometabólicas hepáticas em favor do estabelecimento de CHC contribui para a indiferenciação dessas amostras (LI *et al.*, 2018; LIN *et al.*, 2016).

Pode-se perceber que 5 miRNA's elevados em tecidos tumorais atuam sincronicamente principalmente em vias metabólicas relacionadas ao ciclo celular de forma integral ou ao longo de seus estágios, como interfase e transição entre as fases G1 e S, bem como, em vias moleculares associadas à suscetibilidade a ocorrer divisão celular, como modificações nas histonas. Recentemente, o miR-125a-5p foi estudado para hepatopatia, porém ainda não há consenso sobre a real função, uma vez que aparentemente a etiologia da hepatopatia crônica modula e é singularmente modulada em cada situação. Li *et al.*, 2019 mostraram que o nível de miR-125a-5p foi drasticamente regulado para baixo em tecidos de VHB-CHC e linhagens celulares como HepG2.2.15 e HepG3X em comparação com seu controle, sugerindo que o baixo nível de miR-125a-5p confere uma proliferação inibida e apoptose induzida. Em relação aos nossos resultados, com base nessa premissa, podemos supor que o alto nível de miR-125a-5p pode atuar na indução da proliferação celular e resistência à apoptose, duas características básicas para o estabelecimento do câncer.

Da mesma forma, miRNA-26a/b também estão ligados à manutenção do ciclo celular. Nossos resultados mostram um nível aumentado de expressão de ambos miRNA-26a e miRNA-26b. Por outro lado, Zhu *et al.*, 2012 mostraram em cultura de células de CHC que os baixos níveis de MicroRNA-26a/b, cooperam para inibir a

transição G1 / S pela ativação da proteína PRb, uma vez que baixos níveis de miRNA-26a/b suprimiu diretamente a expressão de CDK6 e ciclina E1. Em estudos anteriores, miR-26a também foi associado a conter eventos de metástases por meio da supressão da expressão de IL6 no CHC (Yang *et al.*, 2013). Também foi relatado que miR-26a/b induz a apoptose celular por direcionamento de MTDH, EZH2 e SLC7A1 em células de câncer de mama e sinalização de NF- $\kappa$ B e aumenta a quimiossensibilidade de células de carcinoma hepatocelular por direcionamento de TAK1 e TAB3 (ZHAO *et al.*, 2014)

Em todos os estudos acima mencionados, onde o microRNA-26a/b foi regulado negativamente, foram realizados em culturas de células. Nossos resultados sugerem um aumento na expressão desses microRNA no tecido CHC e tecido tumoral adjacente. Dessa forma, podemos projetar que as vias citadas anteriormente podem ser ativadas e, assim, contribuir para o estabelecimento da hepatocarcinogênese. Indo de acordo com o que foi proposto por (WANG *et al.*, 2016) que níveis elevados de miRNA-26b estão associados à regulação positiva da proliferação, angiogênese induzida e resistência à apoptose no carcinoma hepatocelular.

Estudos anteriores relataram que let-7 é amplamente visto como um miRNA supressor de tumor (ALY *et al.*, 2020; BALZEAU *et al.*, 2017; BÜSSING *et al.*, 2008). Além disso, também foi demonstrado que os membros da família let-7 são regulados positivamente em certos tipos de câncer. Por exemplo, Lawrie *et al.*, 2009 mostraram que let-7b e let-7i foram regulados positivamente e associados à transformação de alto grau no linfoma, enquanto Lu *et al.*, 2007 descobriram que let-7a-3 estava elevado no câncer epitelial de ovário. Esses dados indicam que a expressão de membros da família de miRNA let-7 difere entre diferentes tecidos e pode ser regulada por muitos fatores.

Nossos resultados mostraram aumento da expressão de dois membros da família let-7 (let-7a e let-7c) em tecido neoplásico. Roush e Slack, 2008 revelaram que os níveis do miRNA let-7c miRNA foram significativamente mais baixos em tecidos de CHC do que em tecidos tumorais adjacentes normais correspondentes, fato reproduzido em nossas amostras também, mas diferentemente quando comparamos com tecido hepático normal, revelamos uma superexpressão de let -7c. Zhu *et al.*, 2015 indica que a superexpressão de let-7c diminuiu os níveis de proteína CyclinD1, CDK6, pRb e E2F2, desta forma, let-7c suprime a progressão de CHC, possivelmente

por direcionar diretamente o regulador do ciclo celular CDC25A e afetar indiretamente suas moléculas alvo a jusante bem como, foi proposto por Liu *et al.*, 2017 que miR-199a-5p e let-7c inibem de forma cooperativa a migração e a invasão ao direcionar MAP4K3 em carcinoma hepatocelular.

O uso terapêutico desse microRNA vem sendo estudado há algum tempo (GILLES E SLACK, 2018). Possivelmente, o alto nível de expressão no tecido adjacente pode ser um reflexo de ser uma zona marginal que está passando por um evento constante de modificação/diferenciação celular. Porém, a menor taxa de expressão do CHC pode indicar uma modulação e/ou controle imunológico a fim de se obter um equilíbrio imune para o estabelecimento crônico da doença ou um mecanismo de defesa do hospedeiro para lidar com essa condição patológica (RINGELHAN *et al.*, 2018).

Atualmente, os genes alvo do let-7a estão envolvidos no crescimento, diferenciação, renovação e tumorigênese celular, incluindo K-RAS, c-myc, CCR7 e HMGA1 (BALZEAU *et al.*, 2017; BÜSSING *et al.*, 2008). Aly *et al.*, 2020 colocaram a expressão diferencial de miRNA let-7a como um potencial biomarcador sérico para CHC induzido por VHC, onde a diminuição dos níveis séricos de let-7a ocorre em pacientes com CHC estabelecido, e sua expressão seria maior em casos de cirrose hepática, resultados semelhantes aos encontrados por Gramantieri *et al.*, 2007, que mostram que let-7a foi regulado negativamente em tecidos de CHC em comparação com tecido não tumoral. Da mesma forma, Shi *et al.*, 2017, em seu estudo com tecidos de CHC, descobriram que os níveis de expressão de let-7 foram menores em tecidos de CHC quando comparados com tecido não tumoral e o nível de expressão foi maior em tecidos de cirrose hepática. Nossos resultados mostram uma diminuição na expressão de let-7a em tecidos de CHC, quando comparados ao tecido adjacente ao tumor, porém maior que o tecido hepático normal, mas sem significância estatística.

A hepatocarcinogênese é um processo complexo associado ao acúmulo de alterações genéticas e epigenéticas que ocorrem durante o início, promoção e progressão da doença (RINGELHAN *et al.*, 2018). As etapas iniciais envolvem a interrupção de um conjunto de vias interdependentes que controlam a homeostase entre o crescimento celular e a apoptose. Em estágios posteriores, as células podem adquirir propriedades angiogênicas, invasivas e metastáticas (BATALLER E BRENNER, 2005; TANA E MUIR, 2018). Como os genes que codificam proteínas,

alguns microRNAs também funcionam como oncogenes ou supressores de tumor no hepatoma humano. Vários estudos relataram as ligações funcionais moleculares entre a expressão anormal de miRNA e as marcas da transformação maligna: crescimento celular aberrante, morte celular, diferenciação, angiogênese, invasão e metástase (INOKAWA *et al.*, 2016; LIN *et al.*, 2016; XU *et al.*, 2016; YAO *et al.*, 2017). Notavelmente, alguns miRNAs podem regular simultaneamente a expressão de vários genes alvo.

Em suma, a diferenciação entre o padrão de expressão dos microRNA's nos próprios tecidos neoplásicos ou diretamente influenciados pelo tumor (borda ou tecido adjacente ao câncer), quando comparado ao tecido hepático saudável. Este trabalho traz de forma inédita a abordagem metodológica aplicada ao tecido humano *ex vivo* de CHC. Acreditamos retratar com mais proximidade a condição clínica do paciente, visto que a maioria dos estudos são feitos em cultura de células, o que pode enviesar os resultados, uma vez que o fator etiológico é capaz de modular o aparecimento e estabelecimento da doença hepática crônica, assim como do câncer primário de fígado células. No entanto, estudos com coorte maior de amostras são indicados para validar os resultados.

