UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

GUILHERME NAVARRO NILO GIUSTI

CARACTERIZAÇÃO DO REPERTÓRIO DE CÉLULAS B EM CRIANÇAS COM LEUCEMIA LINFOIDE AGUDA

CAMPINAS

2025

GUILHERME NAVARRO NILO GIUSTI

CARACTERIZAÇÃO DO REPERTÓRIO DE CÉLULAS B EM CRIANÇAS COM LEUCEMIA LINFOIDE AGUDA

Tese apresentada ao Instituto de Biologia da Universidade Estadual de Campinas como parte dos requisitos exigidos para a obtenção do título de Doutor em Genética e Biologia Molecular, na Área de Concentração de Bioinformática.

Orientador: DR. JOSÉ ANDRÉS YUNES

ESTE ARQUIVO DIGITAL CORRESPONDE À VERSÃO FINAL DA TESE DEFENDIDA PELO ALUNO GUILHERME NAVARRO NILO GIUSTI E ORIENTADA PELO DR. JOSÉ ANDRÉS YUNES.

CAMPINAS

2025

Ficha catalográfica
Universidade Estadual de Campinas (UNICAMP)
Biblioteca do Instituto de Biologia
Mara Janaina de Oliveira - CRB 8/6972

G449c

Giusti, Guilherme Navarro Nilo, 1993-
Caracterização do repertório de células B em crianças com leucemia linfoide aguda / Guilherme Navarro Nilo Giusti. – Campinas, SP : [s.n.], 2025.

Orientador: José Andrés Yunes.
Tese (doutorado) – Universidade Estadual de Campinas (UNICAMP), Instituto de Biologia.

1. Leucemia linfoide aguda. 2. Recombinação V(D)J. 3. Linfócitos B. 4. Genes de cadeia pesada de imunoglobulina. I. Yunes, José Andrés, 1967-. II. Universidade Estadual de Campinas (UNICAMP). Instituto de Biologia. III. Título.

Informações complementares

Título em outro idioma: Characterization of B cell repertoire in children with acute lymphoblastic leukemia
Palavras-chave em inglês:
Acute lymphoblastic leukemia
V(D)J recombination
B-Lymphocytes
Genes, Immunoglobulin heavy chain
Área de concentração: Bioinformática
Titulação: Doutor em Genética e Biologia Molecular
Banca examinadora:
José Andrés Yunes [Orientador]
Carlos Alberto Scrideli
Marcelo Falsarella Carazzolle
Marcelo Mendes Brandão
Ricardo Camargo
Data de defesa: 14-03-2025
Programa de Pós-Graduação: Genética e Biologia Molecular

Objetivos de Desenvolvimento Sustentável (ODS)
ODS: 3. Saúde e bem-estar

Identificação e informações acadêmicas do(a) aluno(a)
- ORCID do autor: https://orcid.org/0000-0002-4251-8153
- Currículo Lattes do autor: http://lattes.cnpq.br/7466469942930928

BANCA DA APROVAÇÃO

Prof. Dr. José Andrés Yunes

Prof. Dr. Carlos Alberto Scrideli

Prof. Dr. Marcelo Falsarella Carazzolle

Prof. Dr. Marcelo Mendes Brandão

Prof. Dr. Ricardo Camargo

A Ata da defesa com as respectivas assinaturas dos membros encontra-se no SIGA/Sistema de Fluxo de Tese e na Secretaria do Programa de Genética e Biologia Molecular da Unidade Instituto de Biologia.

AGRADECIMENTOS

Primeiramente, agradeço a minha mãe Viviane pelo constante apoio incondicional em todas as aventuras e desventuras da minha vida. Você foi, é e sempre será o meu maior exemplo e orgulho.

À minha namorada, Caroline, por ser a minha eterna companheira, celebrando ao meu lado as conquistas e bons momentos; e, ainda mais importante, sempre me ajudando a me levantar nos momentos de queda.

A toda a minha família, que sempre se fez presente quando a minha maior necessidade foi Casa e Raiz. No mais, quando a família é tão importante, é sempre essencial lembrar das origens que nos proporcionaram tudo isso. Por isso, o mais especial dos agradecimentos aos meus avós, Dalva e Filadelfo. Não menos importante, aos meus primos: Lu, Bruneira, Duda, Guide, Dani, Leo, Rafa, Nando, Vick e Bruna. Quem disse que a vida não me deu irmãos? E, para não esquecer, agora também tenho uma nova geração para agradecer: João e Gabriel, obrigado por alegrarem ainda mais essa nossa família.

Agradeço muito também ao meu pai, Carlos, que apesar da distância nunca deixou com que eu me sentisse menos amparado ou amado. A sua postura e caráter me inspiram hoje e sempre.

Um grande agradecimento aos amigos verdadeiros que fiz durante essa minha nômade jornada, irmãos que escolhemos para compartilhar uma vida. Em especial, gostaria de agradecer àqueles que se fizeram especialmente presentes ao longo desse processo, tornando-o menos pesado. Ao Boris e ao KIQ, companheiros da mais longa data que semanalmente alegram minha semana. Aos bagaceiros Diego, Zeni, Léo e Victor pelas anedotas e discussões filosóficas. Ao Abel, por me lembrar que para a vida ser simples e alegre basta um bom amigo para discutir futebol. Ao Fabrício, sem quem minhas mesas de queijos e vinhos nunca ficam completas.

Agradeço também ao meu orientador, o Dr. Andrés, pela disponibilidade e ensinamentos ao longo desses anos de pós-graduação. O seu carinho ao realizar o seu trabalho sempre foi uma inspiração. Aos demais colegas do Centro Infantil Boldrini, que das mais diversas maneiras contribuíram para esse trabalho e para um bom convívio de dia a dia. Em especial, agradeço à toda equipe da DRM: Dr. Meidanis, Dra. Jotta, Dra. Ganazza, Dra. Migita e Samara.

Um enorme agradecimento às crianças e famílias que cederam amostras para a realização desse estudo. A capacidade de empatia e boa vontade com o próximo das pessoas que estão passando pelos momentos mais complicados é emocionante.

Por fim, agradeço às instituições de fomento a pesquisa que ajudaram a financiar o projeto referente a essa tese de doutorado. O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001 (PROEX 88887.342097/2019-00). Também contou com o apoio do Programa Nacional de Apoio à Atenção Oncológica (PRONON, NUP 25000.057709/2015 e NUP 25000.211174.2019-45).

RESUMO

A Leucemia Linfoide Aguda (LLA) é o câncer mais comum na criança. Enquanto atualmente a taxa de cura dessa doença aproxima-se a 90%, as crianças que apresentam recaída ainda possuem um prognóstico ruim. A LLA é caracterizada pela proliferação clonal de precursores linfoides de linhagem B ou T (85% e 15% dos casos, respectivamente), células do sistema imunológico adaptativo responsáveis pela produção de moléculas receptoras de antígenos, como as imunoglobulinas (IG). A produção das cadeias que compõem essas moléculas depende de um processo de recombinação somática denominado rearranjo V(D)J, que gera sequências gênicas virtualmente únicas para cada nova célula linfoide formada. Assim, o sequenciamento V(D)J de uma determinada amostra celular permite discriminar os clones de células linfoides, como, por exemplo, no monitoramento da Doença Residual Mínima (DRM) na LLA. Exames de DRM baseados no sequenciamento V(D)J geram informação acerca do conjunto de clonótipos presentes na medula óssea do paciente, tanto os correspondentes às células leucêmicas quanto os correspondentes às células saudáveis da linhagem B e T. Nesse contexto, este trabalho teve como objetivo explorar o repertório de células linfoides de linhagem B em pacientes de LLA-B pediátrica e buscar associações dessas características com o desenvolvimento clínico dos pacientes. Para tal, sequenciou-se a região VDJ do gene *IGH* de amostras de medula óssea colhidas ao diagnóstico e após o término da terapia de indução de 204 pacientes consecutivos com LLA-B pediátrica. Como controle, foram utilizadas 8 amostras de medula óssea de doadores saudáveis da mesma faixa etária. O processamento e análise dos dados de rearranjos VDJ de *IGH* foi realizado por meio de uma *pipeline* desenvolvida nesta tese. Clonótipos associados a subtipos genéticos específicos de LLA-B apresentaram diversas diferenças na estrutura da junção das cadeias V-D-J em relação aos clonótipos de doadores, como, por exemplo, enriquecimento de certos segmentos VH; baixa utilização de segmentos localizados entre os segmentos *IGHV7-34-1* e *IGHV1-58* e maior frequência de nucleotídeos GC no subtipo *ETV6::RUNX1*; além de encurtamento da região CDR3 e das inserções de nucleotídeos N realizadas durante o rearranjo. Também comparou-se os clonótipos de células leucêmicas que desapareceram após terapia de indução (sensíveis) em relação àqueles das células leucêmicas resistentes. Aqui, observou-se que células leucêmicas resistentes ao tratamento tendem a apresentar rearranjos *IGH* não produtivos. Por fim, pacientes com menor diversidade de repertório *IGH* ao término da terapia de indução tendem a ser pacientes de alto risco pelo sistema NCI e apresentam uma

pior sobrevida global e livre de eventos. Essa associação foi encontrada também quando analisando apenas os pacientes com DRM indetectável nesse ponto do tratamento, gerando informação clinicamente relevante para esse grupo de pacientes. De modo geral, esse conjunto de resultados contribuem para o avanço do entendimento do processo de rearranjo V(D)J em células de LLA-B, bem como da importância da diversidade da medula óssea no prognóstico dos pacientes.

ABSTRACT

Acute Lymphoblastic Leukemia (ALL) is the most common cancer in children. While current cure rates for this disease approach 90%, children who experience relapse still present a poor prognosis. ALL is characterized by the clonal proliferation of lymphoid progenitors of B or T lineage (85% and 15% of cases, respectively), cells of the adaptive immune system that are responsible for producing antigen receptor molecules, such as immunoglobulins (IG). The production of the chains that make up these molecules depends on a somatic rearrangement process known as V(D)J recombination, which generates virtually unique genetic sequences for each newly formed lymphoid cell. As such, the V(D)J sequencing of a cell sample allows the discrimination of lymphoid cell clones, as exemplified by Minimal Residual Disease (MRD) monitoring in ALL. MRD assays based on V(D)J sequencing also provide information about the whole clonotype set present in the patient's bone marrow, including both those corresponding to leukemia cells and those corresponding to healthy B and T lineage cells. In this context, the aim of this study was to explore the B-lineage lymphoid cell repertoire in pediatric B-Cell Precursor ALL (BCP-ALL) patients and to seek associations between these characteristics and the clinical development of the patients. To this end, the VDJ region of the *IGH* gene was sequenced in bone marrow samples from the diagnosis and end of induction therapy time points, collected from 204 consecutive pediatric patients with BCP-ALL. Eight bone marrow samples from healthy donors of the same age range were used as control samples. The processing and analysis of the VDJ rearrangement data of *IGH* were performed using a pipeline developed in this thesis. Clonotypes associated with specific BCP-ALL genetic subtypes showed various differences in the structure of their V-D-J chain junction in comparison to donor clonotypes, such as enrichment in some VH segments; low usage of segments located between segments *IGHV7-34-1* and *IGHV1-58* and higher frequencies of GC nucleotides in the *ETV6::RUNX1* subtype; in addition to shortening of the CDR3 region and of the N nucleotides inserted during rearrangement. Additionally, patients with BCP-ALL from the *ETV6::RUNX1* subtype presented low usage of segments located between segments *IGHV7-34-1* and *IGHV1-58* and higher frequencies of GC nucleotides. Clonotypes associated with leukemia cells that disappeared after induction therapy (sensitive to treatment) were also compared to those from resistant leukemia cells. Here, it was observed that treatment-resistant leukemia cells tend to have non-productive *IGH* rearrangements. Finally, patients with a lower *IGH* repertoire

diversity at the end of induction therapy tend to be classified as high risk in the NCI system and present worse overall and event-free survival. This association was also found when analyzing only patients with undetectable MRD at this time point, providing clinically relevant information for this group of patients. Overall, these results contribute to advancing the understanding of the V(D)J rearrangement process in BCP-ALL cells and the importance of bone marrow diversity in the patient's prognosis.

LISTA DE FIGURAS E TABELAS

INTRODUÇÃO

METODOLOGIA

CAPÍTULO I: MANUSCRITO

SUMÁRIO

1 INTRODUÇÃO

1.1 Leucemia Linfoide Aguda

A Leucemia Linfoide Aguda (LLA) é uma neoplasia maligna cuja principal característica é a proliferação clonal na medula óssea, de células precursoras de linfócitos, denominadas linfoblastos (MALARD e MOTHY, 2020). A LLA é o câncer mais comum na infância, correspondendo a quase 30% dos diagnósticos de neoplasias em pessoas de até 14 anos (SIEGEL *et al*., 2023). Em função dos diversos esforços da comunidade científica internacional, a taxa de pacientes pediátricos que sobrevivem à doença vem sendo elevada constantemente. Conforme demonstrado pela **Figura 1**, a Sobrevida Livre de Eventos (SLE) dos pacientes tratados via protocolos do Grupo Brasileiro de Tratamento das Leucemias na Infância (GBTLI) elevou-se de menos de 30% para cerca de 75% desde a década de 80. Apesar desse sólido avanço, ainda há um longo caminho a ser percorrido, uma vez que apesar das altas taxas de cura já atingidas, aproximadamente 20% dos pacientes apresentam recaída da doença e nesses casos a probabilidade de cura cai para apenas 50% em média (NGUYEN *et al*., 2008; TALLEN *et al*., 2010; SUN *et al*., 2018).

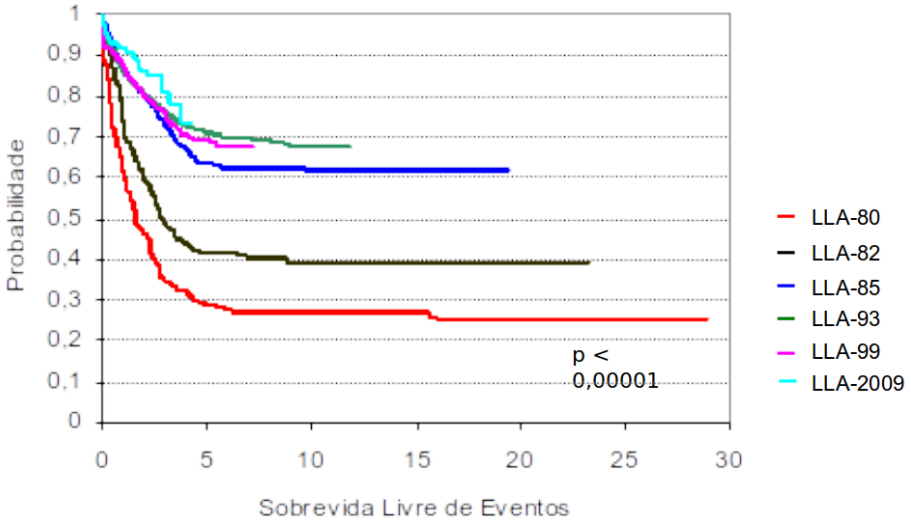

Figura 1. SLE em 29 anos de pacientes pediátricos com LLA tratados via protocolos GBTLI ao longo das décadas. Abaixo das curvas de Kaplan-Meier, encontra-se o número total de pacientes em cada protocolo (N) e o número desses que atingiram remissão clínica completa (RCC). Fonte: Provido ao autor pelo GBTLI.

O processo de surgimento da LLA em um indivíduo, denominado leucemogênese, é caracterizado pelo acúmulo de mutações oncogênicas em precursores de células linfoides, uma vez que usualmente mutações individuais não são capazes de gerar um quadro leucêmico por si só (IACOBUCCI e MULLIGHAN, 2017). Essas mutações podem ser consideradas primárias, quando contribuem diretamente para o surgimento de um clone de linfoblasto pré-leucêmico, ou secundárias, que contribuem para o estabelecimento da leucemia em si, ou seja, quando surgem em células pré-leucêmicas que já possuem uma mutação primária (MULLIGHAN, 2012).

A LLA pode acometer tanto células precursoras de linfócitos de linhagem B (LLA-B), responsáveis pela imunidade adaptativa humoral, quanto de linhagem T (LLA-T), ligadas à imunidade adaptativa celular (MURPHY e WEAVER, 2017). A LLA-B, objeto de estudo desta tese, apresenta maior prevalência em relação a LLA-T, correspondendo a 85% das LLA pediátricas (RAETZ e TEACHEY, 2016). Adicionalmente, ambos os tipos de LLA podem ser mais uma vez subdivididos em função do seu perfil de alterações genéticas (**Figura 2**). Muitas vezes, a presença de alguma dessas alterações informa não apenas o prognóstico da doença, como também a possibilidade de tratamento alvo específico (INABA e GREAVES e MULLIGHAN, 2013). A **Tabela 1**, contém um resumo dos subtipos moleculares de LLA-B pediátrica analisados neste trabalho, bem como suas frequências (ROBERTS, 2018).

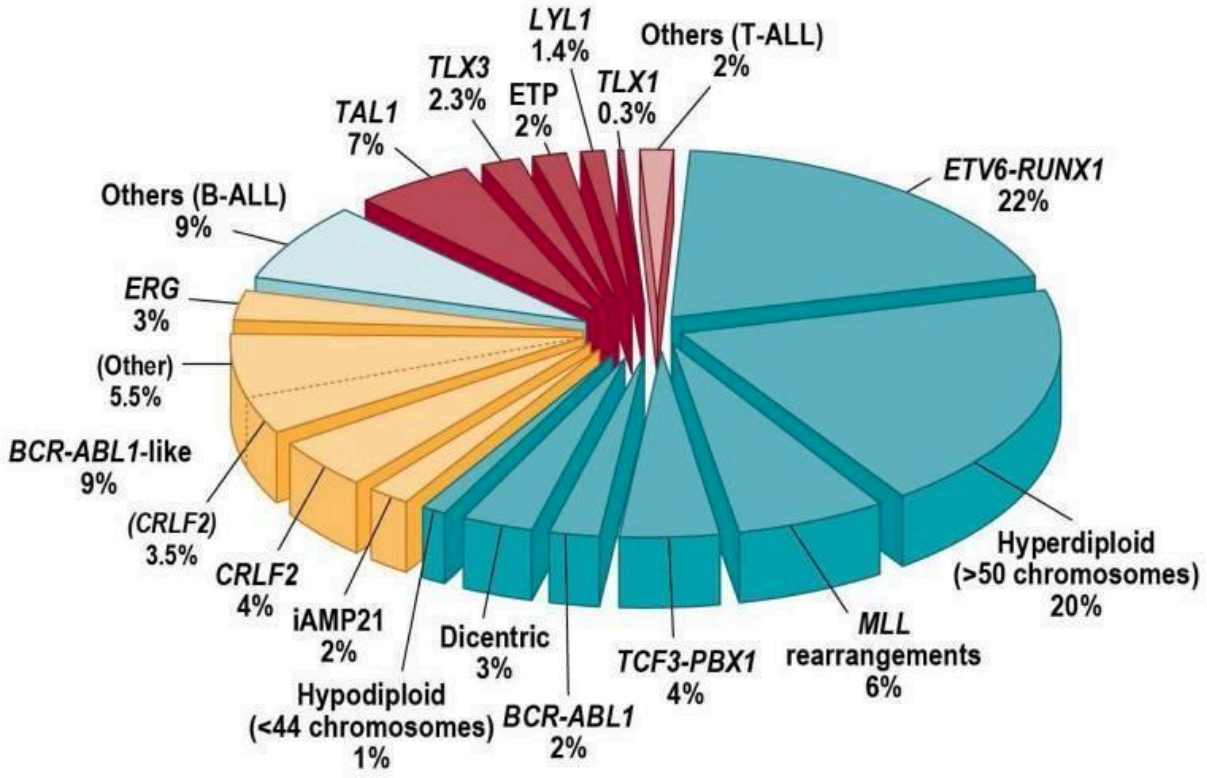

Figura 2. Alterações genéticas na LLA pediátrica. As cores azul e amarela representam alterações clássicas e novas associadas à LLA-B respectivamente, enquanto a cor vermelha representa a LLA-T. As alterações genéticas podem afetar a expressão de um único gene (*CRLF2*, *ERG*, *LYL1*, *TLX1, TLX3, TAL1*), levar a formação de proteínas quiméricas (*ETV6-RUNX1*, *BCR-ABL1*, *TCF3-PBX1*, rearranjos *MLL*) ou causar aneuploidias (hiperdiploidia >50 cromossomos, hipodiploidia <44 cromossomos, iAMP21). Atualmente as alterações subjacentes às LLA-B antigamente classificadas como BCR-ABL1-like já são em sua maior parte conhecidas e incluem fusões com participação de genes como *ABL1*, *ABL2*, *CSFR1*, *PDGFRB, EPOR, JAK2*. Nessa figura, translocações entre dois genes são representadas por "-", enquanto no restante desta tese é utilizado "::", conforme as recomendações mais recentes da *International System for Human Chromosome Nomenclature* (ISCN). Fonte: Inaba et al. (2013).

Tabela 1. Subtipos moleculares de LLA-B pediátrica. Tabela adaptada de Roberts (2018), mantendo apenas os subtipos moleculares de interesse dessa tese. *Nos últimos anos, alguns casos anteriormente classificados como *B-other* têm sido reclassificados em novos subtipos moleculares, como *PAX5alt*, *CRLF2*, *DUX4*, entre outros. Esses novos subtipos moleculares não serão abordados nesta tese.

Subtipo Molecular	Tipo da Alteração	Prognóstico	Observação
Hiperdiploidia (>50 cromossomos)	Aneuploidia	Excelente	
Hipodiploidia (<46 cromossomos)	Aneuploidia	Ruim	
ETV6*::*RUNX1	Rearranjo cromossômico	Excelente	
TCF3*::*PBX1	Rearranjo cromossômico	Bom	
BCR*::*ABL1	Rearranjo cromossômico	Ruim	
***KMT2A* rearranjado**	Rearranjo cromossômico	Ruim	Também denominado *MLL* rearranjado
B-other		Intermediário	Casos sem subtipo molecular definido*

De modo geral, o tratamento da LLA é realizado por meio de uma terapia de quatro etapas (indução, consolidação, intensificação e manutenção) e costuma durar até 3 anos (INABA e MULLIGHAN, 2020). A etapa de indução dura cerca de um mês e tem como o seu principal objetivo eliminar a carga de células leucêmicas na medula óssea do paciente, de modo a restabelecer a sua função hematopoiética normal. É importante ressaltar, no entanto, que a eliminação completa dessas células nessa etapa nem sempre é atingida. Na terapia de consolidação, que costuma durar de 3 a 4 meses, há um aumento das doses de quimioterápicos administradas, buscando reforçar a eliminação de células leucêmicas previamente atingida. A etapa de intensificação é uma espécie de repetição dos 3 primeiros meses do tratamento, porém com substituição de algumas drogas por outras da mesma classe. Por fim, a etapa de manutenção, que pode durar cerca de 2 anos, objetiva manter o paciente em estado de remissão clínica, e faz uso somente de metotrexato e de um análogo de purina (CANCER RESEARCH UK, 2018; MALARD e MOHTY, 2020; TEACHEY e HUNGER e LOH. 2021).

