UNIVERSIDADE ESTADUAL PAULISTA "JÚLIO DE MESQUITA FILHO"
FACULDADE DE MEDICINA

Renata Zugaib

Perfil transcricional da infecção crônica pelo vírus da Hepatite C (VHC) por sequenciamento de nova geração

Dissertação apresentada à Faculdade de Medicina, Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de Botucatu, para obtenção do título de Mestre(a) em Pesquisa e Desenvolvimento (Biotecnologia Médica)

Orientadora: Profª. Dra. Adriana Camargo Ferrasi

Botucatu
2017

Renata Zugaib

Perfil transcricional da infecção crônica pelo vírus da Hepatite C (VHC) por sequenciamento de nova geração

Dissertação apresentada à Faculdade de Medicina, Universidade Estadual Paulista "Júlio de Mesquita Filho", Câmpus de Botucatu, para obtenção do título de Mestre(a) em Pesquisa e Desenvolvimento (Biotecnologia Médica)

Orientadora: Profª. Dra. Adriana Camargo Ferrasi

Botucatu
2017

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSEMEIRE APARECIDA VICENTE-CRB 8/5651

Zugaib, Renata.
Perfil transcricional da infecção crônica pelo vírus da hepatite C (VHC) por sequenciamento de nova geração / Renata Zugaib. - Botucatu, 2017

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu
Orientador: Adriana Camargo Ferrasi
Capes: 20205007

1. Hepatite C. 2. Hepatite por vírus. 3. Fígado - Câncer. 4. Transcrição genética. 5. Expressão gênica.

Palavras-chave: Expressão gênica diferencial ; Hepatocarcinoma; Sequenciamento de nova geração; Transcriptoma; Vírus da hepatite C .

Renata Zugaib

Perfil transcricional da infecção crônica pelo vírus da Hepatite C (VHC) por sequenciamento de nova geração

Dissertação apresentada à Faculdade de Medicina de Botucatu Universidade Estadual Paulista "Júlio de Mesquita Filho", Campus de Botucatu, para obtenção do título de Mestre.

Orientadora: Profa. Dra Adriana Camargo Ferrasi

Comissão examinadora

Profa. Dra Adriana Camargo Ferrasi

Universidade Estadual Paulista "Júlio de Mesquita Filho", FMB

Prof. Dr Giovanni Faria Silva

Universidade Estadual Paulista "Júlio de Mesquita Filho", FMB

Profa. Dra Silvia Helena Barem Rabenhorst

Universidade Federal do Ceará, Faculdade de Medicina

Botucatu, 13 de janeiro de 2017.

Agradecimentos

Agradeço primeiramente a Deus, por estar sempre ao meu lado, por iluminar meu caminho, pelas conquistas e aprendizados, por todas as pessoas maravilhosas que tem colocado ao meu redor que, direta ou indiretamente, me ajudaram a realizar este mestrado.

A toda minha família e amigos que sempre estiveram ao meu lado, apoiando-me em todas as dificuldades encontradas e realizações alcançadas, motivando-me e torcendo sempre pelo meu sucesso.

A toda equipe do Laboratório de Biologia Molecular do Hemocentro de Botucatu, por compartilharem comigo suas experiências, pelo acolhimento, amizade e carinho, principalmente à Aline Faria Galvani, Aline Del'Vescovo, Nathalia Almeida Souza, Caroline Nunes, Francielle Martins e Kelly Colussi por todos os momentos compartilhados, apoio e motivação durante esses anos de laboratório.

À minha orientadora, Profa. Dra. Adriana Camargo Ferrasi, pela oportunidade e confiança, pela paciência, carinho, conselhos, ensinamentos e grande apoio durante todos esses anos, sempre será minha inspiração.

À Dra. Maria Inês de Moura Campos Pardini, pela oportunidade de fazer parte de sua equipe, e adquirir grande aprendizado na área de Biologia Molecular.

Epígrafe

“O que sabemos é uma gota, o que não sabemos é um oceano.”

Isaac Newton

Resumo

ZUGAIB, R. Perfil transcricional da infecção crônica pelo vírus da Hepatite C (VHC) por sequenciamento de nova geração. [Dissertação]. Botucatu: Faculdade de Medicina de Botucatu, Universidade Estadual Paulista – FMB/UNESP; 2016.

O vírus da hepatite C (VHC) constitui a principal causa de doença hepática crônica, que representa um dos maiores problemas de saúde pública. O carcinoma hepatocelular (CHC), altamente associado a infecção crônica pelo vírus da hepatite C (VHC), é um dos tumores malignos mais comuns no mundo, com um alto índice de causa de óbito. Com o avanço das técnicas moleculares, tornou-se possível uma nova abordagem nos estudos gênicos para um melhor entendimento molecular de processos infecciosos e crônicos. Estudos evidenciando uma associação da transcrição gênica ao processo patológico e a importância de análises mais abrangentes. O sequenciamento de nova geração fornece uma maneira poderosa para a avaliação global do transcriptoma com alta resolução e um menor custo, possibilitando uma análise do perfil transcricional da doença. Assim, este estudo teve por objetivo avaliar o perfil de expressão gênica diferencial de pacientes infectados pelo VHC com CHC e comparar com amostras de tecidos não tumorais através do sequenciamento em larga escala do transcriptoma, a fim de identificar potenciais biomarcadores de diagnóstico e prognóstico de CHC. Foram analisados três fragmentos de tecido tumoral e três fragmentos de tecido hepático não tumoral como controle através do sequenciamento de RNAs (RNA-Seq). Os resultados obtidos demonstraram uma expressão diferencial de 4.792 genes. Avaliando os 10 genes mais e menos expressos, foi observada uma grande associação de variações nesses genes em diversos tipos de tumores. Também foram observados, entre os menos expressos, genes intimamente relacionados a função hepática ou relacionados a componentes produzidos pelo fígado. Esses achados sugerem que *COL11A1, SFRP4, SFRP2, LRRC15, CCL18, ADAMDEC1, COL1A1, COL10A1, CTHRC1 e OLR1*, superexpressos, possam atuar juntos no processo tumoral servindo como marcadores moleculares tumorais, e que a presença do tumor possa provocar uma desregulação nos genes associados ao fígado aqui encontrados, contudo, estudos mais específicos devem ser conduzidos para a confirmação dessa hipótese.

Palavras-chave: Carcinoma Hepatocelular, Sequenciamento de nova geração, Transcriptoma.

Abstract

ZUGAIB, R. Transcriptional profile of chronic hepatitis C virus (HCV) infection by next generation sequencing. [Thesis - Master]. Botucatu: Botucatu Medical School, Universidade Estadual Paulista –UNESP; 2016.

Hepatitis C virus (HCV) is the leading cause of chronic liver disease, one of the major public health problems. Hepatocellular carcinoma (HCC), highly associated with chronic hepatitis C virus (HCV) infection, is one of the most common malignant tumors in the world, with a high cause of death. With the advancement of molecular techniques, a new approach in gene studies has become possible for a better molecular understanding of infectious and chronic processes. Studies evidencing an association of the gene transcription to the pathological process and the importance of more comprehensive analyzes. Next generation sequencing provides a powerful way for the global evaluation of the transcriptome with high resolution and a lower cost, allowing an analysis of the transcriptional profile of the disease. Thus, this study aimed to evaluate the differential gene expression profile of HCV infected patients in their highest degree of chronicity (HCC) and to compare with non-tumor tissue samples through large-scale sequencing of the transcriptome in order to identify potential biomarkers for diagnosis and prognosis of HCC. Three fragments of tumor tissue and three fragments of non-tumor liver tissue were analyzed through the sequencing of RNAs (RNA-Seq). The results obtained demonstrated a differential expression of 4.792 genes. Evaluating the 10 over and down regulated genes, a high association of variations in these genes was observed in several types of tumors. Among the least expressed, genes closely related to liver function or related to components produced by the liver were also observed. These findings suggest that COL11A1, SFRP4, SFRP2, LRRC15, CCL18, ADAMDEC1, COL1A1, COL10A1, CTHRC1 and OLR1, overexpressed, may act together in the tumor process serving as molecular tumor markers, and that the presence of the tumor may lead to dysregulation in the genes associated with the liver found here, however, more specific studies should be conducted to confirm this hypothesis.

Keywords: Hepatocellular carcinoma, Next generation sequence, transcriptome.

Lista de Figuras

LISTA DE FIGURAS

Figura 1: Representação da preparação da biblioteca através do kit TruSeqTM Stranded Total RNA sample preparation (Illumina). 31

Figura 2 - Imagem ilustrativa formação dos *clusters* e sequenciamento das bibliotecas.. 33

Figura 4 - Representação gráfica da qualidade média por base dos dados de sequenciamento Illumina no programa FastQC.. 38

Figura 5 - Representação ilustrativa do gráfico gerado pelo programa FastQC da distribuição do tamanho das *reads* (eixo x) pela quantidade de sequências geradas (eixo Y). 39

Figura 6 - Representação gráfica gerada pelo programa FastQC da distribuição por base do conjunto de sequencias geradas pelo sequenciamento Illumina. 40

Figura 7 - Representação gráfica gerada pelo programa FastQC da distribuição do conteúdo de GC ao longo da extensão de cada sequência comparada a uma distribuição normal teórica do conteúdo GC do genoma humano.....41

Figura 8 – Diagrama de dispersão do *log2 fold-change* pela média da contagem normalizada dos genes, evidenciando os genes diferencialmente expressos 42

Figura 9 – Diagrama de dispersão do *log2 fold-change* pela média observada para o gene, evidenciando os genes diferencialmente expressos em pelo menos uma vez quando comparado ao controle. 43

Figura 10 – Categorização dos genes diferencialmente expressos entre os grupos carcinoma hepatocelular e fígado normal de acordo com GO (*Gene Ontology Consortium*) e distribuição conforme sua função molecular. 44

Figura 11 – Categorização dos genes diferencialmente expressos entre os grupos carcinoma hepatocelular e fígado normal de acordo com GO (*Gene Ontology Consortium*) e distribuição conforme seu processo biológico.....44

Figura 12 – Categorização dos genes diferencialmente expressos entre os grupos carcinoma hepatocelular e fígado normal de acordo com GO (*Gene Ontology Consortium*) e distribuição conforme seu componente celular. 45

Lista de Tabelas

LISTA DE TABELAS

Tabela 1. Composição dos *pools* de RNA total para a construção das bibliotecas, agrupados por condição biológica.....30

Tabela 2. Dados gerados pelo sequenciamento RNAseq (Illumina) das bibliotecas de cDNA produzidas a partir de tecidos hepáticos com e sem tumor. 37

Tabela 3. Comparações entre as bibliotecas de cDNA provenientes de amostras de tecido hepático com diferentes condições biológicas. 42

Tabela 4. Análise funcional dos genes mais e menos expressos no grupo carcinoma hepatocelular em relação ao grupo controle, divididos em três categorias de acordo com: processo biológico, função molecular e componente celular através do *Gene Ontology Consortium* (GO). 45

Tabela 5. Top 10 genes mais e menos expressos entre os grupos CHC e fígado normal 46

Sumário

1 INTRODUÇÃO 18
2 OBJETIVOS 25
2.1. Objetivo Geral 25
2.2. Objetivos Específicos 25
3 CASUÍSTICA E MÉTODOS 26
3.1 Casuística 27
3.2 Critérios de inclusão 27
3.3 Critérios de exclusão 28
3.4 Coleta e processamento inicial dos materiais biológicos 28
3.5 Extração de ácido ribonucléico (RNA) 29
3.6 Quantificação e análise qualitativa do RNA total 29
3.7 Montagem dos pools de amostras 29
3.8 Preparo da Biblioteca de cDNA e Sequenciamento 30
3.9 Análise dos dados do sequenciamento 33
4 RESULTADOS E DISCUSSÃO 36
5 CONCLUSÃO 52
6 REFERÊNCIAS 54
7 ANEXOS 65
7.1 Anexo 1 66

Introdução

01 INTRODUÇÃO

A infecção pelo vírus da hepatite C (VHC) é importante causa da doença hepática crônica. Apesar de progressos na terapêutica, representa significante problema de saúde pública, conduzindo a quadros de cirrose hepática e carcinoma hepatocelular (CHC) devido ao seu potencial de evolução para a forma crônica[1]. Segundo a Organização Mundial da Saúde (OMS) existem de 130 a 150 milhões de pessoas com infecção crônica pelo VHC em todo o mundo[2]. Em análise realizada pela Faculdade de Medicina da Universidade de São Paulo (FMUSP) estima-se que existam no Brasil cerca de 1,4 a 1,7 milhões de pessoas infectadas[3].

O VHC, descrito nos Estados Unidos em 1989 por Choo e colaboradores[4], apresenta um diâmetro de, aproximadamente, 50-60nm, pertence à família Flaviviridae, com genoma constituído por uma fita simples de RNA de polaridade positiva, encapsulado por um envelope lipoproteico. O RNA viral apresenta 9,6 kb, com uma única região aberta de leitura (*Open Reading Frame*, ORF) que codifica proteínas estruturais e não estruturais, sendo flanqueada por duas regiões não traduzidas, 5´NCR (*Non-Coding Region*) e 3´NCR[5]. A região 5´NCR é altamente conservada, pois possui um sítio de entrada ribossomal (*Internal Ribosome Entry Site*, IRES), que está envolvido no processo de síntese proteica viral. Durante a expressão do genoma viral é sintetizada uma única poliproteína de, aproximadamente, 3000 aminoácidos, que é posteriormente clivada pelas enzimas virais e celulares constituindo as proteínas estruturais (*core*, E1, E2) e as não estruturais (NS2, NS3, NS4A, NS4B, NS5A, NS5B) [6,7].

O genoma altamente heterogêneo quanto à sequência de nucleotídeos (devido à alta atividade replicativa do vírus e ausência de atividade de reparo das enzimas virais), distingue sete genótipos, nomeados de 1 a 7 e mais de 60 subtipos, os quais são designados por letras do alfabeto, além das quasispecies[8]. As sequências de nucleotídeos entre os diferentes genótipos diferem em, aproximadamente, 30-35%, e entre os subtipos de 20-25% [9].

O VHC é um vírus hepatotrópico que infecta hepatócitos. No entanto, o vírus já foi encontrado em linfócitos T e B[10], monócitos, macrófagos[11], células

dendríticas[12] e plaquetas[13]. Estas últimas foram associadas ao carreamento do vírus pela corrente circulatória em indivíduos infectados[14].