## 8 CONCLUSÃO

Em síntese, os resultados desta pesquisa possibilitam afirmar que:

- Os genes *PTX3*, *MBL2*, *TNF $\alpha$*  e *MPO* podem desempenhar funções significativas na evolução do processo fibrótico em pacientes acometidos pela Hepatite C.
- A construção de um algoritmo computacional com base em dados genéticos, o qual apresenta uma ferramenta promissora para a classificação de pacientes em diferentes grupos de risco.
- Apesar dos experimentos abordarem distintos subgrupos da base de dados original, algumas características mostraram-se recorrentes, como é o caso dos SNPs Pentraxin 3 rs1840680 e Interleukin-10 -592 que aparecem três vezes cada como os melhores SNPs entre os experimentos analisados, evidenciando sua relevância para a evolução da doença hepática;
- No que diz respeito aos classificadores Random Forests e SVM, ambos apresentaram performances parecidas. Apesar do SVM estar presente em três dos cinco melhores experimentos, não é seguro afirmar que este classificador é o melhor, porém se apresenta como o mais promissor;
- A seleção de características de modo geral aumentou o valor da métrica AUC\ROC dos classificadores, porém em contrapartida diminuiu a métrica de acurácia. Pode-se concluir que o One Class SVM é o melhor classificador, utilizando a abordagem Leave One Out, foi obtida uma Acurácia de 0,754 e AUC\ROC de 0,688 com a configuração do conjunto de dados completa.
- Os microRNA's 26a/b, let-7a/c e 125a-5p se mostraram importantes no processo de estabelecimento e surgimento do CHC, sofrendo modulações durante este processo. Porém permanecendo detectável no decorrer do estabelecimento mais agressivo da doença hepática. Sendo assim, promissores biomarcadores para prognóstico.

## REFERÊNCIAS

ALMEIDA-CARVALHO, M. *et al.* Practical Considerations of Real Life of Hepatocellular Carcinoma in a Tertiary Center of Brazil. **Annals of hepatology**, México, v. 16, p. 255-262, 2017.

ANSTEE, Q *et al.* From NASH to CHC: current concepts and future challenges. **Nature Reviews Gastroenterology & Hepatology**, Londres, v. 16, p. 411-428, 2019.

ARAÚJO, N. M., *et al.* High proportion of subgroup A' (genotype A) among Brazilian isolates of Hepatitis B virus. **Archives of virology**, Nova York v. 149, n. 7, p. 1383-1395, 2004.

AROUCHA DC, et. Al. TNF- $\alpha$  and IL-10 polymorphisms increase the risk to hepatocellular carcinoma in VHC infected individuals. **Journal of Medical Virology**. Nova York, v. 88, p. 1587-1595. 2016.

ASSOCIAÇÃO BRASILEIRA DE TRANSPLANTES DE ÓRGÃOS, Dimensionamento dos transplantes no Brasil e em cada estado. **Registro Brasileiro de Transplantes**, São Paulo, v. 4, p. 1-88, 2019.

AUSTIN, P. *et al.* Using methods from the data-mining and machine-learning literature for disease classification and prediction: a case study examining classification of heart failure subtypes. **Journal of Clinical Epidemiology**, Nova York, v.66, n.4, p.398–407, 2013.

BERGER, S. *et al.* The complex language of chromatin regulation during transcription. **Nature**. Londres, v. 447, p. 407–412, 2007.

BONILLA, G. R., ROBERTS, L. The role of hepatitis B virus integrations in the pathogenesis of human hepatocellular carcinoma. **Journal of Hepatology**, Copenhagen, v.42, p.760–77, 2005.

BRASIL. Manual técnico para o diagnóstico das hepatites virais. **Ministério da Saúde**. Brasília, 2018.

CAO, M. *et al.* miR-182-5p promotes hepatocellular carcinoma progression by repressing FOXO3a. **Journal of hematology & oncology**, Londres, v. 2, p.11-19, 2018.

CARRILHO, F. *et al.*, Clinical and epidemiological aspects of hepatocellular carcinoma in Brazil. **Clinics**, São Paulo, v. 65, p.1285-1290, 2010.

CARMO, RF. Genetic variation in PTX3 and plasma levels associated with hepatocellular carcinoma in patients with VHC. **Journal of Viral Hepatitis**, Oxford, v. 23, p. 116–122, 2016.

CHAWLA, N. V. *et al.* SMOTE: synthetic minority over-sampling technique. **Journal of Artificial Intelligence Research**, San Francisco, v.16, p.321–357, 2002.

CHEN, X.; ISHWARAN, H. Random forests for genomic data analysis. **Genomics**, San Diego, v.99, p.323–329, 2012.

CHOI, B, LEE, E, PARK, Y *et al.* Pentraxin-3 silencing suppresses gastric cancer-related inflammation by inhibiting chemotactic migration of macrophages. **Anticancer Research**. Atenas, v. 35, p. 2663–2668, 2015.

CLARK, T., *et al.* Hepatocellular Carcinoma: Review of Epidemiology, Screening, Imaging Diagnosis, Response Assessment and Treatment. **Diagnostic Radiology**, Chicago, v. 44, p. 479 – 486, 2015.

CLARK, T., MCNEELEY, M., MAKI, J. Design and Implementation of Handheld and Desktop Software for the Structured Reporting of Hepatic Masses Using the LI-RADS Schema. **Academic radiology**. Reston, v.21, n.4, p.491-506, 2014.

CRUZ, J., WISHART, D. Applications of Machine Learning in Cancer Prediction and Prognosis. **Cancer Informatics**, Auckland, N.Z, v.2, p.1176-1193, 2006.

DAVI, C. *et al.* Severe Dengue Prognosis Using Human Genome Data and Machine Learning. **IEEE Transactions on Biomedical Engineering**, Nova York, v. 66, p. 2861–2868, 2019.

DAYYEH, B. *et al.* A functional polymorphism in the epidermal growth factor gene is associated with risk for hepatocellular carcinoma. **Gastroenterology**. Baltimore, v. 141, p. 141–149, 2011.



DE LIMA, RE. *et al.*, Two sides of a coin: GG genotype of C7 provides protection against fibrosis severity while showing a higher risk for hepatocellular carcinoma in patients with hepatitis C. **Human Immunology**. Nova York, v. 79, p. 702-707, 2018.

DENG, X. *et al.* Feature selection for text classification: a review. **Multimedia Tools and Applications**, Hingham, MA, v.78, p.3797–3816, 2018.

DHAREL, N. *et al.* MDM2 promoter SNP309 is associated with the risk of hepatocellular carcinoma in patients with chronic hepatitis C. **Clinic Cancer Research**. Secaucus, NJ, v. 12, p. 4867–4871, 2006.

DO CARMO, R. *et al.* Myeloperoxidase gene polymorphism predicts fibrosis severity in women with hepatitis C. **Human Immunology**. Nova York, v. 8, p. 766-70, 2014.

DONATO, F. *et al.* Alcohol and hepatocellular carcinoma: the effect of lifetime intake and hepatitis virus infections in men and women. **American journal of epidemiology**, Baltimore, v. 155, p. 323–331, 2002.

DUREJA, H. *et al.* Topological Models for Prediction of Pharmacokinetic Parameters of Cephalosporins using Random Forest, Decision Tree and Moving Average Analysis. **Scientia Pharmaceutica**, Wien, v.76, p.377–394, 2008.

EL-SERAG, H. Epidemiology of Viral Hepatitis and Hepatocellular Carcinoma. **Gastroenterology**. Baltimore, v. 142, p. 1264–73, 2012.

EL-SERAG, H. Hepatocellular carcinoma. **The New England journal of medicine**. Massachusetts, v. 365, p. 1118–1127, 2011.

EURICH D. *et al.* Association of mannose-binding lectin-2 gene polymorphism with the development of hepatitis C-induced hepatocellular carcinoma. **Liver international**, Oxford, v. 7, p. 1006-1012, 2011.

ESTRADA, J. *et al.* Response to Comment on “Predicting reaction performance in C–N cross-coupling using machine learning”. **Science**, Nova York, v.362, p.8763, 2018.

FAN, J. *et al.* Understanding receiver operating characteristic (ROC) curves. **Canadian Association of Emergency Physicians**, Ottawa, v.8, p.19–20, 2006.

FATTOVICH, G. *et al.* Hepatocellular carcinoma in cirrhosis: incidence and risk factors. **Gastroenterology**, Baltimore, v. 127, p. 35–50, 2004.

FERBER, M., MONTOYA, D., YU, C. Integrations of the hepatitis B virus (HBV) and human papillomavirus (HPV) into the human telomerase reverse transcriptase (hTERT) gene in liver and cervical cancers. **Oncogene**. Basingstoke, v.22, p. 3813–3820, 2003.

FERLAY, J. *et al.* GLOBOCAN 2008 v2.0, Cancer incidence and mortality worldwide: IARC **CancerBase**. Lyon v.4, p. 325-330, 2013.

FERRAZ, A. *et al.* Liver transplantation cost analysis at the Hospital das Clínicas da Universidade Federal de Pernambuco. **Revista Brasileira de Saúde Materno Infantil**. São Paulo, v. 2, p, 313–318, 2002.