No cenário brasileiro, para cada ano do triênio de 2023 a 2025, o Instituto Nacional de Câncer (INCA) estima a ocorrência de aproximadamente 704.000 novos casos de câncer no país, dos quais 7.930 correspondem a casos em crianças e adolescentes de até 19 anos (INSTITUTO NACIONAL DE CÂNCER, 2023). Desses, cerca de 2.380 devem

corresponder a casos de LLA, segundo a conjectura percentual dessa doença. De acordo com as estimativas previamente apresentadas, pode-se esperar que cerca de 475 desses casos apresentarão um cenário de recaída da LLA. Estes números demonstram o relevante impacto social ainda gerado pela LLA, evidenciando que, apesar de melhoras nas taxas de cura, ainda há necessidade de estudos que busquem melhor compreender e combater essa doença.

1.2 Recombinação V(D)J

Os linfócitos B e T, como as células efetoras do sistema imune adaptativo, devem ser capazes de reconhecer com alto grau de especificidade uma vasta gama de antígenos, demandando uma grande diversidade de moléculas receptoras de antígenos, que são as imunoglobulinas (IG) e os receptores de células T (TR), respectivamente. Essa vasta diversidade de receptores é formada pelo conjunto de linfócitos presentes em um organismo, uma vez que, via de regra, cada linfócito individual é capaz de expressar apenas um tipo de molécula IG/TR (ABBAS e LICHTMAN, 2005).

Estruturalmente, receptores IG/TR são compostos por cadeias proteicas que contém regiões conservadas e regiões hipervariáveis. É a combinação das regiões hipervariáveis que determina a sua afinidade antigênica. As moléculas de IG, conforme a **Figura 3A**, são compostas por 4 cadeias: um par de cadeias leves e um par de cadeias pesadas (*immunoglobulin heavy locus* ou *IGH*; MURPHY e WEAVER, 2017). As cadeias *IGH* (localização cromossômica 14q32.33), em específico, são o objeto de estudo deste trabalho.

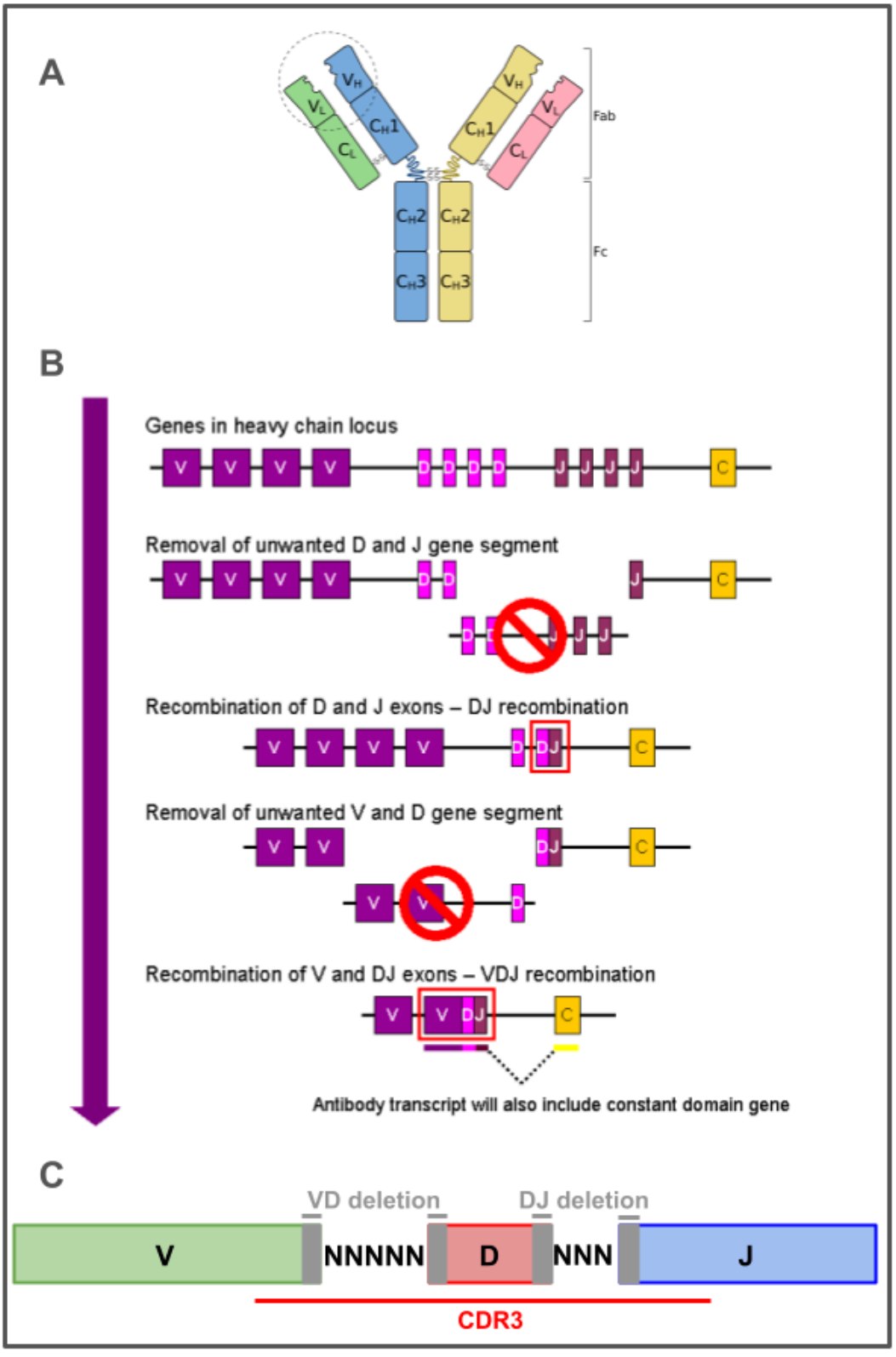

Figura 3. Recombinação V(D)J. (**A**) Representação estrutural de uma molécula de IG. Cadeias pesadas representadas em azul e amarelo, cadeias leves em verde e vermelho. Sítio de ligação de antígeno destacado por um círculo pontilhado. V_H/V_L: regiões variáveis; C_H/C_L: regiões constantes; Fab: fragmento de ligação do antígeno; Fc: fragmento cristalizável. Fonte: Wrochna (2020). (**B**) Diagrama esquemático da recombinação V(D)J em *IGH*. Fonte: Carra (2008). (**C**) Representação estrutural da região variável do *IGH* após recombinação V(D)J em detalhe. Nucleotídeos N em preto. Fonte: elaborado pelo autor.

A grande variabilidade das IG/TR se dá por um mecanismo de rearranjo somático nos *loci* gênicos dessas moléculas, durante o processo de maturação de células linfoides, denominado recombinação V(D)J ou somática. Há diversos segmentos de genes ditos Variáveis (V), Diversidade (D) e Junção (J), que são escolhidos e combinados de maneira única em cada novo linfócito durante o rearranjo. Em seguida, esse rearranjo V(D)J é unido à região Constante (C), responsável pela função efetora da cadeia formada (TONEGAWA, 1983).

A seleção e combinação dos segmentos V, D e J é realizada por um complexo proteico, do qual participam as proteínas RAG1 e RAG2 (*recombination-activating gene*; OETTINGER *et al*., 1990). Essas moléculas são capazes de reconhecer e inserir quebras em regiões do DNA chamadas de sequências de sinal de recombinação (*recombination signal sequences*, RSS), que flanqueiam os segmentos previamente mencionados, iniciando o processo de recombinação (AKIRA e OKAZAKI e SAKANO, 1987; RAMSDEN e BAETZ e WU, 1994). As regiões RSS são compostas por um heptamêro e um nanômero de nucleotídeos conservados, espaçados por 12 ou 23 nucleotídeos degenerados. O complexo RAG1/2 é capaz de rearranjar e unir apenas regiões RSS com espaçamentos distintos, fenômeno denominado regra 12/23 (SCHATZ e SWANSON, 2011).

Conforme a **Figura 3B**, a recombinação em si inicia-se através da junção de um segmento J a um segmento D (quando aplicável, uma vez que nem todos os *loci* que passam por recombinação V(D)J possuem segmentos D), seguida da ligação de um segmento V ao segmento DJ formado. O grau de diversidade das cadeias de IG/TR gerados por esse processo é elevado ainda mais pela deleção e inserção de nucleotídeos N nas interfaces VD e DJ do rearranjo formado. A porção mais variável desses rearranjos, composta pela interface V(D)J em si, é denominada região determinante de complementaridade 3 (*complementarity-determining region* 3, CDR3) e é a principal responsável pela especificidade antigênica de uma cadeia de IG/TR (ABBAS e LICHTMAN, 2005). A **Figura 3C** ilustra a porção variável de uma cadeia *IGH* após a recombinação V(D)J, destacando regiões de interesse analisadas nesta tese.

O processo de inserção de N previamente mencionado é realizado pela enzima desoxinucleotidil transferase terminal (*terminal deoxynucleotidyl transferase*, TdT), ocorrendo sem a utilização de uma fita molde de DNA, o que faz com que quaisquer uma das 4 bases nitrogenadas possíveis (adenina, timina, guanina ou citosina) possam ser utilizadas a cada nucleotídeo incorporado à sequência (DESIDERIO *et al*., 1984). Isso não quer dizer, no entanto, que todos os tipos de nucleotídeos apresentam a mesma chance de serem utilizados.

A TdT apresenta um viés por guaninas/citosinas (GC), fazendo com que a região N de rearranjos V(D)J seja enriquecida para essas bases nitrogenadas (MOTEA e BERDIS, 2010).

De modo semelhante, apesar da escolha dos segmentos V, D e J a serem rearranjados em cada processo de recombinação V(D)J ser aleatória, também existem fatores que enviesam as taxas de utilização de cada um desses segmentos. Por exemplo, a sequência específica de cada região RSS, que usualmente não são perfeitamente conservadas, alteram a sua afinidade pelo complexo RAG e consequentemente sua taxa de recombinação (FEENEY *et al*., 2004). Outro fator na escolha dos segmentos a serem recombinados é o grau de acessibilidade da cromatina para cada uma dessas sequências (JI *et al*., 2010; CHRISTIE e FIJEN e ROTHENBERG, 2022).

1.3 A Recombinação V(D)J na LLA

1.3.1 A Recombinação V(D)J em Células Leucêmicas

Apesar da transformação maligna de um linfoblasto impedir o prosseguimento do seu processo de maturação, a maioria das células de LLA encontra-se em um estado de desenvolvimento no qual pelo menos uma de suas cadeias IG/TR já passou pelo processo da recombinação V(D)J. Como a LLA é uma doença clonal, onde o conjunto total de células leucêmicas presentes no organismo de um paciente se origina da reprodução do linfoblasto neoplásico originário, essa população de células em geral compartilha a(s) mesma(s) sequencia(s) V(D)J. Assim sendo, rearranjos V(D)J podem ser utilizados como uma assinatura molecular para identificar e quantificar células de LLA em meio ao repertório linfocítico do paciente (VAN DONGEN *et al*., 2003).

Nesse contexto, sequências V(D)J vêm sendo utilizadas para acompanhar a carga leucêmica em pacientes com LLA desde pelo menos 1989 (HANSEN-HAGGE e YOKOTA e BARTRAM, 1989). A métrica utilizada para aferir essa carga é denominada Doença Residual Mínima (DRM), sendo essa representada como a frequência de linfoblastos leucêmicos no total de células mononucleares (linfócitos e monócitos) da medula óssea ou sangue periférico do paciente após determinado tempo de tratamento (POTTER, 1992). Desde então, a DRM foi estabelecida como importante fator prognóstico na LLA pediátrica, conforme exemplificado na **Figura 4**, sendo informação essencial para a alocação dos pacientes nos diferentes braços de tratamento (BOROWITZ *et al*., 2015; SCHRAPPE *et al*., 2023).

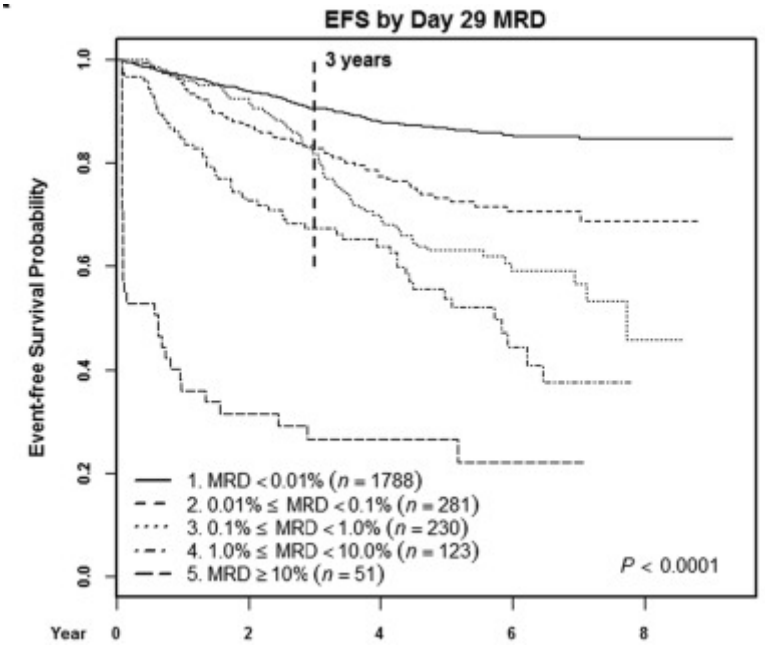

Figura 4. Sobrevida livre de eventos de pacientes pediátricos com LLA agrupados por diferentes níveis de DRM ao vigésimo nono dia de tratamento. MRD, ou *minimal residual disease*, é o termo para DRM em inglês. Adaptado de: Borowitz et al. (2015).

Atualmente, o padrão-ouro para a quantificação de DRM consiste em amplificar via reação de polimerização em cadeia em tempo real (qPCR) os rearranjos de IG/TR contidos em uma amostra de DNA da medula óssea do paciente (HEIKAMP e PUI, 2018). Com a popularização das técnicas de Sequenciamento de Nova Geração (NGS) durante a última década, diversos grupos desenvolveram métodos para aferir a DRM utilizando essa tecnologia, incluindo o grupo do qual faz parte o autor desta tese (FAHAM *et al*., 2012; LADETTO *et al*., 2013; KOTROVA *et al*., 2015; SEKIYA *et al*., 2017; SHIN *et al*., 2017; CHENG *et al*., 2018; KNECHT *et al*., 2019; THEUNISSEN *et al*., 2019; GIUSTI *et al*., 2020).

O estabelecimento desses métodos de DRM por NGS possibilitaram superar algumas das principais limitações enfrentadas pelas técnicas baseadas em qPCR (VAN DONGEN *et al*., 2015). Por exemplo, apesar da natureza clonal da LLA, pacientes frequentemente apresentam mais de um rearranjo V(D)J associado às células leucêmicas, geralmente por conta de um processo de evolução clonal. Cada rearranjo V(D)J, que pode estar presente em uma ou mais de uma célula, é denominado um clonótipo (SOFOU *et al*., 2023). Uma vez que as células de LLA responsáveis pela recaída de um paciente podem apresentar rearranjos distintos em relação ao diagnóstico, é importante monitorar o comportamento da maioria dos clones V(D)J ao longo do tratamento (MULLIGHAN *et al*., 2008). O NGS permite a análise das sequências de DNA com rearranjos V(D)J de

virtualmente todas as células linfoides de uma amostra, ao contrário das técnicas de qPCR, onde há a necessidade de desenhar *primers* específicos para a sequência do linfoblasto leucêmico de cada paciente (VAN DONGEN *et al*., 2015).

Estudos anteriores identificaram que o padrão de rearranjos V(D)J da LLA é afetado por vieses ligados ao seu genótipo e histórico clínico (VAN DER VELDEN *et al*., 2003; GENG *et al*., 2015). O subtipo molecular da LLA, por exemplo, têm influência na probabilidade desses rearranjos serem produtivos ou não (GENG *et al*., 2015). Nesse contexto, a capacidade dos ensaios baseados em NGS de identificar mais clonótipos de LLA permite uma caracterização biológica mais profunda dos seus rearranjos. Mesmo com a utilização desse método, no entanto, a identificação de clonótipos leucêmicos raros presentes em um paciente ao diagnóstico ainda é problemática. Não é possível dizer se um determinado clonótipo é leucêmico ou não, a não ser pela sua abundância acima de um *cut-off* arbitrário de 5% na amostra de medula óssea do diagnóstico (DARZENTAS *et al*., 2023). Assim, um maior entendimento dos possíveis vieses de recombinação V(D)J de células leucêmicas poderia auxiliar nesse processo de identificação dos clonótipos leucêmicos. Outra possibilidade interessante para a detecção desses clonótipos de baixa frequência está ligado ao advento do sequenciamento de células únicas. Nesse ensaio, possibilita-se a detecção de determinado rearranjo V(D)J juntamente com as mutações *driver* da leucemia, permitindo assim a identificação de clonótipos com frequências abaixo de 5%. Detecção de clonótipos leucêmicos através de metodologias semelhantes já foi realizada para Leucemia Mieloide Aguda (LMA; VELTEN *et al*., 2021).

1.3.2 O Repertório V(D)J na LLA

Evidências epidemiológicas têm apontado para possíveis associações entre a leucemogênese e repertórios V(D)J desregulados. Padrões de infecção tardia, na qual indivíduos são expostos apenas tardiamente a infecções comuns nos primeiros anos da infância, parecem estar associados a maiores taxas de desenvolvimento de LLA-B (GREAVES, 2018). Crianças expostas precocemente a ambientes com maior disseminação de infecções, como creches e domicílios que já contém outras crianças, apresentam menor risco de desenvolvimento de LLA (MA *et al*., 2002; GILHAM *et al*., 2005; KAMPER-JØRGENSEN *et al*., 2007; AJROUCHE *et al*., 2015; RUDANT *et al*., 2015). Em concordância com essa hipótese, camundongos com mutações em heterozigose no gene *Paired Box 5* (*PAX5*+/-), uma mutação frequentemente presente na LLA-B, criados em ambientes livres de patógenos e transferidos tardiamente para ambientes de maior exposição

demonstraram maiores níveis de desenvolvimento de LLA (MARTÍN-LORENZO *et al*., 2015). A **Figura 5** apresenta uma representação esquemática desse modelo de leucemogênese, denominado hipótese da infecção tardia.

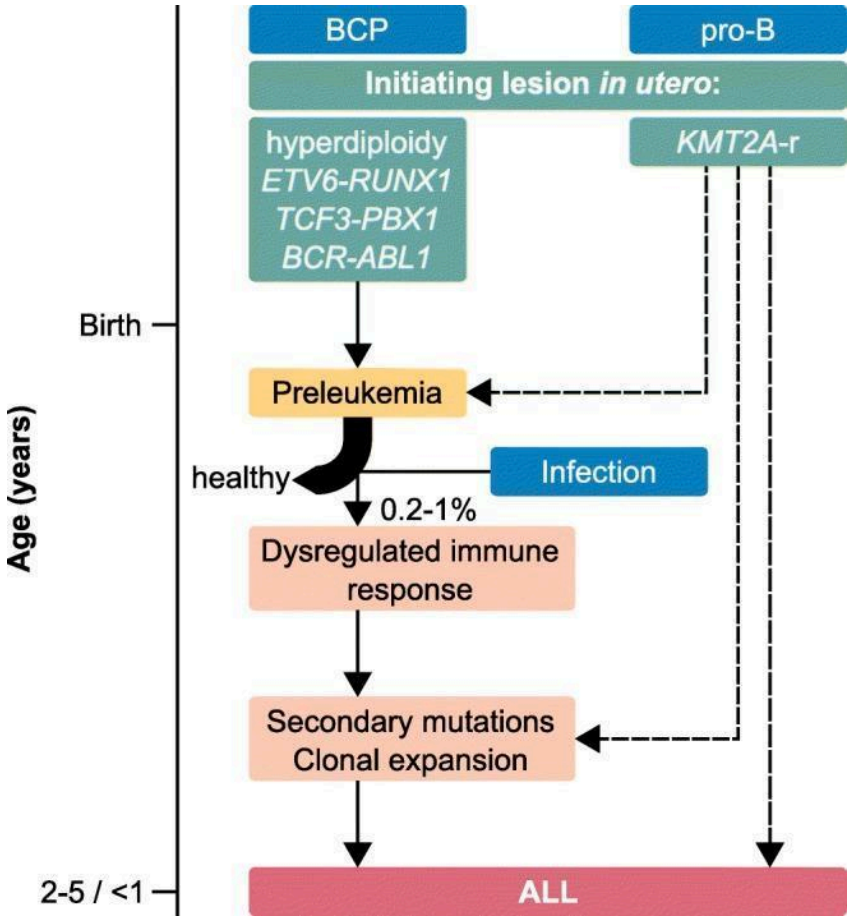

Figura 5. Representação esquemática da hipótese da infecção tardia como explicação da leucemogênese. Aqui, o subtipo de LLA *KMT2A* rearranjado (*KMT2A*-r) é apresentado como uma exceção à hipótese, podendo possivelmente desencadear diretamente lesões secundárias ou até mesmo um fenótipo de LLA. ALL (*acute lymphoblastic leukemia*) é a sigla para LLA em inglês; BCP e pro-B significam precursores de células B e células pró-B, respectivamente. No eixo das ordenadas, apresenta-se a mediana das idades de surgimento da LLA. Fonte: HEIN.e BORKHARDT e FISCHER, 2020.

Outro estudo verificou que a linfopenia de células T pode facilitar a progressão de clones pré-leucêmicos e o surgimento do fenótipo leucêmico, provendo evidência adicional da influência do estado do repertório imunológico na leucemogênese. Nesse, evidenciou-se que a falta de competição durante a maturação de timócitos no timo está relacionada ao aumento do desenvolvimento de LLA-T (MARTINS *et al*., 2014).

Por fim, o estado geral do repertório V(D)J em pacientes com LLA após a terapia de indução, que reflete o estado de regeneração da medula óssea após esse tratamento, parece ter influência em sua evolução clínica. Como exemplo, pacientes com baixos níveis de DRM,

mas com baixa diversidade de clonótipos ao fim da indução, possuem menor SLE do que aqueles com uma maior diversidade, indicativo de uma melhor regeneração da linfopoiese (KOTROVA *et al*., 2015). Em outro estudo, observou-se que pacientes com clonótipos de leucemia não produtivos (com códon de parada ou *out-of-frame*) apresentam melhores prognósticos (KATSIBARDI *et al*., 2011). No entanto, o efeito da produtividade do repertório como um todo (com a produtividade do repertório sendo definida para este trabalho como a frequência de *reads* de rearranjo V(D)J produtivos em uma amostra) ainda não é conhecido. Nesse contexto, a capacidade dos ensaios de NGS de produzir dados sobre o estado do repertório V(D)J do paciente como um todo surge como uma excelente técnica na busca por elucidar associações entre padrões desse repertório e a LLA.

2. JUSTIFICATIVA

O desenvolvimento da tecnologia de NGS possibilitou grandes avanços na caracterização da clonalidade da LLA, com uso imediato na análise da DRM. Permitiu também a caracterização da diversidade e da dinâmica ao longo do tratamento de clones leucêmicos e normais em resoluções muito maiores do que as anteriormente disponíveis. Na última década, o maior foco dos trabalhos nessa área esteve relacionado ao estabelecimento, padronização e análise de ensaios de DRM por NGS, com um menor enfoque em estudos que buscam explorar as características dos clonótipos de LLA e do repertório imunológico.