A infecção pelo VHC é altamente dinâmica com uma meia vida viral de poucas horas e produção/liberação estimada de 1012 vírions por dia em um indivíduo infectado. No entanto, a entrada do vírus nos hepatócitos constitui uma fase essencial para que ocorra infecção, processo que se encontra na dependência de interações entre componentes virais, principalmente proteínas do envelope e fatores de superfície celulares do hospedeiro[15]. Várias moléculas expressas na superfície celular dos hepatócitos já foram associadas à entrada do VHC em sua célula alvo como a tetraspanina CD81[16], principal receptor envolvido neste processo; o receptor *scavenger* tipo B classe I (SR-BI/Cla1)[17], a claudina-1 (CLDN-1)[18] e ocludinas[19].

Além dessas, algumas glicosaminoglicanas também vêm sendo associadas à adesão e entrada do VHC nas células[20,21]. Outros estudos sugerem a participação de proteínas de adesão celular na entrada do VHC. No entanto, o mecanismo exato de entrada do vírus nas células permanece, ainda, desconhecido[15].

Quando se considera a evolução natural da infecção pelo VHC, aproximadamente 20% dos infectados erradicam o vírus em semanas ou meses[22], caracterizando o denominado clareamento viral espontâneo. Aproximadamente 70-80% dos indivíduos portadores de VHC evoluem para a forma crônica da doença com progressão da fibrose hepática, e destes, em torno de 20% evoluem para cirrose em 20 anos de infecção[23,24] e, dos cirróticos, 1 a 4% desenvolvem carcinoma hepatocelular (CHC) ao ano[25].

Tanto a fibrose como a cirrose são consequências de uma resposta sustentada de cicatrização à injúria hepática crônica[26], e no caso da infecção pelo VHC, fatores do hospedeiro mais do que fatores virais parecem estar correlacionados com a progressão da fibrose. Outros fatores como sexo, idade à infecção, consumo de álcool e coinfecção com outros vírus já foram relacionados a essa progressão[27,28].

Frequentemente associado com a infecção pelo vírus da hepatite C (VHC) e o vírus da hepatite B (VHB), o carcinoma hepatocelular (CHC), também conhecido

como hepatocarcinoma, é o sexto tipo de câncer mais incidente em todo o mundo (782.451 casos) e o segundo mais mortal (745.533 casos). No Brasil, foram previstos 10610 novos casos de CHC em 2015[29]. Raramente ocorre antes dos 40 anos, e possui um pico de incidência aos 70 anos, com um índice de duas a quatro vezes maior em homens[30].

Acredita-se que o CHC seja mais frequente em homens pois estes são mais propensos a infeção por vírus como VHB e VHC, consomem mais álcool e tabaco, entre outros fatores ligados ao estilo de vida. Entretanto, nos dias atuais, devido ao aumento da exposição de mulheres a estes fatores de risco, não se pode afirmar que estes fatores sejam a causa dessa incidência elevada em homens. Experimentos demonstraram um aumento no desenvolvimento de CHC em ratos machos, sugerindo assim que andrógenos possam influenciar na progressão do CHC. Alguns fatores endógenos que podem afetar a adversidade no risco de desenvolvimento de CHC no sexo masculino incluem alto índice de massa corporal e altos níveis de hormônios androgênicos[30].

O CHC é um câncer primário do fígado com origem nos hepatócitos, células do fígado que compõe 2/3 de sua massa e são responsáveis pela sua função biológica como a metabolização de substâncias. Constitui cerca de 70 a 85% das neoplasias hepáticas primárias e é um dos tumores malignos mais comuns no mundo. Muito agressivo, o CHC possui um alto índice de causa de óbito após o início dos sintomas, dando ao paciente nesta fase uma expectativa média de vida inferior a um mês[31].

Casos de doenças crônicas do fígado, como infecção crônica pelo VHB e VHC e cirrose, estão presentes em cerca de 70% a 90% de todos os casos de CHC detectados. Embora o CHC seja um exemplo típico de câncer relacionado a infecção viral, também está fortemente associado a fatores de estilo de vida como o alcoolismo crônico e a obesidade[32]. Menos comumente está relacionado a casos de esteatose hepática não alcoólica, hepatite autoimune, dentre outros[30].

A infecção crônica pelo VHC é o maior fator de risco para o desenvolvimento de CHC, conferindo um aumento deste risco pela promoção de fibrose e cirrose, podendo então evoluir para um carcinoma hepatocelular[30].

Tanto nos casos relacionados a infecção viral quanto nos casos relacionados a fatores do estilo de vida, é conhecido que o acúmulo de danos no DNA, causado pela longa duração da inflamação e regeneração, tem um grande papel no processo de hepatocarcinogênese[32].

Entre os fatores do hospedeiro, ocupa lugar de destaque o estudo da expressão gênica dos pacientes portadores de VHC com diferentes níveis de comprometimento hepático, o que pode revelar biomarcadores relevantes no desenho de kits/chips de valor prognóstico, bem como, informações inovadoras para aplicação no desenvolvimento de novas estratégias terapêuticas e/ou vacinas[27,28].

O dogma central da biologia molecular, proposto por Francis Crick (1958)[33], descreve o fluxo de informação genética e atribui ao RNA mensageiro (RNAm) a função de levar informação para a síntese proteica. Estudos tem revelado que apenas uma pequena fração (1,5 a 2%) do genoma é composta por partes codificadoras de proteínas, contudo, mais que 70% do genoma humano é transcrito em RNA[34,35].

Com o avanço das técnicas moleculares, tornou-se possível uma nova abordagem nos estudos gênicos, como a genômica funcional. A fim de entender como os genes desempenham suas funções, a genômica funcional tem como objetivo avaliar a expressão de genes através da análise de transcritos (RNAm), microRNAs ou proteínas para um melhor entendimento molecular de processos como a tumorigênese e descoberta de potenciais biomarcadores[36,37].

Estudos demonstraram diferenças na expressão de genes entre tecidos tumorais e normais, evidenciando assim uma associação da transcrição gênica ao processo patológico e a importância de análises mais abrangentes[38,39].

A técnica de RNA-seq, ou sequenciamento de nova geração de RNAs, também conhecido como *RNAs next-generation sequencing* (*NGS*) em inglês, fornece uma maneira poderosa para a avaliação global do transcriptoma com alta resolução e um menor custo, possibilitando uma análise do perfil transcricional da doença[40].

Assim, este estudo tem por objetivo avaliar o perfil de expressão gênica de pacientes infectados pelo VHC no seu mais alto grau de cronicidade (Carcinoma

Hepatocelular) e comparar com amostras de tecidos não tumorais e de tecido hepático de pacientes com infecção não crônica pelo VHC através do sequenciamento em larga escala do transcriptoma, a fim de identificar potenciais biomarcadores de diagnóstico e prognóstico de carcinoma hepatocelular.

Objetivos

2 OBJETIVOS

2.1. Objetivo Geral

Avaliar o perfil de expressão gênica de pacientes infectados pelo VHC com carcinoma hepatocelular e comparar com amostras de tecidos não tumorais através do sequenciamento em larga escala do transcriptoma, traçando o perfil transcricional da doença, a fim de identificar potenciais biomarcadores de diagnóstico e prognóstico de carcinoma hepatocelular.

2.2. Objetivos Específicos

- Sequenciar em larga escala o transcriptoma de amostras de pacientes infectados pelo VHC no seu mais alto grau de cronicidade (carcinoma hepatocelular) e amostras de tecidos não tumorais;
- Avaliar o perfil de expressão gênica de pacientes infectados pelo VHC com carcinoma hepatocelular e amostras de tecidos não tumorais;
- Comparar o perfil de expressão gênica de pacientes infectados pelo VHC com carcinoma hepatocelular com o perfil de expressão gênica de amostras de tecidos não tumorais;
- Traçar o perfil transcricional diferencial no carcinoma hepatocelular;
- Identificar potenciais biomarcadores de diagnóstico e prognóstico de carcinoma hepatocelular.

Casuística e Métodos

3 CASUÍSTICA E MÉTODOS

3.1 Casuística

Foram analisados três fragmentos de tecido tumoral provenientes de dois pacientes portadores de CHC e infectados pelo vírus da hepatite C, sendo dois deles replicatas técnicas (amostras do mesmo paciente) e um replicata biológica (mesma condição biológica mas de pacientes diferentes) e três fragmentos de tecido hepático não tumoral (histológicamente normal) utilizado como controle provenientes de paciente acidentado e de tecido hepático transplantado não relacionados a qualquer tipo de câncer ou outra doença, sendo duas replicatas técnicas e uma replicata biológica. As amostras foram coletadas de pacientes atendidos pelo Ambulatório de Hepatites da Disciplina de Gastroenterologia Clínica do Departamento de Clínica Médica da Faculdade de Medicina de Botucatu-FMB-UNESP.

As amostras foram divididas em dois grupos:

- **Carcinoma Hepatocelular (CHC)**: Amostras dos pacientes portadores de CHC e infectados pelo vírus da hepatite C;
- **Controle (CT):** Amostras de paciente acidentado e de tecido hepático transplantado não relacionado a qualquer tipo de câncer ou outra doença.

3.2 Critérios de inclusão

Carcinoma Hepatocelular (CHC):

Os critérios de inclusão no estudo para o grupo CHC foram:

a) A presença de carcinoma hepatocelular;

b) Infecção pelo VHC genótipo 1;

c) Indivíduos maiores de 18 anos;

d) Possibilidade de coleta de sangue total do paciente durante o procedimento cirúrgico sem prejuízo ao paciente ou a exames que por ventura foram necessários;

e) Expresso entendimento e concordância em participar do estudo, manifestada pelo paciente ou pelo seu responsável legal previamente à cirurgia através do Termo de Consentimento Livre e Esclarecido.

Controle (CT):

Os critérios de inclusão no estudo para o grupo controle foram:

a) Amostra de tecido hepático considerado histológicamente normal;

b) Infecção pelo VHC negativa;

c) Indivíduos maiores de 18 anos;

d) Proveniente de enxerto de órgãos elegíveis para serem transplantados e necropsias cujo óbito não estivesse relacionado a nenhuma patologia hepática;

3.3 Critérios de exclusão

a) Pacientes que reconhecidamente apresentaram neoplasias de origem diversa às previamente estabelecidas;
b) Pacientes previamente tratados com radioterapia e/ou quimioterapia de qualquer modalidade;
c) Pacientes gestantes;
d) Pacientes que apresentaram qualquer outra coinfecção (Hepatite B, HIV, tuberculose, entre outros);
e) Pacientes ou responsáveis legais que não tenham compreendido as diretrizes da pesquisa ou que tenham se negado a participar do estudo, conforme estabelecido no Termo de Consentimento Livre e Esclarecido.

3.4 Coleta e processamento inicial dos materiais biológicos

As amostras de fragmentos tumorais hepáticos e de tecido hepático não tumoral foram coletadas durante biópsia hepática, acondicionadas e transportadas em criotubos estéreis com tampa de rosca, devidamente identificados, contendo 700µL de solução conservante *RNAlater*™. Amostras de sangue do mesmo paciente foram coletadas em tubos de tampa roxa contendo anticoagulante (EDTA). Foram

separadas alíquotas de 700µL de sangue total em dois tubos tipo eppendorf (1,5mL). Posteriormente, amostras restantes no tubo contendo EDTA, foram centrifugadas a 2.500 rpm por 15 minutos, e, separadas em dois tubos tipo eppendorf (1,5mL), cerca de 500µL de plasma, e, um tubo tipo eppendorf (1,5mL) de *buffy coat* (creme leucocitário). Identificadas através de códigos, as amostras tumorais, e respectivas amostras de sangue total, plasma e *buffy coat,* foram armazenadas em sorotecas devidamente identificadas em freezer à -80°C no Laboratório de Biologia Molecular do Hemocentro de Botucatu (FMB/UNESP).

3.5 Extração de ácido ribonucléico (RNA)

A extração de RNA dos tecidos tumorais e não tumorais foi realizada primeiramente através da maceração do tecido por lise mecânica utilizando o homogeneizador Precellys® 24-Dual e posteriormente foi realizada a extração do RNA total pelo método de uso do isotiocianato de guianidina fenol-clorofórmio com TRIzol® (Invitrogen). As condições utilizadas na extração encontram-se em anexo (ANEXO 1).

3.6 Quantificação e análise qualitativa do RNA total

Para uma pré-quantificação e a determinação do grau de pureza foi utilizado o Nanodrop ND-1000 (*NanoDrop Tecnologies*).

A concentração de RNA total extraído foi quantificada com o fluorômetro Qubit® 2.0 (Invitrogen) com o kit Qubit® RNA HS Assay.

A qualidade/integridade dos RNAs foi avaliada utilizando o Bioanalyzer 2100 (Agilent) com o kit RNA 6000 Nano LabChip; Amostras com relação RNAr 28S/18S próxima a 2 e RIN (*RNA integrity number*) próximo a 8,0 foram consideradas adequadas para o a construção das bibliotecas.

3.7 Montagem dos pools de amostras

Um *pool* de amostras para cada grupo foi realizado de acordo com a tabela a seguir:

Tabela 1. Composição dos *pools* de RNA total para a construção das bibliotecas, agrupados por condição biológica.

CONDIÇÃO	GRUPO	CONTEÚDO BIOLÓGICO
Condição 1	CHC	RNA total de 2 CHC; VHC positivos
Condição 2	CT	RNA total de 2 fígados normais; VHC negativo

3.8 Preparo da Biblioteca de cDNA e Sequenciamento

Com as amostras em *pool* então iniciamos o preparo das bibliotecas. Realizamos então a remoção de RNAs ribossômicos, a fragmentação do RNA, a síntese da primeira e segunda fita de cDNA, reparo das fitas de cDNA, adenilação, indexação e ligação de adaptadores através do kit TruSeqTM Stranded Total RNA sample preparation (Illumina), segundo recomendações do fabricante.

