



**UNIVERSIDADE ESTADUAL PAULISTA “JÚLIO
DE MESQUITA FILHO”
FACULDADE DE MEDICINA**

Cristiane Nonato da Silva

**Avaliação da Variabilidade de uma Biblioteca
de Anticorpos construída a partir de sangue
de pacientes com Hepatite C Crônica com
diferentes Graus de Fibrose**

Dissertação apresentada à Faculdade de
Medicina, Universidade Estadual Paulista “Júlio
de Mesquita Filho”, Câmpus de Botucatu, para
obtenção do título de Mestre em Patologia.

Orientador(a): Prof(a). Dr(a). Rejane Maria Tommasini Grotto
Coorientador(a): Dr(a). Flávia Hebelers Barbosa Trovão

**Botucatu
2019**

Cristiane Nonato da Silva

Avaliação da Variabilidade de uma Biblioteca de Anticorpos construída a partir de sangue de pacientes com Hepatite C Crônica com Diferentes Graus de Fibrose

Dissertação apresentada à Faculdade de Medicina, Universidade Estadual Paulista “Júlio de Mesquita Filho”, Câmpus de Botucatu, para obtenção do título de Mestre em Patologia.

Orientador(a): Prof(a). Dr(a). Rejane Maria Tommasini Grotto
Coorientador(a): Dr(a). Flávia Hebelers Barbosa Trovão

Botucatu
2019

Ficha Catalográfica

FICHA CATALOGRÁFICA ELABORADA PELA SEÇÃO TÉC. AQUIS. TRATAMENTO DA INFORM.
DIVISÃO TÉCNICA DE BIBLIOTECA E DOCUMENTAÇÃO - CÂMPUS DE BOTUCATU - UNESP
BIBLIOTECÁRIA RESPONSÁVEL: ROSANGELA APARECIDA LOBO-CRB 8/7500

Silva, Cristiane Nonato da.

Avaliação da variabilidade de uma biblioteca de anticorpos construída a partir de sangue de pacientes com hepatite C crônica com diferentes graus de fibrose / Cristiane Nonato da Silva. - Botucatu, 2019

Dissertação (mestrado) - Universidade Estadual Paulista "Júlio de Mesquita Filho", Faculdade de Medicina de Botucatu

Orientador: Rejane Maria Tommasini Grotto
Coorientador: Flávia Hebler Barbosa Trovão
Capes: 40105008

1. Cirrose hepática. 2. Hepatite C. 3. Imunoglobulinas.

Palavras-chave: Fibrose Hepática; Hepatite C; Imunoglobulina.

Dedicatória

Dedico este trabalho...

...A minha amada família, pela oportunidade de uma boa formação, por todo amor e paciência durante essa jornada...

Agradecimientos

Aos meus pais, sem vocês, nada disso seria possível.

À professora e orientadora Dra. Rejane, pela oportunidade, apoio e pela confiança até mesmo nos momentos que duvidei da minha capacidade. Muito obrigada pela oportunidade e ensinamentos.

À minha coorientadora Dra. Flávia, pelo auxílio técnico e disponibilidade para resolução de problemas, para que conseguíssemos chegar ao objetivo proposto.

À Dra. Maria Inês de Moura Campos Pardini, por disponibilizar o laboratório de Biologia Molecular do Hemocentro, para realização deste trabalho.

Aos queridos funcionários e amigos do laboratório (Rotina e Pesquisa) pelo nosso cafezinho de todos os dias, por todo apoio e suporte.

Ao CNPq pelo auxílio financeiro.

Com vocês, divido minha alegria.

Epígrafe

*Bem-aventurado o homem que acha sabedoria, e o homem que
adquire conhecimento... (Provérbios 3:13)*

Prefácio

Neste volume é apresentada a Dissertação de Mestrado de Cristiane Nonato da Silva.

*Este trabalho foi subdividido em capítulos da forma que se segue: (I) Levantamento Teórico: breve descrição dos conceitos e bibliografia para contextualizar o cenário de desenvolvimento do trabalho; Objetivos do trabalho; (II) Versão preliminar do artigo: *Profile of Expressed Antibodies in Patients with Chronic Hepatitis with Different Degrees of Hepatic Fibrosis*, nas normas da revista *Frontiers in Immunology*; (III) Conclusão: breve conclusão do trabalho realizado evidenciando sua contribuição para a academia.*

Resumo

Resumo:

A progressão da fibrose hepática somada à infecção pelo vírus da hepatite C (VHC) tem sido associada à resposta imunológica permanente. O estudo no repertório de anticorpos Anti-VHC na progressão da fibrose hepática foi explorado através de ferramentas de sequenciamento em larga escala (NGS) possibilitando uma análise de repertórios altamente variáveis como as porções variáveis VH (cadeia pesada) e V κ (cadeia leve) das imunoglobulinas, e a determinação de famílias e subfamílias dos genes V-D-J associadas à resposta humoral encontrada nas diferentes fases da doença proporcionam uma ferramenta importante no entendimento da resposta imune frente à infecção viral pelo VHC. As porções VH e V κ das imunoglobulinas foram obtidas a partir da amplificação de RNA de sangue de paciente VHC positivos e com diferentes graus de fibrose, e sequenciadas na plataforma Illumina® Miseq, fornecendo uma grande variabilidade de seqüências que foram pré-processadas por ferramentas de bioinformática e analisadas em dois bancos de anticorpos diferentes: IgBlast (NCBI) e IMGT® quanto às famílias e subfamílias mais expressas. A expressão restrita de algumas famílias e subfamílias: IGHV1, IGHV3, IGHV4 e subfamílias já descritas em vários estudos associados ao VHC corrobora com nossos achados de que existe uma tendência do uso de algumas subfamílias como: IGHV1-2, IGHV1-8, IGVH1-69, IGHV3-11, IGHV3-21, IGHV3-23, IGVH3-30, IGHV4-4, IGHV4-34 IGHV4-39 na cadeia pesada; assim como IG κ V3-15 e IG κ V3-20 na cadeia leve, porém as subfamílias: IGHV1-8, IGHV3-11, IGHV4-39, IG κ V1-5, IG κ V1-12, IG κ V1-39 também estavam entre as mais expressas, indicando que estas subfamílias também podem estar relacionadas a expressão preferencial induzida pelo vírus em detrimento de outras; como encontramos na família IGVH3 que estava menos expressa no grupo F3/F4 de fibrose avançada; resultado esse inovador na literatura relacionada ao VHC o sugere uma mudança no padrão de resposta imunológica à medida que a doença progride.

Palavras-chave: Fibrose Hepática, Hepatite C, Imunoglobulina.

Abstract

Abstract:

The progression of hepatic fibrosis in the hepatitis C virus (HCV) infection has been linked with permanent immune response. The study in the repertoire of Anti-HCV antibodies in the progression of hepatic fibrosis was analysed for large-scale sequencing (NGS) tools enabling a highly variable analysis such as the VH (heavy chain) and V κ (light chain) portions of immunoglobulins and the determination of families and subfamilies of the V-D-J genes in the humoral response found in the different phases of the disease, a great tool in the understanding of the immune response to HCV viral infection. The VH and V portions of the immunoglobulins were obtained from the amplification of HCV positive HCV patient blood with different degrees of fibrosis and sequenced on the Illumina® Miseq platform, providing a large sequence variability that was preprocessed by tools of bioinformatics and analyzed in two different antibody banks: IgBlast (NCBI) and IMGT® for the most expressed families and subfamilies. The restricted expression of some families: IGHV1, IGHV3, IGHV4 and subfamilies already described in several HCV-related studies confirm our findings that there is a tendency to use some subfamilies, such as: IGHV1-2, IGHV1-8, IG κ V1-69, IGHV3-11, IGHV3-21, IGHV3-23, IG κ V3-30, IGHV4-4, IGHV4-34 IGHV4-39 in the heavy chain; as well as IG κ V3-15 and IG κ V3-20 in the light chain, but the subfamilies: IGHV1-8, IGHV3-11, IGHV4-39, IG κ V1-5, IG κ V1-12, IG κ V1-39 were also among the most expressed, indicating that these subfamilies may also be related to preferential expression induced by the virus over others; as found in the IG κ V3 family, which was less expressed in the advanced fibrosis group F3/F4; result this groundbreaking finding in the HCV literature suggests a change in the pattern of immune response as the disease progresses.

Key words: Hepatic Fibrosis, Hepatitis C, Immunoglobulin.

Sumário

Sumário

Resumo	
Abstract	
Capítulo I: Revisão de Literatura	18
1.0 A infecção pelo vírus da Hepatite C	19
1.1 Fibrose Hepática	20
1.2 Tratamento da infecção pelo VHC	21
1.3 Anticorpos	21
1.4 Sequenciamento de Nova Geração (NGS) e Ferramentas de Análise	23
1.5 Software de Análise: IgBlast e IMGT	24
1.6 Alvos para o VHC	25
2. Referências Bibliográficas	27
3. Objetivo	31
Capítulo II: Artigo a ser submetido na <i>Frontiers Immunology</i>	32
1. Título: PERFIL DE ANTICORPOS EXPRESSOS EM PACIENTES PORTADORES DE HEPATITE C CRÔNICA COM DIFERENTES GRAUS DE FIBROSE HEPÁTICA	33
Capítulo III: Conclusão	62
1. Conclusão	63
2. Referências Bibliográficas	64

1. Capítulo I: Revisão de Literatura

1.0 A infecção pelo Vírus da Hepatite C

O vírus da hepatite C (VHC), pertence à família *Flaviviridae*, possui tropismo preferencial por células hepáticas e é atualmente uma das maiores causas de doença hepática no mundo. Dados recentes da Organização Mundial de Saúde mostram uma alta taxa de infecção pelo VHC com estimativas de aproximadamente 71 milhões de pessoas em todo mundo. A infecção pelo VHC causa distúrbios hepáticos e cerca de 70% dos indivíduos infectados desenvolvem uma infecção crônica que pode levar a cirrose e carcinoma hepatocelular (CHC). A infecção possui caráter silencioso na fase inicial da doença, sendo inicialmente assintomática na maioria dos casos, e os infectados geralmente desconhecem a infecção até os sintomas aparecerem nas fases tardias da doença, quando o dano hepático já na maioria das vezes é irreversível. A necessidade de diagnóstico nas fases iniciais da doença determina o sucesso na resposta ao tratamento para eliminação do vírus (1).

O vírus da hepatite C é um vírus de RNA, de polaridade positiva, pertence ao gênero *Hepacivirus*. Atualmente são descritos sete genótipos e 67 subtipos que possuem pelo menos 33 ou 15% de variação na sequência de nucleotídeos, e é constituído por um capsídeo que contém o genoma viral envolto por uma membrana derivada do retículo endoplasmático coroado pelas proteínas do envelope E1-E2. As proteínas E1 e E2 são glicoproteínas tipo I com domínios transmembranares C-terminais que formam um heterodímero no envelope viral que promove a entrada do vírus nas células hospedeiras, o que envolve interações do vírus com glicosaminoglicanos de superfície celular da célula hospedeira e com outras proteínas necessárias à ligação e internalização do vírus, como a tetraspanina CD81, membro de classe B do receptor de limpeza 1 (SR-B1), claudina-1, e ocludina que são essenciais para entrada do vírus (1,10).

A proteína E2 atua como a proteína de ligação ao receptor do VHC e interage diretamente com o CD81 e o SR-B1. Enquanto a proteína E1 parece ajudar a modular as interações do receptor E2 e a fusão com a membrana da célula hospedeira. A proteína E2 tem sido amplamente estudada como alvo de anticorpos neutralizantes (bNAbs) sugerindo um papel importante no mecanismo de neutralização viral e o bloqueio das interações entre E2 e seu receptor CD81 (27).

1.1 Fibrose Hepática

A infecção causada pelo VHC no tecido hepático promove uma agressão tecidual que culmina em uma resposta irreversível do organismo frente à injúria causada pelo vírus. É um processo de cicatrização que ocorre após uma agressão tecidual e que reflete um equilíbrio entre o dano e a tentativa de reparação do fígado com consequente formação de cicatriz. Nas fases iniciais da infecção os danos hepáticos podem ser reversíveis, mas com a progressão para lesão crônica, há substituição progressiva do parênquima hepático pelo tecido cicatricial. Devido à alta capacidade regenerativa do fígado as lesões hepáticas só apresentam estágios avançados depois de muitos anos de injúria tecidual. Um importante fator na progressão da fibrose é a resposta inflamatória persistente que ocorre nos casos crônicos da doença. O sistema imune na tentativa de eliminação viral contribui para aumentar os danos celulares e consequentemente a fibrose hepática, e atua também favorecendo a formação de tecido fibroso no processo de retroalimentação positiva no qual células inflamatórias e fibrogênicas como as células estreladas hepáticas (HSCs) estimulam-se mutuamente amplificando a fibrose (21).

Além do VHC promover dano celular durante o ciclo celular, as proteínas do core e proteínas não estruturais parecem estar relacionadas diretamente com a ativação de HSCs e, assim, com a iniciação da fibrogênese. As proteínas do core ativam vias promitogênicas dentro de HSCs, e as proteínas NS3 e NS5 ativam as vias pró-inflamatórias NF- κ B e JNK. As proteínas do núcleo e NS3 aumentam os níveis de cálcio [Ca^{2+}] intracelular e as espécies reativas de oxigênio (ROS). O que favorece a produção de osteopontina e ROS que juntas contribuem para a transição epitelial para mesênquima de hepatócitos. A ativação da metaloproteínase de matriz 2 também ocorre com a ligação da proteína E2 ao CD81 das HSCs, promovendo a degradação de matriz extracelular normal e infiltração de células inflamatórias no local. As proteínas NS3 e NS5A contribuem com o dano hepático através da indução do estresse oxidativo em hepatócitos e monócitos via ativação da NADPH oxidase e repressão da heme oxigenase 1. Proteínas do core e NS3 atuam na modulação da resposta imune inata pela ativação da inflamação através do receptor *Toll-like 2* em monócitos (21,22).

Além de outros sinais pró-fibróticos relacionados à presença do VHC, a ativação da transformação fator de crescimento β 1 contribui para a transformação fibrótica no fígado. A progressão da fibrose culmina na perda de função do órgão com consequências importantes no desenvolvimento da patologia do carcinoma

hepatocelular (CHC). O grau de fibrose indica o aparecimento progressivo da doença, que pode levar a cirrose e complicações hepáticas terminais em estágios avançados da doença. A ausência ou leve presença de fibrose tem um risco relativamente baixo (25% - 30%) de desenvolver cirrose nos próximos 20 anos, porém a presença de fibrose portal e fibrose septal, ambas causam cirrose, mesmo com diferentes taxas de progressão (18-20 anos para pacientes com fibrose portal e 8-10 anos para pacientes com fibrose septal). A importância do diagnóstico nas fases iniciais da doença são importantes na eficácia do tratamento e não progressão da fibrose. A fibrose é classificada em diferentes estágios, sendo F0 a ausência de fibrose, F1 fibrose portal sem septos, F2 expansão fibrosa com poucos septos, F3 numerosos septos podendo visualizar esboços de nódulos e, F4 cirrose. Além de indicação para tratamento antiviral, um estágio de fibrose hepática avançado exige intervenções para controlar os efeitos negativos, como modificações no estilo de vida (dieta, perda de peso, exercício físico), abstinência de álcool e drogas, encaminhamento a especialistas (22).

1.2 Tratamento da Infecção pelo VHC

O tratamento atualmente engloba antivirais de ação direta (DAAs) contra a infecção pelo vírus da hepatite C (VHC) e podem atingir cura em, aproximadamente, 95% dos casos, sugerindo que tem o potencial de eliminar a infecção que afeta mais de 71 milhões de pessoas em todo o mundo, porém o caráter silencioso da doença contribui para a disseminação da infecção e diagnóstico tardio, permitindo que a infecção pelo VHC permaneça assintomática por anos, durante os quais muitos não são diagnosticadas (23).

1.3 Anticorpos

O estudo de anticorpos como alvos terapêuticos, e no estudo da elucidação de mecanismos de resposta frente a infecções virais, tem sido amplamente difundido no estudo de outras patologias, como o HIV (17).