Há alguns anos, o autor dessa tese e o grupo de pesquisa que ele integra desenvolveram um ensaio para quantificação da DRM baseado na análise por NGS dos rearranjos de *IGH*. Desde então, essa metodologia tem sido utilizada com sucesso, concomitantemente às técnicas padrão-ouro de qPCR, no monitoramento da progressão clínica de pacientes com LLA do Centro Infantil Boldrini. Conforme descrito anteriormente, essa técnica também gera dados sobre todo o repertório imunológico do paciente, isto é, dos precursores de células linfoides normais. Esta tese buscou realizar uma análise desses dados.

Os resultados obtidos são apresentados em dois capítulos. O primeiro capítulo apresenta um manuscrito que buscou caracterizar o perfil de rearranjos VDJ de *IGH* na LLA-B. Essas características foram analisadas de maneira comparativa a clonótipos de crianças saudáveis, doadores de medula óssea. O segundo capítulo apresenta resultados relativos à análise do repertório *IGH* desses pacientes, incluindo tanto os clonótipos de LLA como aqueles referentes a linfócitos sadios, buscando associações com a progressão clínica da doença.

3. OBJETIVOS

3.1 Objetivo Geral

Analisar o repertório VDJ do gene *IGH* em pacientes pediátricos portadores de LLA-B, buscando correlacionar padrões desse repertório às características clínicas desses pacientes.

3.2 Objetivos Específicos

1. Sequenciar o rearranjo VDJ do gene *IGH* de amostras de medula óssea do diagnóstico e do fim da terapia de indução em casos consecutivos de LLA-B pediátrica;
2. Sequenciar o rearranjo VDJ do gene *IGH* de amostras de medula óssea de doadores saudáveis de medula óssea;
3. Desenvolver uma *pipeline* para processamento dos dados de sequenciamento, a fim de: calcular o valor de DRM por clonótipo, remover clonótipos correspondentes aos controles externos *spike-in*, e agrupar automaticamente clonótipos altamente semelhantes;
4. Caracterizar os rearranjos VDJ associados a células leucêmicas em relação ao comprimento das regiões CDR3 e N, conteúdo de GC, produtividade e segmentos gênicos utilizados;
5. Analisar a diversidade de clonótipos *IGH* nas amostras do diagnóstico e do final da indução;
6. Buscar associações entre as características de repertórios VDJ citadas acima e dados clínicos dos pacientes.

4. METODOLOGIA

4.1 Seleção das Amostras

O trabalho para esta tese envolveu amostras de medula óssea de 204 pacientes pediátricos portadores de LLA-B com idades entre 0 e 19 anos, atendidos no Centro Infantil Boldrini (n = 183), Hospital São José (n = 2), HEMOAM (n = 10), CETOHI (n = 7) e GRENDACC (n = 2). Cada participante contribuiu com 2 amostras: uma colhida no diagnóstico (D0) e outra colhida após o término da terapia de indução (33 dias após o diagnóstico ou posterior em caso de medulas aplásicas, denominada *follow-up* ou Fup). Os pacientes foram tratados entre 2012 e 2021, de acordo com protocolos para LLA-B pediátrica do Grupo Brasileiro para Tratamento da Leucemia Linfoide Aguda de 2009 (GBTLI LLA-2009, 97 pacientes) ou da *Associazione Italiana di Ematologia e Oncologia Pediatrica* e do grupo Berlim-Frankfurt-Münster, também de 2009 (AIEOP BFM ALL-2009, 107 pacientes). As amostras foram analisadas de forma retrospectiva, com os participantes sendo selecionados de maneira sequencial. Pacientes para os quais não havia disponibilidade de pelo menos uma amostra do término da terapia de indução, seja por óbito ou abandono do tratamento, foram excluídos deste trabalho. Amostras retrospectivas de medula óssea de 8 doadores de medula óssea saudáveis, com idades entre 6 e 18 anos, foram utilizadas como controle.

Este trabalho foi submetido ao Comitê de Ética em Pesquisa do Centro Infantil Boldrini e aprovado sob o identificador CAAE: 57280616.4.0000.5376.

4.2 Isolamento das Células Mononucleares

As amostras de medula óssea dos pacientes foram colhidas em EDTA e então diluídas em solução salina na proporção de 1:1. Em seguida, elas foram centrifugadas em um gradiente de Ficoll Hypaque Plus (GE Healthcare), com as células mononucleares obtidas através desse processo sendo então aliquotadas em solução de isotiocianato de guanidina 4M.

4.3 Preparo da Biblioteca de NGS

Para realizar a extração do DNA utilizou-se o Blood genomicPrep Mini Spin Kit (Cytiva). As amostras foram quantificadas utilizando o ensaio HS dsDNA no fluorímetro Qubit (Life Technologies). A amplificação da região de rearranjo VDJ de *IGH* e preparo da biblioteca para NGS foi realizada conforme descrito por Giusti *et al*. (2020). Enquanto esse

protocolo objetiva aferir a DRM de um paciente através dessas amostras, ele também fornece os dados de todas as sequências VDJ *IGH* da amostra, o que permite aferir o grau de diversidade da linfopoiese de células B na medula.

Resumidamente, a amplificação e o sequenciamento dos rearranjos VDJ foram realizados via um processo de *nested* PCR utilizando cerca de 50 ng (D0 e doadores) ou 600 ng (Fup) de DNA. Desse modo, levando em conta que uma célula diploide humana contém cerca de 6 pg de DNA (GILLOOLY e HEIN e DAMIANI, 2015), analisou-se o material genético de cerca de 8.333 ou 100.000 células mononucleares, respectivamente. A maior quantidade de DNA utilizada para as amostras Fup visa tornar teoricamente possível a detecção de 1 célula leucêmica a cada 100,000 células mononucleadas saudáveis sequenciadas (correspondendo, portanto, a uma DRM com sensibilidade 10^{-5})

Na primeira etapa de PCR, foram utilizados um *multiplex* de *primers* do grupo EuroClonality/Biomed2 para rearranjos completos de *IGH*. Esse sistema conta com um *multiplex* de *primers* que têm como alvo a região *framework* 2 (FR2) da porção V do rearranjo *IGH* (VH) e com um *primer* consenso para a região J desse mesmo rearranjo (JH; VAN DONGEN *et al*., 2003). Essa primeira reação foi catalisada pela enzima GoTaq G2 *Hot Start* DNA *polymerase* (Promega) por 25 ciclos, incluindo 5 de *touchdown* no início da reação. A segunda reação decorreu por 15 ciclos, utilizando *primers* Nextera XT (Illumina) que tem como alvo regiões de *overhangs* presentes nos *primers* da etapa anterior. Os *primers* Nextera XT contém não apenas os adaptadores para o sequenciamento (P5 e P7), mas também indexadores únicos para cada amostra, de modo a possibilitar a análise de amostras de diversos pacientes em apenas uma única corrida do sequenciador (**Figura 6**). O material genético obtido, composto de *amplicons* de cerca de 350 pares de base, foi então purificado utilizando uma razão de 0.8 Agencourt Ampure XP Beads (Beckman Coulter) e em sequência quantificado utilizando o fluorímetro Qubit (Life Technologies).

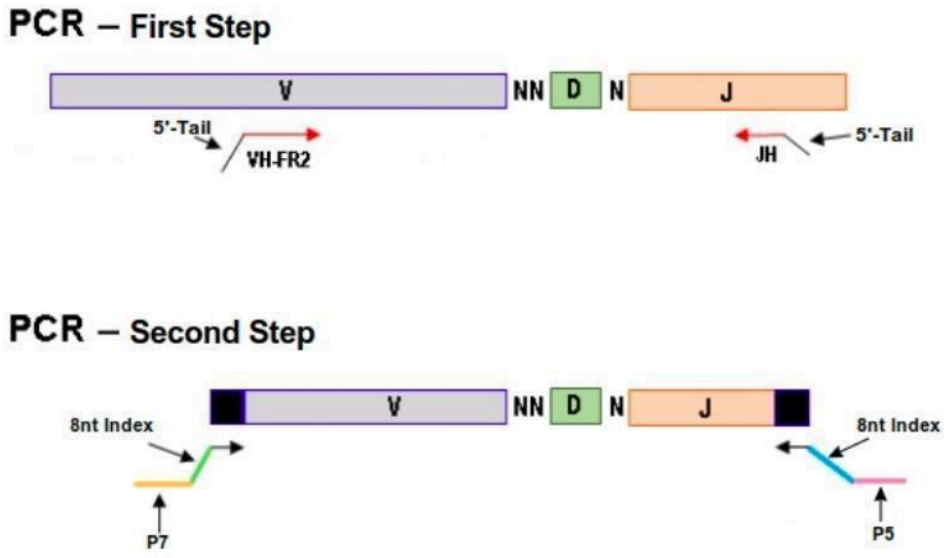

Figura 6. Resumo esquemático do processo de *nested* PCR dos rearranjos de IGH. A primeira reação utiliza um multiplex de *primers* para as diversas famílias de segmentos VH e um *primer* consenso JH. Os *primers* Nextera XT, da segunda etapa, contém os adaptadores para sequenciamento (P5 e P7) e os indexadores únicos. Fonte: elaborado pelo autor.

Os sequenciamentos foram realizados no sequenciador MiSeq (Illumina). Para o sequenciador MiSeq, os *kits* utilizados foram o Miseq Reagent Nano Kit v2 (300-cycles, 1,000,000 *reads*) e o Miseq Reagent Kit v3 (150-cycles, 25,000,000 *reads*). O primeiro *kit* produz *paired-end reads* de 300 pares de base (*base pairs*, bp) e foi utilizado majoritariamente para amostras D0 (apesar de também ter sido usado para algumas amostras Fup), enquanto o segundo gera *single-end reads* de 150 bps. Independente do método de sequenciamento utilizado, todas as análises apresentadas ao longo deste trabalho foram realizadas utilizando *reads* únicos de 150 bps, de modo a garantir uniformidade do tipo de dado de sequenciamento analisado dentre as amostras.

Na *flow cell*, superfície onde o sequenciamento do DNA em si ocorre, o material genético obtido se atrela a essa lâmina em local aleatório. Cada um desses fragmentos é então amplificado por um processo denominado *bridge PCR*, com cada molécula de DNA individual gerando um *cluster* de fragmentos clonais. Os *clusters* são então sequenciados por um método de sequenciamento por síntese, com a realização de ciclos adicionais para sequenciar também o indexador de amostra previamente adicionado (**Figura 7**). Resumidamente, o sequenciamento (nesse exemplo, *single-end*) é realizado em três etapas: (1) utilização do *primer* para sequenciamento a partir da extremidade JH, obtendo a sequência JH-N(D)N-VH (2) lavagem da fita formada e adição do *primer* para sequenciamento do primeiro indexador e (3) adição do *primer* para sequenciamento do segundo indexador.

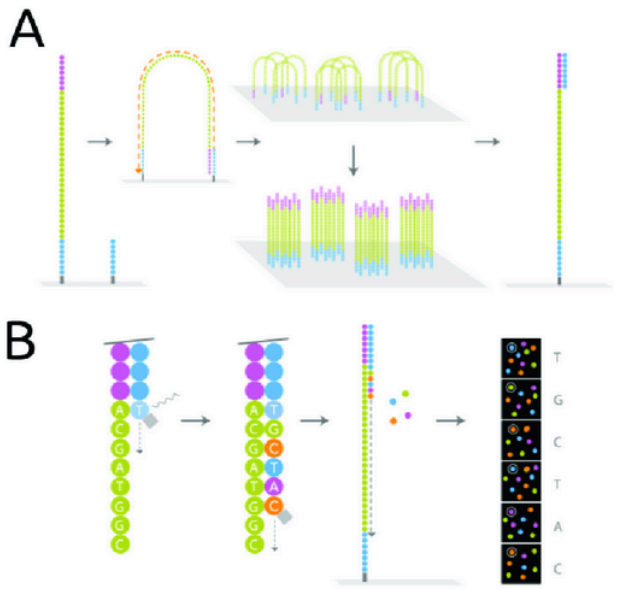

Figura 7. Resumo esquemático do processo de sequenciamento por síntese. (**A**) Os fragmentos de DNA que se ligam à *flow cell* são amplificados por *bridge* PCR. Adaptadores P5 e P7 representados em roxo e azul, respectivamente. (**B**) Os *clusters* gerados pelo processo de *bridge* PCR são sequenciados por síntese, com um nucleotídeo marcado com fluorescência sendo incorporado à leitura por ciclo. Cada tipo de nucleotídeo adicionado emite uma frequência de radiação específica. Para garantir a incorporação de apenas um nucleotídeo por ciclo, cada resíduo contém um bloqueador 5' removível (representado em cinza). Fonte: Westbury (2018).

O número almejado de leituras (*reads*) para as amostras D0, Fup e de doadores foi de 50.000, 500.000 e 300.000 respectivamente. As amostras D0 tiveram um menor número de *reads* sequenciados uma vez que a sua principal função no ensaio para DRM é apenas permitir a identificação dos clonótipos leucêmicos, que possuem número de *reads* altos. Às amostras Fup, foram adicionados DNA de plasmídeos contendo sequências controle de rearranjos *IGH* conhecidos, denominadas controles *spike-in*. Esses controles externos servem para realizar o cálculo da DRM do paciente, e foram subsequentemente removidos das demais análises aqui apresentadas. Assim, o número de *reads* produzidos que de fato correspondem a rearranjos VDJ da amostra sequenciada é bem menor do que o número total de *reads* gerados no sequenciamento, uma vez que a mediana do número de *reads* correspondentes a controles externos por amostra Fup foi de 251,511.5.

Enquanto esse foi o modo de preparo das bibliotecas de *IGH* de uma maneira geral, o sequenciamento das amostras D0 em específico apresentou também as seguintes peculiaridades: (1) Primeiramente, 10 ng de DNA genômico de linfócitos de sangue periférico de doadores (*peripheral blood lymphocytes*, PBL) foram adicionados às amostras de 152 dos 204 pacientes. Essa adição teve como objetivo confirmar o funcionamento das reações de PCR no preparo da biblioteca de sequenciamento, em casos nos quais a LLA não apresenta

rearranjo (assim, mesmo reações de PCR nas quais a amostra do paciente não possui rearranjos irão gerar bandas no gel de agarose). (2) No mais, outros rearranjos V(D)J foram sequenciados com o *IGH* em 92 de 204 pacientes. Mais especificamente, rearranjos de *immunoglobulin kappa locus* (IGK), *immunoglobulin lambda locus* (IGL), T *cell receptor gamma locus* (TRG) e T *cell receptor delta locus* (TRD). Nesses casos, o número de *reads* de *IGH* almejados por amostra foi de 20,000 em vez de 50,000, e apenas esses *reads* foram usados nesta tese.

4.4 Processamento dos Dados de NGS

Os dados de sequenciamento gerados tiveram sua qualidade analisada através do *software* FastQC (Babraham Bioinformatics). O agrupamento (*clustering*) de *reads* de rearranjos *IGH* idênticos foi realizado através do algoritmo `vidjil` (GIRAUD *et al*., 2014; DUEZ *et al.*, 2016), que foi executado segundo as configurações abaixo:

```
vidjil-algo -c clones -g homo-sapiens.g -3 -z 1000 -r 2 -w 80
--label-json Boldrini-spikes.json [path-to-fastq-file]
```

O `vidjil` primeiramente determina se um *read* corresponde a uma sequência de rearranjo V(D)J através de uma janela deslizante, que percorre o *read* até encontrar a sua região CDR3. O tamanho dessa janela, que é então utilizada como a sequência identificadora (id) daquele *read*, é controlado pelo argumento de linha de comando `-w`, que neste caso foi de 80 nucleotídeos. Em seguida, *reads* com id idênticos são agrupados em clonótipos, com pelo menos 2 *reads* sendo necessários para sustentar um clonótipo, conforme o argumento `-r 2`. Por fim, o `vidjil` realiza uma caracterização das sequências dos clonótipos identificados, gerando uma gama de dados que incluem as famílias V, D e J utilizadas, o comprimento do segmento CDR3, a sequência e comprimento das regiões de inserção N, entre outros. Essa última etapa é computacionalmente custosa, e por isso é restrita aos 1000 clonótipos mais frequentes em cada amostra (`-z 1000`). Apenas para as amostras de doadores de medula óssea esse parâmetro foi de 5,000, uma vez que os seus clonótipos passaram por um processo posterior de amostragem aleatória nesse estudo. Por fim, o argumento `--label-json` identifica o arquivo que contém as sequências dos controles *spike-in*, o que possibilita tanto a sua utilização no cálculo da DRM dos pacientes como também a subsequente remoção dessas sequências dos repertórios gerados.

Em seguida, os arquivos `vidjil` gerados são processados por uma *pipeline* desenvolvida nesta tese. O seu primeiro passo consiste na remoção das sequências correspondentes aos controles *spike-in*. Enquanto o `vidjil` é capaz de identificar clonótipos relativos a esses controles, o fato dele levar em consideração alinhamentos perfeitos com as sequências presentes no arquivo `Boldrini-spikes.json` faz com que clonótipos que apresentem pequenas variações, decorrentes de erros de sequenciamento, não sejam removidos. Para resolver essa questão, antes da remoção dos *spike-ins* em si, a *pipeline* utiliza o algoritmo `blastn` para alinhar clonótipos contra essas sequências de *spike-in* (ALTSCHUL *et al*., 1990). O clonótipo cuja sequência apresenta um comprimento de alinhamento maior que 25 nucleotídeos e uma identidade maior que 96% é marcado como *spike-in* e removido.

O passo seguinte da *pipeline* é responsável por agrupar clonótipos muitos semelhantes entre si, que também são artefatos da demasiada estringência do `vidjil` no processo de identificação das sequências de DNA. Nessa etapa, dois clonótipos são agrupados desde que eles não difiram mais do que 2 nucleotídeos entre si, e o id do clonótipo mais abundante é utilizado para representar o grupo gerado. Esse processo foi em partes executado utilizando o *software* VSEARCH (ROGNES *et al*., 2016).

Por fim, a *pipeline* também é responsável por identificar clonótipos correspondentes à LLA. Em amostras D0, essa identificação é realizada com base na frequência, com clonótipos que correspondem a pelo menos 1% do repertório em questão sendo considerados como leucêmicos. Em amostras Fup, essa identificação mantém o que se determinou na amostra D0, de modo que todos os clonótipos de LLA presentes ao diagnóstico que se mantém presentes após o fim da terapia de indução são marcados como tal. Uma breve representação esquemática da *pipeline* aqui descrita é apresentada na **Figura 8**.

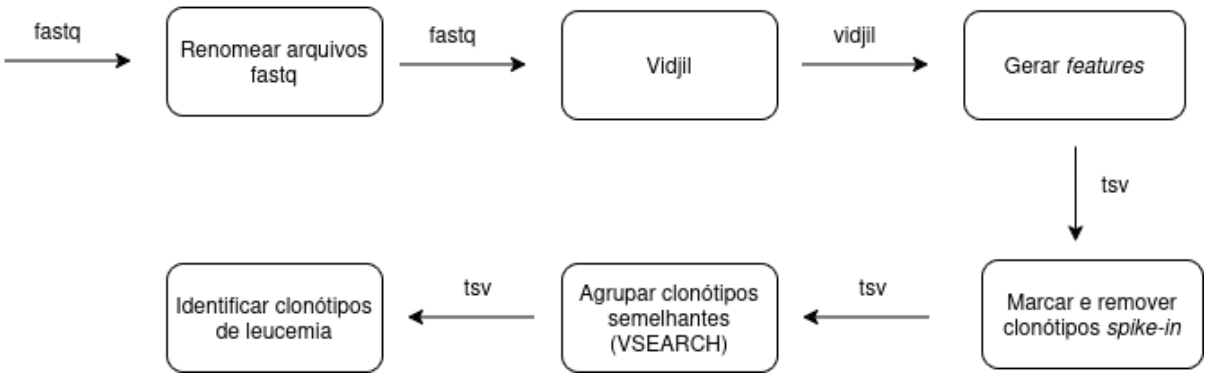

Figura 8. Diagrama esquemático da *pipeline* para o processamento dos dados de repertório V(D)J gerados por NGS. Caixas representam os passos da *pipeline*, enquanto as setas representam o formato dos arquivos de entrada e saída. Fonte: elaborado pelo autor.

4.5 Cálculo da DRM

A DRM é um parâmetro tradicionalmente expresso como a frequência de células de leucemia na população de células mononucleadas da medula óssea do paciente. Ensaios de NGS, por outro lado, produzem resultados expressos na forma de *reads* de rearranjos V(D)J. Somente células da linhagem linfoide realizam rearranjos VDJ de *IGH*, e, portanto, seriam alvo da quantificação por NGS. Porém, uma vez que a população linfocitária corresponde a apenas cerca de 18% das células mononucleadas da medula óssea, a utilização direta do número de *reads* de clones leucêmicos pelo total de *reads* de uma amostra representa apenas a frequência da leucemia em relação à população linfocítica, e não em relação ao conjunto total de células mononucleadas. Assim sendo, para obter valores de DRM a partir desses ensaios, é necessário converter essa frequência da carga leucêmica de *reads* para células, de modo a levar em consideração o número total de células presentes na amostra de medula óssea analisada. Neste trabalho, essa conversão se deu utilizando um sistema desenvolvido pelo grupo do qual faz parte o autor desta tese. (GIUSTI *et al*., 2020).

Nesse sistema, amostras Fup recebem, previamente ao início do preparo da biblioteca, material genético (plasmídeos) contendo 21 rearranjos *IGH* de sequência conhecida, e em quantidades (número de cópias) pré-determinadas. Mais especificamente, esse conjunto de sequências, que são denominadas controle *spike-in*, é composto por 3 sequências de cada família de segmentos VH (VH1-7), que são adicionadas à amostra em massa correspondente a 10, 40 e 160 cópias respectivamente.

Após o sequenciamento, os *reads* correspondentes a esses controles *spike-in* são identificados conforme explicado na seção anterior. Uma vez que tanto o número de cópias inicial e o número de *reads* gerados desses controles é conhecido, é possível determinar um

modelo de regressão linear para inferir o número de cópias de qualquer clonótipo identificado pela análise. Essa regressão linear é usualmente realizada utilizando apenas os três controles *spike-in* que compartilham a mesma família VH do clonótipo analisado, mas também pode usar todo o conjunto de controles em casos no qual a amplificação de algum dos controles família-específico falha. Considerando que o número total de células mononucleadas analisado no ensaio é previamente determinado pela quantidade de DNA utilizada no preparo da biblioteca (no caso, 600 ng ou 100.000 células), é possível calcular a frequência de clonótipos presentes na medula óssea do paciente em função do número total dessas células. Havendo mais de um clonótipo leucêmico, o valor de DRM do paciente corresponde ao maior valor entre todos os clonótipos analisados. Esse cálculo também é realizado pela *pipeline* mencionada na seção anterior do trabalho. Um diagrama esquemático desse processo é apresentado na **Figura 9**.