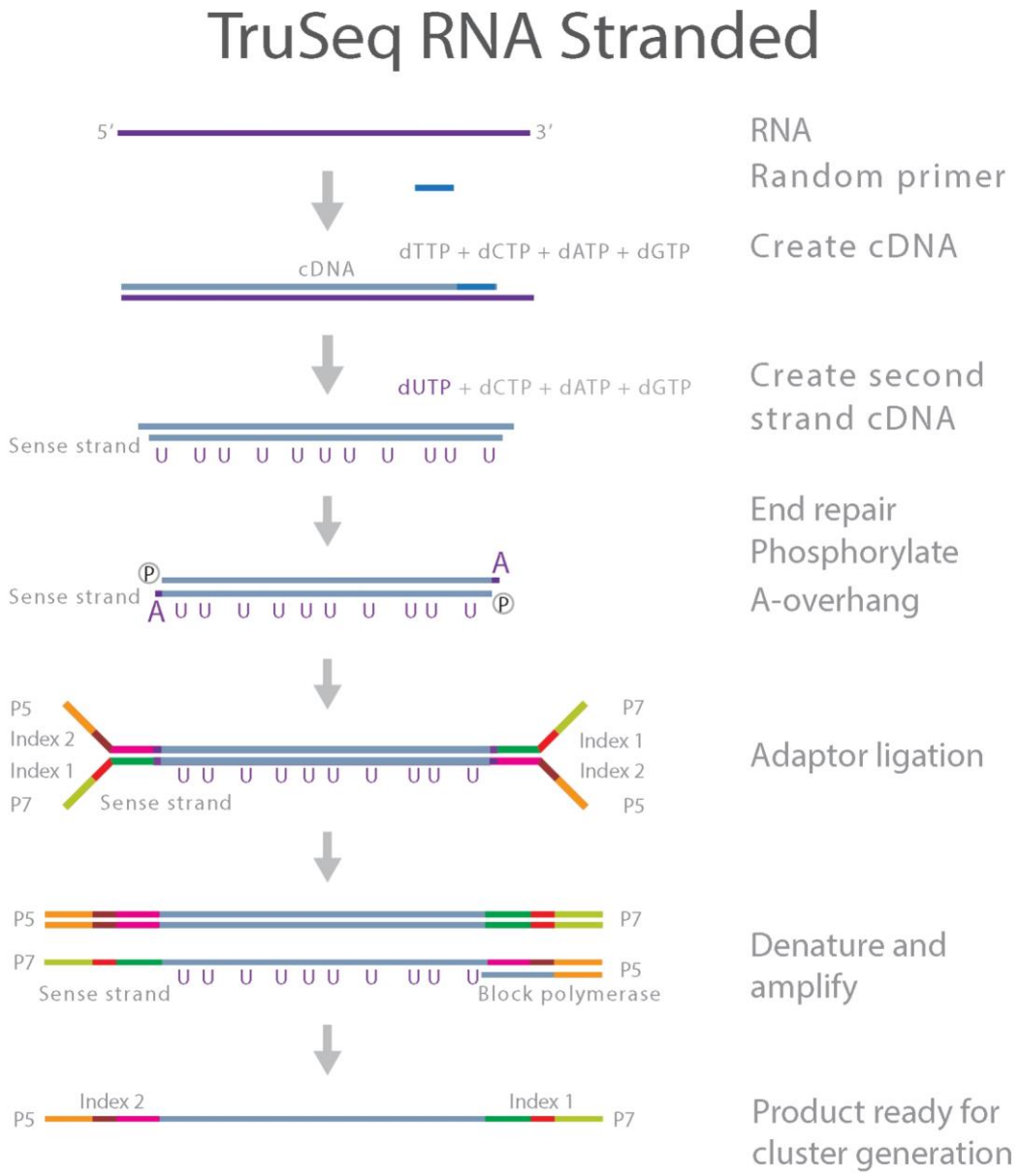
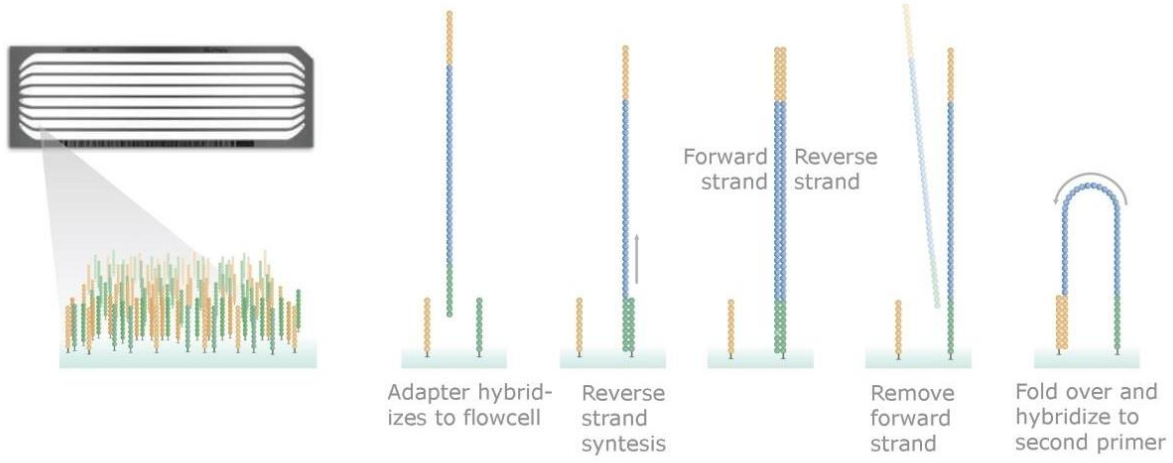
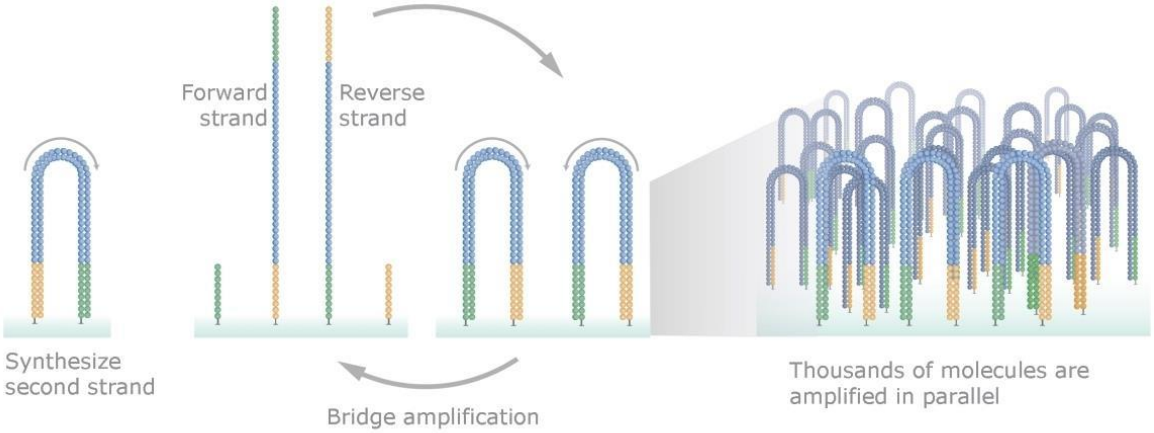
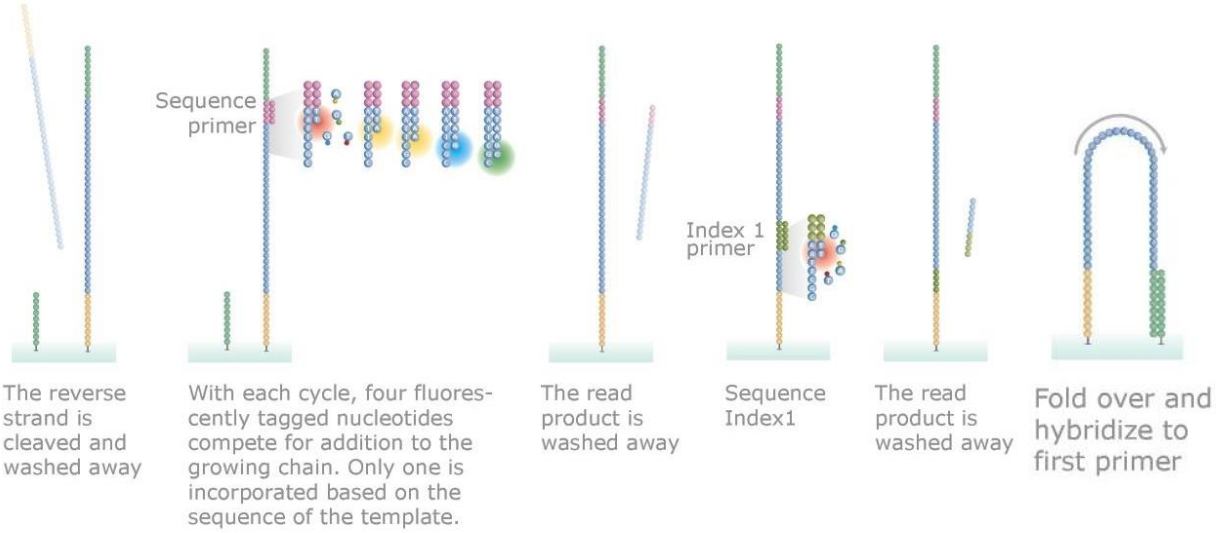


Figura 1: Representação da preparação da biblioteca através do kit TruSeqTM Stranded Total RNA sample preparation (Illumina). Primeiramente é realizada a remoção de RNAs ribossômicos e a fragmentação do RNA. Então é realizada a síntese da primeira e segunda fita de cDNA, na segunda fita são adicionados dUTPs ao invés de dTTPs possibilitando a identificação da fita de RNA de origem a qual aquele cDNA corresponde, bem como sua direção (sense ou antisense), favorecendo a identificação de RNAs antisense (reguladores); o reparo do final das fitas de cDNA com a adenilação e indexação e ligação de adaptadores. Depois é feita a desnaturação e amplificação do fragmento, e então a biblioteca está pronta para geração dos clusters e sequenciamento.

As bibliotecas foram inseridas na *flowcell* para formação de clusters e posterior sequenciamento, como ilustrado na figura 2. O sequenciamento (RNA-Seq) foi realizado na plataforma NextSeq 500 System (Illumina) utilizando o kit de sequenciamento NextSeq 500/550 High Output v2 150 ciclos (Illumina). Cada molécula foi sequenciada em ambas as extremidades (*pair-end*) e os fragmentos obtidos foram de 160 pares de bases (2x80pb). Os dados gerados foram carregados automaticamente para a nuvem da Illumina, Basespace, onde as análises foram realizadas.




Forward strand
Reverse strand
Synthesize second strand
Bridge amplification
Thousands of molecules are amplified in parallel


Sequence primer
Index 1 primer
The reverse strand is cleaved and washed away
With each cycle, four fluorescently tagged nucleotides compete for addition to the growing chain. Only one is incorporated based on the sequence of the template.
The read product is washed away
Sequence Index1
The read product is washed away
Fold over and hybridize to first primer

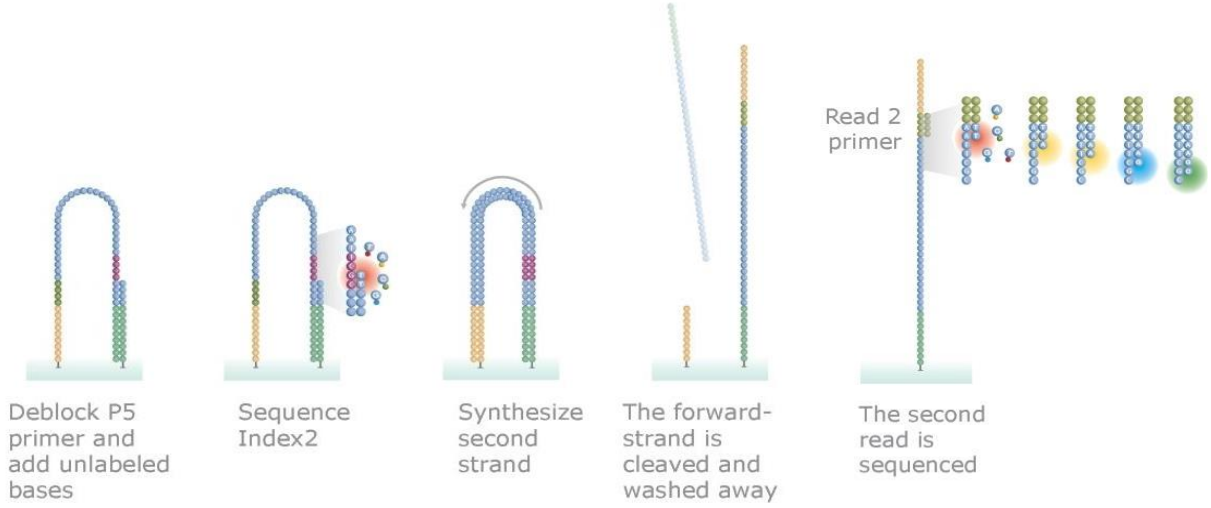

Read 2 primer
Deblock P5 primer and add unlabeled bases
Sequence Index2
Synthesize second strand
The forward-strand is cleaved and washed away
The second read is sequenced

Figura 2 - Imagem ilustrativa formação dos *clusters* e sequenciamento das bibliotecas. No primeiro quadro vemos representada uma *flowcell* utlizada nas plataformas de sequenciamento de nova geração da Illumina (NextSeq, HiSeq, MiSeq, MiniSeq). Nela encontram-se oligos já conhecidos que promovem a hibridização dos fragmentos da biblioteca através de seus adaptadores. Com os fragmentos hibridizados na flowcell inicia-se então a síntese de novas fitas através da amplificação por ponte, formando assim milhares de moléculas amplificadas em paralelo. Nucleotídeos marcados são adicionados por complementariedade, tornando possível assim o sequenciamento de cada fragmento.

3.9 Análise dos dados do sequenciamento

Para análise dos dados gerados foi utilizado o BaseSpace, uma plataforma em nuvem da Ilumina composta por diversos aplicativos baseados em softwares utilizados na bioinformática para análise de sequências. Nesta plataforma é possível transmitir os dados do sequenciamento diretamente para a nuvem, possibilitando assim a análise dos dados através do uso dos aplicativos[41].

Com o sequenciamento concluído, utilizamos o aplicativo *FastQC* (https://basespace.illumina.com/apps/706706/start) para analisar a qualidade dos dados gerados. Assim, este aplicativo utiliza a indicação de valores em gráfico onde o eixo Y apresenta a qualidade *Phred* e o eixo X à posição das bases nas *reads*. Os valores abaixo de *Phred* 20 correspondem à qualidade ruim, entre 20 e 28 qualidade intermediária e acima de 28 qualidade alta, em relação à posição das bases nas *reads*. O programa também permite a verificação da variação do tamanho das sequências geradas, a análise da distribuição das quatro bases ao longo das sequências e a distribuição de conteúdo de GC ao longo das sequências geradas.