Os anticorpos são proteínas que atuam no reconhecimento de moléculas potencialmente nocivas (antígenos), e que são produzidos por células B e são secretados agindo como uma membrana recobrimo um antígeno, ou como um receptor para outras células do sistema imune. Em humanos, são constituídos por quatro cadeias

polipeptídicas sendo duas pesadas (*Heavy*) e duas leves (*Light*) que contém frações variáveis e constantes quanto à sequência de aminoácidos (Figura 1a). No que se refere as porções variáveis (Fv) os anticorpos são compostos de dois tipos de cadeias polipeptídicas, conhecidas como leve (VL= *variable light*) e pesadas (VH= *variable heavy*) as quais possuem cada uma região variável de aproximadamente 110 aminoácidos de comprimento, composta de uma região de estrutura (*frameworks* =FRs) e três alças hipervariáveis denominadas regiões determinantes de complementaridade (CDRs) (Figura 1b, c). As sequências germinativas são limitadas, porém mutações somáticas, deleções e inserções de nucleotídeos nas CDRs aumentam a variabilidade do repertório de anticorpos que determinam o reconhecimento de antígenos. Estima-se que o ser humano possa produzir cerca de 10^{10} moléculas distintas de anticorpos (20).

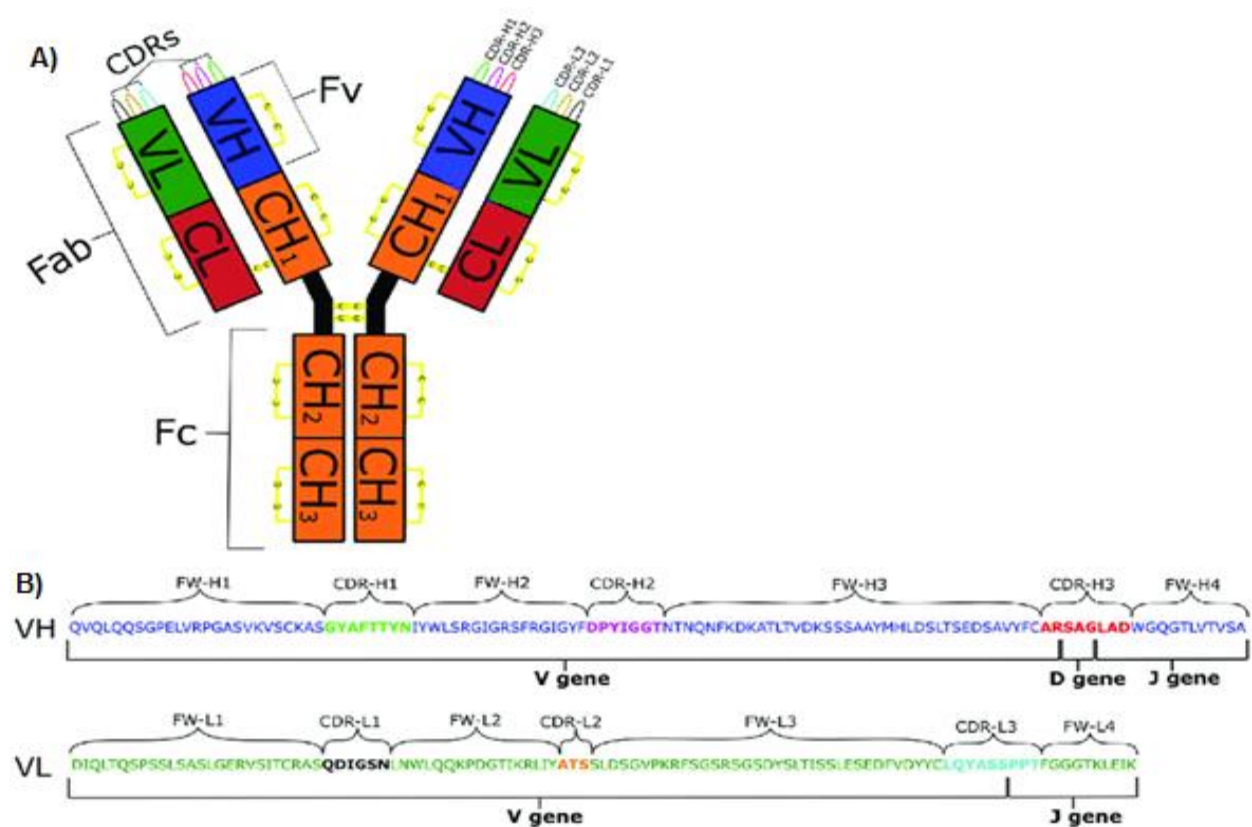


Figura 1: (A) Representação esquemática de uma estrutura de anticorpo IgG. (B) Composição genética das cadeias VH e VL; e CDRs. Adaptado de Kovaltsuk e colaboradores (2017).

Repertórios de anticorpos humanos constituem um amplo campo de estudo nas respostas humorais frente a infecção pelo VHC. Os genes das imunoglobulinas humanas que codificam as cadeias pesadas e leves humanas são IGVH, IGHD, IGHD, IGHD e IGHC presentes no *locus* de cadeia no cromossomo 14; e IGKV, IGKJ e IGKC presentes

no locus da cadeia leve *kappa* no cromossomo 2; e os segmentos genéticos de IGLV, IGLJ e IGLC do locus da cadeia leve *lambda* no cromossomo 22 (18).

A alta diversidade dos anticorpos estão presentes nos locais de ligação ao antígeno que são obtidos por combinações das cadeias H e L. Essa combinação nos seres humanos pode gerar milhões de anticorpos diferentes que podem ligar-se ao antígeno quando são rapidamente selecionados (18).

O estudo de Bibliotecas de anticorpos selecionadas a partir de amostras de pacientes com infecção ativa com ferramentas de sequenciamento de nova geração (NGS) permite analisar grandes e diversificadas sequencias de anticorpos que podem produzir ligantes para muitos antígenos diferentes (18).

1.4 Sequenciamento de Nova Geração (NGS) e Ferramentas de Análise

A ampla variabilidade encontrada nos anticorpos pode ser analisada com ferramentas de sequenciamento de DNA que geram grandes conjuntos de dados, e que estão sendo empregados no estudo da composição de repertórios de imunoglobulinas, de doenças infecciosas, alérgicas e doenças auto-imunes e no câncer. Com essa tecnologia é possível analisar repertórios imunológicos, incluindo características específicas de anticorpos como seus alvos, e sequencias únicas e altamente variáveis de anticorpos. A determinação de genes de linhagens germinativas de imunoglobulina é possível utilizando análise de bancos dados de genes conhecidos da linhagem germinativa; os quais compreendem a composição de sequências germinativas de genes de imunoglobulinas, as quais estão depositadas nestes bancos e que funcionam como parâmetro para determinação da origem das sequências analisadas e avaliar o conjunto de genes da linha germinativa (25,5).

O tamanho das abordagens NGS requer o uso de ferramentas de bioinformática, que permitem avaliar a qualidade das sequencias obtidas com ferramentas como o FastQC, e ferramentas como o Trimmomatic para retirada dos *primers* das PCRs, bem como as sequencias dos adaptadores Illumina. E ferramentas de montagem dos *reads* como o programa Velvet, permitindo assim filtrar dados de qualidade para análise nos bancos de dados de imunoglobulinas como IgBlast (NCBI) e IMGT / HighV-QUEST (7,19,30,2,8,16,29).

As ferramentas IgBlast e IMGT usam no alinhamento diferentes algoritmos, e restringem os repertórios dos genes aqueles definidos pelo banco de anticorpos, o que se faz necessário um pós-processamento dos dados de saída (8,16,29).

1.5 Softwares de Análise: IgBlast e IMGT

O software IgBlast (NCBI) é um algoritmo que utiliza a similaridade para comparar as sequências de anticorpos da linhagem germinativa contidos em seu banco de dados e em outros bancos como IMGT / V-QUEST, o repertório de cadeias pesadas humanas UNSWlg e o banco de genes da linha germinativa do NCBI com as sequências analisadas; isso contribui para que a análise do gene da linhagem germinativa seja o mais preciso possível e para diminuir perdas da melhor correspondência germinal do gene V (8,16,29).

IgBLAST analisa sequências do domínio IGV identificando os genes V, D e J, analisa também o rearranjo de junção e as regiões de framework (FR) e regiões determinantes de complementaridade (CDR) para o V gene (29).

Particularmente a análise do gene D é mais difícil por serem muito curtos e terem alta similaridade com outros genes existentes na linhagem germinativa, e também devido a adições aleatórias de nucleotídeos no rearranjo das junções, e as mutações somáticas contribuem para dificultar determinar o gene D da linhagem germinativa mais próxima da sequência analisada (29).

O banco IMGT/V-QUEST também pode realizar a identificação das linhagens germinativas dos genes V, D e J e alelos por alinhamento com a linhagem germinativa. A análise tanto de sequências de imunoglobulinas como de receptores pode ser realizada com este programa. O IMGT/V-QUEST utiliza sequências publicadas nas bases de dados DDBJ, EMBL e GenBank como fonte de dados para seu banco (8,16).

Para determinar as sequências V, D e J rearranjadas, o programa compara as sequências analisadas com os conjuntos de sequências de referência da linhagem germinativa de genes e alelos IG e TR, definidos de acordo com os critérios do IMGT e agrupados no *Conjuntos de Diretórios de Referência* do IMGT. Além de determinar a linhagem germinativa mais próxima o software identifica regiões de estruturas (FR) e

regiões determinantes de complementaridade (CDR) importantes na determinação do sítio de ligação ao antígeno. Identifica também mutações somáticas das sequências rearranjadas de imunoglobulinas e analisa a sequência de aminoácidos traduzidos da variável IG e TR da região com base na numeração única do IMGT (8,16).

A análise do IMGT mostra resultados de cinco genes e alelos V e D mais próximos, e três genes e alelos J mais próximos da referência de linhagem germinativa do banco, sendo a melhor pontuação de alinhamento indicativa do gene e alelo com maior probabilidade de estar envolvido no rearranjo analisado. Exibe a sequência de aminoácidos da região de junção delimitada pela V-REGION 2^a-CYS na posição 104 e a J-REGION J-PHE ou J-TRP na posição 118. O alinhamento com o gene D por padrão realizado pelo IMGT/ *Junction Analysis*, que faz o alinhamento com o melhor gene D e alelo, permitindo uma análise mais detalhada da região de diversidade. Também fornece a sequência de nucleotídeos e aminoácidos das regiões de estrutura FRs e CDRs (FR1-, CDR1-, FR2-, CDR2, FR3- e CDR3) das sequências analisadas, permitindo e localizar mutações somáticas (8,16).

1.6 Alvos para VHC

Na busca do entendimento sobre os mecanismos patológicos associados ao VHC, o estudo de anticorpos constitui uma ferramenta importante que permite ampliar o conhecimento de genes específicos que são expressos na infecção pelo VHC, o que impulsiona o desenvolvimento de vacinas e outros tratamentos direcionados contra o vírus (4,6).

Os anticorpos neutralizantes virais (bNAbs) foram descritos principalmente contra a proteína E2. Alguns anticorpos descritos demonstraram inibir a infecção, bloqueando a interação com o RBS ou inibindo o mecanismo de fusão pós-entrada direcionando regiões conservadas dentro das glicoproteínas virais envolvidas nesses processos (12,15).

Acredita-se que o aparecimento precoce dos bNAbs estão envolvidos no processo de eliminação espontânea da infecção pelo VHC, pois estes se ligam às proteínas do envelope do VHC (E1 e E2) e, são capazes de neutralizar múltiplas variantes virais (15).

O estudo de várias famílias e subfamílias de genes das imunoglobulinas foram encontradas predominantemente em infecções virais, e foram associados a infecção pelo VHC e outras infecções virais; como a subfamília do gene VH 1-69 que foi relatada como associada à neutralização cruzada de VHC, vírus influenza e HIV-1. As famílias e subfamílias específicas estão mais expressas nessas infecções e podem caracterizar um modelo de resposta humoral presente nas fases iniciais e tardias da doença (28).

O mecanismo pelo qual o VHC desencadeia expansões de células B não é totalmente conhecido, porém parece que o VHC possa causar um efeito mutagênico nas células B e promover estimulação antigênica específica contra o VHC. Essa estimulação de famílias específicas caracterizam um repertório próprio de cadeia pesada de imunoglobulina (VH) e cadeia leve (VL) presentes na infecção, e que podem caracterizar a resposta específica contra o vírus, tornando esses anticorpos alvos importantes para estudo e aplicações clínicas (26).

O processo de triagem direta de anticorpos específicos fornece conhecimento diferenciado sobre o papel do sistema humoral na resolução ou na progressão da doença causada pelo VHC. Outros estudos nesta área já estão mais avançados e permitem visualizar como no caso do HIV, como essas tecnologias de análises proporcionaram um aumento à geração de anticorpos monoclonais amplamente neutralizantes, anticorpos com potência, amplitude e especificidade contra o vírus. Sabendo que o repertório de anticorpos de um indivíduo armazena informações sobre patógenos e processos atuais e progressos que o indivíduo se expôs, com o potencial de desenvolvimento de anticorpos de triagem, desenhos de vacinas, e importantes alvos terapêuticos, seu estudo torna-se uma ferramenta importante no processo diagnóstico, prognóstico e terapêutico para a infecção pelo VHC (3,6,9,11,12,14,15,23,24).

2. Referências Bibliográficas:

1. Alazard-Dany N, Denolly S, Boson B, Cosset FL. Overview of HCV Life Cycle with a Special Focus on Current and Possible Future Antiviral Targets. *Viruses*. 2019; 11(1): 1-18. doi: 10.3390/v11010030.
2. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
3. Arcaini, L et al. Splenic marginal zone lymphoma: Clinical clustering of immunoglobulin heavy chain repertoires. *Blood Cells, Molecules, And Diseases*. 2009; 42 (3):286-291. doi:10.1016/j.bcmd.2009.01.004.
4. Bailey JR. et al. Broadly neutralizing antibodies with few somatic mutations and hepatitis C virus clearance. *Jci Insight*. 2017; 2 (9):1-17. <http://dx.doi.org/10.1172/jci.insight.92872>. American Society for Clinical Investigation.
5. Barreto K. et al. Next-generation sequencing-guided identification and reconstruction of antibody CDR combinations from phage selection outputs. *Nucleic Acids Research*. 2019; 47 (9): 1-13. doi: 10.1093/nar/gkz131. Oxford University Press (OUP).
6. Bazzill JD. et al. Interrogation of Antigen Display on Individual Vaccine Nanoparticles for Achieving Neutralizing Antibody Responses against Hepatitis C Virus. *Nano Letters*, 2018; 18 (12): 7832-7838. doi: 10.1021/acs.nanolett.8b03601. American Chemical Society (ACS).
7. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30 (15): 2114-2120. doi: 10.1093/bioinformatics/btu170. Oxford University Press (OUP).
8. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Research*. 2008; 36: W503-508. doi: 10.1093/nar/gkn316. Oxford University Press (OUP).
9. Buonaguro L. et al. Immune signatures in human PBMCs of idiotypic vaccine for HCV-related lymphoproliferative disorders. *Journal Of Translational Medicine*. 2010; 8 (1): 1-11. doi: 10.1186/1479-5876-8-18. Springer Nature.
10. Charles ED. et al. Clonal expansion of immunoglobulin M+CD27+ B cells in HCV-associated mixed cryoglobulinemia. *Blood*. 2007; 111 (3): 1344-1356. doi: 10.1182/blood-2007-07-101717. American Society of Hematology.
11. Chaudhary N, Wesemann DR. Analyzing Immunoglobulin Repertoires. *Frontiers In Immunology*. 2018; 9: 1-18. doi: 10.3389/fimmu.2018.00462.