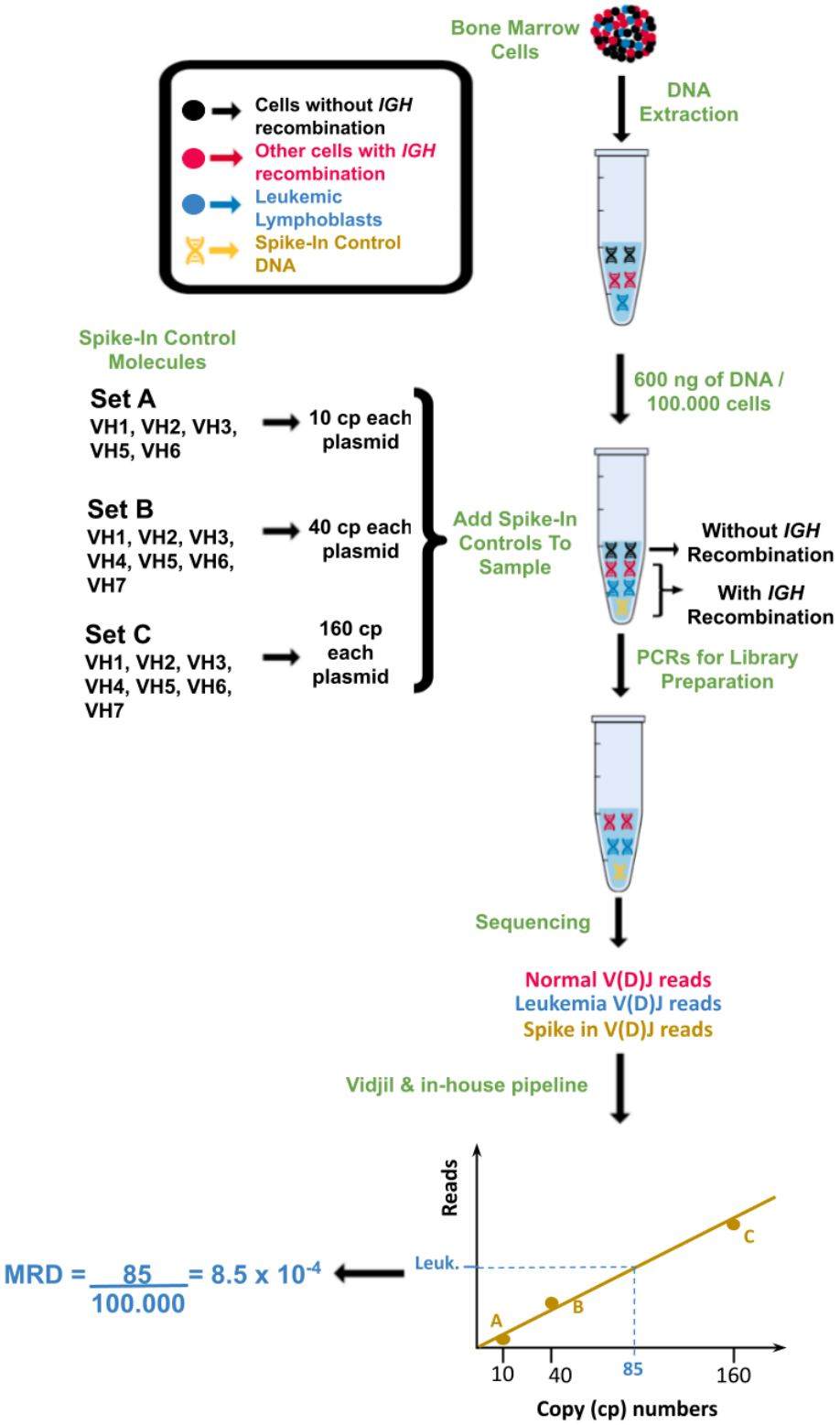

Figura 9. Diagrama esquemático do funcionamento do sistema de controles *spike-in* para determinação da DRM em amostras Fup. A regressão linear apresentada como exemplo utiliza apenas controles *spike-in* específicos para a família VH do clonótipo analisado. Fonte: elaborado pelo autor.

4.6 Análise Estatística

As análises de frequência de clonótipos produtivos e de enriquecimento de segmentos V e J foram realizadas através do teste exato de Fisher, com as análises de enriquecimento em particular havendo sido tratadas através da correção de Bonferroni. A comparação de características quantitativas, como comprimentos de sequência, frequência de nucleotídeos GC e diversidade do repertório de células B foi realizada através dos testes de Mann-Whitney U (comparações dois a dois) ou Kruskal-Wallis (comparações entre mais de dois grupos). Essa diversidade foi quantificada utilizando o índice de Gini-Simpson, que representa a probabilidade de dois *reads* amostrados aleatoriamente em uma amostra pertencerem a clonótipos distintos (JOST, 2006). Esse índice é representado por valores entre 0 e 1, com valores maiores correspondendo a maiores diversidades. Análises de sobrevida foram realizadas através do teste log-rank. Todas as análises estatísticas realizadas neste trabalho foram implementadas em `Python`, utilizando as bibliotecas `scipy` e `lifelines` (DAVIDSON-PILON, 2019; VIRTANEN *et al*., 2020).

4.7 Descrição do Conjunto de Dados

O gene *IGH* rearranjado foi sequenciado em amostras pareadas de medula óssea de 204 pacientes pediátricos com LLA-B diagnosticados entre 17/10/2012 e 03/09/2021 e em 8 amostras de medula de doadores de medula óssea. Uma breve descrição desse conjunto de dados é apresentada abaixo na **Tabela 1**. Para os doadores de medula, o único parâmetro clínico disponível é a idade.

Tabela 2. Sumário de dados clínicos dos 204 pacientes pediátricos com LLA-B incluídos neste trabalho. Também são apresentados os dados disponíveis para os 8 doadores de medula óssea. A seção Protocolo refere-se ao protocolo de tratamento a que o paciente foi submetido. Contagens de *reads* aferidas após a remoção dos controles *spike-in*.

	Pacientes LLA-B (N=204)	**Doadores (N=8)**
Sexo, n (%)		
Feminino	115 (56,3%)	-
Masculino	89 (43,6%)	-
Idade ao diagnóstico (anos)		
Média (Desvio Padrão)	6,2 (4,5)	11,3 (3,9)
Mínimo	0,2	6
Máximo	19	18
Subtipo molecular, n (%)		
Hiperdiploidia (>50 cromossomos)	67 (32,8%)	-
Hipodiploidia (<46 cromossomos)	3 (1,5%)	-
ETV6::RUNX1	35 (17,2%)	-
TCF3::PBX1	18 (8,8%)	-
KMT2A rearranjado	6 (2,9%)	-
BCR::ABL1	3 (1,5%)	-
B-other	72 (35,3%)	-
Protocolo, n (%)		
GBTLI LLA-2009	90 (44,1)	-
BFM LLA-2009	60 (29,4)	-
BFM LLA-2009/2017	39 (19,1)	-
Contagem de leucócitos/mm^3		
Média (Desvio Padrão)	31,5 (45,62)	-
Mínimo	1,1	-
Máximo	317,9	-
Estratificação DRM em Fup, n (%)		
DRM < 0,0001	155 (76)	-

0,001 > DRM ≥ 0,0001	28 (13,7)	-
DRM > 0,001	21 (10,3)	-
Contagem de *read* em D0 ou controle		
Média (Desvio Padrão)	$4,17 \times 10^4$ ($4,2 \times 10^4$)	$4,12 \times 10^5$ ($4,44 \times 10^4$)
Mínimo	$7,38 \times 10^3$	$3,7 \times 10^5$
Máximo	$3,01 \times 10^5$	$5,07 \times 10^5$
Contagem de *read* em Fup		
Média (Desvio Padrão)	$3,06 \times 10^5$ ($1,87 \times 10^5$)	-
Mínimo	$8,33 \times 10^3$	-
Máximo	$1,22 \times 10^6$	-
***Status* recaída clínica**, n (%)		
Sem recaída	187 (91,6)	-
Recaída	17 (8,3)	-
***Status* óbito**, n (%)		
Sem óbito	186 (91,1)	-
Óbito	18 (8,8)	-

5. CAPÍTULO I: MANUSCRITO

Characterization of Immunoglobulin Heavy Locus rearrangements in molecular subtypes of childhood B-Cell Precursor Acute Lymphoblastic Leukemia

Guilherme Navarro Nilo Giusti[1,2], Patrícia Yoshioka Jotta[1], Caroline de Oliveira Lopes[1,3], Natacha Azussa Migita[1], Amilcar Cardoso de Azevedo[1], Sílvia Regina Brandalise[1], João Meidanis[1,4] and José Andrés Yunes[1,3]

1 Centro Infantil Boldrini, Campinas 13083-210, SP, Brazil;
2 Graduate Program in Genetics and Molecular Biology, Institute of Biology, University of Campinas, Campinas, 13083-862, SP, Brazil;

3 Department of Translational Medicine, Faculty of Medical Sciences, University of Campinas, Campinas, São Paulo, SP, Brazil;

4 Computer Theory Department, Institute of Computing, University of Campinas, Campinas, 13083-862, SP, Brazil.

Short title: *IGH* rearrangements in BCP-ALL subtypes

Correspondence: José Andrés Yunes, Laboratório de Biologia da Leucemia, Centro Infantil Boldrini, Rua Dr Gabriel Porto 1270, Campinas, São Paulo, 13083-210, Brazil; e-mail: andres@boldrini.org.br.

Abstract

Biased *IGH* VDJ recombination has been previously described in childhood B-cell precursor acute lymphoblastic leukemia (BCP-ALL), although its causes are not yet fully understood. This study assesses differential features in 565 *IGH* clonotypes from BCP-ALL molecular subsets against 560 clonotypes from bone marrow donors. Leukemia clonotypes were enriched for *IGHV6-1* segments in the KMT2A rearranged and B-other subtypes, while *IGHV3-23* was enriched in TCF3::PBX1. ETV6::RUNX1 presented a topological gap in the usage of central IGHV segments. BCP-ALL also presented shorter CDR3 regions, higher GC content and lower productivity. Interestingly, productive clonotypes tended to be absent after induction therapy.

Short report

The Immunoglobulin Heavy Locus (*IGH*) gene has long been used as a marker for Minimal Residual Disease (MRD) evaluation in childhood B-Cell Precursor Acute Lymphoblastic Leukemia (BCP-ALL)[1]. This gene undergoes a process called VDJ recombination, in which one variable (IGHV), one diversity (IGHD) and one joining (IGHJ) gene segment are rearranged to form the non-constant portion of *IGH*. The resulting sequence, also called complementarity-determining region 3 (CDR3), is highly diverse, owing both to the multitude of gene segments available for recombination, as well as to the addition of nontemplated indels in the VD (N1) and DJ (N2) interfaces. As VDJ recombination is carried out independently for each newly-formed B-cell, it can be used as a molecular signature that uniquely identifies clonal B-cell populations originating from a common ancestor[2]. As BCP-ALL is caused by clonal populations of defective B-cell precursors, their *IGH* sequences serve to track and quantify residual leukemia cells along treatment[1].

Recently, our group has developed an assay for MRD assessment via *IGH* Next-Generation Sequencing (NGS)[3]. In addition to MRD, this method also allows the in-depth evaluation of the *IGH* sequence profile for the identified leukemia clonotypes. While traditionally *IGH* is seen mostly as a clonal marker for BCP-ALL, there is evidence that points to biased VDJ rearrangements in leukemia. Both patient age and leukemia genotype have been previously associated with differential leukemia *IGH* features[4,5]. The reasons for these biases are not fully understood, but may relate to potential differences in the VDJ recombination machinery (such as PAX5 controlled *IGH* accessibility or TdT and RAG activity levels)[6,7]. Therefore, in this study we describe the VDJ characteristics in *IGH* clonotypes from the different molecular subtypes of BCP-ALL, and in comparison to clonotypes present in healthy children who were bone marrow (BM) donors. To our knowledge, this is the most comprehensive look to date into *IGH* VDJ recombination patterns at DNA level for different childhood BCP-ALL subtypes.

In order to explore the *IGH* profile of BCP-ALL leukemia cells, BM samples from 204 children with BCP-ALL (age mean 6.2 years, range 0.2-19) and 8 healthy children donors (11.3, 6-18) had their *IGH* repertoire sequenced via NGS, as previously described[3]. The molecular subtypes for the BCP-ALL patients include hyperdiploidy (n=67), ETV6::RUNX1 (n=35), TCF3::PBX1 (n=18), KMT2A rearranged (n=6), hypodiploidy (n=3) and

BCR::ABL1 (n=3). The remainder of the dataset, with no classical chromosomal alterations, were called B-other (n=72). BM samples from two time points, diagnostic and post induction therapy (33-40 days), were sequenced for each patient. Patients were treated according to the Brazilian GBTLI ALL-2009 (n=97) or AEIOP BFM ALL-2009 protocols (n=107). The diagnostic samples were used to identify the leukemia clonotypes. Follow-up samples were used to evaluate leukemia clonotype persistence post induction. The number of *IGH* reads obtained were 411,546.13 ± 44,406.13 for donor samples, 41,761.76 ± 41,246.06 for BCP-ALL samples at diagnostic, and 306,247.10 ± 186,530.38 at follow-up.

Analysis of the NGS data for clonotype identification was performed using Vidjil[8]. The data was then further processed using an in-house pipeline, which is responsible for (1) removing spiked-in control clonotypes from the post induction repertoires, (2) clustering highly similar clonotypes using VSEARCH, which are likely to result from sequencing errors, and (3) identifying BCP-ALL clonotypes in diagnostic and follow-up samples[9]. Clonotypes which were present at diagnosis in frequencies higher than 1% were flagged as BCP-ALL. In total, we obtained data for 565 *IGH* sequences, pertaining to hyperdiploidy (n=192), ETV6::RUNX1 (n=83), TCF3::PBX1 (n=34), KMT2A rearranged (n=23), hypodiploidy (n=9), BCR::ABL1 (n=6) and B-other (n= 218) BCP-ALL. A group composed of 560 *IGH* sequences obtained by randomly sampling 70 clonotypes from each of the 8 healthy BM samples was used as control, being hereafter referred to as the Random Control *IGH* group. Additionally, a group called Frequent Control *IGH* was also created by gathering the 70 most frequent *IGH* sequences from each healthy BM sample in order to investigate possible repertoire biases in highly frequent clonotypes. Further detail on library preparation, sequencing and repertoire analysis, as well as rationale for automated clonotype clustering, can be found in the Supplemental Materials (Figures S1 and S2).

The KMT2A rearranged and B-other BCP-ALL subgroups presented an enrichment of the *IGHV6-1* when compared to the random *IGH* clonotypes. This gene segment is the most proximal one in relation to the *IGH* constant region, indicating an IGHV proximal segment bias in these groups. This corroborates with a recent study which observed higher expression of *IGHV6-1* transcripts in BCP-ALL adult and child patients with KMT2A rearrangements[10]. Additionally, there is a topological gap in IGHV usage for ETV6-RUNX1 *IGH* sequences, with segments ranging from *IGHV7-34-1* to *IGHV1-58* presenting lower frequencies (p=0.0016). ETV6::RUNX1 BCP-ALL has been previously linked to increases in RAG

activity levels[7], which could be investigated as a possible cause for this effect. Finally, the *IGHV3-23* was significantly enriched in TCF3::PBX1 BCP-ALL *IGH* clonotypes. No subtype presented differential frequencies of IGHJ in relation to the random *IGH* clonotypes (Figures 1A and 1B).

The groups ETV6::RUNX1, hypodiploidy and KMT2A rearranged presented shorter CDR3 sequences and higher GC content than the randomly sampled control *IGH* clonotypes (Figures 1C and 1D). No BCP-ALL subtype presented differential insertion length in the N region between V-D (Figure 1E). As for the N region between D-J, shorter insertions were observed for the ETV6::RUNX1, hyperdiploidy and B-other subsets (Figure 1F). Shorter N2 insertions did not necessarily determine an overall decrease in CDR3 length, as only the ETV6::RUNX1 clonotypes showed shortening for both of these features. These VDJ profile variations between BCP-ALL subtypes, including productivity (below), did not entail cluster formation in unsupervised learning analysis via PCA (Figure S3).

Unproductive sequences were vastly more common among BCP-ALL clonotypes when compared to those from the Random Control *IGH* group, with the noted exception of the TCF3::PBX1 subtype (Figure 1G). This reflects the overall tendency of BCP-ALL to elude clonal selection by hijacking pre-BCR expression, in contrast with TCF3::PBX1 where this fusion upregulates genes that are components of the pre-BCR, as shown by prior studies[5].

We also compared the *IGH* clonotypes which were randomly sampled from healthy BMs against the most frequent clonotypes retrieved from these same samples. Surprisingly, the *IGH* profile for these Frequent Control *IGH* clonotypes differed from the Random Control *IGH* group considerably more than any of the BCP-ALL subsets. The *IGHV1-2*, *IGHV1-18, IGHV1-46* and *IGHV1-69D* sequences were enriched in this group, while there was a decrease in *IGHV3-7*, *IGHV3-33*, *IGHV3-64*, *IGHV3-71* and *IGHV4-39* (Figure 1A). As for J, all segments presented differential usage in relation to the random control *IGH*, with the exception of *IGHJ1* (Figure 1B). Interestingly, the *IGHJ6* segment is completely absent among the Frequent Control *IGH* sequences, while being very common in both the Random Control *IGH* group and in all BCP-ALL subtypes. This group also presented shorter CDR3 sequences, accompanied by shortening of both N1 and N2, higher GC content, as well as a slightly higher frequency of productive clonotypes (Figures 1C-G).

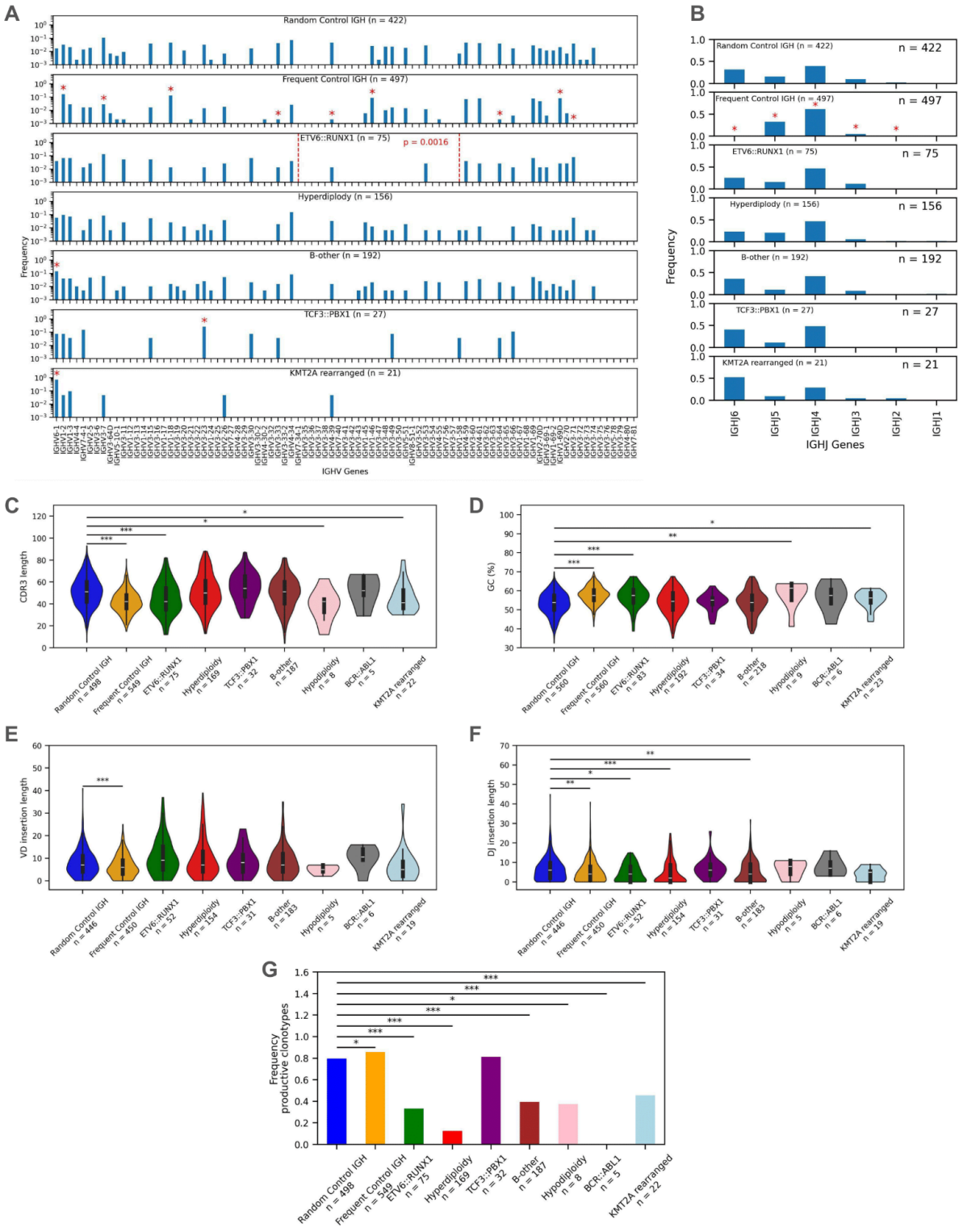

Figure 1. Characterization of *IGH* rearrangements in BCP-ALL molecular subtypes. Includes randomly sampled and most frequent clonotypes from healthy donors. All statistical tests performed in relation to the random control *IGH* group. The number of clonotypes per group varies across panels depending on data availability. (A) Distribution of *IGHV* gene segments, in log scale. ETV6::RUNX1 IGHV gap region depicted by the dashed red lines. (B) Distribution of *IGHJ* gene segments. (A and B) IGHV and IGHJ gene segments are ordered according to chromosomal position. (C) CDR3 length in nucleotides. (D) GC content, in percentages. (E) N1 insertion length in nucleotides. (F) N2 insertion length in nucleotides. (G) Frequency of productive CDR3 clonotypes.

Finally, we studied differences between those BCP-ALL *IGH* clonotypes that were still detectable in the patient's BM post induction therapy and those that were rendered undetectable by it. No differences were observed between these groups in relation to V or J segment usage, CDR3 and insertion lengths and GC content (Figures S4-9). However, clonotypes that were still detectable in the patient post induction therapy showed a lower proportion of productivity. To control for different productivity levels across BCP-ALL subtypes, this analysis was also performed in separate for three subtypes with at least 50 clonotypes available. Lower productivity at post induction was similarly observed for the ETV6::RUNX1 and B-other subsets (Figure 2A). The pre-BCR has been shown to play a tumor suppressor role in some BCP-ALL cases, which could explain the depletion of productive *IGH* in treatment-resistant sequences[,11].

In contrast, previous studies have noted that patients with unproductive BCP-ALL *IGH* presented better leukemia-free survival (LFS) than those with productive *IGH* clonotypes[12,13]. In order to explore these opposing observations, we analyzed whether *IGH* productivity status would relate to post induction MRD levels. As shown in Figure 2B, *IGH* clonotypes with higher MRD tended toward unproductiveness. However, when it comes to the clonotypes with the highest MRD values ($\geq 10^{-3}$), productivity goes up again. This productivity enrichment in a group of bad responders may help to explain the previously reported worse LFS for patients with productive BCP-ALL. Regardless, these results point to a connection between *IGH* productivity and outcome in BCP-ALL, indicating the need for additional investigation regarding the causes of this effect.