Assim, este programa permite a verificação de diversos parâmetros de qualidade das bibliotecas, facilitando a determinação do tipo de limpeza que cada biblioteca deverá ser submetida, como: a retirada das sequências dos adaptadores usados na construção da biblioteca; retirada das caudas poli-A dos RNAs mensageiros; retirada de trechos que foram mal sequenciados e que apresentam baixa qualidade; retirada dos trechos sequenciados, mas de tamanho muito reduzido, entre outros.

Após a verificação da qualidade do sequenciamento gerado, os dados foram submetidos à análise no aplicativo *BaseSpace® Express RNA*

(https://basespace.illumina.com/apps/529529/start). O aplicativo combina os recursos do alinhador STAR 2.3.1s e as ferramentas de análise de expressão gênica diferencial DESeq2 1.0.17 em um fluxo de trabalho simples.

O programa STAR (*Spliced Transcripts Alignment to a Reference*) (STAR) alinha as *reads* a um genoma de referência com a finalidade de identificar junções *exon-exon*.

DESeq2 1.0.17[42], desenvolvido em linguagem R[43] baseia no princípio de que a contagem de *reads* que se encontram associadas a uma classe (por exemplo genes e transcritos) está linearmente relacionada à abundância dessa classe[44]. Visa realizar testes estatísticos para verificar se o efeito de diferentes condições biológicas sobre a contagem de leituras ocorreu ou não ao acaso[42].

Este programa utiliza um método estatístico paramétrico que se baseia na relação entre a média dos níveis de expressão do gene e a dispersão dos seus dados. A dispersão representa o quadrado do coeficiente de variação biológica, isso é o quadrado da diferença da expressão do gene entre as replicatas biológicas[45].

O pacote *DESeq* informa o valor da razão de expressão (*fold-change*) em logaritmo de base 2 (*log2 fold-change*) de uma condição em relação à outra. *Fold-change* é a razão entre uma condição final (teste) e a sua condição inicial (controle), FC = teste/controle. Assim, se a condição teste for maior que a condição controle, *fold-change* será maior que 1 (FC>1); se a condição controle for maior que a teste, *fold-change* será menor que 1 (FC<1). Para uma representação gráfica mais simétrica e informativa, calcula-se o logaritmo de base 2 desse valor (*log2 fold-change*) e assim, para FC superiores a 1 (expressão teste>expressão controle) *log2 fold-change* terá valores positivos; para FC inferiores a 1 (expressão teste<expressão controle) *log2 fold change* terá valores negativos.

Especificamente, o gene é considerado diferencialmente expresso, se o valor de *log2 fold change* exceder um valor de corte arbitrário pré-estabelecido ou determinado empiricamente[46]. No presente estudo, foram considerados genes diferencialmente expressos aqueles cujo *log2 fold change* seja igual ou superior a 1 (*log2 fold change* >1; pelo menos duas vezes mais expresso que o controle) e *log2 fold change* seja igual ou inferior a -1 (*log2 fold change*<1; pelo menos duas vezes

menos expresso que o controle) e, foram considerados somente os resultados com significância estatística (*p-valor*) menor que 0,05.

A análise de expressão diferencial foi realizada comparando o grupo CHC com o grupo CT com a finalidade de se comparar genes diferencialmente expressos no estágio mais crônico da infecção pela hepatite C (carcinoma hepatocelular), podendo assim ter um melhor entendimento da evolução da doença, possibilitando futuramente traçar um perfil transcricional da doença, auxiliando na descoberta de biomarcadores de diagnóstico e prognóstico da infecção.

Com os resultados obtidos nas análises de expressão diferencial, elegemos os 10 genes mais e menos expressos (top 10) e avaliamos suas funções em processos biológicos e moleculares através da plataforma online *Gene Ontology Consortium* (GO).

A plataforma GO, fundada em 1998, é um projeto em esforço colaborativo para lidar com a necessidade de descrições consistentes de produtos de genes entre bancos de dados. O projeto desenvolveu três estruturas ontológicas que descrevem produtos de genes em termos de seus processos biológicos associados, componentes celulares e funções moleculares de uma maneira independente de espécies. Há três aspectos distintos: em primeiro lugar, o desenvolvimento e manutenção das próprias ontologias; segundo, a anotação de produtos de genes, o que implica fazer associações entre as ontologias e os genes e produtos de genes nos bancos de dados de colaboração; e terceiro, o desenvolvimento de ferramentas que facilitam a criação, manutenção e uso de ontologias. Possibilitando assim uma análise funcional dos dados[47].

Resultados e Discussão

4 RESULTADOS E DISCUSSÃO

As amostras de CHC e controle foram submetidas ao sequenciamento *paired-end*, ou seja, os fragmentos foram sequenciados a partir das duas porções terminas (*end*), sendo assim então nas duas direções da molécula de cDNA fragmentada, permitindo uma maior cobertura das amostras em número de *reads*, o que reduz a chance de erros.

No total obtivemos 420.817.394 *reads*, sendo 165.649.536 de *reads* sequenciadas no tecido tumoral e 255.167.858 de *reads* sequenciadas no tecido normal. Após o alinhamento das *reads* geradas com o genoma referência (hg. 19), obtivemos um total de 152.258.390 *reads* alinhadas no tecido tumoral e 240.834.876 no tecido normal, fazendo um total de 393.093.266 *reads* alinhadas.

Os dados gerados proporcionaram uma cobertura de aproximadamente 300 vezes o transcriptoma humano, o qual possui aproximadamente 110 milhões de pares de base[40]. Os resultados detalhados encontram-se na tabela a seguir:

Tabela 2. Dados gerados pelo sequenciamento RNAseq (Illumina) das bibliotecas de cDNA produzidas a partir de tecidos hepáticos com e sem tumor.

AMOSTRA*	GRUPO	N.*CLUSTERS*	N.*READS* (R1+R2)	N.*READS* ALINHADAS
B7	**CHC**	16.021.277	32.042.554	30.703.175 (95,82%)
B8	**CHC**	8.370.532	16.741.064	15.078.676 (90,07%)
5	**CHC**	58.432.959	116.865.918	106.476.537 (91,11%)
B2	**CT**	23.282.365	46.564.730	45.093.284 (96,84%)
B6	**CT**	12.296.976	24.593.952	23.361.795 (94,99%)
4	**CT**	92.004.588	184.009.176	172.379.796 (93,68%)
	TOTAL	**210.408.697**	**420.817.394**	**393.093.263**

CHC = Grupo CHC; CT = Grupo controle. *Código identificador das bibliotecas produzidas com as amostras analisadas.

Passamos então para a análise da qualidade do sequenciamento realizado utilizando o aplicativo *FastQC*, já descrito anteriormente na metodologia.

Foram analisadas todas as sequências geradas a partir das 6 bibliotecas sequenciadas, possibilitando verificar a qualidade e composição das sequências obtidas. Podemos ver no gráfico a seguir uma ilustração da qualidade média por base dos dados de sequenciamento Illumina:

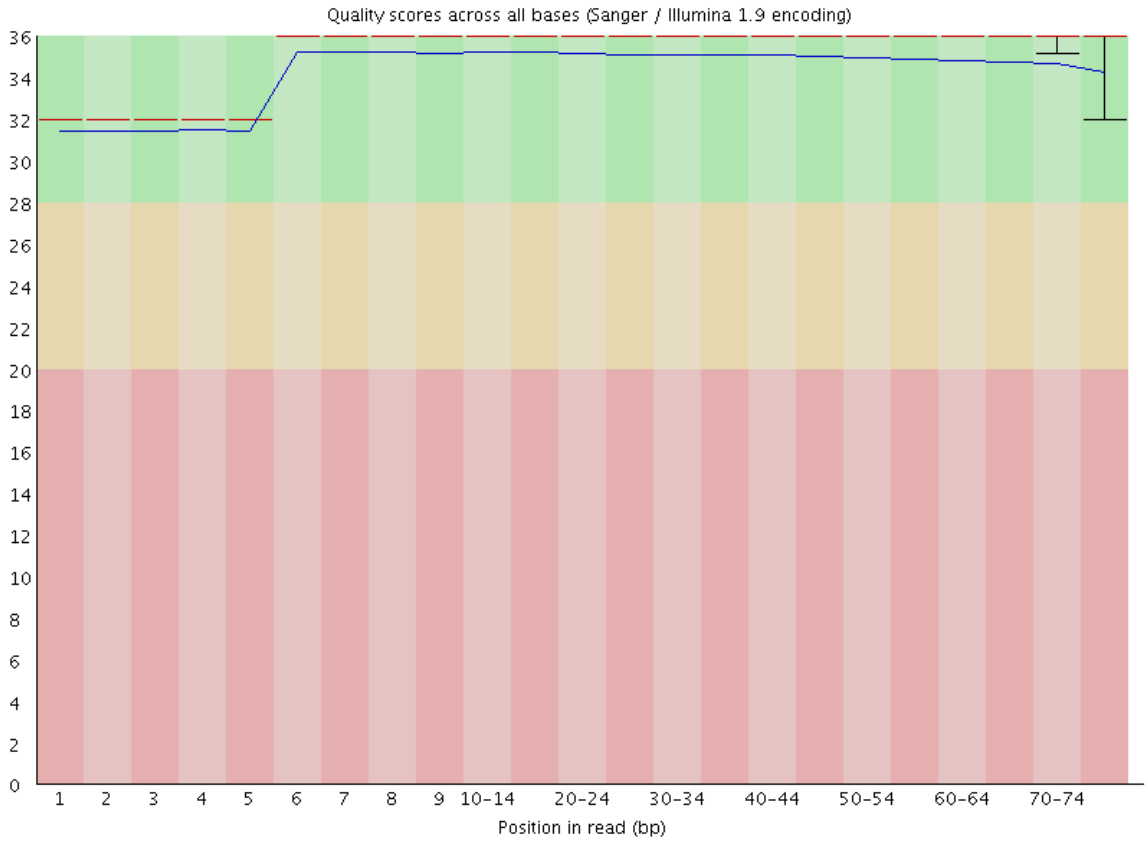


Figura 4 - Representação gráfica da qualidade média por base dos dados de sequenciamento Illumina no programa *FastQC*. O eixo Y apresenta a qualidade *Phred* e o eixo X a posição das bases nas reads. Faixa verde: boa qualidade, laranja: qualidade intermediária, rosa: baixa qualidade; a linha vermelha representa a mediana das leituras e a azul a média. Dados da análise da biblioteca B6.

Todas as bibliotecas apresentaram valores de alta qualidade (Q>30), concentrando-se na região da faixa verde. O resultado de *Phred* Q30 representa 99.9% ou mais de acurácia, ou seja, a probabilidade de que uma vez em 1000 vezes uma base seja atribuída incorretamente. Sequências qualificadas com um *phred quality* Q30, praticamente exibem zero erros e ambiguidades[48]. Assim, pode-se inferir que nossos resultados possuem um alto grau de confiança para prosseguir com as análises.

Analisando o tamanho das reads geradas, observamos uma variação entre 34 e 80 pb, representada no gráfico a seguir:

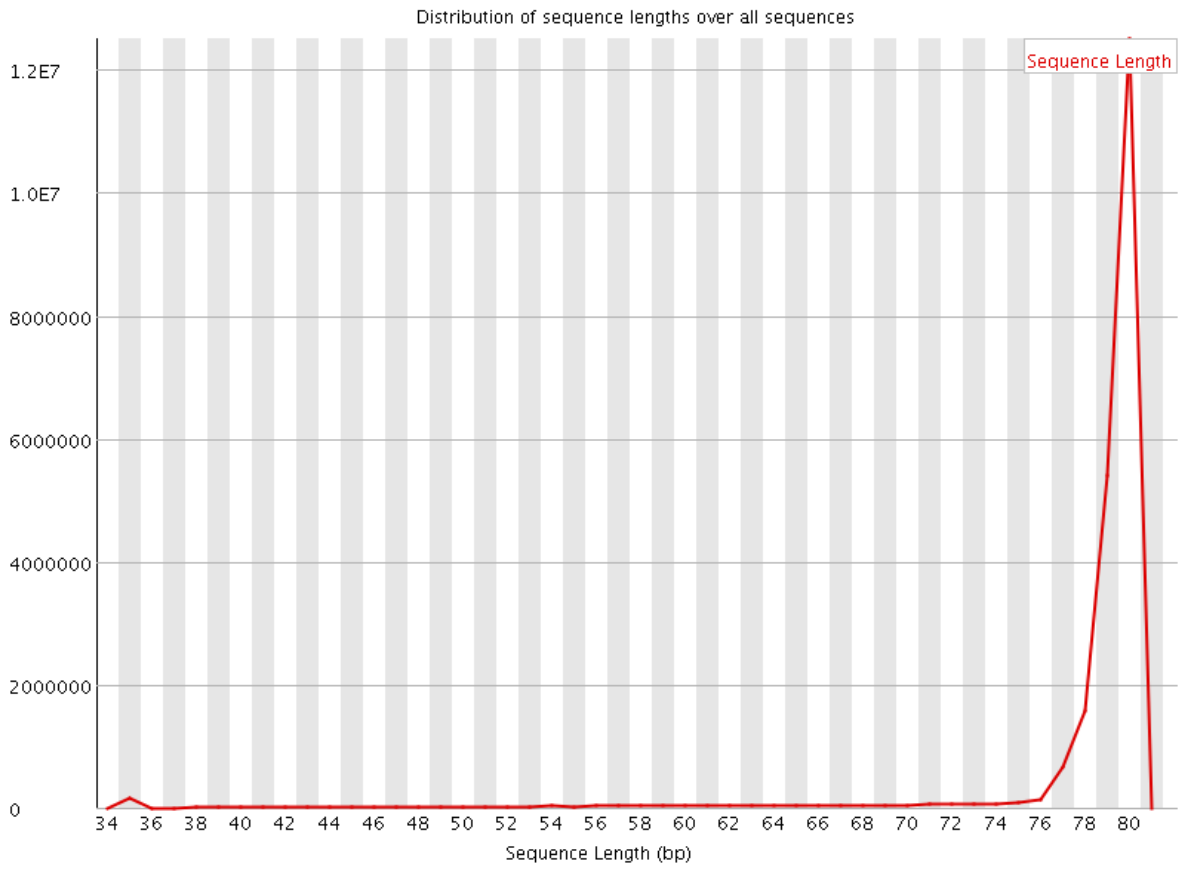


Figura 5 - Representação ilustrativa do gráfico gerado pelo programa *FastQC* da distribuição do tamanho das *reads* (eixo x) pela quantidade de sequências geradas (eixo Y). Dados da análise da biblioteca 4.