FORNER, A. *et al.* Hepatocellular carcinoma. **Lancet**. Londres, v. 379, p. 1245–1255, 2012.

FREEMAN T. *et al.* Recent advances in alcohol-induced adduct formation. **American Medical Society on Alcoholism**, Nova York, v. 29, p.1310-1316, 2005.

GADDIKERI, S., McNEELEY, M., WANG, C. Hepatocellular carcinoma in the noncirrhotic liver. **American journal of roentgenology**. Leesburg, VA, v. 203, p. 34-47, 2014.

GÎRLEANU, I. *et al.* Platelet indices and liver fibrosis evaluation in chronic hepatitis C. **Revista medico-chirurgicală a Societății de Medici și Naturaliști din Iași**, Iasi, v. 120, p. 55-61, 2016.

GLOBAL BURDEN OF DISEASE CANCER. Global, regional, and national cancer incidence, mortality, years of life lost, years lived with disability, and disability-adjusted life-years for 32 cancer groups, 1990 to 2015: a systematic analysis for the global burden of disease study. **American Medical Association, issuing body**, Chicago, v. 3, p. 524–548. 2017.

GOLDSTEIN, B., POLLEY, E., BRIGGS, F. Random Forests for Genetic Association Studies. **Statistical Applications in Genetics and Molecular Biology**, Berlin, v.10, p. 1532-1544, 2011.

GOLUB, T. *et al.* Molecular Classification of Cancer: class discovery and class prediction by gene expression monitoring. **Science**, Nova York, v.286, p.531–537, 1999.

GOMES, M., *et al.* Carcinoma hepatocelular: epidemiologia, biologia, diagnóstico e terapias. **Revista da Associação Médica Brasileira**, São Paulo, v. 59, p. 514-524, 2013.

GREWAL, P., VISWANATHEN, V. Liver cancer and alcohol. **Clinic of Liver Disease**, Philadelphia v.16, p. 839–850, 2012.

GÜNES C, RUDOLPH K. The role of telomeres in stem cells and cancer. **Cell**. Cambridge, v.152, p. 390–398, 2013.

HALLA, M. *et al.*, Association of hepatitis C virus infection and liver fibrosis severity with the variants alleles of MBL2 gene in a Brazilian population. **Human Immunology**, Nova York, v. 71, p. 883-887, 2010.

HALLIDAY, J., KLENERMAN, P., BARNES, E. Vaccination for hepatitis C virus: closing in on an evasive target. **Expert review of vaccines**, Londres, v. 10, p. 659–672, 2011.

HAN, T. *et al.* The Epigenetic Regulation of CHC Metastasis. **International journal of molecular sciences**. Switzerland, MDPI, v. 19, p. 3978, 2018.

HAROUAKA, D. *et al.* Diminished viral replication and compartmentalization of hepatitis C virus in hepatocellular carcinoma tissue. **Proceedings of the National Academy of Sciences of the United States of America**. Washington, DC, v. 113, p.1375–80, 2016.

HERCEG, Z. PALIWAL, A. Epigenetic mechanisms in hepatocellular carcinoma: How environmental factors influence the epigenome. **Mutation research**. Amsterdam, v. 727, p.55–61, 2011.

HERNANDEZ-GEA, V., TOFANIN, S., FRIEDMAN, S. L. & LLOVET, J. M. Role of the microenvironment in the pathogenesis and treatment of hepatocellular carcinoma. **Gastroenterology**, Baltimore, v. 44, p. 512–527, 2013.

HUAN, L. *et al.* Role of microRNAs in inflammation-associated liver cancer. **Cancer biology & medicine**. Tianjin, v. 13, p. 407-425, 2016.

HUANG, S. *et al.* Applications of Support Vector Machine (SVM) Learning in Cancer Genomics. **Cancer genomics & proteomics**, Kapandriti, v.15, p.41–51, 2018.

IOANNOU, G. *et al.* HCC surveillance after SVR in patients with F3/F4 fibrosis. **Journal of hepatology**, Amsterdam, v. 74, p. 458-465, 2021.

KARAGOZ, E. *et al.* Mean platelet volume and red cell distribution width to platelet ratio for predicting the severity of hepatic fibrosis in patients with chronic hepatitis C. **European Journal of Gastroenterology & Hepatology**, Londres, v. 28, p. 744-748, 2016.

KIM, H. *et al.* Hepatitis C Virus Impairs Natural Killer Cell-Mediated Augmentation of Complement Synthesis. **British journal of virology**, Lydney, V. 88, p. 2564–2571, 2014.

KSIAZEK, W. *et al.*, A novel machine learning approach for early detection of hepatocellular carcinoma patients. **Cognitive Systems Research**, Londres, v.54, p.116–127, 2019.

KUDO, M., *et al.* Survival Analysis over 28 Years of Patients with Hepatocellular Carcinoma in Japan. **Liver cancer**, Brisbane, v. 5, p.190–197, 2016.

KULIK, L., EL-SERAG. H., Epidemiology and Management of Hepatocellular Carcinoma. **Gastroenterology**. Baltimore, v. 156, p. 477–491, 2019.

KUMAR, V., ABBAS, A., ASTER, J. Pathologic Basis of Diseases. **Clinic of Liver Disease**, Philadelphia, v. 9. p. 1190-1200, 2014.

LAFARO, K., DEMIRJIAN, A., PAWLIK, T. Epidemiology of Hepatocellular Carcinoma. **Surgical oncology clinics of North America**, Philadelphia, v.24, p. 1-17, 2015.

LAMONTAGNE, R., BAGGA, S., BOUCHARD, M. Hepatitis B virus molecular biology and pathogenesis. **Hepatoma research**, Mumbai, v. 2, p. 163-186, 2016.

LAURET, E. *et al.* HFE gene mutations in alcoholic and virus-related cirrhotic patients

with hepatocellular carcinoma. **The American journal of gastroenterology**, Nova York, v. 97, p. 1016–1020, 2002.

LG, A., AT, E. Using Three Machine Learning Techniques for Predicting Breast Cancer Recurrence. **Journal of Health & Medical Informatics**, Londres, v. 4, p. 124-127, 2013.

LI, M. *et al.* Hepatocellular carcinoma associated with attenuated familial adenomatous polyposis: a case report and review of the literature. **Clinical colorectal cancer**, Dallas, v. 11, p. 77–81, 2012.

LIU, J. *et al.* A Viral Exposure Signature Defines Early Onset of Hepatocellular Carcinoma. **Cell**, Cambridge, v. 182, p. 317-328, 2020.

LIU, L., JIN, G., ZHOU, X. Modeling the relationship of epigenetic modifications to transcription factor binding. **Nucleic acids research**, Oxford, v. 43, p. 3873–3885, 2015.

LONG, X. *et al.* The polymorphism of XRCC3 codon 241 and AFB1-related hepatocellular carcinoma in Guangxi population, China. **Annals of epidemiology**, Nova York, v. 18, p. 572–578, 2008.

LOVET, J. *et al.* Hepatocellular carcinoma. **Nature reviews. Disease primers**, Londres, v. 2, p. 16-18, 2016.

LU, H., WANG, H., YOON, S. A dynamic gradient boosting machine using genetic optimizer for practical breast cancer prognosis. **Expert Systems with Applications**, Nova York, v.116, p.340–350, 2019.

LUO, P. *et al.* Current Status and Perspective Biomarkers in AFP Negative HCC: Towards Screening for and Diagnosing Hepatocellular Carcinoma at an Earlier Stage. **Pathology oncology research**. Budapest, v. 26, p. 599-603, 2020.

MANNS, M. *et al.* Hepatitis C virus infection. **Nature Review. Immunology**, Londres, v. 3, p. 1-19, 2017.

MELLO, F. *et al.* Hepatitis B virus genotypes circulating in Brazil: molecular characterization of genotype F isolates. **BMC microbiology**, Londres, v. 7, p. 103,

2007.

MINAMI, M. *et al.* Hepatitis B virus-related insertional mutagenesis in chronic hepatitis B patients as an early drastic genetic change leading to hepatocarcinogenesis. **Oncogene**. Basingstoke, v.24, p. 4340–4348, 2005.

MITTAG, F. *et al.* Influence of Feature Encoding and Choice of Classifier on Disease Risk Prediction in Genome-Wide Association Studies. **PLoS ONE**, San Francisco, v.10, p. 1-18, 2015.

MORADPOUR, D., PENIN, F., RICE, C. Replication of hepatitis C virus. **Nature Reviews Microbiology**, Londres, v. 5, p. 453–463, 2007.