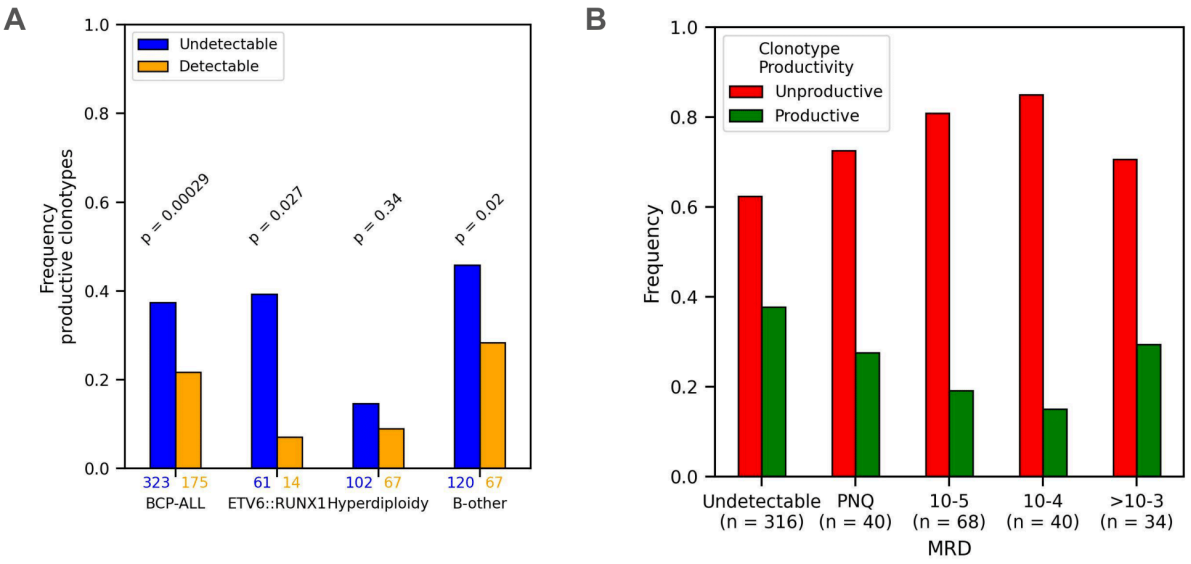

Figure 2. Analysis of IGH productivity by MRD at the end of induction therapy. (A) Comparison of the frequency of productive BCP-ALL *IGH* clonotypes which were undetectable or still detectable in the patients' BM post induction therapy. Colored numbers below bars indicate the total number of clonotypes included in each group. (B) Comparison of BCP-ALL *IGH* productivity status by clonotype abundance. Abundance measured by the order of magnitude of the MRD value for each *IGH* sequence. PNQ = positive not quantifiable, defined as *IGH* clonotypes with $0 < MRD < 10^{-5}$.

Acknowledgements

Guilherme N N Giusti was financed by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finance Code 001 (PROEX 88887.342097/2019-00). João Meidanis received a grant from São Paulo Research Foundation (FAPESP, 2024/01200-8). José A Yunes received a productivity fellowship from the National Counsel of Technological and Scientific Development (CNPq, 308399/2021-8). This work was supported by research funding from PRONON (Programa Nacional de Apoio à Atenção Oncológica, NUP 25000.057709/2015 and NUP 25000.211174.2019-45).

Authorship

Contribution: GNNG and JAY conceived and designed the study; GNNG, PYJ, COL and NAM performed all sequencing experiments; SRB and ACA was responsible for the diagnosis and treatment of patients; GNNG, JM and JAY analyzed results; GNNG and JAY wrote the manuscript.

Data availability: Data concerning the clonotypes analyzed in this study available in the supplemental file `clonotypes.tsv`. Raw `fastq` files available upon request.

Conflict-of-interest disclosure: The authors have no conflicting financial interests.

Keywords: Acute Lymphoblastic Leukemia, VDJ recombination, Immunoglobulin Heavy Locus, Minimal Residual Disease, ETV6-RUNX1

References

1. van Dongen JJM, Langerak AW, Brüggemann M, Evans PAS, Hummel M, Lavender FL, et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 Concerted Action BMH4-CT98-3936. Leukemia. 2003 Dec;17(12):2257–317.

2.Tonegawa S. Somatic generation of antibody diversity. Nature. 1983 Apr;302(5909):575–81.

3.Giusti GNN, Jotta PY, Lopes C de O, Ganazza MA, Azevedo AC, Brandalise SR, et al. Test trial of spike‐in immunoglobulin heavy‐chain (*IGH*) controls for next generation sequencing quantification of minimal residual disease in acute lymphoblastic leukaemia. British Journal of Haematology. 2020 Mar 18;189(4).

4.van der Velden VHJ, Szczepanski T, Wijkhuijs JM, Hart PG, Hoogeveen PG, Hop WCJ, et al. Age-related patterns of immunoglobulin and T-cell receptor gene rearrangements in precursor-B-ALL: implications for detection of minimal residual disease. Leukemia. 2003 Sep;17(9):1834–44.

5.Geng H, Hurtz C, Lenz K, Chen Z, Baumjohann D, Thompson SK, et al. Self-Enforcing Feedback Activation between BCL6 and Pre-B Cell Receptor Signaling Defines a Distinct Subtype of Acute Lymphoblastic Leukemia. Cancer Cell. 2015 Mar 9;27(3):409–25.

6.Hill L, Ebert A, Jaritz M, Wutz G, Nagasaka K, Tagoh H, et al. Wapl repression by Pax5 promotes V gene recombination by Igh loop extrusion. Nature. 2020 Jul 1;584(7819):142–7.

7.Jakobczyk H, Jiang Y, Debaize L, Soubise B, Avner S, Sérandour AA, et al. ETV6-RUNX1 and RUNX1 directly regulate RAG1 expression: one more step in the understanding of childhood B-cell acute lymphoblastic leukemia leukemogenesis. Leukemia [Internet]. 2022 Feb 1 [cited 2022 Nov 7];36(2):549–54. Available from: https://www.nature.com/articles/s41375-021-01409-9

8.Duez M, Giraud M, Herbert R, Rocher T, Mikaël Salson, Florian Thonier. Vidjil: A Web Platform for Analysis of High-Throughput Repertoire Sequencing. PLOS ONE. 2016 Nov 11;11(11):e0166126–6.

9.Rognes T, Flouri T, Nichols B, Quince C, Mahé F. VSEARCH: a versatile open source tool for metagenomics. PeerJ. 2016 Oct 18;4:e2584.

10.Müller H, Dicker F, Bär C, Walter W, Hutter S, Nadarajah N, et al. Proximally biased V(D)J recombination in the clonal evolution of IGH alleles in *KMT2A::AFF1* BCP‐ALL of all age classes. HemaSphere. 2024 Apr 1;8(4).

11.J Eswaran, Sinclair P, Heidenreich O, Irving J, Russell LJ, Hall A, et al. The pre-B-cell receptor checkpoint in acute lymphoblastic leukaemia. Leukemia. 2015 May 6;29(8):1623–31.

12.Katsibardi K, Braoudaki M, Papathanasiou C, Karamolegou K, Tzortzatou-Stathopoulou F. Clinical significance of productive immunoglobulin heavy chain gene rearrangements in childhood acute lymphoblastic leukemia. Leukemia & Lymphoma. 2011 Jun 8;52(9):1751–7.

13.Li A, Rue M, Zhou J, Wang H, Goldwasser MA, Neuberg D, et al. Utilization of Ig heavy chain variable, diversity, and joining gene segments in children with B-lineage acute lymphoblastic leukemia: implications for the mechanisms of VDJ recombination and for pathogenesis. Blood. 2004 Jun 15;103(12):4602–9.

Supplemental data for:

Characterization of Immunoglobulin Heavy Locus rearrangements in molecular subtypes of childhood B-Cell Precursor Acute Lymphoblastic Leukemia

Methods

Patients and samples

The study was approved by the institutional ethical committee (CAAE 57280616.4.0000.5376) and included retrospective samples from children with BCP-ALL, aged 1 to 19 years, treated according to the Brazilian GBTLI ALL-2009 (n = 97) or the European AIEOP BFM ALL-2009 (n=107) protocols at Boldrini Center. Samples from 2 time points, diagnostic and post induction therapy (33-40 days) were used for each patient. Patients that died during induction or which did not have available post induction samples were excluded. This study also included retrospective samples from 8 healthy children, aged 6 to 18, who were donors for bone marrow transplantation. Informed consent was obtained from all patients or their parents.

NGS library preparation

NGS library preparation for diagnostic (D0) and follow-up (Fup) BCP-ALL BM samples was carried out as previously described[1]. Briefly, 50 ng of D0 gDNA or 600 ng of Fup gDNA, which correspond to about 8,333 and 100,000 genomes respectively, underwent a two-step nested PCR. For the first step, BIOMED-2 *IGH* primers with 5' overhangs for Nextera XT (Illumina) were used for amplification. Second-step PCR reactions were carried out using Nextera XT primers. Libraries were purified using a 0.7 ratio of Agencourt Ampure XP Beads (Beckman Coulter). Sequencing of D0 and Fup samples were carried out mostly using the Miseq Reagent Nano Kit v2 (300-cycles) and the Miseq Reagent Kit v3 (150-cycles) respectively, although some Fup samples were also sequenced using the 300 cycles kit. However, analysis for both types of samples was carried out using 150-cycle single read data.

The target read count per sample was 50,000 and 500,000 for D0 and Fup samples respectively. Fup samples are spiked-in with known IGH control sequences during library preparation, but these are only relevant for Minimal Residual Disease (MRD) calculation and thus are removed prior to any analysis for this study. For this reason, the count of reads actually belonging to this kind of sample is usually lower than the sequenced 500,000. As for the donor samples, 50 ng of gDNA and a target read count of 300,000 were used.

Generally, NGS libraries were prepared as described above, but there were slight variations during preparation of some D0 patient samples, as follows. Genomic DNA from peripheral blood lymphocytes (PBL) was added at 10 ng to 152 out of the 204 sequenced samples. This addition does not affect BCP-ALL *IGH* sequence identification in these samples. Additionally, amplicons for Immunoglobulin Kappa Locus (*IGK*), Immunoglobulin Lambda Locus (IGL), T Cell Receptor Gamma Locus (TRG) and T Cell Receptor Delta Locus (TRD) were also sequenced together with IGH in 92 out of the 204 BM samples, although only reads related to *IGH* were used for this study. Target *IGH* read count for these samples was 20,000 and the Agencourt Ampure XP Beads ratio was 0.8

NGS data processing

As mentioned, *IGH* clonotype identification and quantification was performed using Vidjil and its output data was processed using an in-house pipeline, for which the main steps are described here. Vidjil was run using the command below. The `-z` argument, which determines how many of the top most frequent clonotypes will be fully analyzed by Vidjil, received a value of 5000 for the donor samples. As the clonotypes from this group undergo random sampling during this study, this ensures that all clonotypes sampled have been fully analyzed. The `Boldrini-spike.json` file contains the sequences for the aforementioned spiked-in sequences, and is used to properly tag their clonotypes as such.

```
vidjil-algo -c clones -g homo-sapiens.g -3 -z 1000 -r 2 -w 80
--label-json Boldrini-spikes.json [path-to-fastq-file]
```

The first step of the pipeline consists in removing the known *IGH* sequences spiked into the Fup samples. As mentioned, Vidjil is already able to tag clonotypes which match particular sequences of interest, which can be exploited to flag these spiked-in clonotypes for removal.

However, this is only done for clonotypes which match the given sequence exactly. Therefore, prior to removal, it is important to identify other spike-in clonotypes which have not been tagged as such due to sequencing errors. In order to do this, BLAST's nblast algorithm was used. With this, clonotypes with alignment lengths > 25 nucleotides and identities > 96% to any spike-in sequence were also flagged as a spike-in sequence and removed from analysis.

Next, another major step of our pipeline involves automatically clustering very similar clonotypes, which are likely to result from sequencing errors, since Vidjil does not allow any degree of mismatches during clonotype identification. A consequence of this strictness can be observed in Figure S1. There, the proportion of productive clonotypes was assessed for both BCP-ALL and background clonotypes. This was carried out for both D0 and Fup samples. While no distortions were observed in Fup samples, in D0 the percentage of productive background clones was unexpectedly low. We hypothesized that this was a result of a multitude of low frequency clonotypes which were related to dominant BCP-ALL clonotypes (frequency > 1%) and that were not considered as such due to small sequencing errors. To assess this, we automatically clustered clonotypes that differed by up to 0, 1, 2, 3, 4, 5, 6, 7, and 8 nucleotides. As shown, clusterization with a threshold as low as 1 nucleotide was able to revert the productivity levels to those expected. A threshold of 2 reached a plateau in productivity levels. In order to better inform the clusterization threshold choice, we also assessed the percentage of clonotypes per sample which had N regions (N1 + N2) shorter than this threshold, as shown in Figure S2. In theory, this would allow two unrelated clonotypes to end up mistakenly clustered together, although for this to happen they would also need to share the exact same IGHV, IGHD and IGHJ segments, and have suffered deletions of the same length and location during recombination. Taking both these analyses together, we decided for a clusterization threshold of 2. This clusterization process is in part carried out using VSEARCH.

Finally, this pipeline also identifies and tags BCP-ALL clonotypes in both D0 and Fup samples. For D0 samples, all clonotypes present in a frequency above 1% are considered as belonging to BCP-ALL. This threshold is lower than the traditionally used 5%, taking into account studies which have demonstrated the presence of BCP-ALL clonotypes with frequencies below this value[1,2]. After induction therapy BCP-ALL cells in the bone marrow are either undetectable or greatly reduced, making it impossible to use this same frequency

approach for Fup samples. Therefore, BCP-ALL clonotypes from paired D0 samples are searched for in Fup samples and tagged as such when found.

Leukemia clonotype Minimal Residual Disease calculation

MRD measures the leukemia burden in BCP-ALL patients along treatment, being expressed as the frequency of detectable leukemia cells among the total mononucleated cell count in a patient's BM sample. If a patient has, for instance, three different leukemia clonotypes, each clonotype will have an independent MRD value, and the highest one will be chosen as the MRD value for the patient. As such, MRD is a feature that can be analyzed not only at the patient level, but also at the clonotype level.

The MRD level for each leukemia clonotype was calculated using the same method referred to in the *NGS library preparation* section above and in our previous publication[1], and as such, MRD analysis did not include the clonotype clustering step described above. This is relevant, because all other IGH features in this study did include the clusterization step.

Statistical analysis

Statistical analyses for *IGHV* and *IGHJ* enrichment, as well as for frequency of productive clonotypes, were performed via Fisher's exact test. *IGHV* and *IGHJ* enrichment results were treated via Bonferroni correction. Comparison of quantitative features, such as CDR3 length, GC content and insertion length were performed via Mann-Whitney U test.

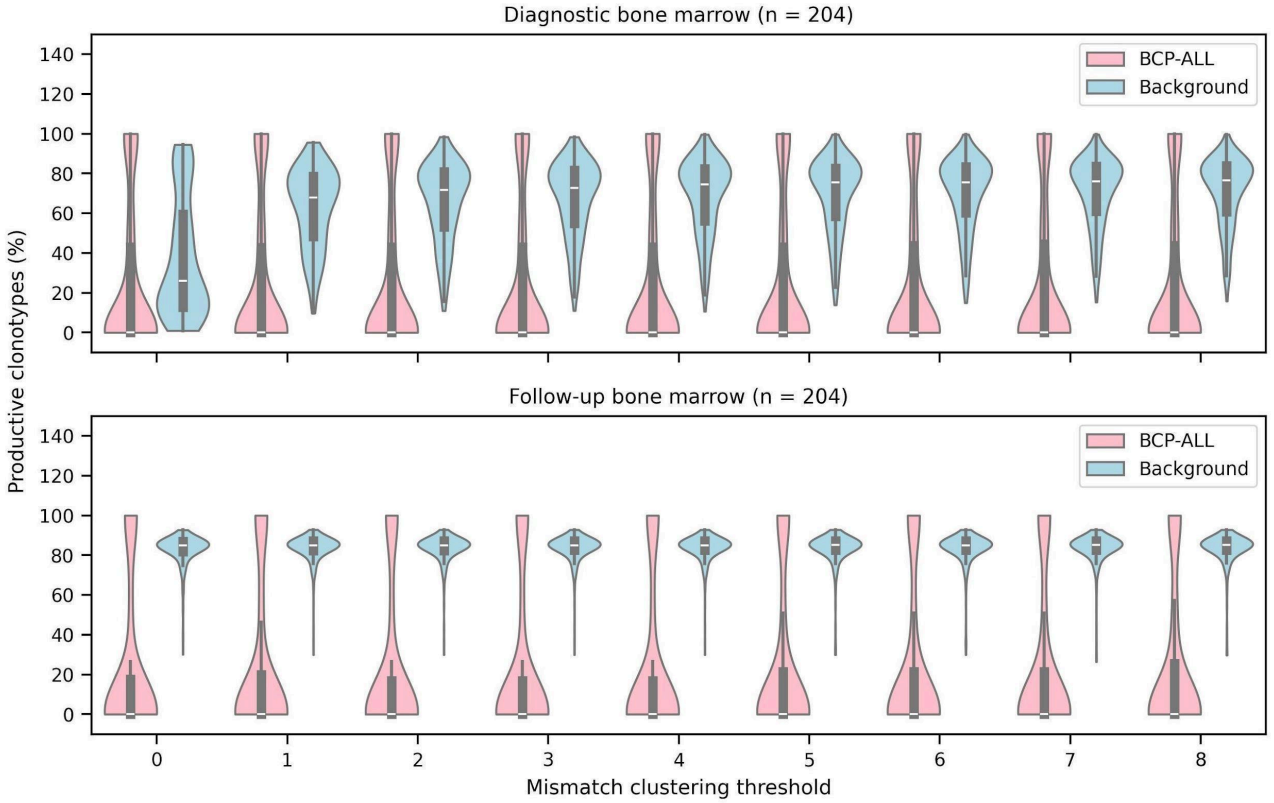

Figure S1. Percentage of productive BCP-ALL and background clonotypes in diagnostic and follow-up samples at several different mismatch clusterization thresholds.

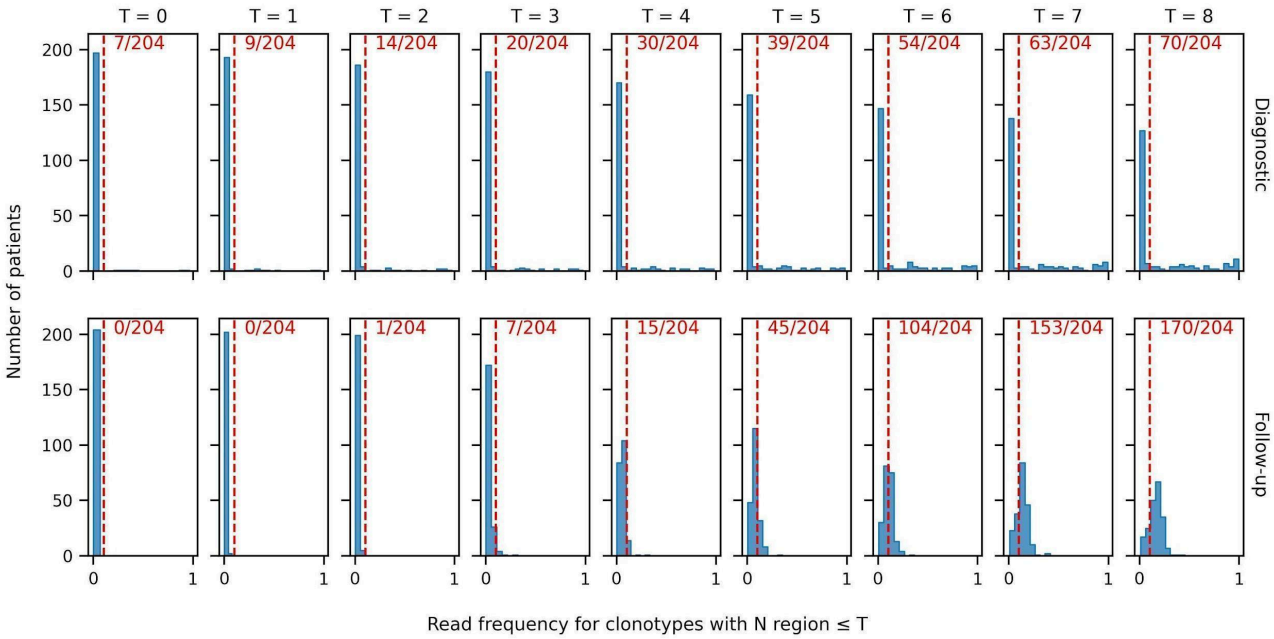

Figure S2. Number of patients where the frequency of *IGH* reads with N regions ≤ mismatch clusterization threshold exceeds 10% in D0 and Fup samples. Clusterization threshold is represented by T. The 10% read frequency threshold is represented by the dashed red line. Number of patients exceeding this threshold is presented in red.

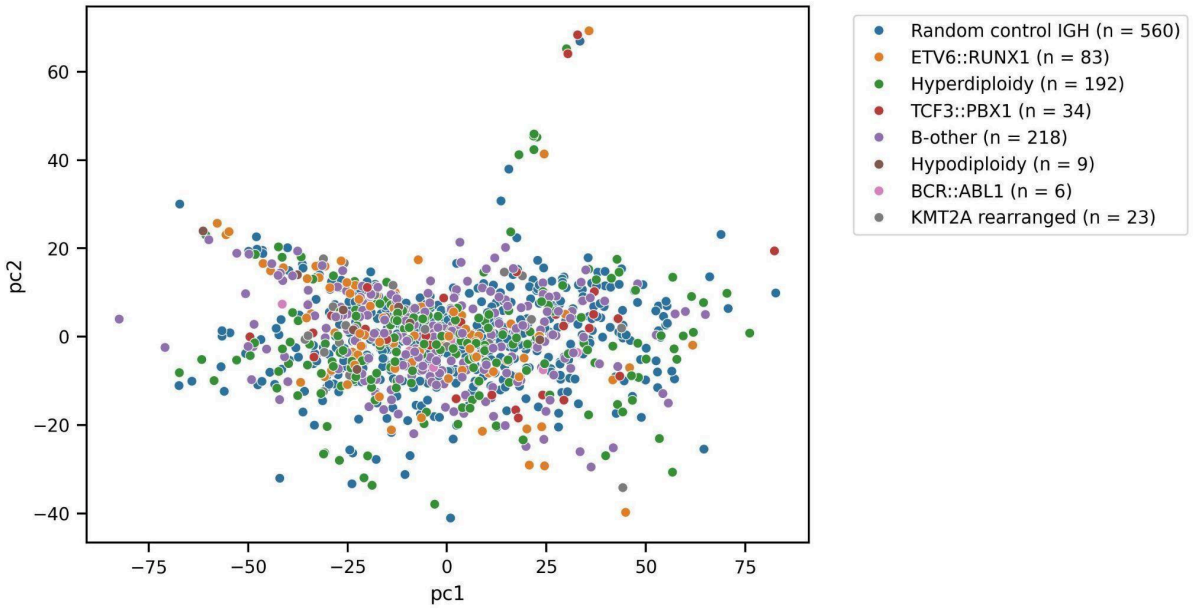

Figure S3. Unsupervised clustering of donor and BCP-ALL *IGH* clonotypes via Principal Component Analysis (PCA). Variables analyzed were insertion lengths, CDR3 (nucleotide length, start and end position), IGHV (end position and deletion length on the IGHV-IGHD interface), IGHD (start and end position, deletion lengths on both ends of the segment), IGHJ (start and end position, deletion length on the IGHD-IGHJ interface) and percents of A, T C and G.