12. Chen F. et al. VH1-69 antiviral broadly neutralizing antibodies: genetics, structures, and relevance to rational vaccine design. *Current Opinion in Virology*. 2019; 34: 149-159. doi: 10.1016/j.coviro.2019.02.004.
13. Cowton VM. et al. Predicting the Effectiveness of Hepatitis C Virus Neutralizing Antibodies by Bioinformatic Analysis of Conserved Epitope Residues Using Public Sequence Data. *Frontiers in Immunology*. 2018; 9:1-13. doi: 10.3389/fimmu.2018.01470. Frontiers Media SA.
14. Eliyahu S. et al. Antibody Repertoire Analysis of Hepatitis C Virus Infections Identifies Immune Signatures Associated With Spontaneous Clearance. *Frontiers in Immunology*. 2018; 9: 1-16. doi:10.3389/fimmu.2018.03004. Frontiers Media SA.
15. Flyak AI. et al. HCV Broadly Neutralizing Antibodies Use a CDRH3 Disulfide Motif to Recognize an E2 Glycoprotein Site that Can Be Targeted for Vaccine Design. *Cell Host & Microbe*. 2018; 24 (5): 703-716. doi: 10.1016/j.chom.2018.10.009. Elsevier BV.
16. Giudicelli V, Duroux P, Kossida S, Lefranc MP. IG and TR single chain fragment variable (scFv) sequence analysis: a new advanced functionality of IMGT/V-QUEST and IMGT/HighV-QUEST. *Bmc Immunology*. 2017; 18 (1): 1-13. doi: 10.1186/s12865-017-0218-8. Springer Nature.
17. Han Q. et al. Difficult-to-neutralize global HIV-1 isolates are neutralized by antibodies targeting open envelope conformations. *Nature Communications*. 2019; 10 (1): 1-15. doi: 10.1038/s41467-019-10899-2. Springer Science and Business Media LLC.
18. Houimel M. The analysis of VH and VL genes repertoires of Fab library built from peripheral B cells of human rabies virus vaccinated donors. *Human Immunology*. 2014; 75 (8): 745-755. doi: 10.1016/j.humimm.2014.05.005. Elsevier BV.
19. Huang X. CAP3: A DNA Sequence Assembly Program. *Genome Research*. 1999; 9 (9): 868-877. doi: 10.1101/gr.9.9.868. Cold Spring Harbor Laboratory.
20. Krawczyk K. et al. Structurally Mapping Antibody Repertoires. *Frontiers In Immunology*. 2018; 9: 1-8. doi: 10.3389/fimmu.2018.01698. Frontiers Media SA.
21. Lee UE, Friedman SL. Mechanisms of hepatic fibrogenesis. *Best Practice & Research Clinical Gastroenterology*. 2011; 25 (2): 195-206. doi: 10.1016/j.bpg.2011.02.005. Elsevier BV.
22. Sebastiani G. Chronic hepatitis C and liver fibrosis. *World Journal of Gastroenterology*. 2014; 20 (32): 11033-11053. doi: 10.3748/wjg.v20.i32.11033. Baishideng Publishing Group Inc.
23. Shoukry NH. Hepatitis C Vaccines, Antibodies, and T Cells. *Frontiers in Immunology*. 2018; 9: 1-7. doi: 10.3389/fimmu.2018.01480. Frontiers Media SA.

24. Swann Re. et al. Broad Anti-Hepatitis C Virus (HCV) Antibody Responses Are Associated with Improved Clinical Disease Parameters in Chronic HCV Infection. *Journal of Virology*. 2016; 90 (9): 4530-4543. doi: 10.1128/jvi.02669-15. . American Society for Microbiology.
25. Thörnqvist L, Ohlin M. Critical steps for computational inference of the 3'-end of novel alleles of immunoglobulin heavy chain variable genes - illustrated by an allele of IGHV3-7. *Molecular Immunology*. 2018; 103: 1-6. doi: 10.1016/j.molimm.2018.08.018.. Elsevier BV.
26. Tucci FA. et al. Biased IGH VDJ gene repertoire and clonal expansions in B cells of chronically hepatitis C virus–infected individuals. *Blood*. 2017; 131 (5): 546-557. doi: 10.1182/blood-2017-09-805762. American Society of Hematology.
27. Tzarum N, Wilson IA, Law M. The Neutralizing Face of Hepatitis C Virus E2 Envelope Glycoprotein. *Frontiers In Immunology*. 2018; 9: 1-8. doi: 10.3389/fimmu.2018.01315. Frontiers Media SA.
28. Tzarum N. et al. Genetic and structural insights into broad neutralization of hepatitis C virus by human VH1-69 antibodies. *Science Advances*. 2019; 5 (1): 1-11. doi: 10.1126/sciadv.aav1882.American Association for the Advancement of Science (AAAS).
29. Ye J. et al. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Research*. 2013; 41 (1): 34-40. doi: 10.1093/nar/gkt382.Oxford University Press (OUP).
30. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*. 2008; 18 (5): 821-829. doi: 10.1101/gr.074492.107. Cold Spring Harbor Laboratory.

3. Objetivo

Objetivo:

Avaliação das regiões variáveis de bibliotecas de anticorpos construídas a partir de sangue de pacientes portadores de Hepatite C Crônica em fases de fibrose diferentes: moderada e avançada.

*1. Capítulo II: Artigo: a ser
submetido na Frontiers in
Immunology*

Título: PERFIL DE ANTICORPOS EXPRESSOS EM PACIENTES PORTADORES DE HEPATITE C CRÔNICA COM DIFERENTES GRAUS DE FIBROSE HEPÁTICA

Cristiane Nonato da Silva^{1,2}, Lucas Farinazzo Marques³, Guilherme Targino Valente³, Francielle Martins Santos^{1,2}, Flávia Hebel Barboza^{1,2}, Leonardo Nazário de Moraes^{1,3}, Maria Inês de Moura Campos Pardini^{1,2}, Adriana Camargo Ferrasi^{1,2}, João Gonçalves⁴, Rejane Maria Tommasini Grotto^{1,3*}

¹ Laboratório de Biologia Molecular do Hemocentro, Botucatu, SP, Brasil.

² Universidade do Estadual de São Paulo (Unesp), Faculdade de Medicina, Botucatu, SP, Brasil.

³ Universidade do Estadual de São Paulo (Unesp), Faculdade de Ciências Agrônômicas, Botucatu, SP, Brasil.

⁴ Instituto de Investigação dos Medicamentos (iMed.Ulisboa), Faculdade de Farmácia, Universidade de Lisboa (ULisboa), Lisboa, Portugal.

Comitê de Ética e Pesquisa: 2.368.179

Financiamento: CNPq – Conselho Nacional de Desenvolvimento Científico e Tecnológico.

*Correspondence: rejane.grotto@unesp.br

Resumo

A progressão da fibrose hepática somada à infecção pelo vírus da hepatite C (VHC) tem sido associada à resposta imunológica permanente. O estudo no repertório de anticorpos Anti-VHC na progressão da fibrose hepática foi explorado através de ferramentas de sequenciamento em larga escala (NGS) possibilitando uma análise de repertórios altamente variáveis como as porções variáveis VH (cadeia pesada) e V κ (cadeia leve) das imunoglobulinas, e a determinação de famílias e subfamílias dos genes V-D-J associadas à resposta humoral encontrada nas diferentes fases da doença proporcional uma ferramenta importante no entendimento da resposta imune frente à infecção viral pelo VHC. As porções VH e V κ das imunoglobulinas foram obtidas a partir da amplificação de RNA de sangue de paciente VHC positivos e com diferentes graus de fibrose, e sequenciadas na plataforma Illumina® Miseq, fornecendo uma grande

variabilidade de sequências que foram pré-processadas por ferramentas de bioinformática e analisadas em dois bancos de anticorpos diferentes: IgBlast (NCBI) e IMGT® quanto às famílias e subfamílias mais expressas. A expressão restrita de algumas famílias e subfamílias: IGHV1, IGHV3, IGHV4 foram encontradas, e podem estar associadas à resposta humoral específica causada pelo VHC, o que confere importante ferramenta o estudo dessas famílias e subfamílias mais expressas durante a infecção. Em especial a família IGVH3 é altamente expressa em vários repertórios imunológicos, e parece estar diminuída nas fases tardias da infecção caracterizada pela progressão da fibrose hepática e perda da função (F3/F4); isso indica o importante papel das famílias de imunoglobulinas principalmente VH3 na resposta contra o vírus nas fases iniciais da doença (F1/F2); e que confere aos achados uma importante indicação de que a progressão da doença pode ser determinada pela diminuição das principais famílias, e que pode ser um importante marcador dessa progressão.

Introdução

A infecção pelo Vírus da Hepatite C (VHC) apresenta abrangência global, com incidência atual de 71 milhões de pessoas infectadas no mundo, das quais cerca de 70% desenvolvem a forma crônica da doença denominada Hepatite C crônica. Embora grandes avanços tenham sido obtidos na última década no tratamento da doença com a introdução dos antivirais de ação direta (DAA) a natureza assintomática da infecção contribui para o diagnóstico tardio, o qual ocorre geralmente quando o dano hepático já é significativo. O diagnóstico tardio, associado à alta capacidade mutagênica do vírus e o dano hepático resultante de um processo fibrótico progressivo nas células do fígado impactam diretamente na resolução e/ou controle da doença (1,2,3,4,5,6).

A fibrose hepática que ocorre durante a Hepatite C crônica é consequência da tentativa de reparo tecidual realizado pelo organismo frente a injúria ocasionada pela presença do vírus. O processo de fibrogênese envolve a substituição do tecido hepático funcional por matriz extracelular, colágeno e outras proteínas relacionadas ao reparo tecidual que ocorre devido a múltiplos eventos, sendo o principal a transformação de células estreladas hepáticas (HSCs) em miofibroblastos contráteis e fibrogênicos (7,8,9,10).

À medida que o tecido hepático é substituído por matriz extracelular a fibrose progride sendo, histologicamente, classificada em quatro estágios, sendo F1 fibrose portal sem septos, F2 expansão fibrosa com poucos septos, F3 numerosos septos podendo visualizar esboços de nódulos e, F4 cirrose (11,12,13).

A progressão para fibrose avançada na Hepatite C crônica parece depender mais de fatores do hospedeiro do que fatores relacionados ao vírus e, além dos fatores genéticos a resposta imunológica humoral, mediada por anticorpos, do paciente infectado pode se relacionar ao desfecho obtido e, prognóstico da doença (1,3,4,5,6,14).

Anticorpos são proteínas secretadas por linfócitos B diferenciados (plasmócitos) estimulados pelo antígeno. Em humanos, eles são compostos de quatro cadeias polipeptídicas, sendo duas pesadas, do inglês, *heavy* (H) e duas leves, do inglês, *light* (L). Cada uma das quatro cadeias polipeptídicas é constituída por uma região variável com, aproximadamente, 110 aminoácidos de comprimento e, uma região constante. As regiões variáveis das cadeias H e L são estruturalmente divididas em três alças hipervariáveis denominadas regiões determinantes de complementaridade (CDRs) e, quatro regiões estruturais denominadas *framework regions* (FR) (15,16,17,18).

A alta variabilidade das VH e V κ ocorrem devido a rearranjos gênicos durante o processo de maturação do linfócito B. Três loci independentes, localizados nos cromossomos 14, 22 e 2 codificam a cadeia pesada (H), cadeia leve do tipo κ e cadeia leve do tipo λ , respectivamente. A cadeia pesada (H) é constituída por três conjuntos de segmentos gênicos (V, D, J) enquanto as cadeias leves por apenas dois (V, J) (18,19).

A diversidade dos anticorpos é resultante de vários processos, os quais ocorrem naturalmente durante os diversos estágios de diferenciação do linfócito B. Inicialmente ocorre justaposição de um segmento D com um J e, em seguida com um V, originando um rearranjo VDJ. Posteriormente durante o processamento do RNA ocorre união do exon rearranjado VDJ com um dos cinco exóons para a cadeia constante (C μ , C γ , C ϵ , C α ou C δ). Para a cadeia leve ocorre processo similar, envolvendo justaposição dos segmentos V e J e posterior união com o exón para cadeia constante (preferencialmente C κ na espécie humana). Essa variabilidade é ainda maior devido a deleções e inserções de nucleotídeos ainda na medula óssea em linfócitos B imaturos e, a hipermutações

somáticas que se acumulam na célula B madura elevando seu potencial de reconhecimento antigênico (19,20,21,22,23).

Anticorpos específicos anti-VHC já foram descritos, sendo que a maioria dos anticorpos neutralizantes descritos são os dirigidos contra os antígenos expressos no envelope viral. Estudos realizados com outros agentes infecciosos mostraram que o perfil de anticorpos expressos pode funcionar como marcador molecular de progressão da doença (referência) (3,16,20,24,25,26,27,28,29).

Anticorpos policlonais específicos podem ser isolados a partir do soro de indivíduos infectados pelo VHC e anticorpos monoclonais podem ser produzidos por hibridomas ou identificados pela triagem de bibliotecas conformacionais de anticorpos. Classicamente estudos dos anticorpos expressos em determinada situação biológica podem ser realizados por técnicas moleculares associadas à triagem imunológica, nas quais bibliotecas de anticorpos são construídas utilizando métodos de biologia molecular e, os anticorpos expressos são posteriormente selecionados utilizando tecnologias fundamentadas na interação antígeno-anticorpo. No entanto, estas tecnologias constituem atualmente métodos onerosos no que se refere a tempo, financiamento, especificidade e sensibilidade (1,30,31,32,33).

Com o advento das tecnologias de sequenciamento em larga escala, do inglês, *Next Generation Sequencing* (NGS) possibilitou a identificação de anticorpos ou frações deles expressos a partir de amostras biológicas com alta sensibilidade e especificidade. O NGS constitui uma metodologia para análise genômica que pode ser utilizada no estudo de imunoglobulinas, permitindo um aumento exponencial no número de dados gerados associada à uma diminuição significativa do custo em comparação com plataformas de sequenciamento convencionais. Gerando aproximadamente 400 bilhões de nucleotídeos em uma corrida a tecnologia conduz a alta sensibilidade no sequenciamento de bibliotecas de anticorpos amplamente variáveis, constituindo um método extremamente sensível no estudo de novos biomarcadores relacionados às patologias humanas (15).

No entanto, a alta variabilidade das imunoglobulinas confere ao processo de sequenciamento um grande desafio frente a complexidade e quantidade de dados gerados. Neste cenário, ferramentas de análises computacionais vêm sendo desenvolvidas para análise de linhagens germinativas e descoberta de novas variantes.

Assim, estratégias para estudo desse repertório imunológico são possíveis associando a geração de dados por NGS e ferramentas de análises de bioinformática, que, em conjunto, permitem avaliar o perfil altamente variável dos anticorpos expressos em determinado momento biológico (34,35).

Neste contexto, considerando que a resposta imunológica humoral do hospedeiro pode se associar a progressão da fibrose hepática no curso da infecção pelo VHC, o presente estudo teve por finalidade construir uma biblioteca das frações VH e V_K de anticorpos expressos em pacientes portadores de Hepatite C crônica com diferentes graus de fibrose hepática e caracterizar as frações expressas por NGS (36,37).

Casuística e Métodos

Seleção das Amostras

Foram incluídas neste estudo 18 amostras de sangue total provenientes de pacientes com anti-VHC IgG reagente, com diagnóstico clínico de Hepatite C crônica, Anti-HIV não reagente, HBsAg, IgM anti-HBc não reagente, sem tratamento anterior para Hepatite C e, com quantificação do RNA do VHC (carga viral) acima de 12 UI/mL armazenadas no repositório de amostras do Laboratório de Biologia Molecular do Hemocentro de Botucatu (Laboratório de Rotinas Diagnósticas em Infecções Virais). De acordo com o estadiamento da fibrose hepática os pacientes foram divididos em dois grupos, sendo Grupo F1/F2: pacientes com fibrose leve a moderada, classificada pelo patologista como F1 ou F2 de acordo com o escore METAVIR (n=9) e, Grupo F3/F4: pacientes com fibrose avançada, classificada pelo patologista como F3 ou F4, segundo escore METAVIR (n=9) (14).

Metodologia

Extração de RNA, constituição dos *pools* e síntese de cDNA

RNA total de cada uma das amostras foi isolado a partir de sangue total e *buffy coat* em tubos separados utilizando *TRIzol™ Reagent* (Invitrogen, Carlsbad, CA, USA) segundo as especificações do fabricante. A integridade do RNA foi analisada por eletroforese em gel de agarose 1,0 % utilizando *Gelred 1X* (GelRed Nucleic Acid Gel Stain-Biotium, Hayward, CA, EUA). A eletroforese foi conduzida à 90 V / 4,0 mA por 60 minutos em tampão Tris-Borato-EDTA 1X (TBE).