MORAN-SALVADOR AND MANN, Epigenetics and Liver Fibrosis. **Cellular and Molecular Gastroenterology and Hepatology**, Philadelphia, v. 4, p. 125-164, 2017.

MORGAN, R., BAACK, B., SMITH, B. Eradication of hepatitis C virus infection and the development of hepatocellular carcinoma: a meta-analysis of observational studies. **Annals of internal medicine**. Philadelphia, v. 158, p. 329 – 337, 2013.

NAHON, P. *et al.* Liver iron, HFE gene mutations, and hepatocellular carcinoma occurrence in patients with cirrhosis. **Gastroenterology**, Baltimore, v. 134, p. 102–110, 2008.

NAHON, P. *et al.* A variant in myeloperoxidase promoter hastens the emergence of hepatocellular carcinoma in patients with *HCV*-related cirrhosis. **Journal of hepatology**, Amsterdam, v. 56, p. 426-432, 2012.

NI, J. *et al.* MicroRNA-197-3p acts as a prognostic marker and inhibits cell invasion in hepatocellular carcinoma. **Oncology letters**, Athens, v. 2, p. 2317-2327, 2019.

NJEI, B. *et al.* Emerging trends in hepatocellular carcinoma incidence and mortality. **Hepatology**. Baltimore, v. 61, p. 191–199, 2015.

PARANAGUÁ-VEZOZZO, *et al.* Epidemiology of CHC in Brazil: Incidence and risk factors in a ten-year cohort. **Annals of Hepatology**, México, v. 13, p. 386-393, 2014.

PARISE, *et al.* Nonalcoholic Fatty Liver Disease Brazilian Society of Hepatology Consensus. **Arquivos de gastroenterologia**. São Paulo. v. 53, p. 118-122, 2016.

PEDREGOSA, F. *et al.* Scikit-learn machine learning in Python. **Journal of Machine Learning Research**, Cambridge, v.12, p.2825–2830, 2011.

PELLICORO, A. *et al.* Liver fibrosis and repair: immune regulation of wound healing in a solid organ. **Nature Reviews Immunology**, Londres, v. 14, p. 181–194, 2014.

PORTELA, M. *et al.* O custo do transplante hepático em um hospital universitário do Brasil. **Revista da Associação Médica Brasileira**, São Paulo, v. 56, p. 322-325, 2010.

PROTZER, U., MAINI, M., KNOLLE, P. Living in the liver: hepatic infections. **Nature Reviews Immunology**, Londres, v. 24, p. 201–213, 2012.

RAMOS-LOPEZ, O. *et al.* Genetic, metabolic and environmental factors involved in the development of liver cirrhosis in Mexico. **World journal of gastroenterology**, Pleasanton, v. 21, p. 11552-11566, 2015.

REIG, M. *et al.* BCLC strategy for prognosis prediction and treatment recommendation: The 2022 update. **Journal of Hepatology**, Copenhagen, v. 76, p. 681-693, 2022.

REIS, J. *et al.* Abordagem clínica da Cirrose Hepática: Protocolos de Atuação. **Serviço de Gastroenterologia do Hospital Prof. Dr. Fernando Fonseca**. 1ª Edição, 2018;

RINGERLHAN, M. *et al.* The immunology of hepatocellular carcinoma. **Nature Reviews Immunology**, Londres, v.19, 222-232, 2019.

RUPAIMOOLE, R., SLACK, F. MicroRNA therapeutics: towards a new era for the management of cancer and other diseases. **Nature Reviews Drug Discovery**, Londres, v. 16, p. 203-222, 2017.

SADRI, N. *et al.* microRNAs: Key players in virus-associated hepatocellular Carcinoma. **Journal of cellular physiology**, Nova York, v. 234, p. 12188-12225, 2019.

SATO, M. *et al.* Machine-learning Approach for the Development of a Novel Predictive Model for the Diagnosis of Hepatocellular Carcinoma. **Scientific reports**, Londres, v.9, p. 7704, 2019.

SETSHEDI, M. *et al.* Acetaldehyde adducts in alcoholic liver disease. **Oxidative**

**Medicine and Cellular Longevity**, Austin, v. 3, p. 178-185, 2010.

SHAFRITZ, D. *et al.* Integration of hepatitis B virus DNA into the genome of liver cells in chronic liver disease and hepatocellular carcinoma. Studies in percutaneous liver biopsies and post-mortem tissue specimens. **The New England journal of medicine**, Boston, v. 305, p. 1067–1073, 1981.

SIA, D. *et al.* Liver cancer cell of origin, molecular class, and effects on patient prognosis. **Gastroenterology**, Baltimore, v. 152, p. 745–761, 2017.

SINGH, H. *et al.* The global burden of diagnostic errors in primary care. **BMJ Quality & Safety**, Londres, v.26, p.484–494, 2016.

SOUSA, V. *et al.* Association of Catalase and Glutathione Peroxidase 1 Polymorphisms with Chronic Hepatitis C Outcome, **Annals of Human Genetics**. Oxford, v. 80, p. 145–153, 2016.

STALLONE, G. *et al.* Pentraxin 3: a novel biomarker for predicting progression from prostatic inflammation to prostate cancer. **Cancer research communications**, Philadelphia, v. 74, p. 4230–4238, 2014.

SZABO, G., CSAK, T. Role of microRNAs in NAFLD/NASH. **Digestive diseases and sciences**, Nova York. v. 61, p. 1314–1324, 2016.

SWEILAM, N. *et al.* Support vector machine for diagnosis cancer disease: a comparative study. **Egyptian Informatics Journal**, Amsterdam, v.11, p.81–92, 2010.

TARHUNI, A. *et al.* Impact of cytokine gene variants on the prediction and prognosis of hepatocellular carcinoma in patients with cirrhosis. **Journal of Hepatology**, Copenhagen, v. 61, p. 342–350, 2014.

THARWAT, E. *et al.* Impact of IL-27p28 (rs153109) and TNF- $\alpha$  (rs1800629) Genetic Polymorphisms on the Progression of VHC Infection in Egyptian Patients. **Immunological investigations**, Londres, v. 11, p. 255-267, 2019.

THE INTERNATIONAL CONSENSUS GROUP FOR HEPATOCELLULAR NEOPLASIA. Pathologic diagnosis of early hepatocellular carcinoma: a report of the



international consensus group for hepatocellular neoplasia. **Hepatology**. Baltimore, v.49, p. 658-64, 2009.

THIELE, G. *et al.* Formation and immunological properties of aldehyde-derived protein adducts following alcohol consumption. **Methods in molecular biology**, Totowa, v. 447, p. 235-257, 2008.

TOMMASO, A. *et al.* Risk of CHC: genetic heterogeneity and complex genetics. **Journal of Hepatology**, Copenhagen, v. 52, p. 252–257, 2010.

TUMA, D., CASEY, C. Dangerous byproducts of alcohol breakdown focus on adducts. **Alcohol research & health: the journal of the National Institute on Alcohol Abuse and Alcoholism**, Washington, DC, v. 27, p. 285-290, 2003.

VERDUCE, L. *et al.* The circRNA-microRNA code: emerging implications for cancer diagnosis and treatment. **Molecular Oncology**. New Jersey, v. 13, p. 669-680, 2019.

WANG, J. *et al.*, miR-21 promotes cell migration and invasion of hepatocellular carcinoma by targeting KLF5. **Oncology Letters**, Athens, v. 2, p. 2221-2227, 2019.

WANG, Y. *et al.* Interleukin-1beta gene polymorphisms associated with hepatocellular carcinoma in hepatitis C virus infection. **Hepatology**, Copenhagen, v. 37, p. 65–71, 2003.

WANG, Y. *et al.* Gene selection from microarray data for cancer classification—a machine learning approach. **Computational Biology and Chemistry**, Oxford, v.29, p.37–46, 2005.

WHITLEY, D. *et al.* genetic algorithm tutorial. **Statistics and Computing**, Nova York, v.4, p.65-85, 1994.

WHO-GLOBOCAN. Global Cancer Observatory. International Agency for Research on Cancer. Disponível em: < <https://gco.iarc.fr/>>

WILLSON, J. *et al.* Primary hepatocellular neoplasms in a MODY3 family with a novel HNF1A germline mutation. **Journal of Hepatology**, Copenhagen, v. 59, p. 904–990, 2013.

WILSON, C. *et al.* Epigenetic reprogramming in liver fibrosis and cancer. **Advanced**

**Drug Delivery Reviews**, Amsterdam, v. 121, p. 124–132, 2017.

WONG, C. *et al.* Non-coding RNAs in hepatocellular carcinoma: molecular functions and pathological implications. **Nature Reviews: Gastroenterology e Hepatology**, Londres, v. 15, p. 137-151, 2018.