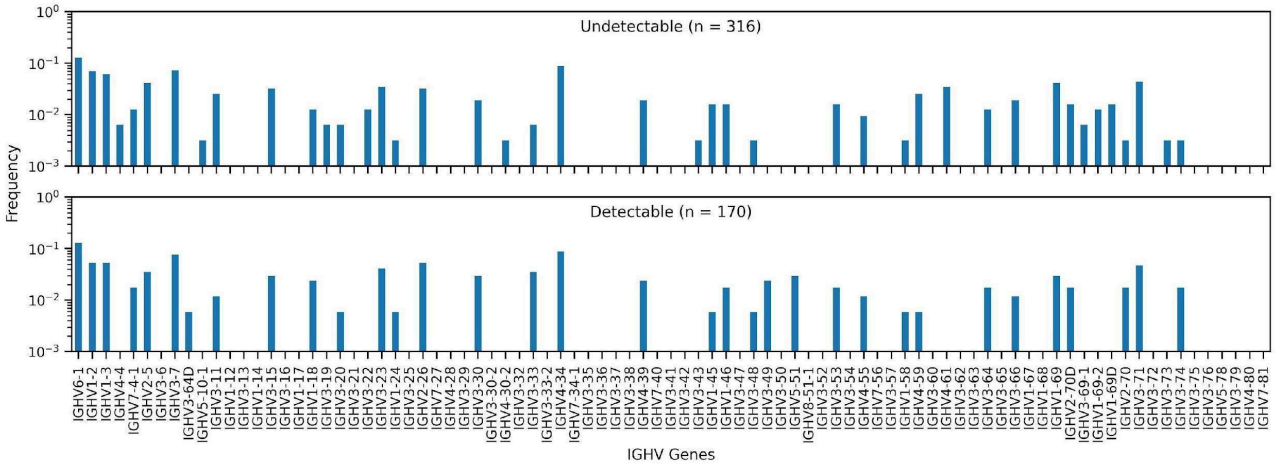

Figure S4. Distribution of *IGHV* gene segments for undetectable and detectable BCP-ALL *IGH* clonotypes post induction therapy, in log scale. No statistically significant differences were found.

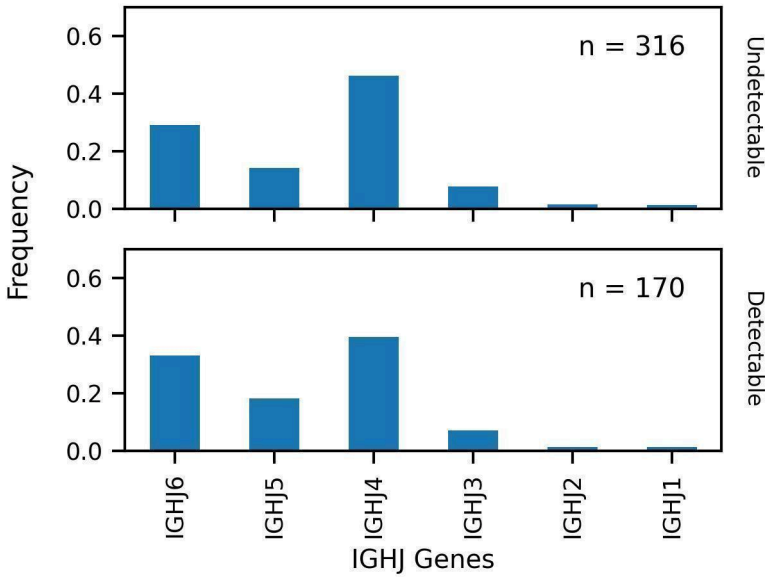

Figure S5. Distribution of *IGHJ* gene segments for undetectable and detectable BCP-ALL *IGH* clonotypes post induction therapy, in log scale. No statistically significant differences were found.

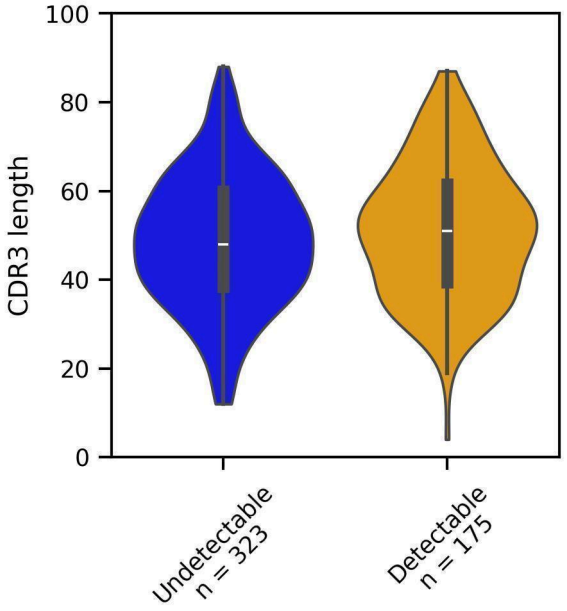

Figure S6. CDR3 length for undetectable and detectable BCP-ALL *IGH* clonotypes post induction therapy. Length in nucleotides. No statistically significant differences were found.

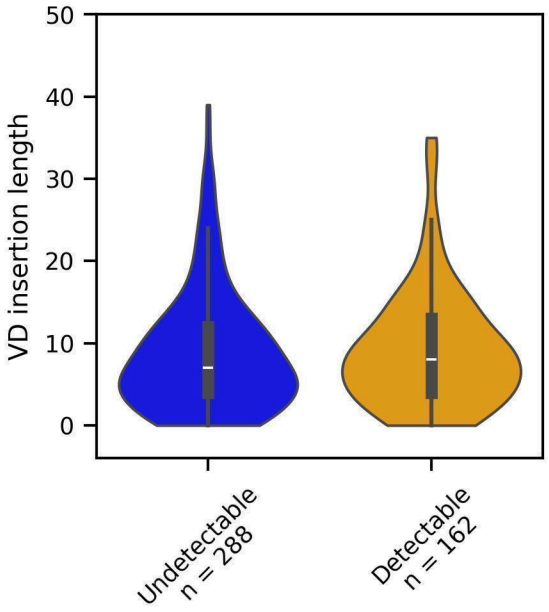

Figure S7. N1 insertion length for undetectable and detectable BCP-ALL *IGH* clonotypes post induction therapy. Length in nucleotides**.** No statistically significant differences were found.

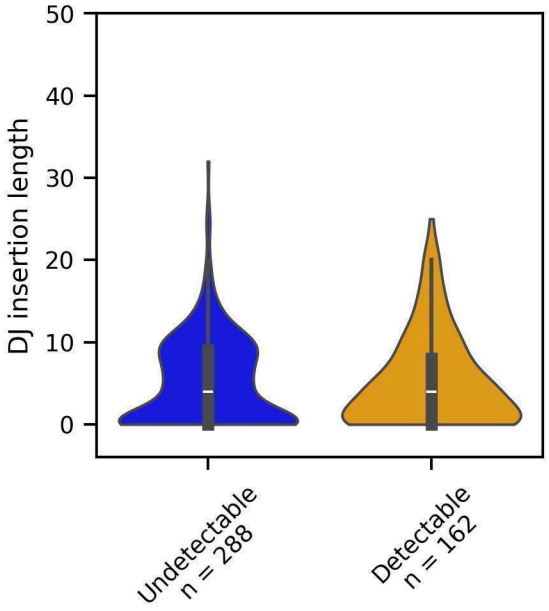

Figure S8. N2 insertion length for undetectable and detectable BCP-ALL *IGH* clonotypes post induction therapy. Length in nucleotides**.** No statistically significant differences were found.

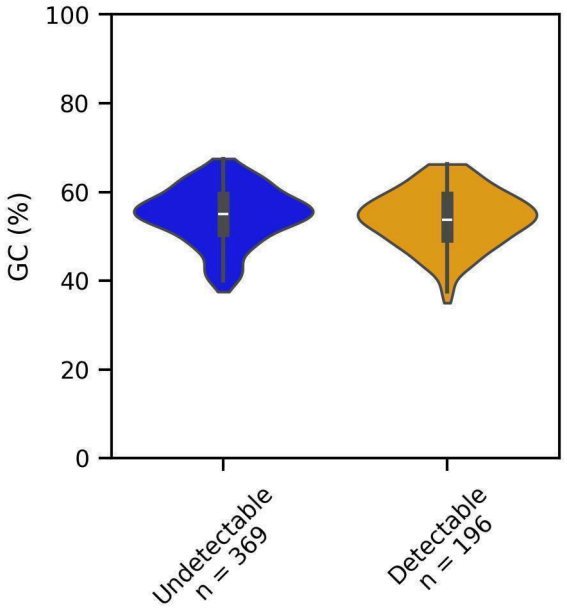

Figure S9. Percentage of GC content for undetectable and detectable BCP-ALL *IGH* clonotypes post induction therapy. No statistically significant differences were found.

Giusti GNN, Jotta PY, Lopes C de O, Ganazza MA, Azevedo AC, Brandalise SR, et al. Test trial of spike‐in immunoglobulin heavy‐chain (IGH) controls for next generation sequencing quantification of minimal residual disease in acute lymphoblastic leukaemia. British Journal of Haematology. 2020 Mar 18;189(4).

Darzentas F, Szczepanowski M, Kotrová M, Hartmann A, Beder T, Gökbuget N, et al. Insights into IGH Clonal Evolution in BCP-all: Frequency, mechanisms, associations, and diagnostic implications. Frontiers in Immunology. 2023 Apr 18;14.

6. CAPÍTULO II: DIVERSIDADE DA MEDULA ÓSSEA NA LLA-B

Conforme anteriormente mencionado nesse trabalho, a avaliação da progressão clínica e efetividade do tratamento na LLA é realizada através da análise de amostras de medula óssea do paciente colhidas ao longo do processo terapêutico. Atualmente, o principal parâmetro avaliado para esse propósito é a citorredução de células leucêmicas, através da DRM. Os ensaios modernos de DRM baseados em NGS também permitem complementar a avaliação do estado geral da medula óssea do paciente ao longo tratamento, possibilidade essa ainda sub explorada no manejo da LLA. Uma vez que a LLA-B, além de suplantar as células normais na medula óssea, também altera o seu microambiente, uma restauração completa dessa medula para um estado saudável passa não apenas pela citorredução das células malignas, mas também pela sua capacidade de restabelecer sua homeostase celular. Nesse contexto, a avaliação da diversidade de linfócitos B nesse tecido, através da análise das suas sequências de *IGH*, pode funcionar como um parâmetro interessante para melhor compreender se a qualidade de regeneração da medula óssea após tratamento impacta a evolução clínica do paciente. Em resumo, a análise da DRM foca na doença em si, quantificando células leucêmicas residuais. Já na análise da diversidade de clonótipos *IGH* na medula óssea realizada neste capítulo, o foco é a "saúde", buscando entender o impacto dessa diversidade na progressão clínica da doença.

Assim sendo, foram analisados os índices de diversidade das medulas ósseas ao diagnóstico (D0) e no final da indução (Fup), dividindo os casos pelos subtipos genéticos de LLA-B. Na **Figura 1A** são apresentados os resultados para os subtipos de LLA que dispunham de pelo menos 10 casos. Conforme esperado, todos os subtipos apresentaram medulas significativamente mais diversas no ponto de tratamento Fup, uma vez que em D0 esse tecido encontra-se tomado por clonótipos leucêmicos altamente dominantes. No entanto, essa baixa diversidade da medula ao diagnóstico variou em função do subtipo de LLA-B ($p = 0.024$), com os casos *ETV6::RUNX1* e *TCF3::PBX1* apresentando menor diversidade que aqueles com hiperdiploidia ou sem subtipo caracterizado (*B-other*). O mesmo efeito não foi observado após o término da terapia de indução. É importante ressaltar que a adição de PBL em algumas amostras D0 não resultou em alteração significativa da sua diversidade, de modo que as diferenças de diversidade observadas não se devem a esse artefato da técnica (**Figura 1B**). Não se observou correlação entre a diversidade medular e outras características

biológico-clínicas dos pacientes, como idade, sexo, tempo para remissão clínica, produtividade do repertório *IGH* e DRM, conforme apresentado na **Tabela 1**.

Em seguida, comparou-se a diversidade da medula de pacientes que apresentaram recaída da doença contra aqueles que mantiveram seu estado de remissão ao longo do período de acompanhamento deste trabalho (**Figura 1C**). Nesse caso, aqueles pacientes que recaíram ao longo do tratamento apresentaram uma menor diversidade de linfócitos B logo após a terapia de indução ($p = 0.012$). Essa diferença é condizente com a tendência de pacientes que recaem possuírem valores de DRM mais altos nesse ponto do tratamento, uma vez que os clonótipos leucêmicos presentes nesses casos muitas vezes são relativamente abundantes e, portanto, reduzem a diversidade da medula óssea. Esse efeito é exemplificado pela maior diversidade em pacientes com DRM indetectável em relação ao total de pacientes analisados (**Figura 1D**).

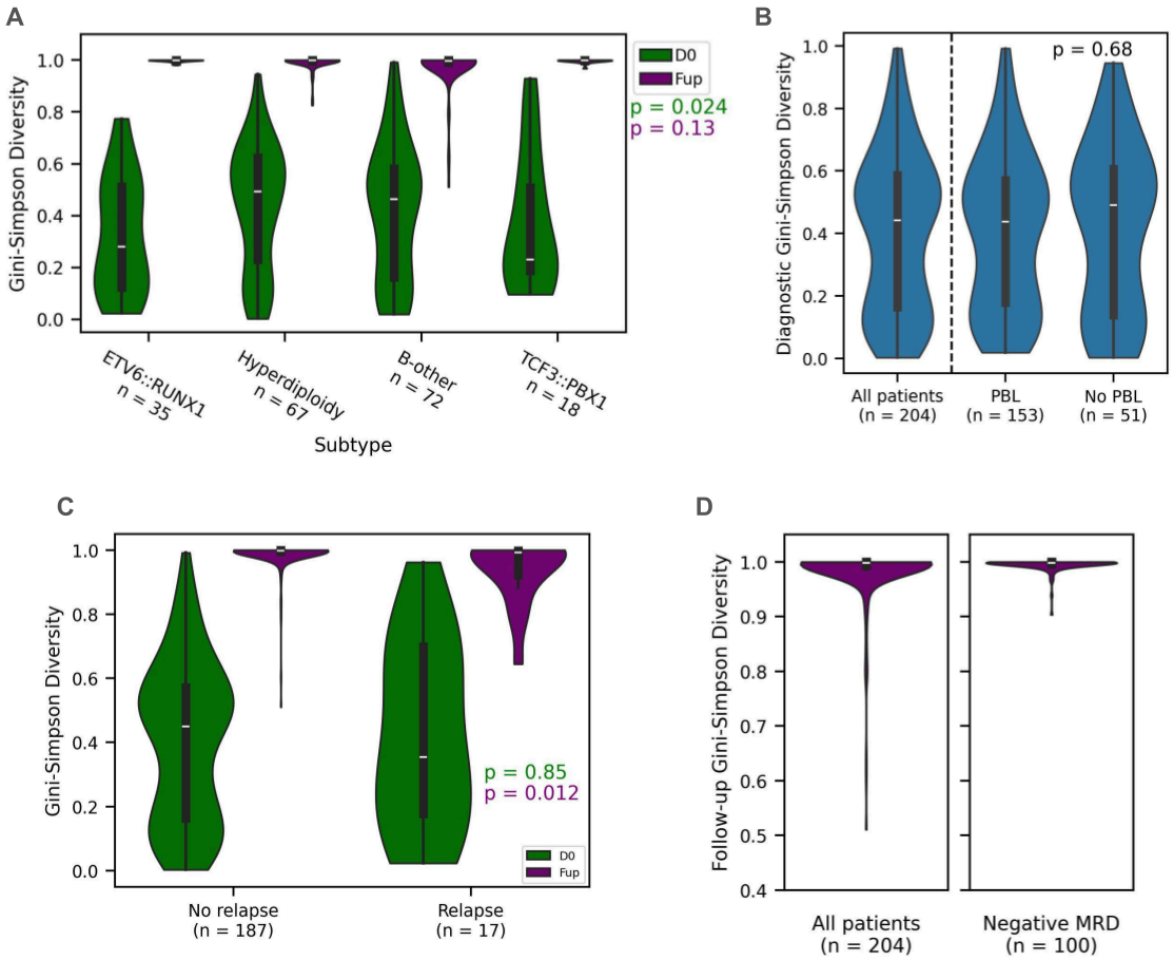

Figura 1. Análise da diversidade da medula óssea no diagnóstico e após a terapia de indução em pacientes de LLA-B. (**A**) Comparação do índice de diversidade Gini-Simpson entre subtipos de LLA-B com pelo menos 10 casos disponíveis. P-valores referentes à teste qui-quadrado comparando a diversidade de todos os subtipos representados em um mesmo ponto do tratamento. (**B**) Comparação do índice de diversidade Gini-Simpson entre amostras de medula óssea D0 que receberam ou não adição de DNA de PBL durante o preparo da biblioteca para sequenciamento. P-valores referentes a teste de Mann-Whitney U. (**C**) Comparação do índice de diversidade Gini-Simpson entre pacientes que apresentaram ou não recaída após ao longo do tratamento. P-valores referentes a teste de Mann-Whitney U comparando a diversidade dos grupos em um mesmo ponto do tratamento. (**D**) Diversidade após a terapia de indução do conjunto total de pacientes estudados em relação aos pacientes com DRM indetectável.

Tabela 1. Variáveis biológico-clínicas sem correlação com a diversidade medular de pacientes pediátricos com LLA-B ao diagnóstico e ao término da terapia de indução.

	Diversidade D0	**Diversidade Fup**
Idade (Coeficiente de Spearman)	-0.26	-0.068
Sexo (P-valor Mann-Whitney U)	0.97	0.57
Tempo até Remissão (Coeficiente de Spearman)	0.02	-0.16
Produtividade do repertório *IGH* (Coeficiente de Spearman)	0.037	0.44
DRM 15º Dia após Início da Indução (Coeficiente de Spearman)	-0.012	-0.25
DRM Fup (Coeficiente de Spearman)	-0.0054	-0.44

Nesse cenário, a análise da diversidade da medula em pacientes com DRM indetectável mostra-se então especialmente interessante, uma vez que (1) esses casos não apresentam interferência da DRM como variável de confusão e (2) possíveis associações com a evolução clínica deles seriam valiosas para eventual direcionamento do tratamento. Para possibilitar essa análise, avaliou-se primeiramente a distribuição dos valores de diversidade da medula após a terapia de indução entre todos os pacientes avaliados nesse estudo (**Figura 2A**). A maioria dos pacientes (127/204) apresentou valores de diversidade extremamente altos nesse ponto do tratamento (> 0.9968), com o restante (77/204) tendo medulas menos diversas, o que pode ser um possível indicador de uma pior regeneração dessas medulas.

Esse limiar foi então utilizado como ponto de corte para definir dois grupos dentre os pacientes sem DRM (DRM indetectável), aqui denominados grupos de alta e baixa diversidade. Em conformidade à ideia de que a menor diversidade seria um fator prognóstico negativo, esse grupo apresentou uma maior proporção de pacientes classificados como de alto risco ($p = 0.00052$) pelo sistema de classificação do *National Cancer Institute* (NCI, **Figura 2B**). Essa classificação de risco é baseada em 2 variáveis clínicas: a idade do paciente ao ato do diagnóstico (onde pacientes de alto risco são aqueles com menos de 1 ano ou mais de 10 anos) e a sua contagem de leucócitos no sangue periférico (valores acima de 50.000 leucócitos por milímetro cúbico indicam alto risco). Desse modo, avaliou-se também a influência

individual de ambas essas variáveis na diversidade da medula (**Figura 2C**). Nesse contexto, houve uma diminuição significativa de diversidade em pacientes com alta leucometria ($p = 0.0000037$), indicando essa variável como a principal responsável pela associação entre baixa diversidade e alto risco. Essa associação entre a baixa diversidade no final da indução e a alta leucometria ao diagnóstico, indicando talvez que a alta leucometria deixa a medula óssea em menores condições de regenerar. Outra possibilidade seria a alta leucometria ser sinal de que, nesse paciente, as células normais da medula óssea não foram capazes de competir e inibir as células leucêmicas, sendo também mais vagarosas em regenerar-se após tratamento de indução.

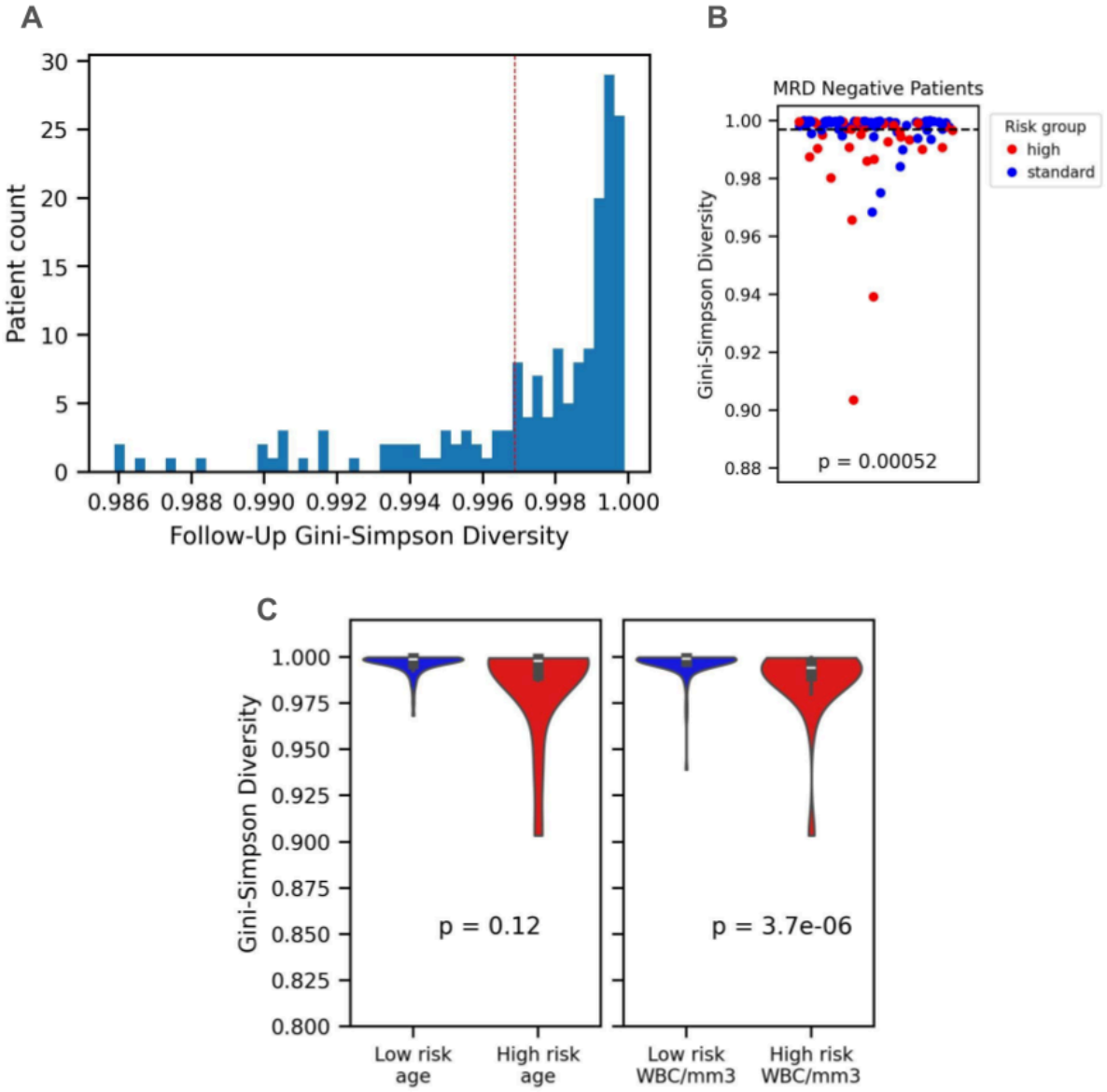

Figura 2. Análise da diversidade da medula óssea após a terapia de indução em pacientes de LLA-B com DRM indetectável. (**A**) Distribuição dos índices de diversidade para as medulas de todos os pacientes analisados no estudo. *Outliers* não representados para melhor visualização da faixa de diversidade que concentra a maioria dos pacientes. Limiar de diversidade selecionado para análises posteriores representado pela linha vermelha tracejada. (**B**) Índices de diversidade em pacientes com DRM indetectável em função do seu grupo de risco NCI. A linha preta tracejada indica o limiar de diversidade que separa os pacientes em dois grupos. P-valor referente a um teste exato de Fisher. (**C**) Índices de diversidade em pacientes com DRM indetectável em função de grupos de risco baseados em idade e leucometria ao diagnóstico. P-valores referentes a teste de Mann-Whitney U.