As amostras de RNA foram tratadas com Turbo™ DNase (Ambion® - AppliedBiosystem, Foster City, CA, USA), segundo especificações do fabricante. A eficiência da degradação de DNA foi verificada por eletroforese em gel agarose 1,0 %, com *Gelred 1X* (GelRed Nucleic Acid Gel Stain-Biotium, Hayward, CA, EUA), à 90 V/4 mA por 50 minutos em tampão Tris-Borato-EDTA 1X (TBE).

O RNA isolado e isento de DNA foi utilizado para constituição de dois *pools* conforme segue: (a) *pool* F1/F2: constituído do RNA extraído a partir de sangue total e *buffy coat* das amostras obtidas de pacientes com fibrose hepática classificada como F1/F2 pelo escore METAVIR; (b) *pool* F3/F4: constituído do RNA extraído a partir de sangue total e *buffy coat* das amostras obtidas de pacientes com fibrose hepática classificada como F3/F4 pelo escore METAVIR (14).

Os *pools* de RNA obtidos foram quantificados utilizando o fluorímetro Qubit® 2.0 (Invitrogen, Carlsbad, CA, USA) e o kit Qubit® RNA HS Assay (Invitrogen, Carlsbad, CA, USA), segundo as especificações do fabricante.

Os *pools* de RNA foram submetidos à reação de transcrição reversa utilizando a enzima *Improm™ II Reverse Transcriptase* (Promega Corporation, WI, USA), segundo as especificações do fabricante. As reações de transcrição reversa foram realizadas em triplicata sendo um tubo com *primer* para IgG (GTCCACCTTGGTGTGCTGGGCTT) Barbas et al. (2001), um com *primer* para IgM (TGGAAGAGGCACGTTCTTTCTTT) Barbas et al. (2001) e um tubo com *random primer* (Random Thermo Fisher Scientific, Waltham, MA, USA).

Depois da reação de transcrição reversa as triplicatas de cada grupo foram utilizadas para constituir um *pool* de cDNA sendo que foram constituídos um *pool* para F1/F2 e um *pool* para F3/F4.

Amplificação das cadeias variáveis pesada (H) e leve (κ) do Anticorpo

O cDNA foi utilizado como molde para amplificação por PCR das porções variáveis dos anticorpos. Devido à alta variabilidade destas cadeias foi necessária a realização de um conjunto de reações de PCR utilizando 12 combinações de *primers* para VH e quatro combinações para V κ para que a maioria das famílias gênicas para VH e V κ pudessem ser detectadas. As combinações de *primers* utilizadas nestas reações foram as descritas por Barbas e colaboradores (2001). As condições das reações foram Master Mix 2X Go *Taq Hot Start* (Promega Corporation, MA, USA), 10,0 μ M de cada *primer*, 1,0 μ L cDNA e água ultrapura. As condições de ciclagem foram 94 °C por 5 minutos, 5 ciclos de 94 °C por 30 segundos, rampa de ciclagem de 30 %, 60 °C por 45 segundos, 72 °C por 1 minuto e 30 segundos, 40 ciclos 94 °C por 30 segundos, 56 °C por 45 segundos, 72 °C por 1 minuto e 30 segundos, 1 ciclo de 72 °C por 10 minutos. Com o objetivo de elevar a concentração do produto de PCR obtido foram realizadas para o grupo F1/F2 36 reações de PCR para VH e 12 para V κ . No grupo F3/F4 foram realizadas 36 reações de PCR para VH e 24 para V κ .

Os produtos das PCRs foram analisados em gel de agarose 2,0 %, com *GelRed 1X* (*GelRed Nucleic Acid Gel Stain-Biotium, Hayward, CA, EUA*), a eletroforese foi realizada à 80 V / 4 mA por 1 hora e 20 minutos em tampão Tris-Borato-EDTA 1X (TBE). Os produtos foram isolados do gel e purificados utilizando o kit *Invisorb® Fragment Clean Up* (*Invitek, STRATEC Molecular, GmbH, Berlin, Germany*) de acordo com as especificações do fabricante. Foram, então, constituídos quatro *pools*, um *pool* VH-F1/F2 constituído por todos os produtos de PCR de VH do grupo com fibrose moderada, um *pool* V κ -F1/F2 constituído por todos os produtos de PCR de V κ do grupo com fibrose moderada, um *pool* VH-F3/F4 constituído por todos os produtos de PCR de VH do grupo com fibrose avançada, um *pool* V κ -F3/F4 constituído por todos os produtos de PCR de V κ do grupo com fibrose avançada.

A detecção/quantificação dos *pools* de produtos de PCR foram realizadas pela técnica de fluorimetria com o kit *Qubit® dsDNA HS Assay* (*Invitrogen, Carlsbad, CA, USA*), fluorímetro *Qubit® 2.0* (*Invitrogen, Carlsbad, CA, USA*).

Sequenciamento em Larga Escala (NGS) das VH e Vκ

Ambas as bibliotecas VH e Vκ de cada grupo F1F2 e F3F4 foram sequenciadas em duplicata a fim de garantir a homogeneidade e representatividade de cada biblioteca na comparação final dos grupos de fibrose.

Inicialmente, visando otimizar os resultados de NGS, optamos por re-amplificar as regiões variáveis com primers contendo adaptadores, a fim de aumentar em aproximadamente 50 pb em cada extremidade das variáveis para que no processo de tagmentação para inserção dos adaptadores da Illumina não fosse perdido parte da região variável e sim a parte dos novos primers inseridos e descritos na Tabela 1.

Tabela 1. Sequência de primers construídos para a cadeia pesada VH e cadeia leve Vκ.

Primers VH	Sequence
NT-VH-F	5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGCTGCCCAACCACCATG3'
NT-VH-R	5'GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGCGATGGGCCCTTGGTGGAG3'
Primers Vκ	Sequence
NT-Vκ-F	5'TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGGGGCCAGGCGGCCGAG3'
NT-Vκ-R	5'GTCRCGTGGGCTCGGAGATGTGTATAAGAGACAGGAAGACAGATGGTGCAGCCAC3'

NT-VH-F *primer forward* e NT-VH-R *primer reverse* construído para a re-amplificação da variável pesada de ambos os grupos de fibrose. NT-Vκ-F *primer forward* e NT-Vκ-R *primer reverse* construído para a re-amplificação da variável pesada de ambos os grupos de fibrose.

Assim, foi realizada a seguinte reação de PCR para ambas as cadeias pesada e para leve: Master Mix 2x Go Taq Hot Start, 2,0 µL a 10 µM de cada *primer*, 1,5 ng do produto Vh ou Vκ e água ultrapura em um volume final de 20 µL; com o protocolo de ciclagem de 94 °C por 5 minutos, 40 ciclos de 94 °C por 30 segundos, *slope* de 30 %, 57 °C por 45 segundos, 72 °C por 1 minuto, 1 ciclo de 72 °C por 5 minutos.

Todas as reações de re-amplificação foram purificadas utilizando o kit *Invisorb® Fragment Clean Up* (*Invitek, STRATEC Molecular, GmbH, Berlin, Germany*).

Preparação das Bibliotecas para o NGS

Para a preparação das bibliotecas seguiu-se o protocolo da Illumina com algumas adaptações. Os produtos re-amplificados das variáveis VH e Vk foram diluídos para a concentração de 0,5 ng/μL. O kit utilizado foi o *Nextera®XT* (Illumina, Inc., San Diego, USA). A reação de tagmentação foi realizada em uma concentração de 5,0 μL (1,0 ng) do DNA, 10,0 μL do tampão TD (*Tagment DNA Buffer*) e 5,0 μL de ATM (*AmpliconTagment Mix*), e depois centrifugada a 280 xg por 1 minuto à temperatura de -4°C e logo depois colocada no termociclador a 55 °C por 5 minutos e 10 °C por 2 minutos. Após finalização do tempo de tagmentação uma reação de neutralização foi realizada adicionando 5,0 μL do tampão NT (*Neutralize Tagment Buffer*) com incubação por 5 minutos a temperatura ambiente. Após a finalização da etapa de tagmentação do produto, seguiu-se a reação de PCR para indexação das bibliotecas. Em cada poço de uma placa de 96 poços foi adicionado 5,0 μL de cada index e 15,0 μL NPM (*Nextera PCR Master Mix*), a reação foi então centrifugada à 280 xg à 20 °C e ciclada por 3 minutos a 72 °C, depois 95 °C por 30 segundos, 12 ciclos de 95 °C por 10 segundos, 55 °C por 30 segundos, 72 °C por 30 segundos e ciclagem final de 72 °C por 5 minutos. Após a PCR foi realizada a purificação dos produtos com 90,0 μL de *beads* magnéticas *AMPure XP* (BeckmanCoulter Agencourt, Indianapolis, IN, USA) com incubação por 5 minutos à temperatura ambiente. A placa de 96 wells com as *beads* foi colocada em uma rack magnética por 2 minutos e o sobrenadante foi então descartado, e o *pellet* de DNA com as *beads* foram lavados duas vezes com etanol 80 % e incubados por 30 segundos a temperatura ambiente. O etanol foi retirado e descartado e as *beads* ficaram secando por 15 minutos fora da rack magnética. Em seguida, as *beads* foram ressuspensas em 52,5 μL de RBS (*Resuspension Buffer*) incubadas à temperatura ambiente por 2 minutos, a placa foi então colocada novamente na rack magnética por 2 minutos e os sobrenadantes foram transferidos para outros poços da placa.

A biblioteca foi então quantificada por detecção de fluorescência na PCR em tempo real utilizando o kit *KAPA SYBR® FAST ABI PrismqPCR* (KapaBiosystems, Wilmington, MA, USA) de acordo com o protocolo do fabricante, utilizando o equipamento *7500 Real Time PCR System* ((AppliedBiosystems™, Foster City, CA, USA).

O tamanho das bibliotecas construídas foi avaliado por eletroforese em gel de agarose 2% com 3,0 μL de *GelRed 1X (GelRedNucleicAcid Gel Stain-Biotium, Hayward, CA, EUA)* à 70 V / 4 mA por 60 minutos em solução tampão de Tris-Borato-EDTA 1X (TBE).

Cada biblioteca foi corrigida para a concentração para 4,0 nM sendo, em seguida, realizado um *pool* de todas as bibliotecas a 4,0 nM. A biblioteca controle *PhixControl V3 (Illumina, San Diego-USA)* foi utilizada na corrida.

Resumidamente, o *pool* das bibliotecas e o *Phix* diluído a 4,0 nM foram desnaturados com 5,0 μL de NaOH 0,2 N, com homogenização por vortex e centrifugação à 280 xg por 1 minuto, incubados à temperatura ambiente por 5 minutos. Desta forma, a concentração é de 2,0 nM. Foi adicionado 990,0 μL de HT1 (*Hybridization Buffer*) gelado à biblioteca e ao controle reduzindo a concentração a 20,0 pM em um volume final de 1,0 mL. A concentração final do *Phix* foi de 12,5 pM, ou seja, adicionamos 375,0 μL do *Phix* a 20,0 pM a 225,0 μL de tampão HT1, em um volume final de 600,0 μL . A concentração da biblioteca foi ajustada para 11,0 pM, ou seja, adicionamos 330,0 μL do *pool* a 20,0 pM a 270,0 μL de HT1, em volume final de 600,0 μL . Para a corrida, o *Phix* foi utilizado na proporção de 10 % do volume total a ser adicionado ao cartucho. A biblioteca foi sequenciada utilizando a plataforma *MiSeq (Illumina, San Diego-USA)* e o kit *MiSeq Reagent v2 300 cycle (Illumina, San Diego-USA)* com a estratégia *ParedEnd*, onde ambas cadeias de DNA são sequenciadas, com parâmetro de qualidade de > Q30 e cobertura de 24 a 30 milhões de *reads*.

Análise dos Dados Gerados

Os *reads* gerados no sequenciamento foram inicialmente pré-processadas utilizando o *FastQC* (versão 0.11.8) para avaliação da qualidade, sendo estabelecido um Q-Score superior a 30 como parâmetro de qualidade, sendo que *reads* com q-score inferior a este valor foram descartadas. O *Trimmomatic* (versão 0.36) foi utilizado para retirar os *primers* das PCRs, bem como as sequencias dos adaptadores *Illumina*. O *FastQC* também foi utilizado para avaliar o conteúdo de CG, comprimento das sequencias, presença de sequências não únicas e conteúdo de adaptador pós processamento (38,39).

O alinhamento entre as sequências *forward* e *reverse* foi realizado utilizando o script *Pairfq*. O programa *Hisat 2* foi utilizado para o alinhamento das sequências obtidas VH e Vκ utilizando como referências as sequências VH ID:GU179345.1 e Vκ ID:NG_000834.1, respectivamente. O *Hisat 2* também foi utilizado para verificação de eventual contaminação com o vírus do VHC utilizando a sequência viral ID:AF009606.1. (40,41).

A montagem dos *reads* foi realizada utilizando o programa *Velvet* com k-mers 33. O programa *Cap3* foi utilizado para remover sobreposições em que a pontuação de diferença foi maior que o valor especificado e para unir os arquivos *contigs*. A comparação entre a porcentagem de alinhamento e os consensos obtidos foi analisada com o programa *Bowtie-2* (20,42,43,44).

As duplicatas foram unidas e as sequências obtidas foram utilizadas para avaliação das frações VH e Vκ expressas nos dos grupos estudados (F1/F2 e F3/F4) utilizando o CD-HIT e, o banco de anticorpos Ig-Blast do *National Center for Biotechnology Information* e o banco de anticorpos IMGT®, The International ImMunoGeneTics Information System® (34,35,45).

Resultados

Dados Brutos do Sequenciamento e Pré-processamento

De forma geral 741K/mm² *clusters* foram gerados no sequenciamento e, depois de realizadas as análises de pré-processamento os resultados obtidos são mostrados na tabela 2. A figura 1 ilustra um dos resultados obtidos no que se refere a qualidade por base dos *reads* obtidas no sequenciamento antes do processamento (Figura 1A) com as ferramentas de bioinformática e depois deste processo (Figura 1B).

Nota-se uma melhora significativa na qualidade dos *reads* após o pré-processamento realizada. Assim, a partir de 8.310.536 *reads* originalmente obtidas quando considerado o dado bruto gerado no sequenciamento, foram obtidas 354.241

pós-processamento sendo que, em média, o comprimento dos *reads* foi de 100 pares de bases.

A tabela 2 mostra o número de *reads* obtidos por grupo e replicatas pré e pós processamento. Na tabela 2 pode ser observado que houve redução do número de *reads* em todos os grupos e replicatas avaliados.

Tabela 2: Resultados obtidos após pré-processamento das bibliotecas de cada grupo avaliado (F1/F2 e F3/F4) e suas replicatas.

Replicate	Average Quality per read (Q-Score)	CC contente (%)	Per base N contente (%)	Sequence length (bp)	Non-unique sequences (%)	Adapter Content (%)
VH a F1/F2	38	56	0	150-151	76,11	0
VH b F1/F2	38	56	0	150-151	75,1	0
Vκ a F1/F2	38	59	0	150-151	40,47	0
Vκ b F1/F2	38	59	0	150-151	39,26	0
VH a F3/F4	38	64	0	150-151	53,87	0
VH b F3/F4	38	64	0	150-151	51,84	0
Vκ a F3/F4	38	59	0	150-151	51,84	0
Vκ b F3/F4	38	55	0	150-151	46,24	0

VH: cadeia variável pesada; Vκ: cadeia variável leve do tipo kappa; a: primeira replicata; b: segunda replicata; F1/F2: grupo com fibrose moderada segundo escore Metavir; F3/F4: grupo com fibrose avançada segundo escore Metavir.