XU, L. *et al.* Genomic and transcriptional heterogeneity of multifocal hepatocellular carcinoma. **Annals of oncology: official journal of the European Society for Medical Oncology**. London, v. 30, p. 990-997, 2019.

YANG, J., HONAVAR, V. Feature Subset Selection Using a Genetic Algorithm. In: Feature Extraction, **Construction and Selection**. Boston, v. 2, p.117–136, 1998.

YANG, J. *et al.* A global view of hepatocellular carcinoma: trends, risk, prevention and management. **Nature Reviews: Gastroenterology e Hepatology**, Londres, v. 16, p. 589–604, 2019.

YOON, Y. *et al.* MDM2 and p53 polymorphisms are associated with the development of hepatocellular carcinoma in patients with chronic hepatitis B virus infection. **Carcinogenesis**. Oxford, v. 29, p. 1192–1196, 2008.

YOUNOSSI, Z. *et al.*, Global burden of NAFLD and NASH: trends, predictions, risk factors and prevention. **Nature Reviews Gastroenterology & Hepatology** Londres, v. 15, p. 11–20, 2018.

YUAN, J. *et al.* Genetic polymorphisms in the methylenetetrahydrofolate reductase and thymidylate synthase genes and risk of hepatocellular carcinoma. **Hepatology**. Copenhagen, v. 46, p. 749–758, 2007.

ZHANG, P. *et al.* MicroRNA-139 suppresses hepatocellular carcinoma cell proliferation and migration by directly targeting Topoisomerase I. **Oncology Letters**, Athens, v. 2, p.1903-1913, 2019.

ZHANG, D., FRIEDMAN, S. Fibrosis-dependent mechanisms of hepatocarcinogenesis. **Hepatology**. Copenhagen, v. 56, p.769–775, 2012.

ZHOU, L. *et al.* Investigation of microRNA expression signatures in HCC via microRNA Gene Chip and bioinformatics analysis. **Pathology, research and**

**practice**, New York, v. 216, p. 1529-1582, 2020.

ZHU, H., JIA, Z., MISRA, H. Oxidative stress and redox signaling mechanisms of alcoholic liver disease: updated experimental and clinical evidence. **Journal of digestive diseases**, Carlton, v.13, p.133–142, 2012.

ZUCMAN-ROSSI, *et al.* Genetic Landscape and Biomarkers of Hepatocellular Carcinoma. **Gastroenterology**. Baltimore, v. 149, p. 1226–1239, 2015.

## ANEXO A - TERMO DE APROVAÇÃO DO COMITÊ DE ÉTICA DA UNIVERSIDADE DE PERNAMBUCO

UNIVERSIDADE DE  
PERNAMBUCO/ PROPEGE/



### PARECER CONSUBSTANCIADO DO CEP

#### DADOS DO PROJETO DE PESQUISA

**Título da Pesquisa:** AVALIAÇÃO DOS FATORES GENÉTICOS DE PREDISPOSIÇÃO À DOENÇA HEPÁTICA GRAVE CAUSADA PELO VÍRUS DA HEPATITE C (HCV)

**Pesquisador:** Patrícia Muniz Mendes Freire de Moura

**Área Temática:** Área 1. Genética Humana.

(Haverá envio para o exterior de material genético ou qualquer material biológico humano para obtenção de material genético.);  
Área 8. Pesquisa com cooperação estrangeira.

**Versão:** 3

**CAAE:** 07283812.1.0000.5207

**Instituição Proponente:** FUNDACAO UNIVERSIDADE DE PERNAMBUCO (FESP-UPE )

**Patrocinador Principal:** Financiamento Próprio  
FUNDACAO DE AMPARO A CIENCIA E TECNOLOGIA - FACEPE  
Conselho Nacional de Desenvolvimento Científico e Tecnológico ((CNPq))

#### DADOS DO PARECER

**Número do Parecer:** 212.100

**Data da Relatoria:** 05/03/2013

#### Apresentação do Projeto:

O presente estudo é do tipo analítico exploratório com grupos de comparação. Este desenho de estudo tem como objetivo avaliar a existência de associação entre transcritos expressos de genes e uma determinada característica (estudo de associação vírus da hepatite C e gravidade da doença hepática), através das diferenças entre os genes expressos entre os grupos, indicando uma possível associação entre as vias diferencialmente expressas e o mecanismo fisiopatológico. A implicação dessas vias no mecanismo da doença será confirmada por análise genética, avaliando a associação dos polimorfismos nos genes dessas vias e o grau da doença hepática. Serão definidos como grupos de comparação para os diversos estágios da doença hepática o Grupo 1 (Fibrose escala METAVIR 0-1-2), Grupo 2 (Fibrose escala METAVIR 3-4) e Grupo 3 (HCC, carcinoma hepatocelular). O projeto contará com as instalações presentes no Laboratório INSERM- UMR 906 dirigido pelo Prof. Alain Dessein - Université Aix-Marseille para as análises de transcriptoma e polimorfismos genéticos.

O estudo da expressão gênica do enxerto pode fornecer evidências sobre os mecanismos envolvidos na lesão hepática durante a recorrência da hepatite pelo vírus C após o transplante

**Endereço:** Av. Agamenon Magalhães, s/n°  
**Bairro:** Santo Amaro **CEP:** 50.100-010  
**UF:** PE **Município:** RECIFE  
**Telefone:** (81)3183-3775 **Fax:** (81)3183-3775 **E-mail:** comite.etica@upe.br

UNIVERSIDADE DE  
PERNAMBUCO/ PROPEGE/



hepático e, ainda, permitir a identificação de pacientes que têm um risco maior de desenvolver doença grave. Um dos possíveis benefícios da pesquisa é a possibilidade de empregar os achados para construir um conjunto de preditores que possam evitar procedimentos invasivos. Além disso, poderia auxiliar no processo de transplante hepático, indicando determinada intervenção no aloenxerto ou no controle da replicação do HCV no receptor e também no fígado enxertado.

**Objetivo da Pesquisa:**

Este projeto terá como objetivos identificar vias de expressão gênica em aloenxertos de fígado relacionadas com o processo fibrogênico e investigar polimorfismos de genes que possam ser marcadores da fibrose em pacientes transplantados; Investigar a diferença nas assinaturas de expressão gênica em tecido hepático de pacientes transplantados que evoluíram com diferentes graus de fibrose hepática; Notificar as vias fisiológicas relacionadas aos genes, cuja expressão está aumentada ou diminuída na doença hepática de paciente transplantado; Verificar a relação entre as assinaturas de expressão gênica em tecido hepático de pacientes transplantados e a recorrência da infecção pelo HCV, e determinar as diferenças de frequências alélicas e genotípicas entre pacientes com estágios de fibrose leve vs. grave.

**Avaliação dos Riscos e Benefícios:**

**Riscos**

A amostra para o estudo do transcriptoma será obtida uma pequena parte do fragmento (aprox. 0,5 cm) que é retirado para o exame histopatológico de rotina requisitado pelo médico para orientar o tratamento do paciente. Este procedimento é obrigatório para que o paciente seja avaliado quanto sua necessidade de tratamento e oferece riscos inerentes ao procedimento, dessa forma não será realizada uma biopsia especificamente para esse estudo. Além disso, será retirada uma única coleta de 10 ml de sangue por punção venosa, por profissionais habilitados e treinados para tal, entretanto, um risco de maior gravidade para o paciente seria a não coleta por dificuldade em puncionar a veia do paciente e esse fato pode representar um pequeno risco de ocasionar um pequeno hematoma no paciente no local da punção, mesmo que esse acontecimento seja muito raro de acontecer.

**Benefícios**

O presente estudo apresenta os seguintes benefícios: Fornecer subsídios para o desenho de estratégias terapêuticas empregando as moléculas estudadas; Fornecer subsídios para estabelecer o melhor manejo clínico dos pacientes infectados por HCV; Incrementar a produção científica do

Endereço: Av. Agamenon Magalhães, s/nº  
Bairro: Santo Amaro CEP: 50.100-010  
UF: PE Município: RECIFE  
Telefone: (81)3183-3775 Fax: (81)3183-3775 E-mail: comite.etica@upe.br

UNIVERSIDADE DE  
PERNAMBUCO/ PROPEGE/



grupo de pesquisa em gastrohepatologia da UPE; Formar novos pesquisadores na área de imunogenética e pesquisa clínica; Consolidar parcerias com outros centros de pesquisa de excelência na área do estudo entre infecção pelo HCV e a doença hepática crônica; Fortalecer a pós-graduação no Estado representada pelos Programas de Mestrado em Ciências da Saúde (FCM/ICB), Biologia Celular e Molecular Aplicada (ICB) da UPE e os Programas de Doutorado em Medicina Tropical e Rede Nordeste de Biotecnologia (RENORBIO).