Todas as bibliotecas apresentaram similaridade na distribuição do tamanho das *reads*, e foi observada uma predominância de sequências de 74 a 80 pares de bases.

Foi possível também analisar a distribuição das quatro bases ao longo das sequências geradas.

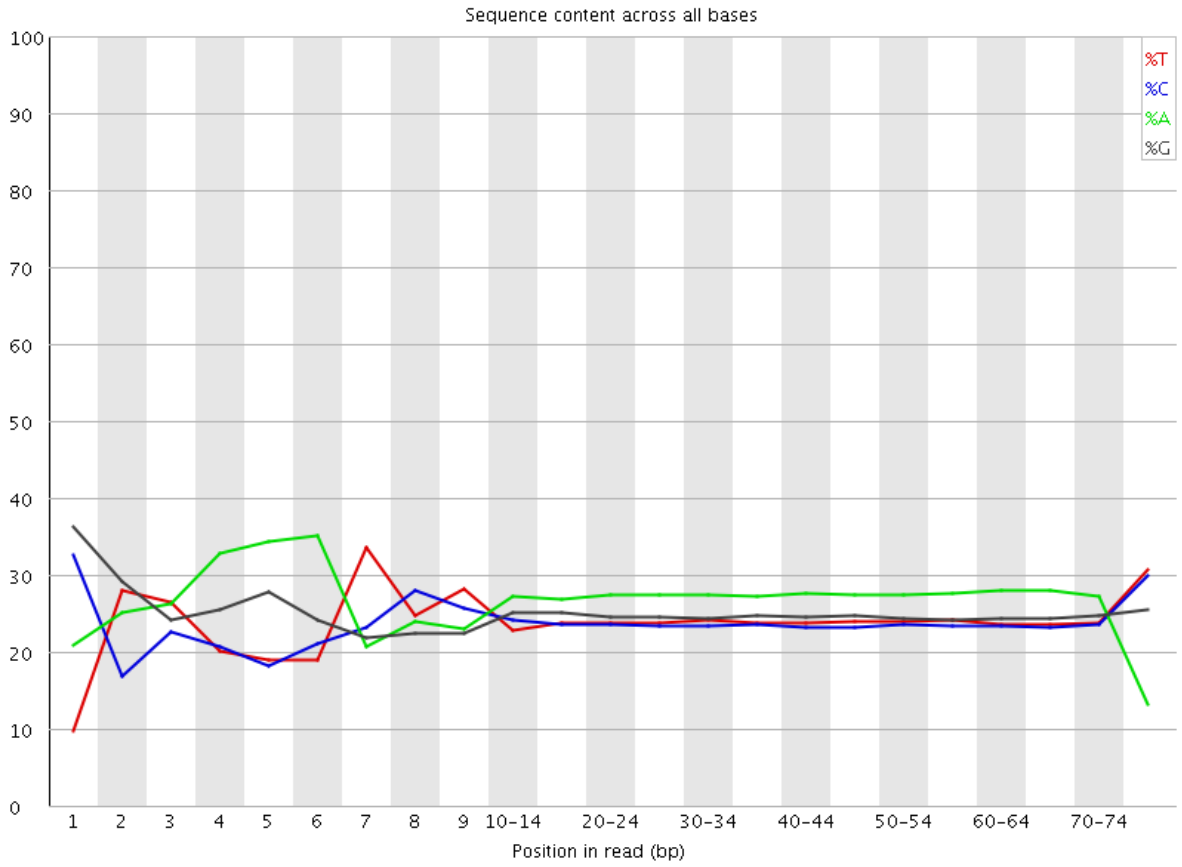


Figura 6 - Representação gráfica gerada pelo programa *FastQC* da distribuição por base do conjunto de sequencias geradas pelo sequenciamento Illumina. Dados da análise da biblioteca B6.

As bases nas leituras apresentaram-se distribuídas homogeneamente, nas dez primeiras bases mostrou-se uma discreta desproporção, mostrando uma super-representação de todas as bases nesta região, comum nesse tipo de sequenciamento. Aqui, a desproporção observada nas dez primeiras bases representa a presença do primer e, por isso, as porcentagens de cada nucleotídeo nessas regiões encontravam-se enviesadas. Em uma biblioteca aleatória é esperado que haja pouca diferença entre as diferentes bases de uma sequência, portanto, as linhas coloridas que representam cada base, devem ocorrer em paralelo. A quantidade relativa de cada base deve refletir a quantidade global destas bases no genoma, assim não deve haver desequilíbrio na proporção entre elas. Se um desbalanço entre as bases for observado ao longo de todas as posições, pode indicar que alguma sequência super-representada está contaminando sua biblioteca.

A seguir pode ser observada a distribuição do conteúdo de GC por sequência gerada e comparar a uma distribuição normal a partir do conteúdo GC do genoma humano.

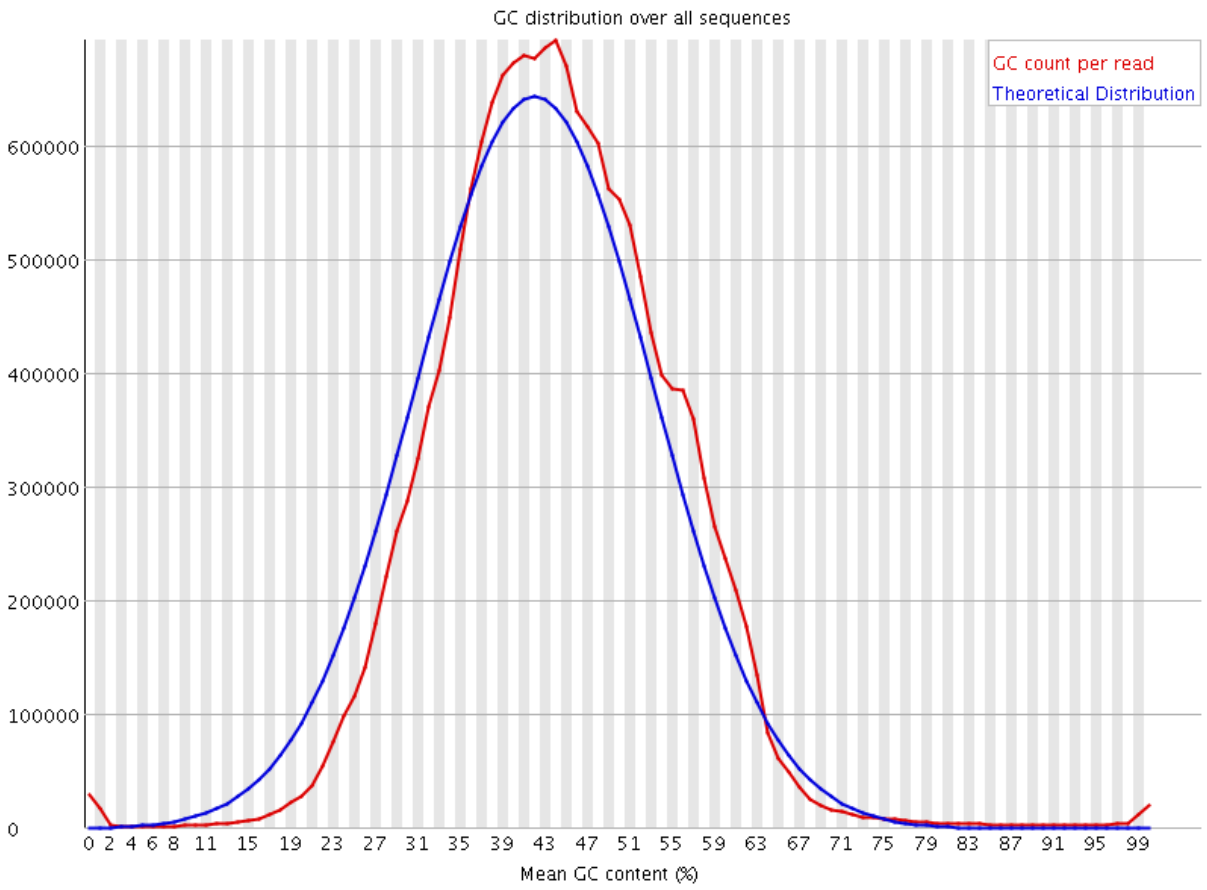


Figura 7 - Representação gráfica gerada pelo programa *FastQC* da distribuição do conteúdo de GC ao longo da extensão de cada sequência (linha vermelha) comparada a uma distribuição normal teórica do conteúdo GC do genoma humano. Dados da análise da biblioteca 5.

A distribuição do conteúdo médio de GC acompanhou o modelo de distribuição normal gerado pelo programa para o genoma humano. Tais resultados indicam não haver contaminação, ou presença de outro tipo de subgrupo enviesado de sequências em nenhuma das bibliotecas analisadas.

Com os parâmetros do *FastQC* analisados, concluímos que as sequências geradas estão com uma boa qualidade, assim seguimos para a análise de expressão diferencial através do aplicativo *BaseSpace® Express RNA*, já descrito anteriormente.

Realizamos então a análise de expressão diferencial das amostras comparando-as em dois pares, representando então a expressão gênica diferencial entre tecido tumoral de infectados pelo VHC e tecido hepático não tumoral (normal) de não infectados pelo VHC, e entre o tecido tumoral de infectados pelo VHC e tecido hepático não tumoral de infectados pelo VHC sem cronicidade.

Tabela 3. Comparações entre as bibliotecas de cDNA provenientes de amostras de tecido hepático com diferentes condições biológicas.

COMPARAÇÃO	BIBLIOTECAS
Carcinoma Hepatocelular x Normal	(4, B2, B6) x (5, B7, B8)

Carcinoma Hepatocelular *versus* Controle

Comparando as amostras de carcinoma hepatocelular com as amostras de tecido hepático não tumoral (grupo controle), identificamos 4.792 genes diferencialmente expressos (GDE) com significância estatística. Podemos observar em vermelho no diagrama de dispersão a seguir os genes diferencialmente expressos no grupo carcinoma hepatocelular *versus* grupo controle.

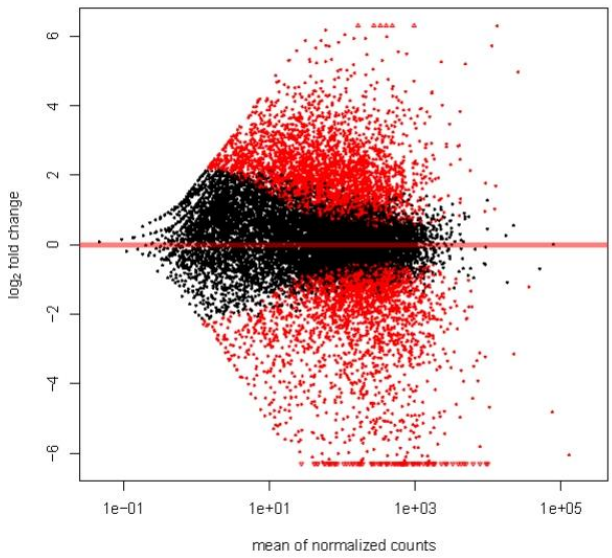


Figura 8 – Diagrama de dispersão do *log2 fold-change* pela média da contagem normalizada dos genes, evidenciando os genes diferencialmente expressos (cor vermelha). Acima de zero estão os genes sobreexpressos e abaixo de zero os genes subexpressos no grupo de amostras de CHC, quando comparado com amostras de fígado normal.

Em análise preliminar, foram selecionados apenas genes que se encontraram diferencialmente expressos (sub ou sobreexpressos) em pelo menos uma vez em relação ao controle (*log2 fold-change* ≤-1 ou ≥1) e um p-value > 0,05 (estatisticamente significante). Assim, foram selecionados 4.299 genes diferencialmente expressos, sendo 2.378 sobreexpressos e 1.921 subexpressos (Figura 9).

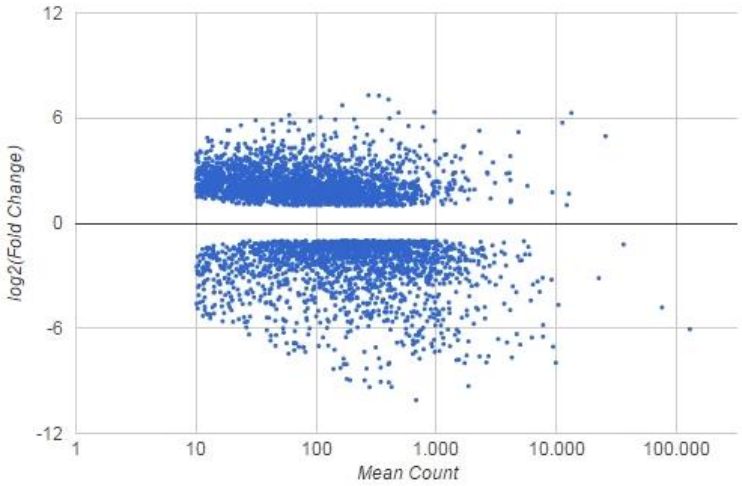


Figura 9 – Diagrama de dispersão do *log2 fold-change* pela média observada para o gene, evidenciando os genes diferencialmente expressos (pontos azuis) em pelo menos uma vez quando comparado ao controle. Acima de zero estão os genes sobreexpressos e abaixo de zero os genes subexpressos no grupo de amostras de CHC, quando comparado com amostras de fígado normal.

Os genes diferencialmente expressos foram classificados de acordo com sua função em três categorias: processos biológicos (atividade biológica com o qual o gene ou seu produto contribui), função molecular (atividade bioquímica do gene ou produto) e componente celular (local na célula onde o gene ou seu produto é ativo), podendo um gene ser encontrado em uma ou mais categorias. Essa avaliação foi realizada através da análise de ontologia gênica pelo *Gene Ontology Consortium (GO)*.