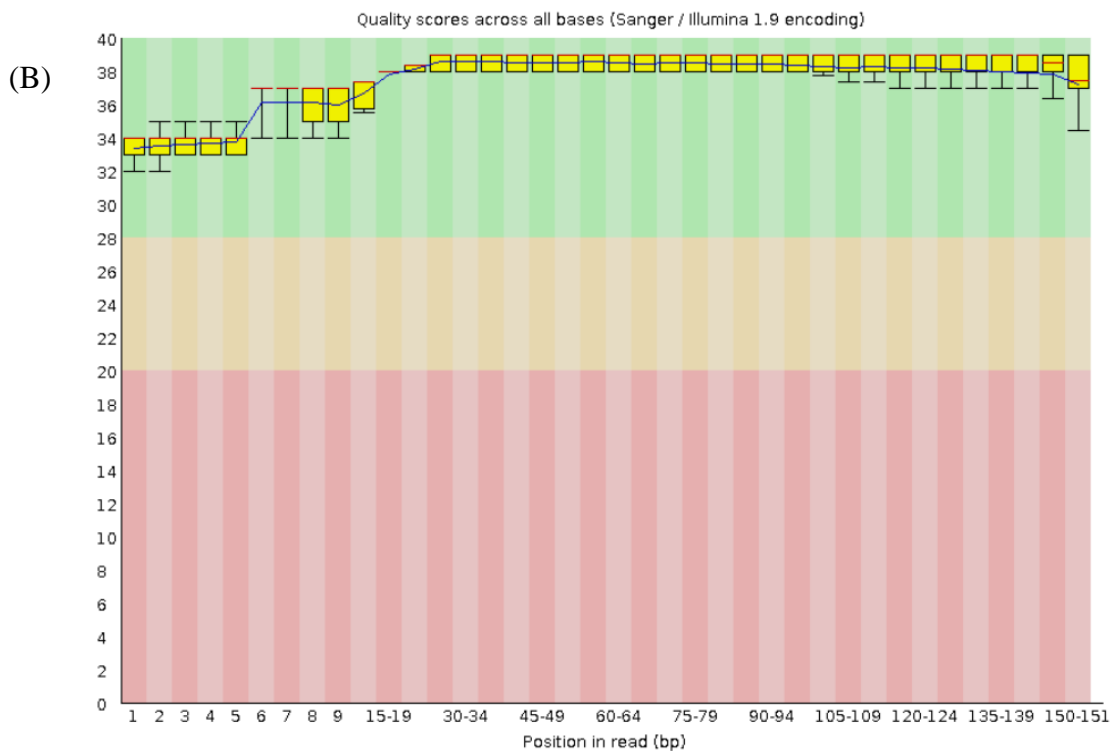
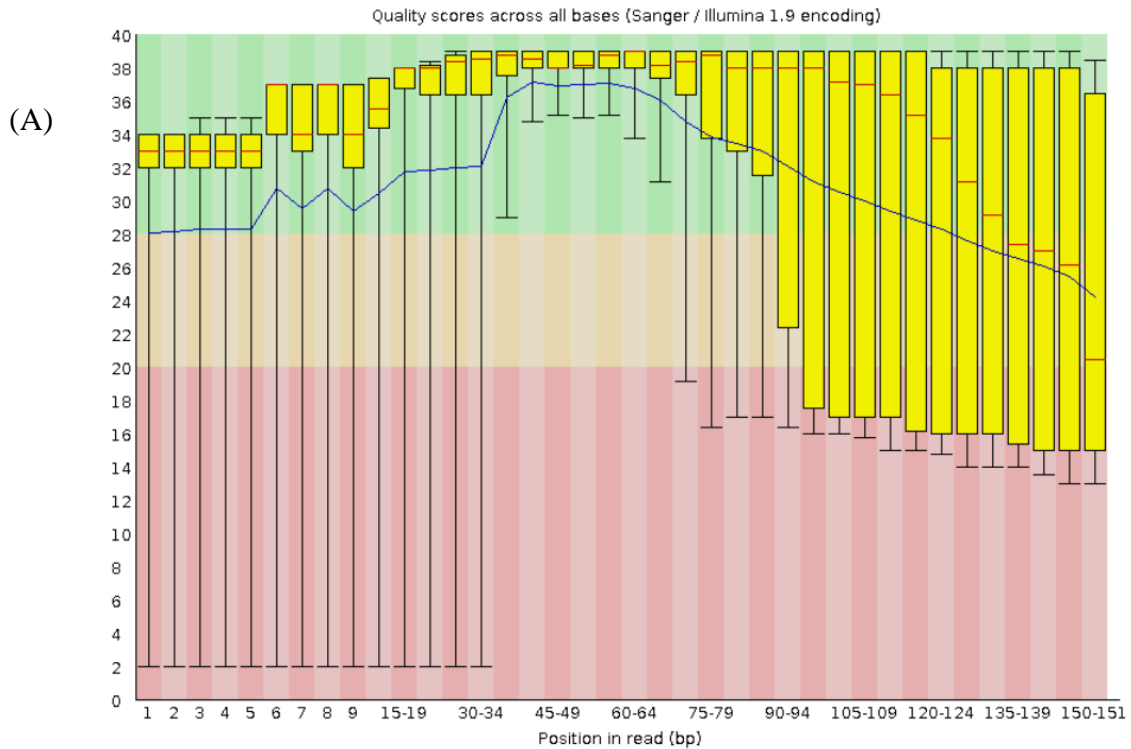


Figura 1: Exemplo de imagens representativas da qualidade por base ao longo das sequências evidenciando valor médio (linhas vermelhas), intervalo interquartil 25-75% (caixas amarelas) e a média da qualidade obtida (linha azul). (A) Imagem antes do processamento dos dados brutos; (B) Imagem pós-processamento dos dados brutos.

Tabela 3: Número de *reads* pré e pós processamento nos grupos (F1/F2 e F3/F4) e replicatas (a e b)

	Número de <i>reads</i> Pré Processamento	Número de <i>reads</i> Pós Processamento
VH a F1/F2	931.100	16.384
VH b F1/F2	950.106	18.504
V κ a F1/F2	908.601	62.906
V κ b F1/F2	1.093.926	73.568
VH a F3/F4	757.015	15.154
VH b F3/F4	820.184	17.509
V κ a F3/F4	1.060.911	51.988
V κ b F3/F4	1.788.693	98.228

VH: cadeia variável pesada; V κ : cadeia variável leve do tipo *kappa*; a: primeira replicata; b: segunda replicata; F1/F2: grupo com fibrose moderada segundo escore Metavir; F3/F4: grupo com fibrose avançada segundo escore Metavir.

As duplicatas dos *reads* obtidos (ver tabela 3) foram unidas e utilizadas para montagem das sequências de VH e V κ por bioinformática (ver metodologia de análise de dados) sendo obtidas 6171 e 3440 sequências de VH no grupo F1/F2 e, F3/F4, respectivamente. Quanto ao número de sequências de V κ foram obtidas 3191 e 7968 sequências nos grupos F1/F2 e F3/F4, respectivamente. Estas sequências foram utilizadas para avaliação das famílias e subfamílias das imunoglobulinas, dos rearranjos VDJ e VJ e CDR3-H encontrados nos grupos estudados.

Frequência das famílias dos segmentos gênicos das frações VH

A figura 2 ilustra a frequência das famílias gênicas da cadeia pesada VH nos grupos estudados F1/F2 e F3/F4 analisadas em dois bancos de sequências de anticorpos diferentes.

Em F1/F2 nota-se que as frequências das famílias V foram: IGHV1 (24 %), IGHV3 (43 %), IGHV4 (30 %), IGHV5 (3 %) de 17.952 linhagens germinativas do gene V da

cadeia pesada, e que apresentaram semelhança com as sequências da biblioteca Anti-VHC, e para o grupo F3/F4 de 6.168 linhagens germinativas alinhadas: IGHV1 (11 %), IGHV3 (23 %), IGHV4 (62 %), IGHV5 (4 %), realizadas com o IgBlast (Figura 2a).

Comparando a análise do banco de dados IMGT encontramos para o grupo F1/F2: IGHV1 (27 %), IGHV3 (37 %), IGHV4 (34 %), IGHV5 (2 %), de 37.281 linhagens germinativas alinhadas com o banco. E para o grupo F3/F4 temos IGHV1 (17 %), IGHV3 (18 %), IGHV4 (63 %), IGHV5 (2 %) do total de 16.347 linhagens germinativas alinhadas (Figura 2b).

As famílias IGHV1, IGHV3 e IGHV4 foram as mais frequentes ($p < 0,0001$) em ambos os grupos (Figura 2a e 2b).

No que se refere à avaliação das subfamílias não foi possível uma análise estatística devido à metodologia utilizada no NGS assim esta análise foi realizada considerando o contexto biológico com uma análise descritiva dos dados.

A família IGHV3 com 7.664 alinhamentos na linhagem germinativa, sendo a mais expressa do gene V encontramos subfamílias mais frequentes no grupo F1/F2, sendo as subfamílias IGHV3-30 (20 %), IGHV3-23 (19 %) e IGHV3-11 (10 %) em relação ao total de VH3 encontrado no IgBlast (Figura 2e).

Enquanto na análise do IMGT com 13.786 alinhamentos da família VH3 encontramos as mesmas subfamílias IGHV3-30 (28 %), IGHV3-23 (17 %), porém, a subfamília IGHV3-11 não estava entre as mais frequentes (Figura 2f).

Comparando ao grupo F3/F4 com 1.390 alinhamentos da família VH3, encontramos as mesmas três subfamílias como as mais expressas no IgBlast: IGHV3-30 (15 %), IGHV3-23 (15 %) e IGHV3-11 (32 %); e também no IMGT do total de 2.921 alinhamentos as subfamílias IGHV3-30 (23 %), IGHV3-23 (13 %) e IGHV3-11 (21 %), foram as mais prevalentes, porém, frequência diferentes entre os bancos (Figura 2e e 2f).

A família IGHV1 também foi uma das mais expressas em ambos os grupos de fibrose, e foram analisadas quanto as subfamílias mais prevalentes no grupo F1/F2, sendo as subfamílias IGHV1-69 (32 %), IGHV1-2 (22 %), IGHV1-18 (16 %) IGHV1-46 (14 %) e IGHV1-8 (11 %) as mais frequentes em relação ao total de 4.356 VH1

encontrado no IgBlast (Figura 2c). Na análise do IMGT encontramos as mesmas subfamílias como as mais expressas, porém, com frequência diferente do banco comparado, e também encontramos uma exceção à subfamília IGHV1-8 que não estava entre as mais frequentes; sendo: IGHV1-69 (27 %), IGHV1-2 (32 %), IGHV1-18 (17 %) IGHV1-46 (11 %) as mais frequentes do total de 10.163 alinhamentos (Figura 2d).

Em relação ao grupo F3/F4 com 687 alinhamentos as subfamílias IGHV1-69 (37 %), IGHV1-2 (11 %), IGHV1-18 (8 %) IGHV1-46 (14 %) e IGHV1-8 (25 %) foram encontradas como as mais prevalentes no IgBlast (Figura 2c), e no IMGT encontramos as subfamílias IGHV1-69 (20 %), IGHV1-2 (32 %), IGHV1-18 (17 %); já as subfamílias IGHV1-46 e IGHV1-8 não estavam como as mais expressas do total de 2.757 alinhamentos (Figura 2d).

Com relação à família IGHV4 com 5.378 alinhamentos, também foi mais expressa em ambos os grupos encontramos as subfamílias IGHV4-30 (22 %), IGHV4-59 (15 %), IGHV4-34 (13 %) IGHV4-39 (12 %) e IGHV4-4 (10 %) mais prevalentes no grupo F1/F2 no IgBlast (Figura 2g). Na análise do IMGT de 12.475 alinhamentos encontramos as subfamílias IGHV4-30 (14 %), IGHV4-59 (17 %), IGHV4-34 (14 %), IGHV4-39 (15 %) como as mais expressas, porém, as subfamílias IGHV4-31 (13 %) e IGHV4-61 (12 %) também aparecem entre as mais frequentes na análise deste banco de anticorpos, enquanto que a IGHV4-4 não estava entre as mais frequentes (Figura 2h).

No grupo F3/F4 com 3.826 alinhamentos as subfamílias IGHV4-30 (20 %), IGHV4-59 (15 %), IGHV4-34 (20 %) e IGHV4-4 (13 %) foram encontradas como as mais prevalentes no IgBlast (Figura 2g), e comparando com o IMGT com 10.266 alinhamentos encontramos as subfamílias IGHV4-30 (14 %), IGHV4-59 (17 %), IGHV4-34 (22 %), IGHV4-39 (9 %), IGHV4-31 (11 %) e IGHV4-61 (11 %) (Figura 2h).

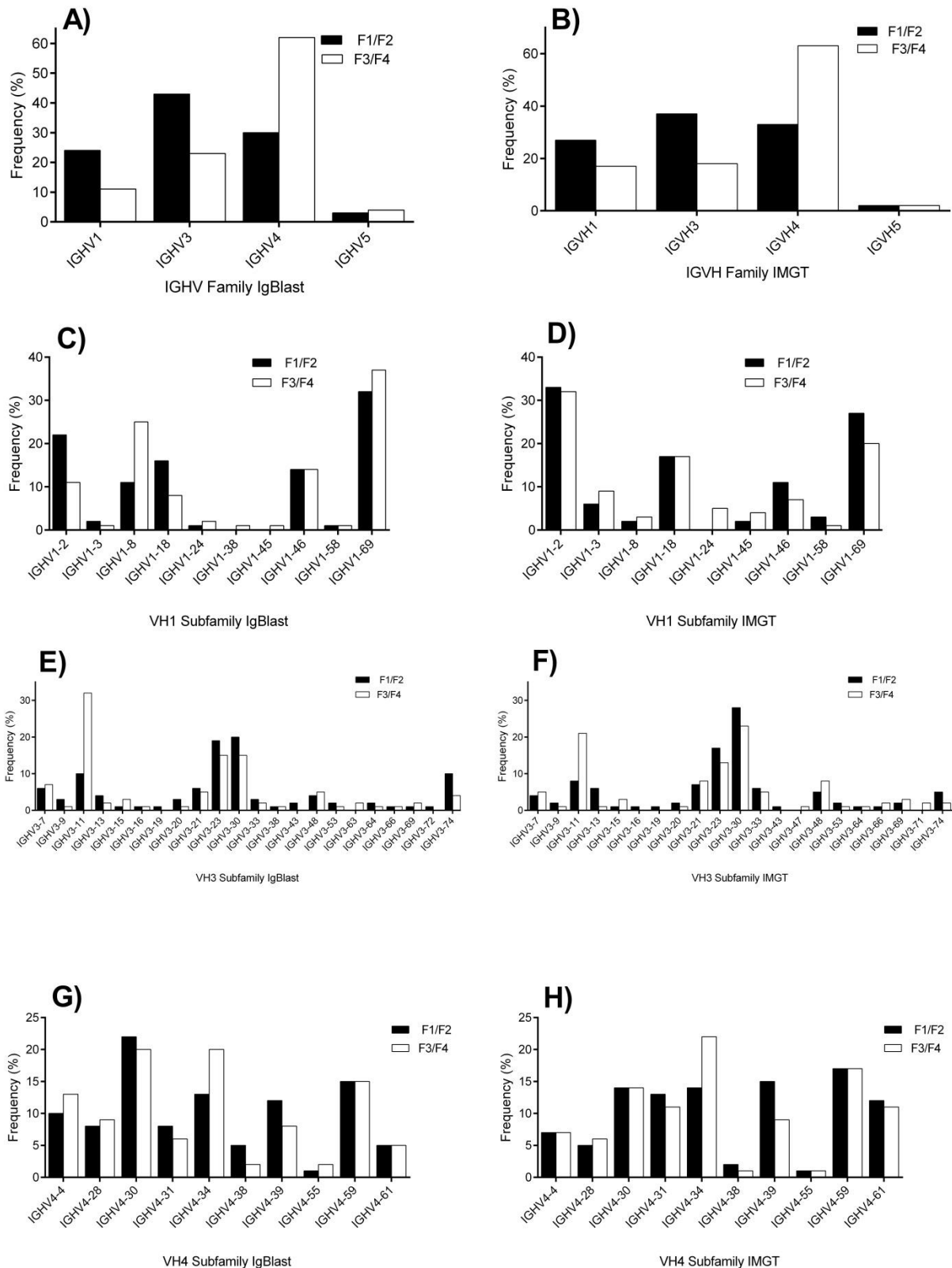


Figura 2. Comparação da frequência das Famílias e Subfamílias do gene V mais expressas na cadeia pesada VH nos grupos F1/F2 e F3/F4. A) Frequência das famílias V em cada grupo no IgBlast (Teste Chi-square $p = <0,0001$). B) Frequência das famílias V em cada grupo no IMGT (Teste Chi-square $p = 0,0006$). C) Frequência das subfamílias VH1 no IgBlast. D) Frequência das subfamílias VH1 no IMGT. E) Frequência das subfamílias VH3 no IgBlast. F) Frequência das subfamílias VH3 no IMGT. G) Frequência das subfamílias VH4 no IgBlast. H) Frequência das subfamílias VH4 no IMGT.