Em seguida, analisou-se a sobrevida dos pacientes agrupados em função da diversidade medular após a terapia de indução. Foram avaliadas: a Sobrevida Livre de Leucemia (SLL), onde o evento em análise é a recaída da LLA após remissão clínica; a Sobrevida Global (SG), onde o evento é o óbito do paciente; e, por fim, a Sobrevida Livre de Eventos (SLE), na qual o evento pode ser tanto uma recaída quanto o óbito. Essas análises foram realizadas tanto para o conjunto total de pacientes desta tese (**Figura 3A**), quanto para apenas aqueles com DRM negativa (**Figura 3B**).

Pacientes com baixa diversidade medular ao final da indução apresentaram prognósticos menos favoráveis de modo geral, enquanto para o universo amostral de pacientes DRM negativa, apenas pioras na SG e SLE foram observadas ($p = 0.0054$ e 0.033, respectivamente). A discordância entre SG e SLL pode ser interpretada como óbitos decorrentes de outros eventos que não a doença em si, como toxicidade associada ao tratamento ou infecções (MEDEIROS, 2018). Uma pior regeneração da medula pode indicar maior toxicidade do tratamento e uma menor capacidade de combater infecções, ambas situações que podem ameaçar a vida dos pacientes e que não necessariamente precisa estar vinculada à recaída da doença para merecer atenção. Futuras análises com maior número de casos poderão avaliar melhor a causa do maior número de óbitos em pacientes sem DRM e com baixa diversidade da medula. É relevante mencionar que um estudo por Kotrova *et al.* de 2015 observou maiores taxas de recaída em pacientes com medulas menos diversas. No entanto, esse achado se deu em um grupo de pacientes diferente (DRM $< 10^{-4}$) e em outro ponto do tratamento (dia 78) em relação às análises aqui realizadas.

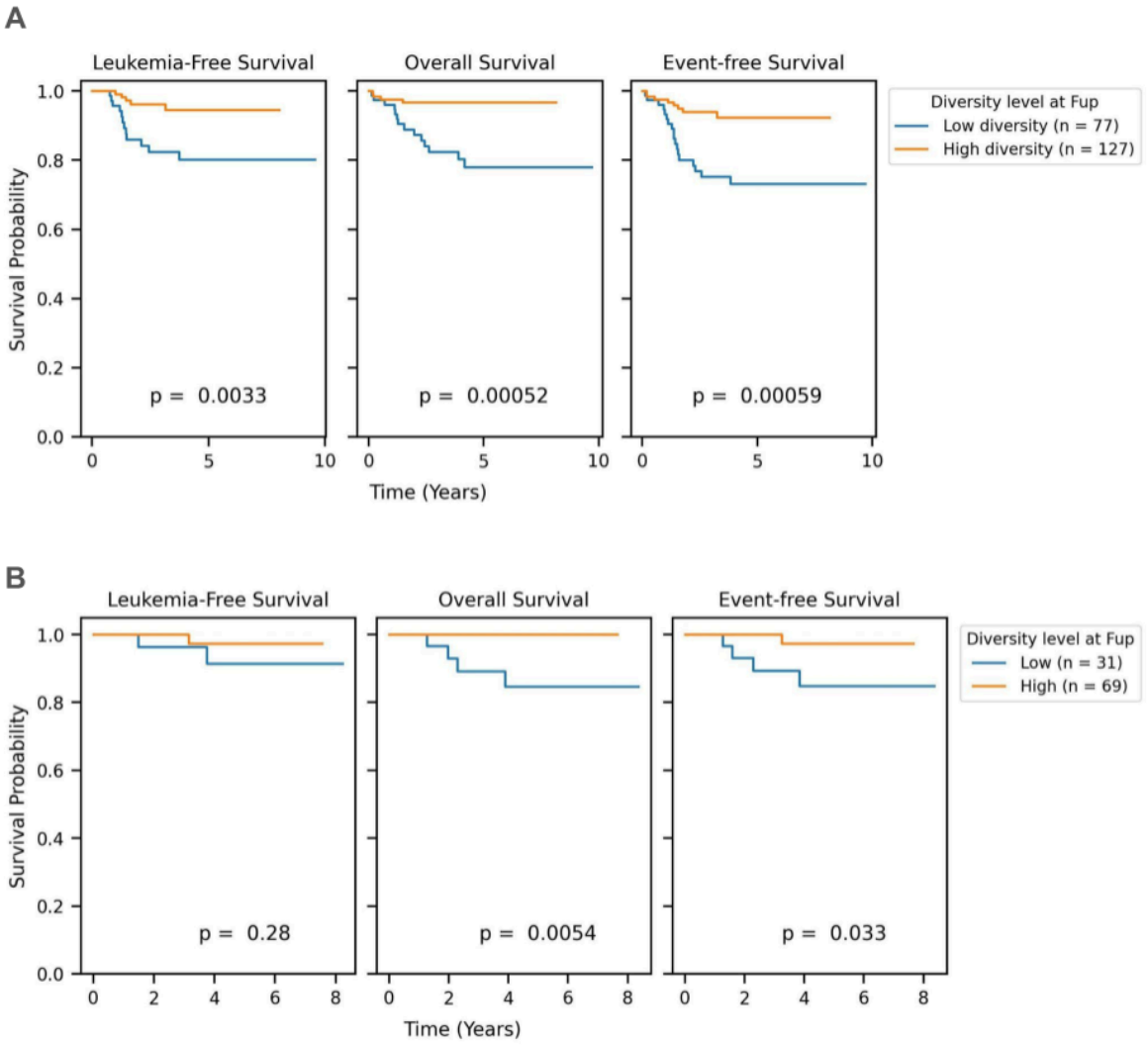

Figura 3. Análise de sobrevida de pacientes pediátricos de LLA-B em função da sua diversidade da medula óssea após a terapia de indução. (**A**) Conjunto total de pacientes analisados. (**B**) Apenas pacientes com DRM indetectável. P-valores relativos a teste de log-rank.

7. CONCLUSÕES

Este trabalho buscou explorar o repertório de *IGH*, representativo de linfócitos B, em pacientes pediátricos com LLA-B, com um enfoque na caracterização dos clonótipos referentes a células de leucemia e da diversidade de clonótipos na medula óssea. Para tal, foram sequenciadas amostras pareadas D0 e Fup de 204 pacientes, cujos dados gerados foram processados para análise via uma *pipeline* desenvolvida nesse trabalho. Abaixo, encontram-se listados os principais achados desta tese:

1. Clonótipos dos subtipos de LLA-B *KMT2A* rearranjado e *B-other* apresentaram enriquecimento do segmento VH topologicamente mais proximal à região constante da cadeia *IGH* (*IGHV6-1*). Já o grupo *TCF3::PBX1*, apresentou enriquecimento do segmento *IGHV3-23*.
2. Clonótipos referentes ao subtipo *ETV6::RUNX1* apresentaram uma baixa utilização dos segmentos VH que se encontram topologicamente localizados entre os segmentos *IGHV7-34-1* e *IGHV1-58*.
3. Clonótipos referentes à LLA-B de modo geral demonstraram possuir rearranjos que tendem a ser improdutivos, seja pela presença de códons de parada ou por mudanças na janela de leitura, com exceção do subtipo *TCF3::PBX1*.
4. Clonótipos de LLA-B apresentaram encurtamento nas regiões CDR3 (*ETV6::RUNX1*, hiperdiploidia e *KMT2A* rearranjado) e encurtamento de inserções N na interface DJ (*ETV6::RUNX1*, hiperdiploidia e *B-other*).
5. Diversas características diferenciais foram encontradas quando comparados os clonótipos de alta frequência provenientes de medulas ósseas de doadores pediátricos em relação àqueles que foram amostrados aleatoriamente dessas mesmas amostras. Observou-se utilização diferencial de diversos segmentos VH e de quase todos os segmentos JH (excetuando *IGHJ1*); encurtamento dos segmentos CDR3 e de inserções N; aumento do conteúdo de GC; e maior frequência de clonótipos produtivos.
6. Clonótipos de LLA-B que se mantiveram detectáveis após o término da terapia de indução apresentaram menores níveis de rearranjos produtivos em relação àqueles que foram aparentemente eliminados pelo tratamento. Esse efeito foi observado tanto para o panorama geral da LLA-B, quanto para especificamente os grupos *ETV6::RUNX1* e

B-other. Esse efeito mostrou-se mais intenso quanto maior a ordem de grandeza da DRM associada ao clonótipo analisado, excetuando clonótipos com DRM $\geq 10^{-3}$.

7. O repertório IGH na medula do diagnóstico apresentou menor diversidade nos subtipos genéticos *ETV6::RUNX1* e *TCF3-PBX1*, e maior diversidade em pacientes com LLA hiperdiploide ou *B-other*.
8. Observou-se associação entre alta leucometria ao diagnóstico e menor diversidade do repertório de IGH na medula do final da indução de pacientes com DRM negativa.
9. Pacientes com medulas menos diversas após a terapia de indução apresentaram pior prognóstico. Esse efeito foi observado tanto quando analisando todos os 204 pacientes incluídos nesse estudo (SLL, SG e SLE), quanto apenas os pacientes com DRM negativa (apenas SG e SLE).

Esse conjunto de observações aprofunda o conhecimento atual acerca do processo de rearranjo V(D)J em células de LLA-B em pacientes pediátricos, e também evidenciam associações entre o estado geral dos seus repertórios de *IGH* e a evolução clínica da doença. Alguns desses achados também contribuem para questionamentos a serem explorados em trabalhos futuros. Dentre esses, podem ser destacados:

1. Uma vez que a identificação de clonótipos de LLA-B é baseada na porcentagem do repertório total que esses clonótipos ocupam, a identificação precoce de clonótipos leucêmicos de baixa frequência é problemática, sendo um dos grandes desafios no acompanhamento do quadro clínico dos pacientes através da DRM. Apesar desta tese ter identificados características biológicas que tendem a diferir entre clonótipos leucêmicos subtipo-específicos e clonótipos de medulas normais, essas distinções são sutis e, portanto, não parecem boas candidatas para contribuir com futuras estratégias para a discriminação dos clonótipos leucemicos. Estudos de *single cell* NGS, por outro lado, podem solucionar o problema, ao detectar na mesma célula um marcador genético da leucemia e um clonótipo qualquer.
2. A associação entre a produtividade da cadeia *IGH* de clonótipos de LLA-B e o seu prognóstico clínico pede uma investigação mais profunda. Enquanto este trabalho demonstrou que clonótipos resistentes à terapia de indução tendem a ser improdutivos, trabalhos anteriores demonstraram melhor prognóstico clínico para pacientes com clonótipos dominantes improdutivos ao diagnóstico (LI *et al*., 2004; KATSIBARDI *et*

al., 2011). Os resultados aqui apresentados ajudam a elucidar essa aparente contradição ao demonstrar que pacientes com altos níveis de DRM ($\geq 10^{-3}$) de fato apresentam maior frequência de sequências *IGH* produtivas. Ainda assim, uma melhor compreensão da dinâmica entre produtividade *IGH* de clonótipos de LLA-B e a resposta clínica do paciente se faz necessária para que esse conhecimento possa ser de fato traduzido em benefícios no monitoramento clínico desses casos.

3. O alto número de diferenças entre os clonótipos mais dominantes e os clonótipos amostrados aleatoriamente de medulas ósseas de doadores pediátricos saudáveis chama a atenção, uma vez que o grau de divergência mostrou-se maior inclusive do que em comparação a clonótipos leucêmicos. Desse modo, há a necessidade de análises exploratórias comparativas que elucidem se essa discrepância reflete um fenômeno biológico ou vieses associados à amplificação de rearranjos V(D)J via NGS.
4. Por fim, existe uma clara necessidade da expansão das análises aqui apresentadas que associam repertórios de linfócitos B menos diversos ao final da indução a um pior prognóstico clínico em pacientes com DRM negativa. Esses pacientes tendem a ser bons respondedores e, portanto, eventos de recaída ou óbito nesse grupo são raros. Assim sendo, a confirmação das associações aqui encontradas em *cohorts* com maior número de pacientes, bem como em outros métodos para detecção de DRM com diferentes níveis de sensibilidade, seria de grande valia para viabilizar a utilização dessa informação no monitoramento clínico dos pacientes.

8. REFERÊNCIAS

ABBAS, A. K.; LICHTMAN, A. H. **Cellular and molecular immunology**. Philadelphia, Penns.: Saunders, 2005.

AJROUCHE, R. et al. Childhood acute lymphoblastic leukaemia and indicators of early immune stimulation: the Estelle study (SFCE). **British Journal of Cancer**, v. 112, n. 6, p. 1017–1026, 12 fev. 2015.

AKIRA, S.; OKAZAKI, K.; SAKANO, H. Two pairs of recombination signals are sufficient to cause immunoglobulin V-(D)-J joining. **Science**, v. 238, n. 4830, p. 1134–1138, 20 nov. 1987.

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of Molecular Biology**, v. 215, n. 3, p. 403–410, out. 1990.

BOROWITZ, M. J. et al. Prognostic significance of minimal residual disease in high risk B-ALL: a report from Children's Oncology Group study AALL0232. **Blood**, v. 126, n. 8, p. 964–971, 20 ago. 2015.

CANCER RESEARCH UK. **Phases of treatment | Acute lymphoblastic leukaemia (ALL) | Cancer Research UK**. Disponível em: <https://www.cancerresearchuk.org/about-cancer/acute-lymphoblastic-leukaemia-all/treatment/phases>.

CARRA, G. **Simplistic overview of V(D)J recombination of immunoglobulin heavy chains**. **Wikimedia Commons**, 6 jun. 2008. Disponível em: <https://en.wikipedia.org/wiki/V(D)J_recombination#/media/File:VDJ_recombination.png>. Acesso em: 4 nov. 2024

CHENG, S. et al. Simple deep sequencing-based post-remission MRD surveillance predicts clinical relapse in B-ALL. **Journal of Hematology & Oncology**, v. 11, n. 1, 22 ago. 2018.

CHRISTIE, S. M.; FIJEN, C.; ROTHENBERG, E. V(D)J Recombination: Recent Insights in Formation of the Recombinase Complex and Recruitment of DNA Repair Machinery. **Frontiers in Cell and Developmental Biology**, v. 10, 29 abr. 2022.

DARZENTAS, F. et al. Insights into IGH clonal evolution in BCP-ALL: frequency, mechanisms, associations, and diagnostic implications. **Frontiers in Immunology**, v. 14, 18 abr. 2023.

DAVIDSON-PILON, C. lifelines: survival analysis in Python. **Journal of Open Source Software**, v. 4, n. 40, p. 1317, 4 ago. 2019.

DESIDERIO, S. V. et al. Insertion of N regions into heavy-chain genes is correlated with expression of terminal deoxytransferase in B cells. **Nature**, v. 311, n. 5988, p. 752–755, 1 out. 1984.

DUEZ, M. et al. Vidjil: A Web Platform for Analysis of High-Throughput Repertoire Sequencing. **PLOS ONE**, v. 11, n. 11, p. e0166126–e0166126, 11 nov. 2016.

ESWARAN, J. et al. The pre-B-cell receptor checkpoint in acute lymphoblastic leukaemia. **Leukemia**, v. 29, n. 8, p. 1623–1631, 6 maio 2015.

FAHAM, M. et al. Deep-sequencing approach for minimal residual disease detection in acute lymphoblastic leukemia. **Blood**, v. 120, n. 26, p. 5173–5180, 20 dez. 2012.

FEENEY, A. J.; GOEBEL, P.; ESPINOZA, C. R. Many levels of control of V gene rearrangement frequency. **Immunological Reviews**, v. 200, n. 1, p. 44–56, ago. 2004.

GENG, H. et al. Self-Enforcing Feedback Activation between BCL6 and Pre-B Cell Receptor Signaling Defines a Distinct Subtype of Acute Lymphoblastic Leukemia. **Cancer Cell**, v. 27, n. 3, p. 409–425, 9 mar. 2015.

GILHAM, C. et al. Day care in infancy and risk of childhood acute lymphoblastic leukaemia: findings from UK case-control study. **BMJ**, v. 330, n. 7503, p. 1294, 22 abr. 2005.

GILLOOLY, J. F.; HEIN, A.; DAMIANI, R. Nuclear DNA Content Varies with Cell Size across Human Cell Types. **Cold Spring Harbor Perspectives in Biology**, v. 7, n. 7, p. a019091, jul. 2015.

GIRAUD, M. et al. Fast multiclonal clusterization of V(D)J recombinations from high-throughput sequencing. **BMC Genomics**, v. 15, n. 1, p. 409–409, 1 jan. 2014.

GIUSTI, G. N. N. et al. Test trial of spike‐in immunoglobulin heavy‐chain (*IGH*) controls for next generation sequencing quantification of minimal residual disease in acute lymphoblastic leukaemia. **British Journal of Haematology**, v. 189, n. 4, 18 mar. 2020.

GREAVES, M. A causal mechanism for childhood acute lymphoblastic leukaemia. **Nature Reviews Cancer**, v. 18, n. 8, p. 471–484, 21 maio 2018.

HANSEN-HAGGE, T. E.; YOKOTA, S.; BARTRAM, C. R. Detection of minimal residual disease in acute lymphoblastic leukemia by in vitro amplification of rearranged T-cell receptor delta chain sequences. **Blood**, v. 74, n. 5, p. 1762–1767, 1 out. 1989a.

HANSEN-HAGGE, T. E.; YOKOTA, S.; BARTRAM, C. R. Detection of minimal residual disease in acute lymphoblastic leukemia by in vitro amplification of rearranged T-cell receptor delta chain sequences. **Blood**, v. 74, n. 5, p. 1762–1767, 1 out. 1989b.

HEIKAMP, E. B.; PUI, C.-H. Next-Generation Evaluation and Treatment of Pediatric Acute Lymphoblastic Leukemia. **The Journal of Pediatrics**, v. 203, p. 14-24.e2, 1 dez. 2018.

HEIN, D.; BORKHARDT, A.; FISCHER, U. Insights into the prenatal origin of childhood acute lymphoblastic leukemia. **Cancer and Metastasis Reviews**, v. 39, n. 1, p. 161–171, 4 jan. 2020.

HILL, L. et al. Wapl repression by Pax5 promotes V gene recombination by Igh loop extrusion. **Nature**, v. 584, n. 7819, p. 142–147, 1 jul. 2020.

IACOBUCCI, I.; MULLIGHAN, C. G. Genetic Basis of Acute Lymphoblastic Leukemia. **Journal of clinical oncology : official journal of the American Society of Clinical Oncology**, v. 35, n. 9, p. 975–983, 2017.

INABA, H.; GREAVES, M.; MULLIGHAN, C. G. Acute lymphoblastic leukaemia. **The Lancet**, v. 381, n. 9881, p. 1943–1955, jun. 2013.

INABA, H.; MULLIGHAN, C. G. Pediatric acute lymphoblastic leukemia. **Haematologica**, v. 105, n. 11, 10 set. 2020.

INSTITUTO NACIONAL DO CÂNCER. **Estimativa | 2023 Incidência de Câncer no Brasil**. Rio de Janeiro: Instituto Nacional de Câncer, 2022.

JAKOBCZYK, H. et al. ETV6-RUNX1 and RUNX1 directly regulate RAG1 expression: one more step in the understanding of childhood B-cell acute lymphoblastic leukemia leukemogenesis. **Leukemia**, v. 36, n. 2, p. 549–554, 1 fev. 2022.

JI, Y. et al. The In Vivo Pattern of Binding of RAG1 and RAG2 to Antigen Receptor Loci. **Cell**, v. 143, n. 1, p. 170–170, 1 out. 2010.

JOST, L. Entropy and diversity. **Oikos**, v. 113, n. 2, p. 363–375, maio 2006.

KAMPER-JØRGENSEN, M. et al. Childcare in the first 2 years of life reduces the risk of childhood acute lymphoblastic leukemia. **Leukemia**, v. 22, n. 1, p. 189–193, 9 ago. 2007.

KATSIBARDI, K. et al. Clinical significance of productive immunoglobulin heavy chain gene rearrangements in childhood acute lymphoblastic leukemia. **Leukemia & Lymphoma**, v. 52, n. 9, p. 1751–1757, 8 jun. 2011.

KNECHT, H. et al. Quality control and quantification in IG/TR next-generation sequencing marker identification: protocols and bioinformatic functionalities by EuroClonality-NGS. **Leukemia**, v. 33, n. 9, p. 2254–2265, 21 jun. 2019.

KOTROVA, M. et al. The predictive strength of next-generation sequencing MRD detection for relapse compared with current methods in childhood ALL. **Blood**, v. 126, n. 8, p. 1045–1047, 20 ago. 2015.

LADETTO, M. et al. Next-generation sequencing and real-time quantitative PCR for minimal residual disease detection in B-cell disorders. **Leukemia**, v. 28, n. 6, p. 1299–1307, 17 dez. 2013.

LI, A. et al. Utilization of Ig heavy chain variable, diversity, and joining gene segments in children with B-lineage acute lymphoblastic leukemia: implications for the mechanisms of VDJ recombination and for pathogenesis. **Blood**, v. 103, n. 12, p. 4602–4609, 15 jun. 2004.