**Comentários e Considerações sobre a Pesquisa:**

O estudo em tela é pertinente e apresenta significativo valor científico e social. O pesquisador apresentou o documento solicitado no Parecer n. 162.533.

**Considerações sobre os Termos de apresentação obrigatória:**

O pesquisador apresentou todos os documentos obrigatórios.

**Recomendações:**

Recomendo a aprovação do projeto.

**Conclusões ou Pendências e Lista de Inadequações:**

O estudo possui pertinência, valor científico e apresentou toda documentação solicitada.

Sendo assim, recomendo a aprovação do mesmo.

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Sim

**Considerações Finais a critério do CEP:**

O presente projeto, seguiu nesta data para análise da CONEP e só tem o seu início autorizado após a aprovação pela mesma.

RECIFE, 05 de Março de 2013

Assinador por:  
Nelson Rubens Mendes Loretto  
(Coordenador)

Endereço: Av. Agamenon Magalhães, s/nº  
Bairro: Santo Amaro CEP: 50.100-010  
UF: PE Município: RECIFE  
Telefone: (81)3183-3775 Fax: (81)3183-3775 E-mail: comite.etica@upe.br

## ANEXO B - TERMO DE APROVAÇÃO DO COMITÊ DE ÉTICA DO INSTITUTO AGGEU MAGALHÃES



**INSTITUTO AGGEU  
MAGALHÃES (IAM)**



### PARECER CONSUBSTANCIADO DO CEP

#### DADOS DO PROJETO DE PESQUISA

**Título da Pesquisa:** Fatores genéticos e epigenéticos envolvidos no surgimento de Carcinoma Hepatocelular em pacientes acometidos por diversas hepatopatias

**Pesquisador:** LUYDSON RICHARDSON SILVA VASCONCELOS

**Área Temática:** Genética Humana:

(Trata-se de pesquisa envolvendo Genética Humana que não necessita de análise ética por parte da CONEP.);

**Versão:** 5

**CAAE:** 34976619.4.0000.5190

**Instituição Proponente:** FUNDACAO OSWALDO CRUZ

**Patrocinador Principal:** MINISTERIO DA CIENCIA, TECNOLOGIA E INOVACAO

#### DADOS DO PARECER

**Número do Parecer:** 4.397.038

#### Apresentação do Projeto:

As informações dos campos Apresentação do Projeto, Objetivo da Pesquisa e Avaliação dos Riscos e Benefícios foram retiradas do documento PB\_INFORMAÇÕES\_BÁSICAS\_DO\_PROJETO\_1247914.pdf de 12.11.2020, de preenchimento pelo pesquisador responsável.

Trata-se de um estudo analítico transversal com grupos de comparação que visa identificar o perfil de expressão de microRNA's e polimorfismos associados ao surgimento do Carcinoma Hepatocelular em pacientes com diferentes hepatopatias. Tem como hipótese que "Existem diferentes assinaturas genéticas e epigenéticas que atuam no desenvolvimento/aparecimento do carcinoma hepatocelular, a depender de sua etiologia". A pesquisa ocorrerá em duas etapas principais, na primeira será realizada a análise de expressão diferencial de miRNAs e em seguida uma análise em ampla corte de pacientes. Para isso, será utilizado amostras de tecidos de biopsias hepáticas e de soro de pacientes e de indivíduos saudáveis maiores de 18 anos.

#### Objetivo da Pesquisa:

Objetivo primário:

Identificar o perfil de expressão de microRNA's e mRNA's candidatos e polimorfismos associados

**Endereço:** Av. Prof. Moraes Rego, s/nº

**Bairro:** Cidade Universitária

**CEP:** 50.670-420

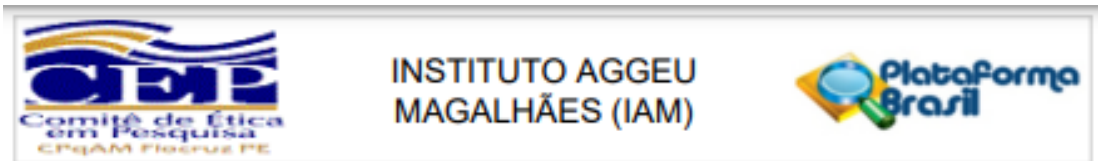
**UF:** PE

**Município:** RECIFE

**Telefone:** (81)2101-2639

**Fax:** (81)2101-2639

**E-mail:** comiteetica@cpqam.fiocruz.br



Continuação do Parecer: 4.397.038

ao surgimento do Carcinoma Hepatocelular em pacientes com diferentes hepatopatias, atendidos em centros de referência do estado de Pernambuco, no período de 2019 a 2022.

Objetivos secundários:

- Determinar a expressão de microRNA's candidatos em tecidos hepáticos (Tumoral e não Tumoral) de pacientes com carcinoma hepatocelular (CHC), por diferentes etiologias;
- Determinar a expressão de RNA's mensageiros (mRNA's) candidatos em tecidos hepáticos (Tumoral e não Tumoral) de pacientes com carcinoma hepatocelular (CHC), por diferentes etiologias;
- Confirmar os microRNAs mais diferencialmente expressos em soro de pacientes com CHC, por diferentes etiologias;
- Determinar a frequência alélica e genotípica de SNPs em genes da resposta imunológica em pacientes com CHC e grupo controle;
- Associar os SNPs detectados com a expressão dos miRNAs em soro e biopsias de pacientes com CHC, por diversas etiologias;
- Correlacionar o perfil de expressão dos mRNA's alvos com o perfil expressão dos miRNAs em biopsias de pacientes com CHC, por diversas etiologias;
- Verificar associações entre haplótipos e genótipos dos SNPs encontradas e os critérios clínicos em pacientes com CHC;
- Verificar a associação entre a variação de expressão de miRNAs candidatos em soro de pacientes com diferentes etiologias de CHC.

#### **Avaliação dos Riscos e Benefícios:**

Riscos:

O principal desconforto relacionado a coleta sanguínea a picada da agulha durante a punção de um vaso periférico realizado para a avaliação de resposta ao tratamento. No tocante a obtenção do tecido hepático, um pequeno fragmento do explante ser encaminhado para a pesquisa, após o transplante hepático.

Entretanto, ambos procedimentos ocorrem rotineiramente, independente do projeto. Informamos que para esse estudo no ser necessária nenhuma outra coleta de amostra, usaremos um fragmento do explante hepático que encaminhado a histopatologia, procedimento comum em situações de transplantes de órgãos sólidos. Esclarecemos que os resultados desse estudo serão repassados ao seu médico, mas poderão não beneficiar o paciente diretamente.

<b>Endereço:</b> Av. Prof. Moraes Rego, s/nº	<b>CEP:</b> 50.670-420
<b>Bairro:</b> Cidade Universitária	
<b>UF:</b> PE	<b>Município:</b> RECIFE
<b>Telefone:</b> (81)2101-2639	<b>Fax:</b> (81)2101-2639
	<b>E-mail:</b> comiteetica@cpqam.fiocruz.br





INSTITUTO AGGEU  
MAGALHÃES (IAM)



Continuação do Parecer: 4.397.038

**Benefícios:**

Fornecer subsídios para o desenho de estratégias terapêuticas empregando as moléculas estudadas. Fornecer subsídios para estabelecer o melhor manejo clínico dos pacientes acometidos pelo câncer hepático. Os resultados do estudo serão divulgados com a garantia de sigilo sobre a identidade dos participantes.

**Comentários e Considerações sobre a Pesquisa:**

A pesquisa é de relevância científica para pesquisa básica uma vez que propõe estudar os polimorfismos genéticos associados ao carcinoma hepatocelular, podendo subsidiar informações relevantes para o desenho de estratégias terapêuticas.

**Considerações sobre os Termos de apresentação obrigatória:**

Todos apresentados sem pendências.

**Recomendações:**

Entregas de relatórios parciais e final devem ser feitas via notificação. Os relatórios parciais devem ser apresentados, pelo menos, semestralmente. Qualquer alteração no protocolo no decorrer da pesquisa, como também mudança de equipe, devem ser informados via emenda. Todas as ações deve ser realizadas via Plataforma Brasil.

**Conclusões ou Pendências e Lista de Inadequações:**

O Comitê avaliou e considera que os procedimentos metodológicos do Projeto em questão estão condizentes com a conduta ética que deve nortear pesquisas envolvendo seres humanos, de acordo com o Código de Ética, Resolução 466/12 do Conselho Nacional de Saúde, de 12 de dezembro de 2012 e complementares.