Quanto aos genes IGHD encontramos 2.457 alinhamentos na linhagem germinativa do IgBlast e a família IGHD3 foi o mais frequente com 33 % no grupo F1/F2 seguida pelo IGHD2 com 26 % de alinhamento. Comparando a família IGHD3 no grupo F3/F4 obteve 39 % de total de 358 alinhamentos na análise IgBlast. Em segundo lugar a família IGHD2 com 23 % (Figura 3a).

A família IGHD3 foi também a mais frequente na análise IMGT para o grupo F1/F2, com 25 % do total de 471 famílias. Já em comparação com F3/F4 a mais frequente foi IGHD2 com (22 %) do total de 36 famílias (Figura 3b).

A análise do gene de junção J apresentou 1.899 alinhamentos na linhagem germinativa do IgBlast com 61 % de alinhamento para a família IGHJ4, sendo está a mais frequente no grupo F1F2, e em comparação o grupo F3/F4 apresentou 64 % de 178 alinhamentos na linhagem germinativa do IgBlast, sendo está família a mais prevalente em ambos os grupos de fibrose para a cadeia pesada dos anticorpos aqui analisados (Figura 3c).

Na análise com o IMGT em ambos os grupos de fibrose a família IGHJ4 foi a mais frequente, com 51 % de 841 para F1/F2 e 34 % de 166 para F3/F4 (Figura 3d).

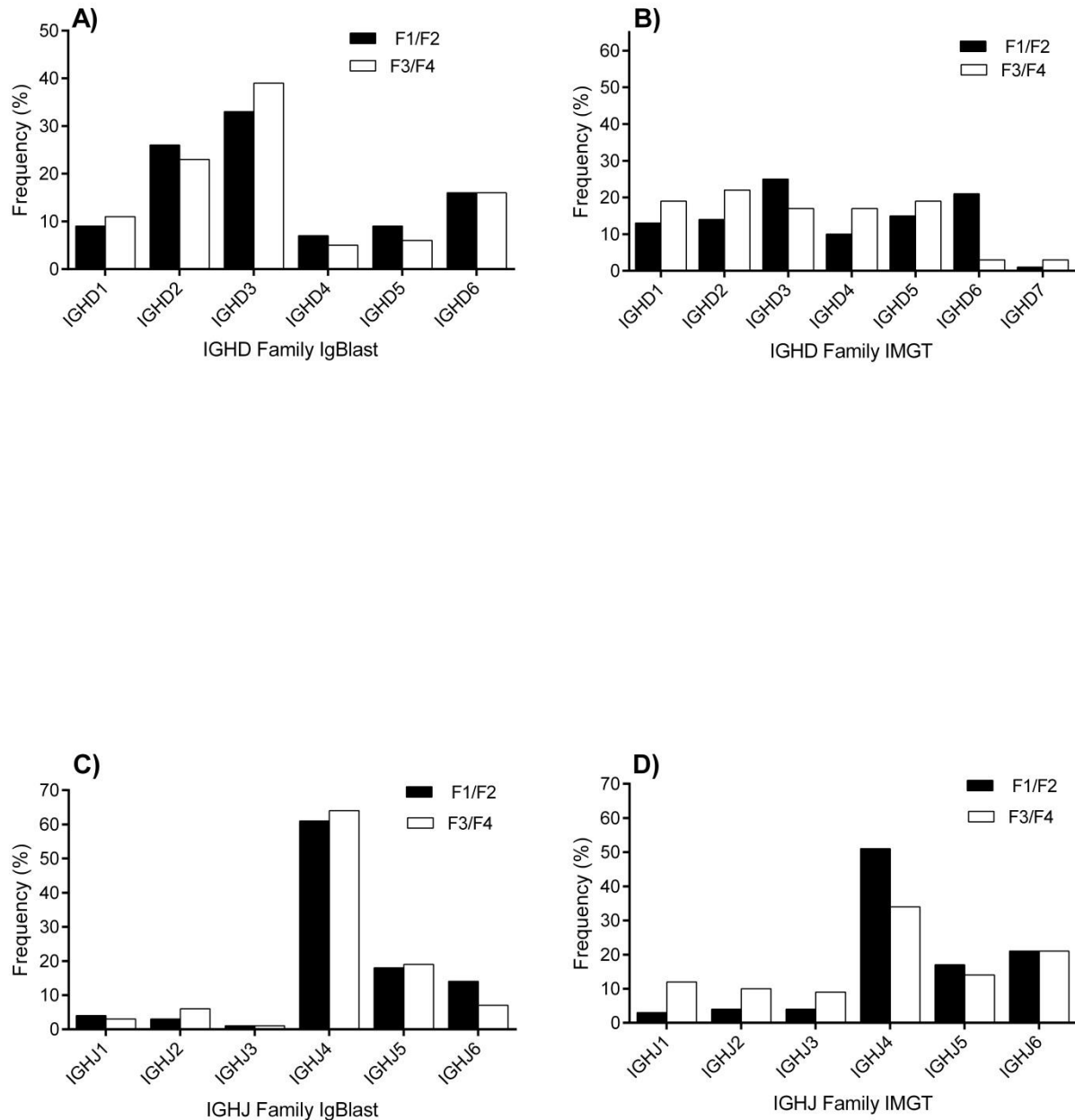


Figura 3. Comparação da frequência das Famílias dos genes D e J mais expressas na cadeia pesada VH dos grupos F1/F2 e F3/F4 nos bancos de anticorpos IgBlast e IMGT. A) Frequência das famílias do gene D no IgBlast (Teste Chi-square $p= 0,8738$). B) Frequência das famílias do gene D no IMGT (Teste Chi-square $p= 0,0011$). C) Frequência das famílias do gene J no IgBlast (Teste Chi-square for trend $p= 0,3054$). D) Frequência das famílias do gene J no IMGT (Teste Chi-square $p= 0,0185$).

Frequência das famílias dos segmentos gênicos das frações V_{κ}

Análise das frequências observadas da cadeia leve V_{κ} do grupo F1/F2 demonstraram o seguinte percentual para as famílias analisadas: $IG_{\kappa}V1$ (28 %), $IG_{\kappa}V2$

(4 %), IG κ V3 (65 %), IG κ V4 (3 %) de 8.543 alinhamentos no IgBlast; e no grupo F3/F4 IG κ V1 (23 %), IG κ V2 (9 %), IG κ V3 (64 %), IG κ V4 (4 %), de 22.393 alinhamentos no IgBlast (Figura 4a).

Em comparação a análise do banco de anticorpos IMGT com 10.188 alinhamentos apresentou o seguinte percentual para o grupo F1/F2: IG κ V1 (28 %), IG κ V2 (2 %), IG κ V3 (68 %), IG κ V4 (2 %), e para o grupo F3/F4 IG κ V1 (25 %), IG κ V2 (6 %), IG κ V3 (66 %), IG κ V4 (3 %), do total de 24.540 (Figura 4b).

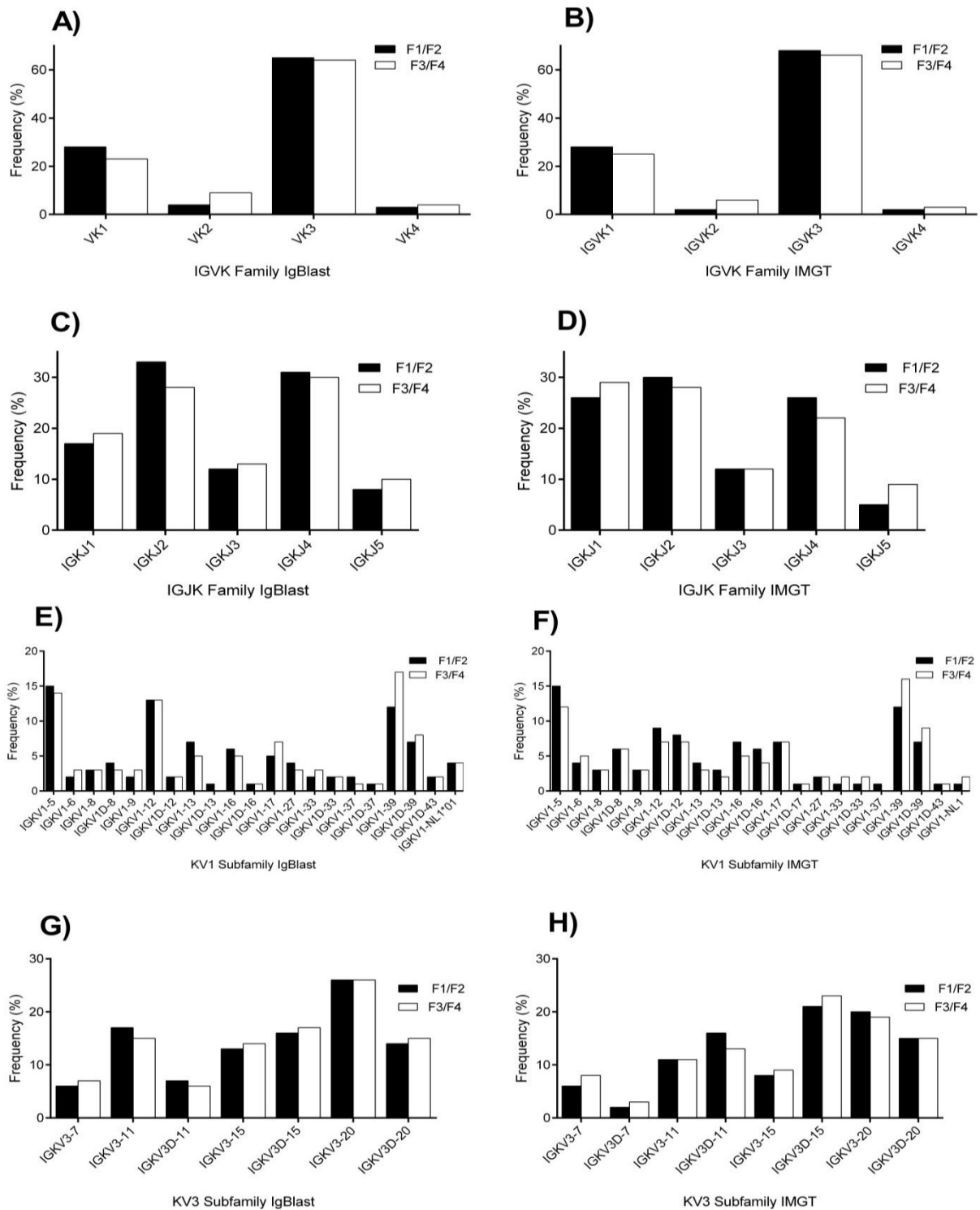


Figura 4. Comparação da frequência das Famílias e Subfamílias do gene V e J da cadeia leve (V_{κ}) F1/F2 e F3/F4 em dois bancos de anticorpos diferentes. A) Frequência das famílias do gene V na cadeia leve no IgBlast (Teste Chi-square $p = 0,4639$). B) Frequência das famílias V da cadeia leve no IMGT (Teste Chi-square $p = 0,4937$). C) Frequência das famílias do gene J na cadeia leve no IgBlast (Teste Chi-square $p = 0,9392$). D) Frequência das famílias do gene J na cadeia leve no IMGT (Teste Chi-square $p = 0,7900$). E) Frequência das subfamílias $V_{\kappa}1$ no IgBlast. F) Frequência das subfamílias $V_{\kappa}1$ no IMGT. G) Frequência das subfamílias $V_{\kappa}3$ no IgBlast. H) Frequência das subfamílias $V_{\kappa}3$ no IMGT.

As famílias do gene V da cadeia leve V_{κ} no grupo F1/F2 mais frequentes foram: $IG_{\kappa}V3$ com 65 %, e 64 % para F3/F4; seguida da família $IG_{\kappa}V1$ com 28 % para F1/F2 e 25 % para F3/F4 no IgBlast e também no IMGT encontramos as mesmas famílias como mais expressas percentualmente (Figura 4a e 4b).

Analisando as subfamílias mais prevalentes de $IG_{\kappa}V1$ com 2.361 alinhamentos, temos: $IG_{\kappa}V1-5$ (15 %), $IG_{\kappa}V1-12$ (13 %) e $IG_{\kappa}V1-39$ (12 %) respectivamente no grupo F1F2; e no grupo F3F4 com 5.188 alinhamentos as subfamílias mais prevalentes eram as mesmas, porém, em porcentagem diferentes de predominância, sendo $IG_{\kappa}V1-39$ (17 %), $IG_{\kappa}V1-5$ (14 %), $IG_{\kappa}V1-12$ (13 %) respectivamente mais expressas no IgBlast (Figura 4e).

Em comparação no IMGT encontramos $IG_{\kappa}V1-5$ (15 %) e $IG_{\kappa}V1-39$ (12 %) no grupo F1/F2 de 2.816 alinhamentos; e as mesmas subfamílias no grupo F3/F4 sendo: $IG_{\kappa}V1-5$ (12 %), $IG_{\kappa}V1-39$ (16 %) do total de 6.072 (Figura 4f).

A subfamília $IG_{\kappa}V3-20$ (26 %) em ambos os grupos foi uma das mais expressas em ambos os grupos e em ambas as análises (Figuras 4g e 4h).

Todas as cinco famílias $IG_{\kappa}J$ foram utilizados na cadeia leve. Sendo em particular $IG_{\kappa}J1$ (17 %), $IG_{\kappa}J2$ (33 %), $IG_{\kappa}J3$ (12 %), $IG_{\kappa}J4$ (31 %), $IG_{\kappa}J5$ (7 %) para o grupo F1/F2 do total de 1.511 alinhamentos na linhagem germinativa e as famílias $IG_{\kappa}J1$ (19 %), $IG_{\kappa}J2$ (28 %), $IG_{\kappa}J3$ (13 %), $IG_{\kappa}J4$ (30 %), $IG_{\kappa}J5$ (10 %) para o grupo F3/F4 do total de 5.138 no IgBlast (Figura 4c).

No IMGT temos $IG_{\kappa}J1$ (26,1 %), $IG_{\kappa}J2$ (30,3 %), $IG_{\kappa}J3$ (12,2 %), $IG_{\kappa}J4$ (26 %), $IG_{\kappa}J5$ (5,4 %) para o grupo F1/F2 do total de 541 alinhamentos na linhagem germinativa; e as famílias $IG_{\kappa}J1$ (29 %), $IG_{\kappa}J2$ (28 %), $IG_{\kappa}J3$ (12 %), $IG_{\kappa}J4$ (22 %), $IG_{\kappa}J5$ (9 %) para o grupo F3/F4 do total de 1.751. Com predominância dos grupos $IG_{\kappa}J1$, $IG_{\kappa}J2$ e $IG_{\kappa}J4$ para ambos os grupos analisados (Figura 4d).

Diversidade de CDR3 de comprimento em composições de aminoácidos

A maior diversidade no repertório de anticorpos é devido ao arranjo V-D-J encontrada na CDR3-H. A variação de aminoácidos quanto ao tamanho da CDR3-H no grupo da cadeia pesada de F1/F2 houve uma variação de 4-29 aminoácidos, com maior porcentagem em cerca de 14 aminoácidos enquanto no grupo VH F3/F4 foi a maior porcentagem foi de 9 aminoácidos no IgBlast (Figura 5).

Em comparação a análise do IMGT a CDR3-H da cadeia pesada mostrou uma variação no tamanho de 1-25 aminoácidos em F1/F2 com uma frequência maior em 2 seguidos de 13 aminoácidos. E em F3/F4 a variação foi de 11-17 aminoácidos de tamanho com uma frequência de 2 seguida de 10 aminoácidos.