MA, X. et al. Daycare attendance and risk of childhood acute lymphoblastic leukaemia. **British Journal of Cancer**, v. 86, n. 9, p. 1419–1424, maio 2002.

MALARD, F.; MOHTY, M. Acute lymphoblastic leukaemia. **The Lancet**, v. 395, n. 10230, p. 1146–1162, 4 abr. 2020.

MARTÍN-LORENZO, A. et al. Infection Exposure Is a Causal Factor in B-cell Precursor Acute Lymphoblastic Leukemia as a Result of *Pax5*-Inherited Susceptibility. **Cancer Discovery**, v. 5, n. 12, p. 1328–1343, 25 set. 2015.

MARTINS, V. et al. Cell competition is a tumour suppressor mechanism in the thymus. **Nature**, v. 509, n. 7501, p. 465–470, 22 maio 2014.

MEDEIROS, B. C. Interpretation of clinical endpoints in trials of acute myeloid leukemia. **Leukemia Research**, v. 68, p. 32–39, maio 2018.

MOTEA, E. A.; BERDIS, A. J. Terminal Deoxynucleotidyl Transferase: The Story of a Misguided DNA Polymerase. **Biochimica et biophysica acta**, v. 1804, n. 5, p. 1151–1166, 1 maio 2010.

MÜLLER, H. et al. Proximally biased V(D)J recombination in the clonal evolution of IGH alleles in KMT2A::AFF1 BCP‐ALL of all age classes. **HemaSphere**, v. 8, n. 4, 1 abr. 2024.

MULLIGHAN, C. G. et al. Genomic Analysis of the Clonal Origins of Relapsed Acute Lymphoblastic Leukemia. **Science**, v. 322, n. 5906, p. 1377–1380, 28 nov. 2008.

MULLIGHAN, C. G. The molecular genetic makeup of acute lymphoblastic leukemia. **Hematology**, v. 2012, n. 1, p. 389–396, 8 dez. 2012.

MURPHY, K.; WEAVER, C. **Janeway's immunobiology**. 9th. ed. New York, NY, USA: Garland Science, Taylor & Francis Group, LLC, 2017.

NGUYEN, K. et al. Factors influencing survival after relapse from acute lymphoblastic leukemia: a Children's Oncology Group study. **Leukemia**, v. 22, n. 12, p. 2142–2150, 25 set. 2008.

OETTINGER, M. et al. RAG-1 and RAG-2, adjacent genes that synergistically activate V(D)J recombination. **Science**, v. 248, n. 4962, p. 1517–1523, 22 jun. 1990.

POTTER, M. N. The detection of minimal residual disease in acute lymphoblastic leukaemia. **Blood Reviews**, v. 6, n. 2, p. 68–82, jun. 1992.

PUI, C.-H. et al. Improved Prognosis for Older Adolescents With Acute Lymphoblastic Leukemia. **Journal of Clinical Oncology**, v. 29, n. 4, p. 386–391, 1 fev. 2011.

RAETZ, E. A.; TEACHEY, D. T. T-cell acute lymphoblastic leukemia. **Hematology**, v. 2016, n. 1, p. 580–588, 2 dez. 2016.

RAMSDEN, D. A.; BAETZ, K.; WU, G. E. Conservation of sequence in recombination signal sequence spacers. **Nucleic Acids Research**, v. 22, n. 10, p. 1785–1796, 1994.

ROBERTS, K. G. Genetics and prognosis of ALL in children vs adults. **Hematology**, v. 2018, n. 1, p. 137–145, 30 nov. 2018.

ROGNES, T. et al. VSEARCH: a versatile open source tool for metagenomics. **PeerJ**, v. 4, p. e2584, 18 out. 2016.

RUDANT, J. et al. Childhood Acute Lymphoblastic Leukemia and Indicators of Early Immune Stimulation: A Childhood Leukemia International Consortium Study. **American Journal of Epidemiology**, v. 181, n. 8, p. 549–562, 1 mar. 2015.

SASAKI, K. et al. Acute lymphoblastic leukemia: A population‐based study of outcome in the United States based on the surveillance, epidemiology, and end results (SEER) database, 1980–2017. **American Journal of Hematology**, v. 96, n. 6, p. 650–658, abr. 2021.

SCHATZ, D. G.; SWANSON, P. C. V(D)J Recombination: Mechanisms of Initiation. **Annual Review of Genetics**, v. 45, n. 1, p. 167–202, 15 dez. 2011.

SCHRAPPE, M. et al. Pediatric Patients with High-Risk B-Cell ALL in First Complete Remission May Benefit from Less Toxic Immunotherapy with Blinatumomab - Results from Randomized Controlled Phase 3 Trial AIEOP-BFM ALL 2017. **Blood**, v. 142, n. Supplement 1, p. 825–825, 2 nov. 2023.

SEKIYA, Y. et al. Clinical utility of next-generation sequencing-based minimal residual disease in paediatric B-cell acute lymphoblastic leukaemia. **British Journal of Haematology**, v. 176, n. 2, p. 248–257, 1 jan. 2017.

SHIN, S. et al. Detection of Immunoglobulin Heavy Chain Gene Clonality by Next-Generation Sequencing for Minimal Residual Disease Monitoring in B-Lymphoblastic Leukemia. **Annals of Laboratory Medicine**, v. 37, n. 4, p. 331–335, 1 jul. 2017.

SIEGEL, R. L. et al. Cancer statistics, 2023. **CA: A Cancer Journal for Clinicians**, v. 73, n. 1, p. 17–48, 12 jan. 2023.

SOFOU, E. et al. Clonotype definitions for immunogenetic studies: proposals from the EuroClonality NGS Working Group. **Leukemia**, v. 37, n. 8, p. 1750–1752, 30 jun. 2023.

SUN, W. et al. Outcome of children with multiply relapsed B-cell acute lymphoblastic leukemia: a therapeutic advances in childhood leukemia & lymphoma study. **Leukemia**, v. 32, n. 11, p. 2316–2325, 15 mar. 2018.

TALLEN, G. et al. Long-Term Outcome in Children With Relapsed Acute Lymphoblastic Leukemia After Time-Point and Site-of-Relapse Stratification and Intensified Short-Course Multidrug Chemotherapy: Results of Trial ALL-REZ BFM 90. **Journal of Clinical Oncology**, v. 28, n. 14, p. 2339–2347, 10 maio 2010.

THEUNISSEN, P. M. J. et al. Next-generation antigen receptor sequencing of paired diagnosis and relapse samples of B-cell acute lymphoblastic leukemia: Clonal evolution and implications for minimal residual disease target selection. **Leukemia Research**, v. 76, p. 98–104, jan. 2019.

TONEGAWA, S. Somatic generation of antibody diversity. **Nature**, v. 302, n. 5909, p. 575–581, abr. 1983.

VAN DER VELDEN, V. H. J. et al. Age-related patterns of immunoglobulin and T-cell receptor gene rearrangements in precursor-B-ALL: implications for detection of minimal residual disease. **Leukemia**, v. 17, n. 9, p. 1834–1844, set. 2003.

VAN DONGEN, J. J. M. et al. Design and standardization of PCR primers and protocols for detection of clonal immunoglobulin and T-cell receptor gene recombinations in suspect lymphoproliferations: Report of the BIOMED-2 Concerted Action BMH4-CT98-3936. **Leukemia**, v. 17, n. 12, p. 2257–2317, dez. 2003.

VAN DONGEN, J. J. M. et al. Minimal residual disease diagnostics in acute lymphoblastic leukemia: need for sensitive, fast, and standardized technologies. **Blood**, v. 125, n. 26, p. 3996–4009, 25 jun. 2015.

VELTEN, L. et al. Identification of leukemic and pre-leukemic stem cells by clonal tracking from single-cell transcriptomics. **Nature Communications**, v. 12, n. 1, 1 mar. 2021.

VIRTANEN, P. et al. SciPy 1.0: fundamental algorithms for scientific computing in Python. **Nature Methods**, v. 17, n. 3, p. 261–272, 3 fev. 2020.

WESTBURY, M. V. **Unraveling evolution through Next Generation Sequencing**. Thesis—Universität Potsdam: [s.n.].

WROCHNA, M. **Schematic structure of an antibody. Wikimedia Commons**, 3 set. 2020. Disponível em: <https://commons.wikimedia.org/wiki/File:Antibody_basic_unit.svg>. Acesso em: 4 nov. 2024

9. ANEXOS

9.1 Parecer do Comitê de Ética em Pesquisa

CENTRO INFANTIL DE INVESTIGAÇÕES HEMATOLÓGICAS DR.

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Seqüenciamento de alto desempenho para quantificação da Doença Residual Mínima em Leucemia Linfóide Aguda

Pesquisador: Jose Andres Yunes

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

Versão: 1

CAAE: 57280616.4.0000.5376

Instituição Proponente: Centro Infantil de Investigações Hematológicas Dr.Domingos A Boldrini

Patrocinador Principal: Ministério da Saúde

DADOS DA NOTIFICAÇÃO

Tipo de Notificação: Outros

Detalhe: Inclusão de nome de pesquisador no projeto

Justificativa: Solicito a inclusão do mestrando Guilherme Navarro Nilo Giusti no projeto de

Data do Envio: 20/06/2018

Situação da Notificação: Parecer Consubstanciado Emitido

DADOS DO PARECER

Número do Parecer: 2.786.804

Apresentação da Notificação:

Adequada.

Objetivo da Notificação:

Inclusão de um novo pesquisador no projeto de pesquisa.

Avaliação dos Riscos e Benefícios:

Não há risco adicionais previstos para os participantes de pesquisa, uma vez que o material que será utilizado na pesquisa será colido para os exames de rotina.

Não haverá benefício direto ao participante de pesquisa por se tratar de um projeto comparativo entre duas técnicas, comparar resultados de DRM por sequenciamento versus RQ- CR do dia 35.

Endereço: Rua Dr. Gabriel Porto, 1270 Cidade Universitária
Bairro: Barão Geraldo **CEP:** 13.083-210
UF: SP **Município:** CAMPINAS
Telefone: (19)3787-5001 **Fax:** (19)3289-3571 **E-mail:** cep@boldrini.org.br

CENTRO INFANTIL DE INVESTIGAÇÕES HEMATOLÓGICAS DR.

Comentários e Considerações sobre a Notificação:

N/A

Considerações sobre os Termos de apresentação obrigatória:

Faltou inserir na Plataforma Brasil o CV do pesquisador Guilherme Navarro Nilo Giusti.

Recomendações:

Adicionar na Plataforma Brasil o CV de Guilherme Navarro Nilo Giusti, podendo ser CV Lattes ou outro modelo. Se for de outro modelo, o mesmo deve ser rubricado em todas as folhas, datado e assinado na última folha.

Apresentar o relatório do projeto do pesquisador Guilherme Navarro Nilo Giusti juntamente com o próximo relatório parcial do projeto geral intitulado "Seqüenciamento de alto desempenho para quantificação da Doença Residual Mínima em Leucemia Linfóide Aguda - Projeto de Pesquisa PRONON – Ministerio da Saúde".

Para um melhor acompanhamento do projeto e evitar trabalho desnecessário, solicitamos que o pesquisador envie via secretaria do CEP na mesma data do envio do relatório parcial os relatórios enviados a FAPESP ou outras instituições envolvidas no projeto.

Conclusões ou Pendências e Lista de Inadequações:

Notificação aprovada.

Considerações Finais a critério do CEP:

Envio do próximo relatório até setembro de 2018.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Outros	ProjetoMestradoGuilhermeNavarro.pdf	20/06/2018 17:45:58	Jose Andres Yunes	Postado
Outros	DespachoFAPESPprojetoGuilherme260218.pdf	20/06/2018 17:46:03	Jose Andres Yunes	Postado

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

Endereço: Rua Dr. Gabriel Porto, 1270 Cidade Universitária
Bairro: Barão Geraldo **CEP:** 13.083-210
UF: SP **Município:** CAMPINAS
Telefone: (19)3787-5001 **Fax:** (19)3289-3571 **E-mail:** cep@boldrini.org.br

CENTRO INFANTIL DE INVESTIGAÇÕES HEMATOLÓGICAS DR.

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: Seqüenciamento de alto desempenho para quantificação da Doença Residual Mínima em Leucemia Linfóide Aguda

Pesquisador: Jose Andres Yunes

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

Versão: 1

CAAE: 57280616.4.0000.5376

Instituição Proponente: Centro Infantil de Investigações Hematológicas Dr.Domingos A Boldrini

Patrocinador Principal: Ministério da Saúde

DADOS DA NOTIFICAÇÃO

Tipo de Notificação: Outros

Detalhe: Inclusão de nome de pesquisador no projeto

Justificativa: Solicito a inclusão do mestrando Guilherme Navarro Nilo Giusti no projeto de

Data do Envio: 20/06/2018

Situação da Notificação: Parecer Consubstanciado Emitido

DADOS DO PARECER

Número do Parecer: 2.786.804

Apresentação da Notificação:

Adequada.

Objetivo da Notificação:

Inclusão de um novo pesquisador no projeto de pesquisa.

Avaliação dos Riscos e Benefícios:

Não há risco adicionais previstos para os participantes de pesquisa, uma vez que o material que será utilizado na pesquisa será colido para os exames de rotina.

Não haverá benefício direto ao participante de pesquisa por se tratar de um projeto comparativo entre duas técnicas, comparar resultados de DRM por sequenciamento versus RQ- CR do dia 35.

Endereço: Rua Dr. Gabriel Porto, 1270 Cidade Universitária
Bairro: Barão Geraldo **CEP:** 13.083-210
UF: SP **Município:** CAMPINAS
Telefone: (19)3787-5001 **Fax:** (19)3289-3571 **E-mail:** cep@boldrini.org.br

CENTRO INFANTIL DE INVESTIGAÇÕES HEMATOLÓGICAS DR.DOMINGOS A BOLDRINI

PARECER CONSUBSTANCIADO DO CEP

DADOS DA EMENDA

Título da Pesquisa: Seqüenciamento de alto desempenho para quantificação da Doença Residual Mínima em Leucemia Linfóide Aguda

Pesquisador: Jose Andres Yunes

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

Versão: 4

CAAE: 57280616.4.0000.5376

Instituição Proponente: Centro Infantil de Investigações Hematológicas Dr.Domingos A Boldrini

Patrocinador Principal: Ministério da Saúde

DADOS DO PARECER

Número do Parecer: 4.089.595

Apresentação do Projeto:

O projeto inical de pesquisa se dispõe a estabelecer um novo método de detecção de Doença Residual Mínima,isto é, detectar por meio de técnicas de Biologia Molecular, pequenas quantidades de células malignas residuais que podem ser detectadas no organismo do paciente em tratamento quimioterápico. A detecção da DRM é extremamente importante, pois permite saber como é que o paciente está respondendo ao tratamento, podendo desta maneira determinar uma outra abordagem terapêutica, passando assim a ter melhores chances de sucesso no tratamento. Por outro lado, os pacientes que
apresentam excelente resposta terapêutica nas primeiras semanas da terapia, podem passam a receber uma quimioterapia menos intensiva, sem comprometer os índices de cura e com diminuição dos efeitos colaterais da quimioterapia. O exame padrão de análise da DRM, implantado desde 2010 pelo Centro Boldrini, demanda muita experiência técnica em biologia molecular, é demorado e oneroso, razões que tem dificultado sua implantação em outras instituições ou centralização em abrangência nacional. Apenas 2% (aproximadamente 60) das crianças diagnosticadas anualmente com leucemia linfóide aguda no Brasil recebem tratamento ajustado pela DRM. Este projeto PRONON pretende padronizar um método novo de detecção da DRM, baseado nos últimos equipamentos e técnicas da genômica, o assim chamado

Endereço: Rua Dr. Gabriel Porto, 1270 Cidade Universitária
Bairro: Barão Geraldo **CEP:** 13.083-210
UF: SP **Município:** CAMPINAS
Telefone: (19)3787-5001 **Fax:** (19)3289-3571 **E-mail:** cep@boldrini.org.br

sequenciamento de DNA de última geração" ou "Next Generation Sequencing (NGS)". Este novo método de análise da DRM tem vantagens em termos de sensibilidade, custo, escalabilidade e possibilidade de automação. A expectativa é implementar um método que permita oferecer este valioso exame laboratorial a todas as crianças brasileiras acometidas pela leucemia.

A nprimeira emenda ao protocolo inicial apresentada aumenta o número para 115 pacientes, incluido os pacientes do estudo AIEOP-BFM ALL 2009 que substituiu o GBTI LLA-2009 a partir do momento em que a Lasparaginase começou a ser adquirida na formulação de PEG-Asparaginase. O Centro Boldrini tem usado esse protocolo para simples tratamento dos pacientes, que assinam o TCLE correspondente.

Esta segunda emenda não altera os objetivos do projeto original. Haverá aumento do número de pacientes (total de 300 pacientes, incluindo LLA-B ou LLA-T) analisados, desta vez com mais marcadores de Doença Residual (rearranjos IGH, IGK, IGL, TRG, TRD, TRA), e os dados obtidos serão também analisados para caracterizar a diversidade do repertorio de células B/T. Como controle, serão analisadas amostras retrospectivas de medula normal (n=20 doadores). Para realizar estas análises, houve extensão do cronograma por mais 3 anos.

Objetivo da Pesquisa:

Objetivo Primário:

Padronizar método de análise da Doença Residual Mínima por sequenciamento massivo (DRM-seq), comparando resultados com os obtidos pelo método de PCR quantitativo.

Objetivo Secundário:

1. Realizar testes de sensibilidade e linearidade de quantificação de rearranjos IgH por sequenciamento de última geração

2. Fazer analise de DRM por RQ-PCR e sequenciamento NGS, em amostras pareadas de DNA da medula óssea do dia 35, em 115casos de LLA.

3. Comparar resultados de DRM por sequenciamento versus RQ-PCR.

4. Analisar a Doença Residual Mínima e o repertório imunológico de pacientes pediátricos portadores de LLA, buscando elucidar padrões de recombinação V(D)J ou de diversidade imunológica em associação às características biológico-clínicas dos pacientes.

Avaliação dos Riscos e Benefícios:

Não há risco para os pacientes, uma vez que o material que será utilizado na pesquisa será colido para os exames de rotina.

Os benefícios possíveis são: diminuir o custo e o tempo de resposta do exame (6 dias versus 23

dias para DRM por RQ-PCR).A expectativa é de que a quantificação da DRM por NGS permitirá oferecer o exame para todas as crianças com LLA do Brasil.

Comentários e Considerações sobre a Pesquisa:

O projeto se propõe a estabelecer uma metodologia inovadora que poderá ser de altíssima importância para o país, uma vez que poderá atender todos os pacientes pediátricos com LLA, o que hoje não é possível.
Os recursos para a realização do projeto serão os obtidos no projeto PRONON, que também foi prorrogado por mais tempo.

Considerações sobre os Termos de apresentação obrigatória:

Ver Conclusões ou Pendências e Lista de Inadequações

Recomendações:

Ver Conclusões ou Pendências e Lista de Inadequações

Conclusões ou Pendências e Lista de Inadequações:

O TCLE (aprovado pelo CEP no parecer 3.776.576) deverá ser aplicado nos casos de pacientes/doadores que não assinaram TCLE para estudo da Doença Residual Mínima.

Considerações Finais a critério do CEP:

Diante do exposto, o Comitê de Ética em Pesquisa, de acordo com as atribuições definidas na Resolução CNS nº 466/2012 e na Norma Operacional CNS nº 001/2013, aguarda o próximo relatório parcial até 8 de Novembro de 2020.

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_1570006_E2.pdf	02/06/2020 20:18:08		Aceito
Outros	Emenda_projDRM_020620.docx	02/06/2020 20:12:44	Jose Andres Yunes	Aceito
Outros	resposta_CEP_271120190001.pdf	27/11/2019 16:31:54	Jose Andres Yunes	Aceito
Projeto Detalhado / Brochura Investigador	Proj_PRONON_DRMseq_140415_com_indice_FINAL_261119.docx	27/11/2019 16:31:35	Jose Andres Yunes	Aceito
Projeto Detalhado / Brochura	Proj_PRONON_DRMseq_140415_com_indice_FINAL_261119.pdf	27/11/2019 16:31:02	Jose Andres Yunes	Aceito

CENTRO INFANTIL DE INVESTIGAÇÕES HEMATOLÓGICAS DR.DOMINGOS A BOLDRINI

Continuação do Parecer: 4.089.595

Investigador	Proj_PRONON_DRMseq_140415_com_indice_FINAL_261119.pdf	27/11/2019 16:31:02	Jose Andres Yunes	Aceito
Declaração de Pesquisadores	termos_CEP_271120190001.pdf	27/11/2019 16:30:44	Jose Andres Yunes	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TA6a10aDRMseqAEIOP050419.docx	21/08/2019 15:39:40	Jose Andres Yunes	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Assentimento11a17aDRMseqAEIOP050419.docx	21/08/2019 15:34:52	Jose Andres Yunes	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLEparaDRMseq050419.docx	21/08/2019 15:34:43	Jose Andres Yunes	Aceito
Folha de Rosto	folha_de_Rosto_DRMseq_assinado.pdf	20/06/2016 16:12:35	Jose Andres Yunes	Aceito
Declaração do Patrocinador	Termo_Acordo_MinisterioSaude_e_Boldrini_proj_PRONON_DRM_NGSseq.pdf	20/06/2016 11:29:39	Jose Andres Yunes	Aceito
Declaração do Patrocinador	Aprova_Readequacao_MinisterioSaude_proj_PRONON_DRM_NGSseq.pdf	20/06/2016 11:28:58	Jose Andres Yunes	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLEs_protocolo_de_tratamento_da_LLA_GBTLI2009.pdf	20/06/2016 11:28:01	Jose Andres Yunes	Aceito
Brochura Pesquisa	Proj_PRONON_DRMseq_260814_resumo_e_sinopse.docx	20/06/2016 11:26:58	Jose Andres Yunes	Aceito

Situação do Parecer:
Aprovado

Necessita Apreciação da CONEP:
Não

CAMPINAS, 16 de Junho de 2020

Assinado por:
Maristela Amaral Palazzi
(Coordenador(a))

Endereço: Rua Dr. Gabriel Porto, 1270 Cidade Universitária
Bairro: Barão Geraldo **CEP:** 13.083-210
UF: SP **Município:** CAMPINAS
Telefone: (19)3787-5001 **Fax:** (19)3289-3571 **E-mail:** cep@boldrini.org.br

9.2 Declaração de Direitos Autorais

Declaração

As cópias de artigos de minha autoria ou de minha co-autoria, já publicados ou submetidos para publicação em revistas científicas ou anais de congressos sujeitos a arbitragem, que constam da minha Dissertação/Tese de Mestrado/Doutorado, intitulada **CARACTERIZAÇÃO DO REPERTÓRIO DE CÉLULAS B EM CRIANÇAS COM LEUCEMIA LINFOIDE AGUDA**, não infringem os dispositivos da Lei n.° 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 19/05/2025

Assinatura : _____
Nome do(a) autor(a): **Guilherme Navarro Nilo Giusti**
RG n.° 56.262.588-4

Assinatura : _____
Nome do(a) orientador(a): **José Andrés Yunes**
RG n.° 711.727.859.53