O projeto está aprovado para ser realizado em sua última formatação apresentada ao CEP.

**Considerações Finais a critério do CEP:**

Vale ressaltar que, em relação ao protocolo de pesquisa, é previsto na Resolução 466/2012 no item XI - DO PESQUISADOR RESPONSÁVEL, que o pesquisador deve apresentar ao CEP relatórios parciais e finais (subitem: d) elaborar e apresentar os relatórios parciais e final). Pede-se que o pesquisador entregue os relatórios como previsto no texto da resolução. Informamos que a entrega dos relatórios parciais precisam ser feitas a cada semestre, conforme cronograma de execução da pesquisa, via Plataforma Brasil.

Endereço: Av. Prof. Moraes Rego, s/nº

Bairro: Cidade Universitária

CEP: 50.670-420

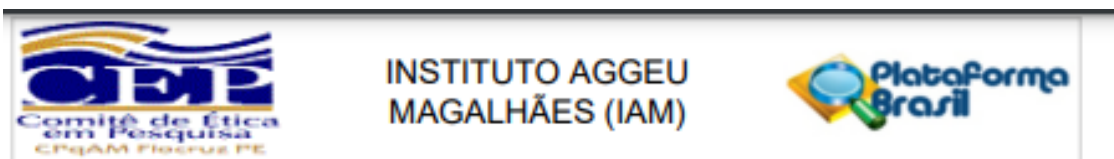
UF: PE

Município: RECIFE

Telefone: (81)2101-2639

Fax: (81)2101-2639

E-mail: comiteetica@cpqam.fiocruz.br



Continuação do Parecer: 4.397.038

**Este parecer foi elaborado baseado nos documentos abaixo relacionados:**

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_DO_PROJETO_1247914.pdf	12/11/2020 14:15:57		Aceito
Outros	Carta_resposta_3.doc	12/11/2020 14:15:24	LUYDSON RICHARDSON SILVA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_V06_com_modificacoes_evidenciadas.docx	09/11/2020 11:43:30	LUYDSON RICHARDSON SILVA VASCONCELOS	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_V06_final.docx	09/11/2020 11:42:33	LUYDSON RICHARDSON SILVA VASCONCELOS	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_CEP_com_modificacoes_evidenciadas.docx	19/10/2020 15:33:30	LUYDSON RICHARDSON SILVA	Aceito
Outros	TCLE_com_modificacoes_evidenciadas.docx	19/10/2020 15:30:46	LUYDSON RICHARDSON SILVA	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_CEP_V05_final.docx	19/10/2020 15:29:21	LUYDSON RICHARDSON SILVA	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_V05_final.docx	19/10/2020 15:28:04	LUYDSON RICHARDSON SILVA VASCONCELOS	Aceito
Declaração de Manuseio Material Biológico / Biorepositório / Biobanco	Biorrepositorio.docx	19/10/2020 15:26:59	LUYDSON RICHARDSON SILVA VASCONCELOS	Aceito
Folha de Rosto	folha_de_Rosto_LUYDSON.pdf	13/07/2020 14:08:15	LUYDSON RICHARDSON SILVA	Aceito

**Situação do Parecer:**

Aprovado

**Necessita Apreciação da CONEP:**

Não

Endereço: Av. Prof. Moraes Rego, s/nº  
 Bairro: Cidade Universitária CEP: 50.670-420  
 UF: PE Município: RECIFE  
 Telefone: (81)2101-2639 Fax: (81)2101-2639 E-mail: comiteetica@cpqam.fiocruz.br

## APÊNDICE A - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO DA UNIVERSIDADE DE PERNAMBUCO

### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Nome da pesquisa: AVALIAÇÃO DOS FATORES GENÉTICOS DE PREDISPOSIÇÃO À DOENÇA HEPÁTICA GRAVE CAUSADA PELO VÍRUS DA HEPATITE C (VHC)

Responsável: Patrícia Muniz Mendes Freire de Moura

**Eu, paciente abaixo assinado, portador do vírus da hepatite C, atendido no ambulatório deste serviço concordo em participar voluntariamente desta pesquisa sobre a AVALIAÇÃO DOS FATORES GENÉTICOS DE PREDISPOSIÇÃO À DOENÇA HEPÁTICA GRAVE CAUSADA PELO VÍRUS DA HEPATITE C (VHC)**

após ter sido informado que:

- a razão deste estudo, segundo o pesquisador, é saber se eu tenho uma alteração genética que poderia piorar a doença do fígado;
- o material da pesquisa será um fragmento do tecido do fígado que é retirado quando eu fizer o exame para a avaliação histopatológica para o tratamento e a retirada de uma única coleta de 10 ml de sangue por punção venosa, por profissionais no acompanhamento ambulatorial;
- essa pesquisa pode trazer benefícios se for aprovado que o problema genético relacionado ao sistema imunológico é importante, mas no momento, esse fato está sendo avaliado;
- os resultados da pesquisa serão utilizados para publicar informações científicas importantes para melhorar o conhecimento da doença no fígado e que meus dados serão mantidos em segredo não sendo divulgada nenhuma informação que possa me identificar;
- tenho a liberdade de retirar meu consentimento a qualquer momento e deixar de participar do estudo, sem que isso traga qualquer prejuízo para mim.
- Estou ciente que o material biológico coletado pode ser enviado para pesquisa fora do Brasil nos laboratórios incluídos neste projeto de colaboração.

Visto que nada tenho contra a pesquisa, concordo em assinar o presente termo de

consentimento. Recife, \_\_\_\_\_ de \_\_\_\_\_ de 20\_\_ .

\_\_\_\_\_  
Assinatura do paciente

\_\_\_\_\_  
Assinatura testemunha

\_\_\_\_\_  
Assinatura do Pesquisador

Patrícia Moura (e-mail: [patmoura@pq.cnp.br](mailto:patmoura@pq.cnp.br)) Fone: 81 3241 4952/ 81 96469951

Contato assistente: Raul Emídio de Lima – Fone: 81 8104 8909/ 9514 2126

CEP UPE

Fone: 3183-3775

E-mail: [comite.etica@upe.pe.gov.br](mailto:comite.etica@upe.pe.gov.br)

## APÊNDICE B - TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO DO INSTITUTO AGGEU MAGALHÃES

INSTITUTO AGGEU MAGALHÃES / FUNDAÇÃO OSWALDO CRUZ /  
MINISTÉRIO DA SAÚDE

### TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

Convidamos o (a) senhor (a) a participar da Pesquisa: “**Fatores genéticos e epigenéticos envolvidos no surgimento de Carcinoma Hepatocelular em pacientes acometidos por diversas hepatopatias**”, sob responsabilidade do Pesquisador Dr. Luydson Richardson Silva Vasconcelos

Se decidir participar dela, por favor, leia este documento com bastante atenção antes de assiná-lo. Caso haja alguma palavra ou frase que o senhor (a) não consiga entender, converse com o(a) pesquisador(a) responsável pelo estudo ou com um membro da equipe desta pesquisa para esclarecê-los. A proposta deste Termo de Consentimento Livre e Esclarecido (TCLE) é explicar o estudo e solicitar a sua permissão para participar do mesmo.

Nesta pesquisa, propomos estudar o Câncer Hepático, tendo como objetivo a identificação de fatores existentes nos genes que possam estar relacionados ao surgimento do mesmo, podendo também contribuir para o desenvolvimento de testes para diagnóstico da doença.

O participante poderá se recusar e/ou retirar este consentimento a qualquer momento que desejar, sem prejuízo a ambas as partes. Todos os seus dados também serão removidos da pesquisa, a não ser que o Sr (a) autorize a permanência desses na nossa análise. A sua contribuição na pesquisa ocorrerá através da participação de uma entrevista, realizada antes da consulta médica para acompanhamento clínico, realizada com a pesquisador(a) responsável ou um membro da equipe de pesquisa. Além disso, será utilizado parte do material biológico coletado no Instituto do Fígado de Pernambuco/ Hospital Oswaldo Cruz. Será coletado dos pacientes com carcinoma hepatocelular 4ml de sangue, em sistema de tubos a vácuo com anticoagulante EDTA com o objetivo de extrair o DNA, também coletado antes da consulta médica para acompanhamento clínico, além de um pedaço do fragmento do fígado retirado em exame de biópsia hepática (0,5mm<sup>2</sup>), que normalmente é solicitada pelo médico para a realização de exame patológico.