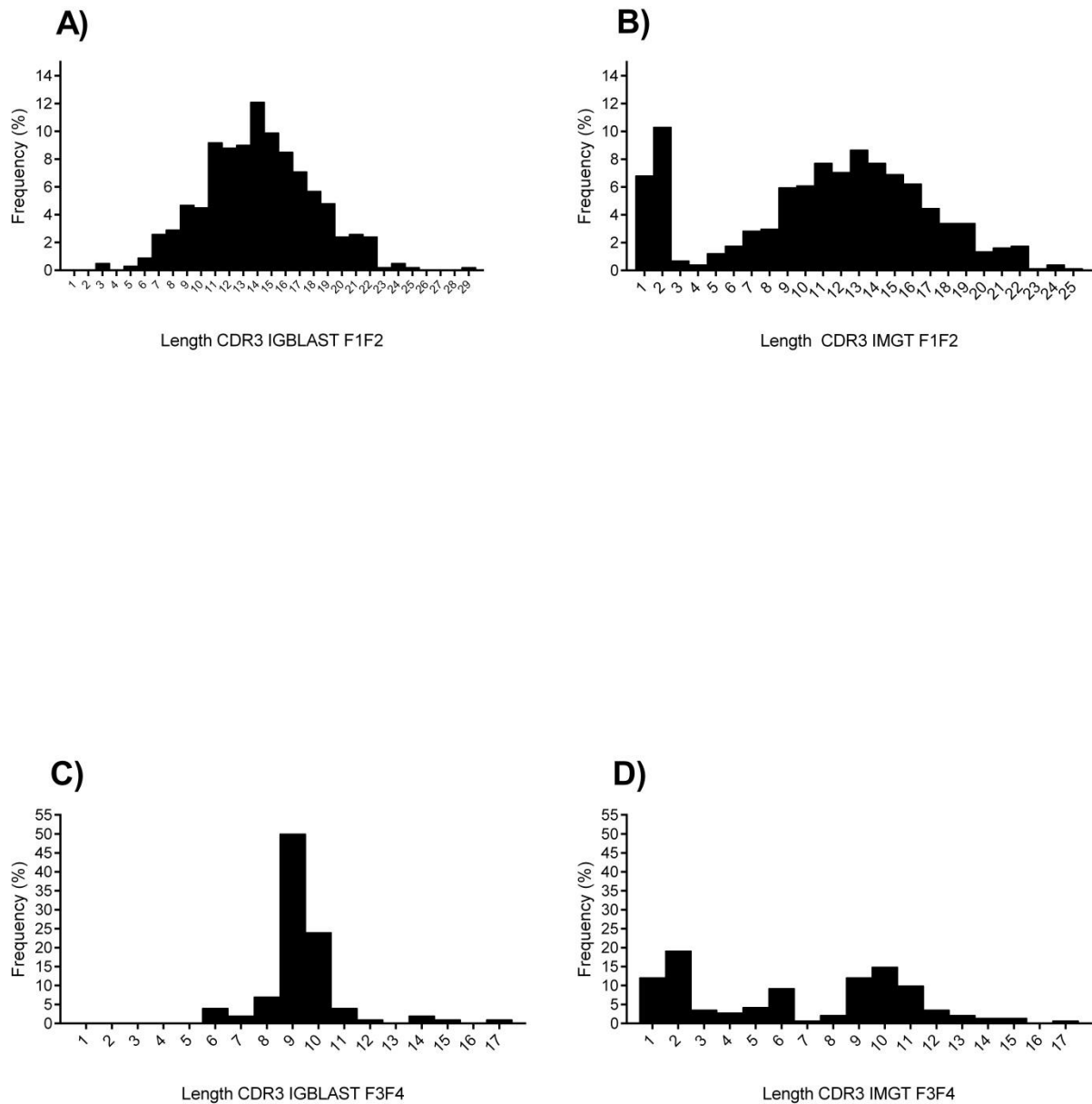


Figura 5. Comprimento em aminoácidos da CDR3-H nos grupos F1F2 e F3F4. A) Comprimento da CDR3 no grupo VH F1/F2 do IgBlast. B) A) Comprimento da CDR3 no grupo VH F1/F2 do IMGT. C) Comprimento da CDR3 no grupo VH F3/F4 do IgBlast. D) Comprimento da CDR3 no grupo VH F3/F4 do IMGT.

Discussão

O estudo do repertório de anticorpos de pessoas infectadas pelo VHC com estadiamento de fibrose moderada (F1/F2) e avançada (F3/F4) permite compreender melhor a evolução da resposta imune frente à infecção crônica causada pelo vírus; e também possui um papel importante na identificação de moléculas com potencial

neutralizante para alvos virais contribuindo para o desenvolvimento de vacinas e medicamentos (11,20,28,46,47).

Através da análise estatística podemos encontrar significância na comparação das famílias VH dos grupos de fibrose em ambos os bancos de anticorpos. Apesar de não podermos comparar cada família separadamente devido à limitação dos testes estatísticos para os grupos analisados, encontramos percentualmente uma diferença nas frequências das famílias analisadas. Os resultados mostraram diferença de frequência para a família IGHV3, IGHV4 e IGHV1 entre os grupos (F1/F2 e F3/F4) sendo a família IGHV3 mais expressa no grupo F1/F2 em ambos os bancos de dados. As maiores frequências encontradas nessas três famílias já foram descritas por Arcaini e colaboradores no estudo de linfomas da zona marginal esplênica associado à infecção pelo VHC; demonstrando que esta família é preferencialmente expressa em pacientes com Hepatite C, resultado concordante com o aqui obtido (48).

Embora não existam na literatura científica estudos que avaliaram a presença dos segmentos gênicos da fração VH durante a progressão da Hepatite C crônica estudos realizados com pacientes portadores de HIV-1 mostraram que a frequência desta família genica é reduzida com a progressão da doença. Assim, os resultados aqui obtidos sugerem que a alteração da frequência de IGHV3 de F1/F2 para F3/F4 constitua um marcador imunológico de progressão da fibrose hepática; visto que é a família mais expressa em repertórios normais (49,50).

Quando analisamos as subfamílias do gene V da cadeia pesada mais em ambos os grupos não encontramos diferença estatística na expressão das subfamílias, porém, percentualmente as subfamílias mais expressas IGHV1-2, IGHV3-23, IGHV3-30 e IGVH4-34 também foram descritas por Arcaini e colaboradores no estudo de linfomas associados ao VHC, o que confirma nossos achados (48).

A subfamília IGVH1-69 foi a mais frequente percentualmente em ambos os grupos na análise do IgBlast e a segunda mais expressa na análise do IMGT, mostrando a importância desta subfamília na infecção pelo VHC, o que é corroborado por estudos com linfomas de células B associado a infecção pelo VHC. Tucci, 2017 e colaboradores mostraram que a IGHV1-69 foi prevalente no grupo com VHC quando comparado a não infectados. A prevalência da subfamília IGVH1-69 já foi associada à resposta imunológica contra a glicoproteína de superfície viral E2, sendo que a estimulação induzida pela proteína E2 pode causar a expansão de células B específicas que expressam anticorpos codificados por IGHV1-69 (26,48,51,52).

Estudos com anticorpos que expressam preferencialmente o gene IGHV1-69 visam regiões hidrofóbicas conservadas de glicoproteínas do envelope viral, o que demonstra seu potencial como anticorpos neutralizantes *Broadly neutralizing antibodies* (bnAbs) o qual é explorado no estudo do desenvolvimento de vacinas (26,52,53,54).

Além da IGHV1-69 a subfamília IGHV3-30 foi, também, uma das mais frequentes no grupo F1/F2 e F3/F4 em ambos os bancos de anticorpos sendo que esta subfamília foi descrita como mais prevalente no estudo com pacientes portadores de VHC tendo, também sido associada à presença a doenças autoimunes e linfomas em pacientes crônicos. (48,51).

Em distúrbios linfoproliferativos relacionados ao VHC as subfamílias IGHV3-21, IGHV3-23, foram também predominantemente encontradas em anticorpos anti-VH1-69, derivados de pacientes VHC positivos e com Crioglobulinemia Mista, indicando uma reação entre anticorpos de famílias diferentes, processo que pode estar envolvido nas reações autoimunes comumente encontradas em pacientes infectados pelo VHC. As subfamílias VH3-33, VH3-7, VH3-43, VH3-30, VH3-20, VH3-74 também foram descritas nesse estudo porém não estavam entre as mais expressas. Estes resultados reforçam a hipótese de que o perfil das famílias e subfamílias das frações VH dos anticorpos possam constituir marcadores da doença (53,54).

No que se refere ao gene D encontramos diferença estatística significativa na análise do IMGT e não no IgBlast, porém todas as famílias foram utilizadas. Na literatura encontramos a família HD3 foi relacionada à infecção pelo VHC, sendo descrita como a mais frequente em um estudo de Racanelli e colaboradores (2011) na comparação de células B de pacientes crônicos, saudáveis e com clareamento espontâneo sendo mais prevalente nas células B de memória dos pacientes crônicos. As maiores frequências do gene D é DH3 seguido do DH2 na análise IgBlast o que corrobora com os achados de Arcaini (2009) e colaboradores no estudo de linfoma de zona marginal esplênica em pacientes VHC positivos. A diferença encontrada em ambos os bancos é devida possivelmente à dificuldade dos bancos de anticorpos em determinar a linhagem germinativa do gene D, pelo fato do gene D ser muito curto e a alta similaridade com outros genes existentes na linhagem germinativa, e também devido a adições aleatórias de nucleotídeos no rearranjo das junções, e a presença de mutações somáticas, todos contribuem para dificultar a determinação da linhagem germinativa (35,48,55).

Em relação à subfamília dos segmentos J da cadeia pesada VH, obtivemos diferença estatística na análise IMGT e não no IgBlast, porém percentualmente a J4 foi a

predominante em ambos os grupos e em ambos os bancos, e também junto a J6 foi uma das mais predominantes no estudo da infecção pelo VHC em pacientes que resolveram a infecção e em pacientes crônicos. Arcaini e colaboradores (2009) também descreveu HJ4, HJ6, HJ5 como as mais prevalentes respectivamente no estudo de linfoma de zona marginal esplênica em pacientes VHC positivos; demonstrando que o uso preferencial dessa família encontrado em nosso estudo pode ser também um importante marcador da expressão preferencial induzida pelo VHC (48,52,55).

No que se refere à cadeia leve do tipo *kappa* estatisticamente não podemos determinar diferença significativa entre os grupos F1/F2 e F3/F4, porém quando analisamos separadamente cada família percentualmente a KV3 seguida de KV1 foi a predominante em ambos os grupos e ambos os bancos, sendo as subfamílias IG κ V3-15 e IGKV3-20 as mais predominantes em ambos os grupos de fibrose, resultado concordando com Re e colaboradores (2006) no estudo de células B produtoras de fator reumatóide (FR) e crioglobulinemia mista tipo II em pacientes VHC positivos (54,56).

A subfamília IGKV3-20 já foi descrita por Buonaguro e colaboradores (2010) como uma molécula usada avaliar a resposta de células apresentadoras de antígeno (APC) e citocinas em pacientes VHC positivos. Re e colaboradores (2009) também descreveram o desenvolvimento de uma vacina com IG κ V3-20 associada a IGHV4-59 derivadas de pacientes VHC positivos com Crioglobulinemia Mista II e Linfoma não-Hodgkin (54,56,57).

O segmento gênico J da cadeia leve *kappa* estatisticamente não apresentou diferença entre os grupos, e analisando individualmente encontramos maior frequência nas famílias IG κ J1 e IG κ J2 e IG κ J4 em ambos os grupos e em ambas as análises, porém as cinco famílias foram utilizadas com pequenas diferenças em cada grupo, porém com significativa expressão frente ao repertório imunológico do gene de junção, indicando ser de potencial importância no estudo da resposta imune frente ao estímulo crônico do vírus conforme encontramos no nosso estudo (53).

As subfamílias IGHV1-8, IGHV3-11, IGHV4-39, IG κ V1-5, IG κ V1-12, IG κ V1-39 também foram encontradas como as mais expressas no nosso estudo podendo indicar um papel preferencial de expressão dessas imunoglobulinas na resposta humoral contra o VHC ainda não descrito na literatura (54).

Estruturalmente, a CDR3-H está localizada no centro do local de ligação ao antígeno, e seu comprimento apresenta uma distribuição aproximadamente normal,

tipo Gaussiano, para repertórios humanos, o que pode ser verificado no grupo VH F1/F2 de ambas as análises, porém o grupo F3/F4 não apresentou tal distinção, mas sim apresentou CDR3s mais curtas que o grupo comparado, resultado concordante com a literatura da área que relata que o processo de maturação de afinidade das células B antígeno-específicas leva a um menor comprimento de CDR3, ou ainda indica deleções por hipermutações somáticas que podem ser sido induzidas pela carga viral. Porém encontramos na literatura que um menor comprimento da CDR3 pode afetar a carga, a hidrofobicidade, o tamanho e a forma do sítio de ligação ao antígeno, o que interfere no processo de reconhecimento antigênico, o que podemos verificar no grupo de fibrose avançada, isso pode indicar uma menor resposta na fase avançada da doença, ou ainda que a diminuição da resposta imune contribuí com a progressão da fibrose (55,60,62).

Apesar de não podermos determinar estatisticamente a diferença preferencial de algumas das famílias, uma análise percentual baseada na literatura indica diferença de expressão entre algumas famílias e subfamílias em relação às outras, e diante de um grande repertório de possível combinações de imunoglobulinas que podem ser formadas à partir da linhagem germinativa, encontramos um número limitado de linhagens expressas nesse estudo, o que indica o uso preferencial desses segmentos gênicos na infecção pelo vírus. Uma análise de comparação do repertório de anticorpos anti-VHC com antígenos VHC-específicos ainda é necessária para elucidação e confirmação dos achados, embora os mecanismos subjacentes para a diversificação de anticorpos sejam particulares de cada patologia, os achados contribuem para o melhor entendimento da resposta imune humoral frente à infecção pelo VHC e também de outras patologias associadas (49,50,54).

Conclusão

Analisando o perfil de imunoglobulinas encontrado na biblioteca fica evidente a tendência do uso de algumas subfamílias em ambas as análises de bancos de dados, esses achados sugerem o uso preferencial de algumas subfamílias na infecção pelo VHC, e, que o vírus pode influenciar a expressão de algumas famílias e subfamílias em detrimento de outras, como encontrado na família IGHV3 que estava menos expressa no grupo F3/F4 de fibrose avançada, o indica uma mudança no padrão de resposta imunológica à medida que a doença progride, visto que essa família é mais

representada em todos os repertórios estudados, sendo indicativa de uma molécula de potencial a ser explorado no estudo e desenvolvimento de terapias direcionadas contra o VHC.

1. Capítulo III- Conclusão

2. Conclusão:

O uso preferencial de algumas subfamílias já descritas em vários estudos associados ao VHC corrobora com nossos achados de que existe uma tendência do uso de algumas subfamílias como: IGHV1-2, IG VH1-69, IGHV3-21, IGHV3-23, IG VH3-30, IGHV4-4, IGHV4-34 IGHV4-39 na cadeia pesada, assim como IG κ V3-15 e IG κ V3-20 na cadeia leve, esses dados evidenciam que o vírus pode influenciar a expressão de algumas famílias e subfamílias em detrimento de outras; como encontramos na família IG VH3 que estava menos expressa no grupo F3/F4 de fibrose avançada; resultado esse inovador na literatura relacionada ao VHC o sugere uma mudança no padrão de resposta imunológica à medida que a doença progride como encontrada em outras patologias virais já amplamente difundido na literatura como no caso do HIV. Outras subfamílias encontradas: IGHV1-8, IGHV3-11, IGHV4-39, IG κ V1-5, IG κ V1-12, IG κ V1-39 também estavam entre as mais expressas indicando que podem estar relacionadas também a uma expressão preferencial induzida pelo vírus e constituem achado novo na literatura relacionado ao VHC.