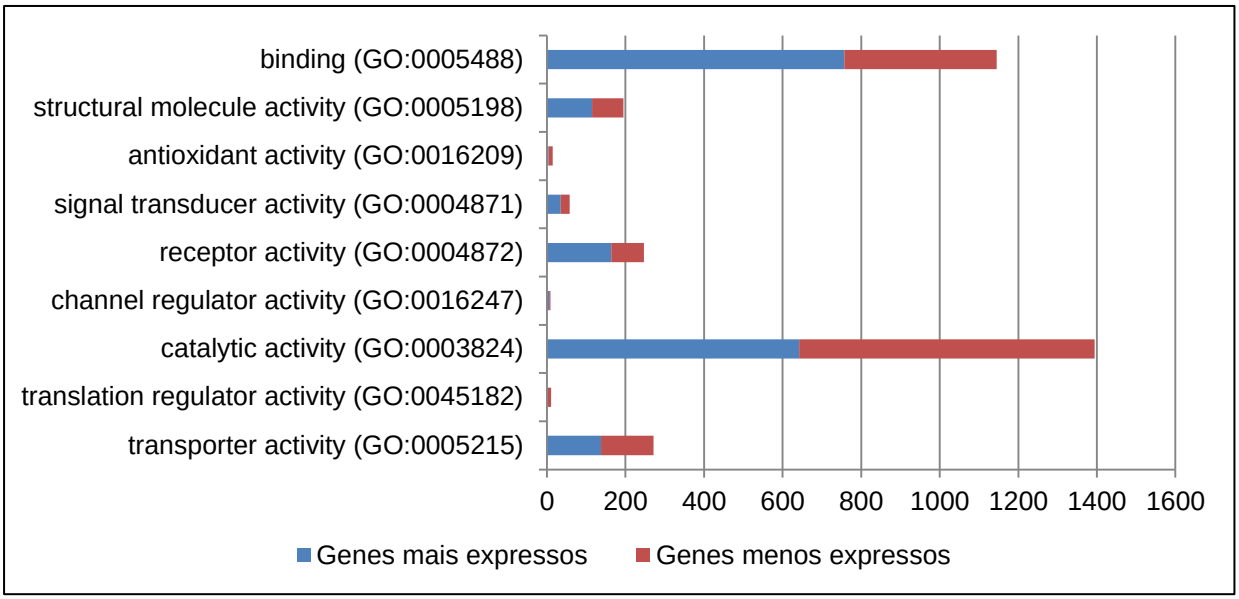


Figura 10 – Categorização dos genes diferencialmente expressos entre os grupos CHC e fígado normal de acordo com GO (*Gene Ontology Consortium*) e distribuição conforme sua função molecular. Função molecular (eixo y) pela quantidade de genes envolvidos (eixo x).

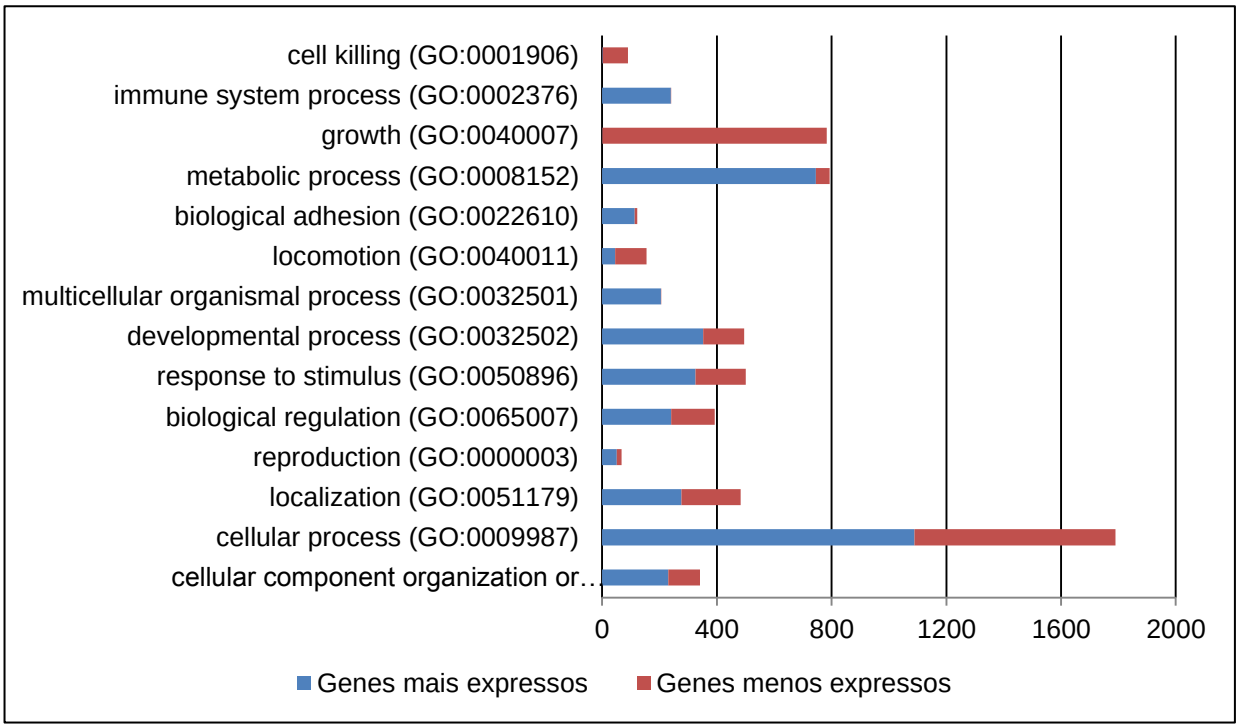


Figura 11 – Categorização dos genes diferencialmente expressos entre os grupos CHC e fígado normal de acordo com GO (*Gene Ontology Consortium*) e distribuição conforme seu processo biológico. Processo biológico (eixo y) pela quantidade de genes envolvidos (eixo x).

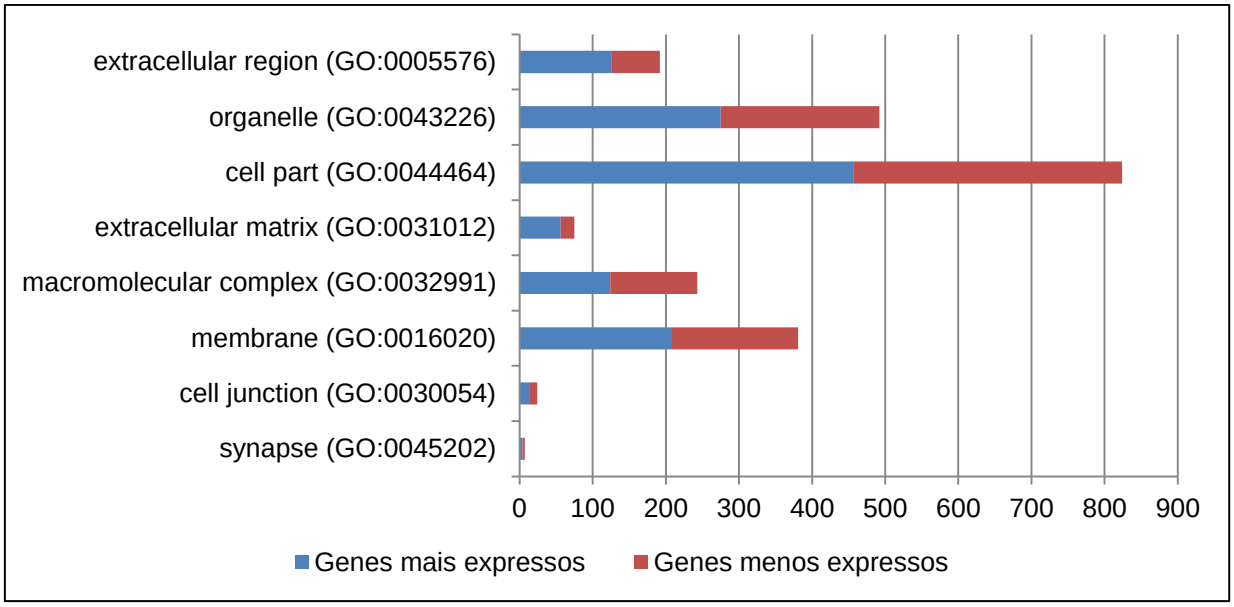


Figura 12 – Categorização dos genes diferencialmente expressos entre os grupos CHC e fígado normal de acordo com GO (*Gene Ontology Consortium*) e distribuição conforme seu componente celular. Componente celular (eixo y) pela quantidade de genes envolvidos (eixo x).

Tabela 4. Análise funcional dos genes mais e menos expressos no grupo CHC em relação ao grupo controle, divididos em três categorias de acordo com: processo biológico, função molecular e componente celular através do *Gene Ontology Consortium* (GO).

Categoria	GO ID	Classificação GO	N.GENES + exp.	N.GENES – exp.
Função Molecular	GO:0005215	transporter activity	138	133
	GO:0045182	translation regulator activity	1	10
	GO:0003824	catalytic activity	642	752
	GO:0016247	channel regulator activity	6	3
	GO:0004872	receptor activity	164	83
	GO:0004871	signal transducer activity	35	23
	GO:0016209	antioxidant activity	4	11
	GO:0005198	structural molecule activity	115	80
	GO:0005488	binding	757	388
Processo Biológico	GO:0071840	cellular component organization or biogenesis	231	110
	GO:0009987	cellular process	1089	701
	GO:0051179	localization	277	206
	GO:0000003	reproduction	51	17
	GO:0065007	biological regulation	241	152
	GO:0050896	response to stimulus	326	175
	GO:0032502	developmental process	353	142
	GO:0032501	multicellular organismal process	205	1
	GO:0040011	locomotion	46	109
	GO:0022610	biological adhesion	114	9

	GO:0008152	metabolic process	745	48
	GO:0040007	growth	1	782
	GO:0002376	immune system process	240	1
	GO:0001906	cell killing	2	88
Componente Celular	GO:0045202	synapse	4	3
	GO:0030054	cell junction	14	10
	GO:0016020	membrane	207	174
	GO:0032991	macromolecular complex	124	119
	GO:0031012	extracellular matrix	56	19
	GO:0044464	cell part	457	367
	GO:0043226	organelle	275	217
	GO:0005576	extracellular region	126	66

Dos genes apresentados como diferencialmente expressos provenientes da análise comparativa entre o grupo carcinoma hepatocelular e grupo controle (fígado normal), foram selecionados os 10 mais expressos e os 10 menos expressos para estudos mais detalhados, como apresentado na Tabela 5.

Tabela 5. Top 10 genes mais e menos expressos entre os grupos CHC e fígado normal

Genes menos expressos			Genes mais expressos		
Gene	***log2 fold-change***	**p-value**	**Gene**	***log2 fold-change***	**p-value**
CFHR5	-10,1	9,56E-42	COL11A1	7,3	4,61E-09
HFE2	-9,36	2,58E-27	SFRP4	7,28	9,18E-05
HPR	-9,35	3,63E-41	SFRP2	7,06	3,58E-03
SLC38A4	-9,3	4,64E-90	LRRC15	6,73	3,94E-05
SLC10A1	-9,09	2,71E-20	CCL18	6,34	5,23E+00
LPA	-9,06	6,28E-27	ADAMDEC1	6,31	2,19E+02
GSTM1	-8,97	2,76E-22	COL1A1	6,29	1,16E-16
ACOT12	-8,96	2,04E-23	COL10A1	6,17	2,19E+01
CYP1A1	-8,89	1,29E-22	CTHRC1	6,04	3,61E-03
SMLR1	-8,33	8,18E-16	OLR1	5,99	8,83E-02

Os genes *COL11A1, COL1A1, COL10A1* e *CTHRC1* são conhecidos por fazer parte da família do colágeno, maior componente intersticial da matriz extracelular. Estudos associam a família do colágeno ao desenvolvimento de fibrose hepática, cirrose e hepatocarcinoma invasivo, e apontam que genes pertencentes a esta família desempenham um importante papel na formação de tumor[49,50,51].

A superexpressão do gene *COL11A1* já foi relacionada com a invasão e progressão tumoral, e foi altamente associada a diversos tipos de câncer, entre eles

o câncer de ovário, pâncreas, fígado, rins, gástrico, pulmonar, oral, colorretal, mamário, glioblastoma, entre outros[51,52,53].

A expressão diferencial do gene *COL1A1* já foi associada a cirrose hepática, câncer renal, de ovário, mamário, entre outros[49,54].

A superexpressão do gene *COL10A1* já foi associado à diversos tipos de tumores, entre eles o câncer de mama, estômago, pulmão, pâncreas, bexiga, ovário, esôfago, colorretal, fígado, próstata, cérebro, rins, entre outros[50].

Ainda, tumores de mama, melanoma invasivo, CHC, câncer gástrico, câncer de pulmão de células não pequenas, tumor estromal gastrointestinal, colorretal apresentaram superexpressão de *CTHRC1*[56,57,58,59,60]. Estudos sugerem que este gene possa contribuir para a reparação tecidual através da deposição de matriz de colágeno e promover a migração celular[61].

Os genes *SFRP4* e *SFRP2* pertencerem a família das proteínas secretadas relacionadas à *frizzled*, glicoproteínas que agem como moduladores da via de sinalização de Wnt[62,63]. Alguns estudos apontaram os genes *SFRPs* como supressores tumorais por estes encontrarem-se hipermetilados em diversos tipos de câncer[63]. No entanto, outros estudos relacionaram *SFRPs* com a proliferação celular em alguns tumores[62]. O gene *SFRP4* tem sido associado a proliferação e diferenciação celular, implicando em um importante papel na carcinogênese[64]. A superexpressão do gene *SFRP2* tem sido associada à promoção do crescimento celular, favorecendo o desenvolvimento de alguns tumores[62]. Estudos verificaram uma expressão alterada desses genes em diversos tipos de câncer, entre eles o câncer de ovário, próstata, pâncreas, rins, endometrial, cervical, bexiga, colorretal, entre outros[62,63,64].

LRRC15 codifica uma proteína transmembrana e está relacionado a invasão celular. Este gene já foi associado a diversos tipos de câncer, principalmente ao câncer de mama e próstata[65,66,67,68].