Quanto aos **riscos e desconfortos**, a metodologia utilizada para esta coleta de dados, até a presente fase e diante dos procedimentos utilizados, nos mostra que o que pode vir a ocorrer, durante ou após o trabalho, é a possibilidade de surgir algum desconforto tais como dor, hematoma ou outro desconforto no local da coleta de sangue. Para evitar os aparentes riscos é necessária uma boa prática de biossegurança durante o procedimento. Caso haja o aparecimento de hematoma, é necessário fazer uma compressa de gelo por quinze minutos a cada hora, durante as primeiras seis horas, caso o hematoma persista ou se houver qualquer outra reação ou dor local, o paciente deve entrar em contato com a pesquisadora para que sejam tomadas as devidas providências.

Os **benefícios** esperados com o resultado desta pesquisa são: possibilitar um maior entendimento a respeito das variações das formas dos genes estudados e sua relação com o carcinoma hepatocelular, podendo contribuir para o diagnóstico diferencial da doença.

O senhor(a) terá os seguintes **direitos**: a garantia de esclarecimento e resposta a qualquer pergunta; a liberdade de abandonar a pesquisa a qualquer momento sem prejuízo para si **ou para seu tratamento (se for o caso)**; a garantia de privacidade à sua identidade e do sigilo de suas informações; a garantia de que caso haja algum dano a sua pessoa (ou o dependente), os prejuízos serão assumidos pelos pesquisadores ou pela instituição responsável inclusive acompanhamento médico e hospitalar. Caso haja gastos adicionais serão absorvidos pelo pesquisador responsável.

Os resultados dessa pesquisa ficarão disponíveis para o Sr (a). Entretanto, os benefícios esperados com o resultado desta pesquisa são: Fornecer subsídios para o desenho de estratégias terapêuticas empregando as moléculas estudadas. Fornecer subsídios para estabelecer o melhor manejo clínico dos pacientes acometidos pelo câncer hepático. Os resultados do estudo serão divulgados com a garantia de sigilo sobre a identidade dos participantes.

Todas as informações colhidas serão analisadas, mantendo-se a confidencialidade (segredo) do paciente a todo o momento, ou seja, em nenhum momento os dados que o identifique serão divulgados, a menos que seja exigido por lei. Os resultados desta pesquisa poderão ser apresentados em reuniões ou publicações, contudo, sua identidade não será revelada nessas apresentações.

Este projeto foi avaliado pelo Comitê de Ética em Pesquisa do Instituto Aggeu Magalhães da Fundação Oswaldo Cruz do Ministério da Saúde. Caso exista qualquer tipo de problema, dúvidas gerais ou questionamentos sobre o estudo, entrar em contato com os pesquisadores responsáveis, **Dr. Luydson Vasconcelos, pelos telefones (81) 2101-2574 ou (81) 99645-2018 / Raul Emídio de Lima, MSc., pelos telefones (81) 98104.8909 ou (81) 99514.2126.**

**Em caso de dúvidas ou preocupações quanto aos seus direitos como participante deste estudo, o (a) senhor (a) pode entrar em contato com o Comitê de Ética em Pesquisa (CEP) deste centro de pesquisas, localizado na Av. Professor Moraes Rego, s/n – Campus da UFPE - Cidade Universitária, Recife/PE, CEP: 50.670-420, através do telefone (81) 2101-2639 ou pelo e-mail: comiteetica@cpqam.fiocruz.br. O horário de funcionamento é das 8 – 12 h e das 13 – 17h. O CEP é responsável pela avaliação e acompanhamento dos aspectos éticos de todas as pesquisas envolvendo seres humanos.**

**Após a leitura desse Termo de Consentimento Livre e Esclarecido, dou meu consentimento de livre e espontânea vontade, para participar como voluntário, deste estudo, assinando esse termo em duas vias, pois uma via pertencerá a mim e a outra a (o) pesquisador(a).**

**O Senhor (a) estará livre para suspender essa autorização de participação no estudo a qualquer momento, sem que haja nenhum prejuízo.**

Rubricas

\_\_\_\_\_ - \_\_\_\_\_

Pesquisador(a)

Participante

Endereço do participante para contato futuro.

Rua \_\_\_\_\_ Nº. \_\_\_\_\_

Bairro \_\_\_\_\_ Cidade \_\_\_\_\_

Telefone \_\_\_\_\_

\_\_\_\_\_  
ASSINATURA DO PARTICIPANTE

\_\_\_\_\_  
ASSINATURA (1º TESTEMUNHA)

Recife, \_\_\_\_ / \_\_\_\_ / \_\_\_\_

\_\_\_\_\_  
PESQUISADOR RESPONSÁVEL

## APÊNDICE C - QUESTIONÁRIO AVALIATIVO ELABORADO PARA O ESTUDO

Investigação sobre a relação de gravidade da doença hepática com polimorfismos genéticos			
Data (coleta):	/ /	Prontuário HUOC:	IFP:
Nome completo:			
Estado atual de tratamento: 1. SEM TRATAMENTO 2. EM TRATAMENTO 3. TRATADO (Follow-Up)			
Idade (momento da biópsia)		Qual a data do seu nascimento. / /	
Sexo:	1. Masc. 2. Fem.	Cor de pele:	1. Branca 2. Parda 3. Negro 4. Índio 5. Outro
Cidade/Estado (mora):		Cidade/Estado (origem):	
Telefone convencional para contato:		Celular:	
Antecedentes:			
1. Hemotransfusão Ano: _____		3. Medicação por seringa de vidro	
2. Drogas Injetáveis		4. Tatuagens	
5. Cirurgias			
Consumo de álcool (antes de descobrir a doença): 1. Sim 2. Não			
Homem		Mulher	
( ) Maior que 40 g/dia		( ) Maior que 20 g/dia	
( ) Menor que 40 g/dia		( ) Menor que 20 g/dia	
Tabagismo: 1. Nunca fumou 2. Ex-fumante 3. Fumou até 1 ano antes da pesquisa 4. Fumante			
<b>Investigação clínica.</b>			
Além da Hepatite C o (a) senhor (a) tem ou teve alguma outra doença.		1. Sim 2. Não 9. Não sabe	
Se sim, qual a outra doença que o (a) senhor (a) tem ou teve.			
1. Hipertensão			
2. Diabetes			
3. Esquistossomose			
4. Outra. Especifique			
<b>Exame físico</b>			
Peso: _____ (quilo, gramas)		Altura: _____ (metro, centímetro)	
IMC [peso/ (altura/m <sup>2</sup> ): _____			
Circunferência abdominal + quadril (cm): _____		P.A.: _____	
<b>Exames Laboratoriais</b>			
Glicemia: _____	TG: _____	HDL: _____	Albumina: _____
Ferro: _____	Ferritina: _____		
CT: _____	LDL: _____	Insulina: _____	Na: _____
K: _____	TP: _____	INR: _____	
<b>Síndrome Metabólica (PA; Glic.; Cintura abdm; TG; HDL):</b> ( ) Sim ( ) Não			
AST: _____	ALT: _____	Alfa-fetoproteína: _____	Bilirr. total: _____
Bilirr. Dir: _____	Bilirr. Índi: _____		
gGT: _____	FA: _____	Hgb: _____	Htc: _____
Plaquetas: _____	Uréia: _____	Creatinina: _____	
<b>Genótipo viral:</b>		<b>Carga viral (absoluta):</b>	
<b>Carga viral (log):</b>			
<b>Histopatológico</b> ( ) Biópsia ( ) Fibroscan → Kpa: _____			
<b>Grau de fibrose (1ª Biópsia)</b>		<b>Grau de fibrose (2ª Biópsia)</b>	
( ) F1 ( ) F2 ( ) F3 ( ) F4 ( ) HCC Data: / /		( ) F1 ( ) F2 ( ) F3 ( ) F4 ( ) HCC Data: / /	
<b>Grau de inflamação (1ª Biópsia)</b>		<b>Grau de inflamação (2ª Biópsia)</b>	
( ) A0 ( ) A1 ( ) A2 ( ) A3 ( ) A4		( ) A0 ( ) A1 ( ) A2 ( ) A3 ( ) A4	
<b>Grau de esteatose (1ª Biópsia)</b>		<b>Grau de esteatose (2ª Biópsia)</b>	
( ) Leve ( ) Moderada ( ) Intensa ( ) Esteatohepatite (NASH)		( ) Leve ( ) Moderada ( ) Intensa ( ) Esteatohepatite (NASH)	
<b>Regime de tratamento</b>			
( ) Sofosbuvir + Simeprevir		( ) Sofosbuvir + Daclatasvir	
( ) Sofosbuvir + Ribavirina			
<b>Resposta ao tratamento</b>			
Week 4:	POS	NEG	
Week 12:	POS	NEG	
Week 24:	POS	NEG	
RVS	SIM	NÃO	
<b>Observações:</b>			
ENTREVISTADOR: _____			