2.0 Referências Bibliográficas:

1. Eliyahu S. et al. Antibody Repertoire Analysis of Hepatitis C Virus Infections Identifies Immune Signatures Associated With Spontaneous Clearance. *Frontiers In Immunology*. 2018; 9: 1-17. doi: 10.3389/fimmu.2018.03004. Frontiers Media SA.
2. World Health Organization. WHO Hepatitis C. [homepage on the internet] World Health Organization; Acesso em: 01 de Agosto de 2019. Disponível em: <<https://www.who.int/news-room/fact-sheets/detail/hepatitis-c>>.
3. Gu J. et al. Escape of Hepatitis C Virus from Epitope I Neutralization Increases Sensitivity of Other Neutralization Epitopes. *Journal Of Virology*. 2018; 92 (9): 1-19. doi: 10.1128/jvi.02066-17. American Society for Microbiology.
4. Irshad M, Gupta P, Irshad K. Immunopathogenesis of Liver Injury During Hepatitis C Virus Infection. *Viral Immunology*. 2019; 32 (3): 112-120. doi: 10.1089/vim.2018.0124. Mary Ann Liebert Inc.
5. Prentoe J, Bukh J. Hypervariable Region 1 in Envelope Protein 2 of Hepatitis C Virus: A Linchpin in Neutralizing Antibody Evasion and Viral Entry. *Frontiers In Immunology*. 2018; 9: 1-11. doi: 10.3389/fimmu.2018.02146. Frontiers Media SA.
6. Yan Q. et al. Next generation sequencing reveals novel alterations in B-cell heavy chain receptor repertoires associated with acute-on-chronic liver failure. *International Journal Of Molecular Medicine*. 2018; 43 (1): 243-255. doi: 10.3892/ijmm.2018.3946. Spandidos Publications.
7. Ajorloo M. et al. Detection of specific antibodies to HCV-ARF/CORE+1 protein in cirrhotic and non-cirrhotic patients with hepatitis C: A possible association with progressive fibrosis. *Arch Iran Med*. 2015; 18 (5): 304 – 307.
8. Bhattacharya D. et al. Isolated Hepatitis B Core Antibody is Associated With Advanced Hepatic Fibrosis in HIV/HCV Infection But Not in HIV Infection Alone. *J aids Journal Of Acquired Immune Deficiency Syndromes*. 2016; 72 (1): 14-17. doi: 10.1097/qai.0000000000000941. Ovid Technologies (Wolters Kluwer Health).
9. Kassela K. et al. High prevalence of antibodies to core+1/ARF protein in HCV-infected patients with advanced cirrhosis. *Journal of General Virology*. 2017; 98 (7): 1713-1719. doi: 10.1099/jgv.0.000851. Microbiology Society.
10. Lee UE, Friedman SL. Mechanisms of hepatic fibrogenesis. *Best Practice & Research Clinical Gastroenterology*. 2011; 25 (2): 195-206. doi: 10.1016/j.bpg.2011.02.005. Elsevier BV.
11. Pedersen J, Lundbo LF, Krarup H, Bukh J, Weis N. Neutralizing antibodies in patients with chronic hepatitis C and correlation to liver cirrhosis and estimated duration of infection. *Journal of Medical Virology*. 2016; 88 (10): 1791-1803. doi: 10.1002/jmv.24537. Wiley.

12. Sebastiani G. Chronic hepatitis C and liver fibrosis. *World Journal Of Gastroenterology*. 2014; 20 (32): 11033-11053. doi: 10.3748/wjg.v20.i32.11033. Baishideng Publishing Group Inc.
13. Smorodin E, Sergeyev B, Kurtenkov O, Kuznetsova T, Geller J. IgG Antibodies to GlcNAc β and Asialo-GM2 (GA2) Glycans as Potential Markers of Liver Damage in Chronic Hepatitis C and the Efficacy of Antiviral Treatment. *Disease Markers*. 2018; 2018: 1-13. doi: 10.1155/2018/4639805. Hindawi Limited.
14. Pickhardt PJ. et al. Multiparametric CT for Noninvasive Staging of Hepatitis C Virus–Related Liver Fibrosis: Correlation With the Histopathologic Fibrosis Score. *American Journal Of Roentgenology*. 2019; 212 (3): 547-553. doi: 10.2214/ajr.18.20284. American Roentgen Ray Society.
15. Barreto K, et al. Next-generation sequencing-guided identification and reconstruction of antibody CDR combinations from phage selection outputs. *Nucleic Acids Research*. 2019; 47 (9): 1-13. doi: 10.1093/nar/gkz131. Oxford University Press (OUP).
16. Re. et al. Type II mixed cryoglobulinaemia as an oligo rather than a mono B-cell disorder: evidence from GeneScan and MALDI-TOF analyses. *Rheumatology*. 2006; 45: 685–693. doi:10.1093/rheumatology/kei278.
17. Ping Y, Song M, Wang M, Li Z, Zhang Y. CDR3 repertoire diversity of CD8+ T lymphocytes in patients with HCV. *Cellular Immunology*. 2019; 336: 34-39. doi: 10.1016/j.cellimm.2018.12.007. Elsevier BV.
18. Hwang JK, Alt FW, Yeap LS. Related Mechanisms of Antibody Somatic Hypermutation and Class Switch Recombination. In: *Mobile DNA III*. American Society of Microbiology. 2015; 3 (1): 325–348. doi: 10.1128/microbiolspec.MDNA3-0037-2014.
19. Jiang Y, Nie K, Redmond D, Melnick AM, Tam W, Elemento O. VDJ-Seq: Deep Sequencing Analysis of Rearranged Immunoglobulin Heavy Chain Gene to Reveal Clonal Evolution Patterns of B Cell Lymphoma. *Journal of Visualized Experiments*. 2015; (106): 1-10. doi: 10.3791/53215. MyJove Corporation.
20. Bailey JR. et al. Broadly neutralizing antibodies with few somatic mutations and hepatitis C virus clearance. *Jci Insight*. 2017; 2 (9): 1-17. doi: 10.1172/jci.insight.92872. American Society for Clinical Investigation.
21. Bazzill JD. et al. Interrogation of Antigen Display on Individual Vaccine Nanoparticles for Achieving Neutralizing Antibody Responses against Hepatitis C Virus. *Nano Letters*. 2018; 18 (12): 7832-7838. doi: 10.1021/acs.nanolett.8b03601. American Chemical Society (ACS).
22. Chaudhary N, Wesemann DR. Analyzing Immunoglobulin Repertoires. *Frontiers in Immunology*. 2018; 9: 1-18. doi: 10.3389/fimmu.2018.00462. Frontiers Media SA.
23. Krawczyk K. et al. Structurally Mapping Antibody Repertoires. *Frontiers in Immunology*. 2018; 9: 1-8. doi: 10.3389/fimmu.2018.01698. Frontiers Media SA.

24. Aleman F. et al. Immunogenetic and structural analysis of a class of HCV broadly neutralizing antibodies and their precursors. *Proceedings of The National Academy Of Sciences*. 2018; 115 (29): 7569-7574. doi: 10.1073/pnas.1802378115. *Proceedings of the National Academy of Sciences*.
25. Bailey JR. et al. Broadly neutralizing antibodies with few somatic mutations and hepatitis C virus clearance. *Jci Insight*. 2017; 2 (9): 1-17. doi: 10.1172/jci.insight.92872. *American Society for Clinical Investigation*.
26. Chen F. et al. VH1-69 antiviral broadly neutralizing antibodies: genetics, structures, and relevance to rational vaccine design. *Current Opinion In Virology*. 2019; 34: 149-159. doi: 10.1016/j.coviro.2019.02.004. *Elsevier BV*.
27. Christiansen D. et al. Immunological responses following administration of a genotype 1a/1b/2/3a quadrivalent HCV VLP vaccine. *Scientific Reports*. 2018; 8 (1): 1-13. doi: 10.1038/s41598-018-24762-9. *Springer Nature*.
28. Kinchen VJ. et al. Broadly Neutralizing Antibody Mediated Clearance of Human Hepatitis C Virus Infection. *Cell Host & Microbe*. 2018; 24 (5): 717-730. doi: 10.1016/j.chom.2018.10.012. *Elsevier BV*.
29. Tzarum N, Wilson IA, Law M. The Neutralizing Face of Hepatitis C Virus E2 Envelope Glycoprotein. *Frontiers In Immunology*. 2018; 9: 1-8. doi: 10.3389/fimmu.2018.01315. *Frontiers Media SA*.
30. Fukasawa M. et al. Monoclonal Antibodies against Extracellular Domains of Claudin-1 Block Hepatitis C Virus Infection in a Mouse Model. *Journal Of Virology*. 2015; 89 (9): 4866-4879. doi: 10.1128/jvi.03676-14. *American Society for Microbiology*.
31. Giang E, Aleman F, Law M. Probing the Antigenicity of HCV Envelope Glycoproteins by Phage Display Antibody Technology. *Methods In Molecular Biology*. 2018; 1911: 381-393. doi: 10.1007/978-1-4939-8976-8_26. *Springer New York*.
32. Keck ZY, Wang Y, Lau P, Fong SKH. Isolation of HCV Neutralizing Antibodies by Yeast Display. *Methods In Molecular Biology*. 2018; 1911: 395-419. doi: 10.1007/978-1-4939-8976-8_27. *Springer New York*.
33. Barbas III CF, Burton DR, Scott JK, Silverman GJ. *Phage display : a laboratory manual*. Cold Spring Harbor Laboratory Press; 2001.
34. Giudicelli V. et al. IG and TR single chain fragment variable (scFv) sequence analysis: a new advanced functionality of IMGT/V-QUEST and IMGT/HighV-QUEST. *Bmc Immunology*. 2017; 18 (1): 1-13. doi: 10.1186/s12865-017-0218-8. *Springer Nature*.
35. Ye J, Ma N, Madden TL, Ostell JM. IgBLAST: an immunoglobulin variable domain sequence analysis tool. *Nucleic Acids Research*. 2013; 41 (1): 34-40. doi: 10.1093/nar/gkt382. *Oxford University Press (OUP)*.

36. Farfán-Arribas DJ. et al. The dynamics of immunoglobulin V-gene usage and clonotype expansion in mice after prime and boost immunizations as analyzed by NGS. *Human Vaccines & Immunotherapeutics*. 2017; 13 (12): 2987-2995. doi: 10.1080/21645515.2017.1379638. Informa UK Limited.
37. Kawajiri-Manako C. et al. Clonal immunoglobulin λ light-chain gene rearrangements detected by next generation sequencing in POEMS syndrome. *American Journal of Hematology*. 2018; 93 (9): 1161-1168. doi: 10.1002/ajh.25213. Wiley.
38. Andrews S. (2010). FastQC: a quality control tool for high throughput sequence data. Acesso em: <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>.
39. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics*. 2014; 30 (15): 2114-2120. doi: 10.1093/bioinformatics/btu170. Oxford University Press (OUP).
40. Kim D, et al. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nature Biotechnology*. 2019; 37 (8): 907-915. Springer Science and Business Media LLC. Doi: 10.1038/s41587-019-0201-4.
41. Stanton E, Chef B. Pairfq: Pairfq version 0.16.1. Zenodo, 2016.
42. Huang X. CAP3: A DNA Sequence Assembly Program. *Genome Research*. 1999; 9 (9): 868-877. doi: 10.1101/gr.9.9.868. Cold Spring Harbor Laboratory.
43. LANGMEAD, Ben; SALZBERG, Steven L. Fast gapped-read alignment with Bowtie 2. *Nature Methods*, [s.l.], v. 9, n. 4, p.357-359, 4 mar. 2012. Springer Nature. <http://dx.doi.org/10.1038/nmeth.1923>.
44. Zerbino DR, Birney E. Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Research*. 2008; 18 (5): 821-829. doi: 10.1101/gr.074492.107. Cold Spring Harbor Laboratory.
45. Li W, Godzik A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*. 2006; 22 (13): 1658-1659. doi: 10.1093/bioinformatics/btl158. Oxford University Press (OUP).
46. Krey T, Meola A, Keck ZY, Damier-Piolle L, Fong SK, Rey FA.. Structural Basis of HCV Neutralization by Human Monoclonal Antibodies Resistant to Viral Neutralization Escape. *Plos Pathogens*. 2013; 9 (5): 1-10. doi: 10.1371/journal.ppat.1003364. Public Library of Science (PLoS).
47. Merat SJ. et al. Hepatitis C virus Broadly Neutralizing Monoclonal Antibodies Isolated 25 Years after Spontaneous Clearance. *Plos One*. 2016; 11 (10): 1-18. doi: 10.1371/journal.pone.0165047. Public Library of Science (PLoS).
48. Arcaini L. et al. Splenic marginal zone lymphoma: Clinical clustering of immunoglobulin heavy chain repertoires. *Blood Cells, Molecules, and Diseases*. 2009; 42 (3): 286-291. doi: 10.1016/j.bcmd.2009.01.004. Elsevier BV.

49. Chang Q. VH3 gene expression in children with HIV infection. *Journal of Infection*. 2004; 49 (4): 274-282. doi: 10.1016/j.jinf.2004.04.007. Elsevier BV.
50. Scamurra RW. et al. Impact of HIV-1 Infection on VH3 Gene Repertoire of Naive Human B Cells. *The Journal Of Immunology*. 2000; 164 (10): 5482-5491. doi: 10.4049/jimmunol.164.10.5482. The American Association of Immunologists.
51. Tucci FA. et al. Biased IGH VDJ gene repertoire and clonal expansions in B cells of chronically hepatitis C virus–infected individuals. *Blood*. 2017; 131 (5): 546-557. doi: 10.1182/blood-2017-09-805762. American Society of Hematology.
52. Marasca R. et al. Immunoglobulin Gene Mutations and Frequent Use of VH1-69 and VH4-34 Segments in Hepatitis C Virus-Positive and Hepatitis C Virus-Negative Nodal Marginal Zone B-Cell Lymphoma. *Molecular Immunology*. 2001; 159 (1): 253-261. Doi: 10.1016/S0002-9440(10)61691-4. American Journal of Pathology.
53. Perotti M. et al. Hepatitis C virus (HCV)-driven stimulation of subfamily-restricted natural IgM antibodies in mixed cryoglobulinemia. *Autoimmunity Reviews*. 2008; 7 (6): 468-472. doi: 10.1016/j.autrev.2008.03.008. Elsevier BV.
54. Sautto G, Mancini N, Solfrosi L, Diotti RA, Clementi M, Burioni R. HCV Proteins and Immunoglobulin Variable Gene (IgV) Subfamilies in HCV-Induced Type II Mixed Cryoglobulinemia: A Concurrent Pathogenetic Role. *Clinical And Developmental Immunology*. 2012; 2012: 1-11. doi: 10.1155/2012/705013. Hindawi Limited.
55. Racanelli V. et al. Antibody Vh Repertoire Differences between Resolving and Chronically Evolving Hepatitis C Virus Infections. *Plos One*. 2011; 6 (9): 1-11. doi: 10.1371/journal.pone.0025606. Public Library of Science (PLoS).
56. Re V. et al. Characterization of Antibodies Directed against the Immunoglobulin Light κ Chain Variable Chain Region (VK) of Hepatitis C Virus-Related Type-II Mixed Cryoglobulinemia and B-Cell Proliferations. *Annals of The New York Academy Of Sciences*. 2009; 1173 (1): 152-160. doi: 10.1111/j.1749-6632.2009.04677.x. Wiley.
57. Buonaguro L. et al. Immune signatures in human PBMCs of idiotypic vaccine for HCV-related lymphoproliferative disorders. *Journal of Translational Medicine*. 2010; 8 (1): 1-11. doi: 10.1186/1479-5876-8-18. Springer Nature.
58. Charles ED. et al. Clonal expansion of immunoglobulin M+CD27+ B cells in HCV-associated mixed cryoglobulinemia. *Blood*. 2007; 111 (3): 1344-1356. doi: 10.1182/blood-2007-07-101717. American Society of Hematology.
59. Brochet X, Lefranc MP, Giudicelli V. IMGT/V-QUEST: the highly customized and integrated system for IG and TR standardized V-J and V-D-J sequence analysis. *Nucleic Acids Research*. 2008; 36: W503-508. doi: 10.1093/nar/gkn316. Oxford University Press (OUP).

60. Boisvert, M. et al. Novel E2 Glycoprotein Tetramer Detects Hepatitis C Virus–Specific Memory B Cells. *The Journal of Immunology*. 2016; 197 (12): 4848-4858. doi: 10.4049/jimmunol.1600763. The American Association of Immunologists.
61. Ping Y, Song M, Wang M, Li Z, Zhang Y. CDR3 repertoire diversity of CD8+ T lymphocytes in patients with HCV. *Cellular Immunology*. 2019; 336: 34-39. doi: 10.1016/j.cellimm.2018.12.007. Elsevier BV.
62. Flyak AI. et al. HCV Broadly Neutralizing Antibodies Use a CDRH3 Disulfide Motif to Recognize an E2 Glycoprotein Site that Can Be Targeted for Vaccine Design. *Cell Host & Microbe*. 2018; 24 (5): 703-716. doi: 10.1016/j.chom.2018.10.009. Elsevier BV.