O gene *CCL18* codifica uma quimiosina produzida por células derivadas de monócitos[69]. Uma produção excessiva de *CCL18* em macrófagos já foi observada em doenças inflamatórias crônicas e fibroses[70,71]. Estudos também verificaram

alteração da sua expressão em câncer de ovário, gástrico e glioma[72,73,74,75]. Outro estudo conduzido em câncer de mama sugere que este gene esteja associado a promoção da adesão a matriz extracelular levando invasão celular[69] e angiogênese[76]. No entanto, o papel do gene *CCL18* na progressão tumoral ainda é controverso. Alguns autores sugerem que este gene funcione como um imunossupressor em câncer de ovário[75,77] enquanto outros associaram este gene com a sobrevida prolongada em pacientes com adenocarcinoma gástrico[73].

O gene *ADAMDEC1* codifica uma proteína pertencente a família das metaloproteinases desintegrinas, expressas por macrófagos e células dendríticas[78], envolvidas em diversas funções biológicas, entre elas a fertilização, adesão, migração e proteólise[79]. Essas proteínas foram relacionadas a progressão tumoral e a tumorigênese de diversos tipos de tumores. Estudos encontraram associações entre *ADAMDEC1* e câncer colorretal, CHC, câncer gástrico em estágios avançados, câncer neurológico, entre outros[78,79].

O receptor *scavanger* semelhante a lectina, codificado pelo gene *OLR1*, é primariamente expresso em células vasculares e órgãos altamente vascularizados[80,81]. Sua ativação pode ocorrer através de estímulos indicativos de dislipidemia, inflamação e danos iniciando diversas cascatas de sinalização como MAPSKs, proteínas quinases e fatores de transcrição como NF-kB e AP-1[82,83]. A superexpressão de *OLR1* relacionada a obesidade foi observada em alguns tipos tumorais como câncer de mama e próstata. Estudos recentes sugerem que a ação do gene *OLR1* através da mediação da sinalização inflamatória pelo NF-kB esteja altamente associada a carcinogênese[80,84].

É digno de nota que os top 10 genes mais expressos observados no presente estudo foram repetidamente associados a diversos tumores, contudo, esses dados são provenientes de estudos individuais, onde apenas um ou poucos genes foram investigados. Nossos resultados apontam superexpressão desses genes simultaneamente, sugerindo que estes possam atuar juntos no processo tumoral.

Dentre os top 10 genes menos expressos, o gene *CFHR5* faz parte da família das proteínas FH (Fator H), agindo como um regulador da ativação do sistema complemento[85,86,87]. Este gene codifica uma glicoproteína plasmática produzida no

fígado[86]. Variações nesse gene foram encontradas em diversas doenças, como a síndrome hemolítico-urêmica atípica e nefropatias[86] e em cânceres como linfoma de células B[85]. No entanto a função da proteína produzida por esse gene ainda não é bem conhecida[86].

HFE2 codifica uma proteína chamada hemojuvelina, sintetizada no fígado, coração e músculos. Esta proteína está envolvida no metabolismo do ferro e também na regulação de hepcidina, a qual está envolvida na manutenção dos níveis regulares de ferro na circulação. Defeitos nesse gene estão associados ao desenvolvimento da hemocromatose juvenil, onde os altos níveis de ferro podem levar ao hipogonadismo hipogonadotrófico, fibrose ou cirrose hepática e cardiomiopatia[88,89].

O gene *HPR* é responsável pela codificação da proteína relacionada a haptoglobina, conhecida por ligar-se à hemoglobina com alta afinidade, podendo contribuir para a depuração da hemoglobina, permitindo a reciclagem hepática do ferro[90,91]. Estudos relacionaram este gene ao câncer de mama[92].

SLC38A4 codifica um transportador de aminoácidos sódio dependente encontrado predominantemente no fígado[93]. Estudos relataram a ocorrência de polimorfismos nesse gene em casos de câncer de pulmão[94].

SLC10A1 codifica uma glicoproteína pertencente a família dos cotransportadores de sódio e ácido biliar e que participa da circulação entero-hepática dos ácidos biliares[95]. Estudos verificaram a presença da proteína gerada por este gene na reabsorção de ácidos biliares na membrana basolateral dos hepatócitos[96].

Codificada pelo gene *LPA*, a serina proteinase possui uma ação inibitória da atividade do ativador de plasminogênio tecidual tipo I e constitui-se de uma porção substancial lipoprotéica clivada proteoliticamente, gerando fragmentos que se fixam a lesões ateroscleróticas e promovem a formação de trombos[97].

O gene *GSTM1* é conhecido por ser membro de uma super família de GSTs (Glutationa-S-tranferases) desenvolvendo um importante papel na biotransformação e conferindo proteção contra uma ampla gama de substâncias tóxicas. Entretanto,

pode ocasionalmente levar a formação de componentes ainda mais tóxicos que o substrato inicial, levando ao desenvolvimento de doenças. Variações nesse gene foram observadas e demonstraram diferença na susceptibilidade a exposição de tóxicos ambientais, resistência a tratamentos quimioterápicos, variabilidade na resposta a droga, manifestação de diversas doenças e desenvolvimento de câncer[98]. Estudos verificaram alterações nesse gene em diversos tipos de câncer, como câncer de cabeça e pescoço e câncer de pulmão[99,100].

ACOT12 codifica proteínas Acyl-CoA thioesterase tipo 12, que fazem parte de um grupo de enzimas que hidrolisam a Coenzima A (CoA). Doenças associadas ao *ACOT12* incluem leucoencefalopatia com substância branca evanescente. Entre suas vias estão a biossíntese e metabolismo de *Fatty* Acyl-CoA. As anotações de GO (*Gene Ontology*) relacionadas com este gene incluem a ligação de lípidos e a atividade de acetil-CoA hidrolase[101].

CYP1A1 codifica uma enzima membro da família *CYP1* e desempenha um papel fundamental no metabolismo de drogas e produtos químicos ambientais. A enzima *CYP1A1* é a mais ativa entre os *CYPs* na metabolização de pró-carcinogênicos, podendo contribuir para a oncogênese[102]. Estudos relacionaram esse gene a vários tipos de câncer, como o carcinoma renal, pulmonar e mamário[102,103,104].

O gene SMLR1 codifica uma proteína rica em leucina e sua função ainda é desconhecida pela falta de estudos relacionados a este gene[105].

Dentre os top 10 genes observados menos expressos na presente casuística, somente *GSTM1*, *SLC38A4, HPR, CFHR5* e *CYP1A1* foram anteriormente associados a processos tumorais. Os demais genes, embora não tenham sido observados em estudos sobre a gênese ou progressão tumoral, estão intimamente associados a funções hepáticas ou relacionadas a componentes produzidos pelo fígado. Esses achados podem sugerir que a presença do tumor provoque a desregulação nesses genes, contudo, estudos mais específicos devem ser conduzidos para a confirmação dessa hipótese.

O NGS (Next-generation sequencing) produz um elevado número de dados, com riqueza quantitativa e qualitativa, estimulando à sua exploração com diferentes

ferramentas computacionais e diferentes abordagens biológicas. Os achados do presente estudo evidenciaram, em uma análise primária, genes com potencial para servirem como biomarcadores em carcinoma hepatocelular. Talvez, a associação de alguns dos top 10 genes considerados em nosso estudo seja uma estratégia a ser considerada nesse sentido, e essa será a próxima etapa do presente estudo, além de outras análises computacionais.

Conclusão

5 CONCLUSÃO

O sequenciamento global do transcriptoma dos grupos carcinoma hepatocelular e controle, e posteriormente a análise de expressão diferencial entre os dois grupos possibilitou a identificação de 4.792 genes diferencialmente expressos;

Os 10 genes mais expressos na casuística analisada já foram associados a diferentes tipos tumorais;

Dentre os 10 genes menos expressos, alguns desempenham papeis importantes na função hepática ou são relacionados a componentes produzidos pelo fígado.

Os resultados do presente estudo podem sugerir que os genes COL11A1, SFRP4, SFRP2, LRRC15, CCL18, ADAMDEC1, COL1A1, COL10A1, CTHRC1 e OLR1, encontrados superexpressos, possam atuar em conjunto no processo tumoral, servindo como possíveis marcadores moleculares da doença.

Referências

6 REFERÊNCIAS

1. Czepiel J, Biesiada G, Mach T. Viral hepatitis C. Pol Arch Med Wewn. 2008;118(12):734-40.

2. Organização Mundial da Saúde. Hepatitis C [acesso em 12 out 2016]. Disponível em: http://www.who.int/mediacentre/factsheets/fs164/en.

3. Brasil Ministério da Saúde [Internet]. Protocolo clínico e diretrizes terapêuticas para hepatite C e coinfecções [acesso em 12 out 2016] Disponível em: http://www.aids.gov.br/sites/default/files/anexos/publicacao/2015/58192/arquivoweb4_pcdt_17_05_2016_pdf_31085.pdf

4. Choo Q, Kuo G, Weiner A, Overby L, Bradley D, Houghton M. Isolation of a cDNA clone derived from a blood-borne non-A, non-B viral hepatitis genome. Science. 1989;244(4902):359-362.

5. Lindenbach B, Rice C. Unravelling hepatitis C virus replication from genome to function. Nature. 2005;436(7053):933-938.

6. Rosen H, Gretch D. Hepatitis C virus: current understanding and prospects for future therapies. Molecular medicine today. 1999;5(9):393-399.

7. Moradpour D, Penin F, Rice C. Replication of hepatitis C virus. Nature Reviews Microbiology. 2007;5(6):453-463.

8. Smith D et al. Expanded classification of hepatitis C virus into 7 genotypes and 67 subtypes: updated criteria and genotype assignment. Hepatology. 2014;59(1):318-327.

9. Simmonds P et al. Consensus proposals for a unified system of nomenclature of hepatitis C virus genotypes. Hepatology. 2005;42(4):962-973.

10. Pavio N, Lai M. The hepatitis C virus persistence: how to evade the immune system?. Journal of biosciences. 2003;28(3):287-304.

11. Müller H et al. Peripheral blood leukocytes serve as a possible extrahepatic site for hepatitis C virus replication. Journal of General Virology. 1993;74(4):669-676.

12. Navas M et al. Dendritic cell susceptibility to hepatitis C virus genotype 1 infection. Journal of medical virology. 2002;67(2):152-161.

13. Hernandez F et al. Autoimmune thrombocytopenia associated with hepatitis C virus infection. Acta haematologica. 1998;99(4):217-220.

14. Hamaia S, Li C, Allain JP. The dynamics of hepatitis C virus binding to platelets and 2 mononuclear cell lines. Blood. 2001;98(8):2293-2300.

15. Burlone M, Budkowska A. Hepatitis C virus cell entry: role of lipoproteins and cellular receptors. Journal of General Virology. 2009;90(5):1055-1070.

16. Pileri P et al. Binding of hepatitis C virus to CD81. Science. 1998;282(5390):938-941.

17. Scarselli E et al. The human scavenger receptor class B type I is a novel candidate receptor for the hepatitis C virus. The EMBO journal. 2002;21(19):5017-5025.

18. Evans M, von Hahn T, Tscherne D, Syder A, Panis M, Wölk B et al. Claudin-1 is a hepatitis C virus co-receptor required for a late step in entry. Nature. 2007;446(7137):801-805.

19. Liu S et al. Tight junction proteins claudin-1 and occludin control hepatitis C virus entry and are downregulated during infection to prevent superinfection. Journal of virology. 2009;83(4):2011-2014.

20. Agnello V, Abel G, Elfahal M, Knight G, Zhang Q. Hepatitis C virus and other Flaviviridae viruses enter cells via low density lipoprotein receptor. Proceedings of the National Academy of Sciences. 1999;96(22):12766-12771.

21. Cormier E, Tsamis F, Kajumo F, Durso R, Gardner J, Dragic T. CD81 is an entry coreceptor for hepatitis C virus. Proceedings of the National Academy of Sciences. 2004;101(19):7270-7274.

22. Halliday J, Klenerman P, Barnes E. Vaccination for hepatitis C virus: closing in on an evasive target. Expert review of vaccines. 2001;10(5): 659-672.

23. Ozaras R, Tahan V. Acute hepatitis C: prevention and treatment. Expert review of anti-infective therapy. 2009;7(3):351-361.

24. Pár A. Prophylaxis and treatment of chronic viral hepatitis as the prevention of hepatocellular carcinoma. Orvosi hetilap. 2008;150(1):19-26.

26. Rockey D, Friedman S. Hepatic fibrosis and cirrhosis. New York. 2007;87-109.

27. Poynard T et al. Natural history of liver fibrosis progression in patients with chronic hepatitis C. The Lancet. 1997;349(9055):825-832.

28. Thomas D et al. The natural history of hepatitis C virus infection: host, viral, and environmental factors. Jama. 2000;284(4):450-456.

29. Cancer Today [Internet]. Estimated number of incidence cases in 2012 [acesso em 15 ago 2016]. Disponível em: http://gco.iarc.fr/today/home

30. El–Serag H, Rudolph K. Hepatocellular Carcinoma: Epidemiology and Molecular Carcinogenesis. Gastroenterology. 2007;132(7):2557-2576.

31. Gomes M et al. Hepatocellular carcinoma: epidemiology, biology, diagnosis, and therapies. Revista da Associação Médica Brasileira. 2013;59(5):514-524.

32. Tateishi R, Omata M. Hepatocellular carcinoma in 2011: genomics in hepatocellular carcinoma—a big step forward. Nature Reviews Gastroenterology and Hepatology. 2012;9(2):69-70.

33. Crick, F. On protein synthesis. Symp Soc Exp Biol. 1958;8.

34. Birney E, Stamatoyannopoulos J, Dutta A, Guigó R, Gingeras T, Margulies E et al. Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature. 2007;447(7146):799-816.

35. Stein L. Human genome: end of the beginning. Nature. 2004;431(7011):915-916.

36. Felix J et al. Genoma Funcional. Biotecnologia Ciência e Desenvolvimento. 2002;(24):60-67.

37. Passos G, Jordan C. Projeto Transcriptoma. Biotecnologia Ciência e Desenvolvimento. 2000;12:34-37.

38. Smith M et al. Identification of novel tumor markers in hepatitis C virus-associated hepatocellular carcinoma. Cancer research. 2003;63(4):859-864.

39. Skawran B et al. Gene expression profiling in hepatocellular carcinoma: upregulation of genes in amplified chromosome regions. Modern Pathology. 2008;21(5):505-516.

40. Zhang FG, He ZY, Wang Q. Transcriptome profiling of the cancer and normal tissues from gastric cancer patients by deep sequencing. Tumor Biology. 2014;35(8):7423-7427.

41. Basespace Sequence Hub Support – Illumina [Internet]. Basespace Sequence Hub [acesso em 09 out 2016]. Disponível em: http://support.illumina.com/ sequencing/sequencing_software/basespace.html

42. Anders SHuber W. Differential expression analysis for sequence count data. Genome Biol. 2010;11(10):R106.

43. Team, R et al. R: A language and environment for statistical computing. 2013.

45. Love M, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome biology. 2014;15(12):1.

47. Gene Ontology Consortium [Internet]. Documentation [acesso em 12 out 2016]. Disponível em: http://geneontology.org/page/documentation

49. Hayashi M et al. Identification of the collagen type 1 alpha 1 gene (COL1A1) as a candidate survival-related factor associated with hepatocellular carcinoma. BMC cancer. 2014;14(1):1.

50. Chapman K, Prendes M, Sternberg H, Kidd J, Funk W, Wagner J et al. COL10A1 expression is elevated in diverse solid tumor types and is associated with tumor vasculature. Future Oncology. 2012;8(8):1031-1040.

51. Wu Y et al. COL11A1 promotes tumor progression and predicts poor clinical outcome in ovarian cancer. Oncogene. 2014;33(26):3432-3440.

52. Simon E et al. Transcriptomic characterization of fibrolamellar hepatocellular carcinoma. Proceedings of the National Academy of Sciences. 2015;112(44):E5916-E5925.

53. Jia D et al. A COL11A1-correlated pan-cancer gene signature of activated fibroblasts for the prioritization of therapeutic targets. Cancer Letters. 2016;382(2):203-214.

54. Kao YH et al. Upregulation of hepatoma-derived growth factor is involved in murine hepatic fibrogenesis. Journal of hepatology. 2010;52(1): 96-105.

55. Lin ZY, Chuang WL. Genes responsible for the characteristics of primary cultured invasive phenotype hepatocellular carcinoma cells. Biomedicine & Pharmacotherapy. 2012;66(6):454-458.

56. Allinen M, Beroukhim R, Cai L, Brennan C, Lahti-Domenici J, Huang H et al. Molecular characterization of the tumor microenvironment in breast cancer. Cancer Cell. 2004;6(1):17-32.

57. Chen Y, Wang T, Hsu H, Yuan R, Jeng Y. Overexpression of CTHRC1 in Hepatocellular Carcinoma Promotes Tumor Invasion and Predicts Poor Prognosis. PLoS ONE. 2013;8(7):e70324.

58. Ma MZ et al. CTHRC1 acts as a prognostic factor and promotes invasiveness of gastrointestinal stromal tumors by activating Wnt/PCP-Rho signaling. Neoplasia. 2014;16(3):265-278.

59. Tang L et al. Aberrant expression of collagen triple helix repeat containing 1 in human solid cancers. Clinical Cancer Research. 2006;12(12):3716-3722.

60. Yang XM et al. CTHRC1 promotes human colorectal cancer cell proliferation and invasiveness by activating Wnt/PCP signaling. International journal of clinical and experimental pathology. 2015;8(10):12793.

61. Leclair R, Lindner V. The role of collagen triple helix repeat containing 1 in injured arteries, collagen expression, and transforming growth factor β signaling. Trends in cardiovascular medicine. 2007;17(6):202-205.

62. Roth W et al. Secreted Frizzled-related proteins inhibit motility and promote growth of human malignant glioma cells. Oncogene. 2000;19(37):4210-4220.

63. Yamamura S et al. Oncogenic functions of secreted Frizzled-related protein 2 in human renal cancer. Molecular cancer therapeutics. 2010;9(6):1680-1687.

64. Pohl S et al. Secreted frizzled-related protein 4 and its implications in cancer and apoptosis. Tumor Biology. 2015;36(1):143-152.

65. O'prey J, Wilkinson S, Ryan K. Tumor antigen LRRC15 impedes adenoviral infection: implications for virus-based cancer therapy. Journal of virology. 2008;82(12):5933-5939.

66. Reynolds P et al. Identification of a DNA-binding site and transcriptional target for the EWS–WT1 (+ KTS) oncoprotein. Genes & development. 2003;17(17):2094-2107.

67. Schuetz C et al. Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. Cancer research. 2006;66(10):5278-5286.

68. Stanbrough M et al. Increased expression of genes converting adrenal androgens to testosterone in androgen-independent prostate cancer. Cancer research. 2006;66(5):2815-2825.

69. Chen J, Yao Y, Gong C, Yu F, Su S, Chen J et al. CCL18 from Tumor-Associated Macrophages Promotes Breast Cancer Metastasis via PITPNM3. Cancer Cell. 2011;19(4):541-555.

70. Martinez F, Helming L, Gordon S. Alternative activation of macrophages: an immunologic functional perspective. Annual review of immunology. 2009;27:451-483.

71. Schutyser E, Richmond A, Van Damme J. Involvement of CC chemokine ligand 18 (CCL18) in normal and pathological processes. Journal of leukocyte biology. 2005;78(1):14-26.

72. Chang C, Lee Y, Leu S, Wang C, Wei C, Hung K et al. CC-chemokine ligand 18/pulmonary activation-regulated chemokine expression in the CNS with special

reference to traumatic brain injuries and neoplastic disorders. Neuroscience. 2010;165(4):1233-1243.

73. Leung S et al. Expression profiling identifies chemokine (CC motif) ligand 18 as an independent prognostic indicator in gastric cancer. Gastroenterology. 2004;127(2):457-469.

74. Schutyser E et al. Identification of biologically active chemokine isoforms from ascitic fluid and elevated levels of CCL18/pulmonary and activation-regulated chemokine in ovarian carcinoma. Journal of Biological Chemistry. 2002;277(27):24584-24593.

75. Zohny S, Fayed S. Clinical utility of circulating matrix metalloproteinase-7 (MMP-7), CC chemokine ligand 18 (CCL18) and CC chemokine ligand 11 (CCL11) as markers for diagnosis of epithelial ovarian cancer. Medical Oncology. 2010;27(4):1246-1253.

76. Lin L et al. CCL18 from tumor-associated macrophages promotes angiogenesis in breast cancer. Oncotarget. 2015; 6(33):34758.

77. Duluc D, Corvaisier M, Blanchard S, Catala L, Descamps P, Gamelin E et al. Interferon-γ reverses the immunosuppressive and protumoral properties and prevents the generation of human tumor-associated macrophages. International Journal of Cancer. 2009;125(2):367-373.

78. Pasini F et al. A gene expression profile related to immune dampening in the tumor microenvironment is associated with poor prognosis in gastric adenocarcinoma. Journal of gastroenterology. 2014;49(11):1453-1466.

79. Xu J et al. Expression and inhibition of ADAMDEC1 in craniopharyngioma cells. Neurological research. 2012;34(7):701-706.

80. Khaidakov M et al. Oxidized LDL receptor 1 (OLR1) as a possible link between obesity, dyslipidemia and cancer. PloS one. 2011; 6(5):e20277.

81. Yamanaka S et al. The human gene encoding the lectin-type oxidized LDL receptor (OLR1) is a novel member of the natural killer gene complex with a unique expression profile. Genomics. 1998;54(2):191-199.

82. Li D, Yang B, Mehta J. Ox-LDL induces apoptosis in human coronary artery endothelial cells: role of PKC, PTK, bcl-2, and Fas. American Journal of Physiology-Heart and Circulatory Physiology. 1998;275(2):H568-H576.

83. Li D et al. Oxidized LDL Upregulates Angiotensin II Type 1 Receptor Expression in Cultured Human Coronary Artery Endothelial Cells The Potential Role of Transcription Factor NF-κB. Circulation. 2000;102(16):1970-1976.

84. Hirsch H et al. A transcriptional signature and common gene networks link cancer with lipid metabolism and diverse human diseases. Cancer cell. 2010; 17(4):348-361.

85. Charbonneau B, Maurer M, Fredericksen Z, Zent C, Link B, Novak A et al. Germline variation in complement genes and event-free survival in follicular and diffuse large B-cell lymphoma. American Journal of Hematology. 2012;87(9):880-885.

86. Csincsi Á, Kopp A, Zöldi M, Bánlaki Z, Uzonyi B, Hebecker M et al. Factor H–Related Protein 5 Interacts with Pentraxin 3 and the Extracellular Matrix and Modulates Complement Activation. The Journal of Immunology. 2015;194(10):4963-4973.

87. Skerka C et al. Complement factor H related proteins (CFHRs). Molecular immunology. 2013;56(3):170-180.

88. National Center for Biotechnology Information – Gene [Internet]. HFE2 hemochromatosis type 2 (juvenile) [Homo sapiens (human)] [acesso em 09 nov 2016]. Disponível em: https://www.ncbi.nlm.nih.gov/gene/148738

89. National Human Genome Research Institute [Internet]. HFE2 gene [acesso em 09 nov 2016]. Disponível em: https://ghr.nlm.nih.gov/gene/HFE2

90. National Center for Biotechnology Information – Gene [Internet]. HPR haptoglobin-related protein [Homo sapiens (human)] [acesso em 10 nov 2016]. Disponível em: https://www.ncbi.nlm.nih.gov/gene/3250

91. Nielsen M et al. Haptoglobin-related protein is a high-affinity hemoglobin-binding plasma protein. Blood. 2006;108(8):2846-2849.

92. Kuhajda F, Piantadosi S, Pasternack G. Haptoglobin-related protein (Hpr) epitopes in breast cancer as a predictor of recurrence of the disease. New England Journal of Medicine. 1989;321(10):636-641.

93. National Center for Biotechnology Information – Gene [Internet]. SLC38A4 solute carrier family 38 member 4 [Homo sapiens (human)] [acesso em 10 nov 2016]. Disponível em: https://www.ncbi.nlm.nih.gov/gene/55089

94. Lee S et al. Polymorphisms in cancer-related pathway genes and lung cancer. European Respiratory Journal. 2016;48(4):1184-1191.

95. National Center for Biotechnology Information – Gene [Internet]. SLC10A1 solute carrier family 10 member 1 [Homo sapiens (human)] [acesso em 10 nov 2016]. Disponível em: https://www.ncbi.nlm.nih.gov/gene/6554

96. Hallen S et al. Organization of the membrane domain of the human liver sodium/bile acid cotransporter. Biochemistry. 2002;41(23):7253-7266.

97. National Center for Biotechnology Information – Gene [Internet]. LPA lipoprotein(a) [Homo sapiens (human)] [acesso em 10 nov 2016]. Disponível em: https://www.ncbi.nlm.nih.gov/gene/4018

98. Bhattacharjee P, Paul S, Banerjee M, Patra D, Banerjee P, Ghoshal N et al. Functional compensation of glutathione S-transferase M1 (GSTM1) null by another GST superfamily member, GSTM2. Sci Rep. 2013;3.

99. Leme, C et al. GSTM1 and GSTT1 genes analysis in head and neck cancer patients. Revista da Associação Médica Brasileira. 2010;56(3):299-303.

100. To-Figueras J et al. Glutathione S-transferase M1 (GSTM1) and T1 (GSTT1) polymorphisms and lung cancer risk among Northwestern Mediterraneans. Carcinogenesis. 1997;18(8):1529-1533.

101. GeneCards [Internet]. ACOT12 Gene [acesso em 10 nov 2016]. Disponível em: http://www.genecards.org/cgi-bin/carddisp.pl?gene=ACOT12

102. Meng FD et al. Association between cytochrome P450 1A1 (CYP1A1) gene polymorphisms and the risk of renal cell carcinoma: a meta-analysis. Scientific reports. 2015;5.

103. Rodriguez M, Potter D. CYP1A1 regulates breast cancer proliferation and survival. Molecular Cancer Research. 2013;11(7):780-792.

104. Shaffi S et al. CYP1A1 polymorphisms and risk of lung cancer in the ethnic Kashmiri population. Asian Pac J Cancer Prev. 2009;10(4):651-656.

105. National Center for Biotechnology Information – Gene [Internet]. SMLR1 small leucine rich protein 1 [Homo sapiens (human)] [acesso em 10 nov 2016]. Disponível em: https://www.ncbi.nlm.nih.gov/gene/100507203

Anexos

7 ANEXOS

7.1 Anexo 1

Extração de RNA com TRIzol®(Invitrogen)

Antes de iniciar a extração, realizamos a maceração do tecido hepático através de lise mecânica com o homogeneizador Precellys®24-Dual, sob agitação com *beads* de vidro a 6500rpm, por 2 ciclos de 15 segundos, com intervalo de 15 segundos entre eles.

Seguiu-se então à extração de RNA total pelo método de uso do isotiocianato de guanidina fenol-clorofórmio, realizada por TRIzol®(Invitrogen), como descrito a seguir:

1. As amostras maceradas foram transferidas para tubos com 1000µl de TRIzol®;
2. Após incubação em temperatura ambiente por cinco minutos, 200µl de clorofórmio foram adicionados a mistura;
3. Procedeu-se à homogeneização vigorosa utilizando-se o *vortex*;
4. Seguiu-se a centrifugação a 11000g por 15 minutos em centrífuga refrigerada (2°C);
5. A solução foi separada em duas fases, uma aquosa (superior e incolor) e outra orgânica (inferior e com tonalidade rosa). A precipitação do RNA ocorreu a partir da fase superior que foi transferida em um novo tubo tipo eppendorf;
6. A esse novo tubo, já com a solução aquosa, adicionados 500µl de isopropanol;
7. A mistura foi incubada por 10 minutos a temperatura ambiente;
8. Seguiu-se a centrifugação a 11000g por 10 minutos em centrífuga refrigerada (2°C);
9. O sobrenadante foi removido e seguiu-se à lavagem do *pellet* de RNA formado no fundo do tubo com 1000µl de etanol 75% ;
10. Após desprender o pellet do fundo do tubo, a amostra foi centrifugada por 5 minutos a 7000g a 2°C;
11. O sobrenadante foi removido e o pellet foi eluído em água tratada com DEPC (dietil pirocarbonato).

OBS.: Os tubos e ponteiras utilizados em todas as etapas da extração estavam livres de DNAses e RNAses;

Após a extração, as amostras de RNA foram submetidas ao tratamento com DNAse I (Ambion®) para eliminação de possível DNA genômico contaminante.