



UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE BIOLOGIA

NATACHA AZUSSA MIGITA

**Mutações genéticas na Leucemia
Linfoide Aguda pediátrica
do subgrupo '*B-other*'**

***Genetic Mutations in pediatric B-other
Acute Lymphoblastic Leukemia***

CAMPINAS - SP

2022

TESE DE DOUTORADO

**Mutações genéticas na Leucemia
Linfoide Aguda pediátrica
do subgrupo ‘B-other’**

***Genetic Mutations in pediatric B-other
Acute Lymphoblastic Leukemia***

Orientador: Dr. José Andrés Yunes

Co-orientadora: Dra. Katlin Brauer Massirer

Tese apresentada ao Instituto de
Biologia da Universidade Estadual de
Campinas como parte dos requisitos exigidos
para a obtenção do título de Doutora em
Genética e Biologia Molecular, na área de
Genética Animal e Evolução.

*Thesis presented to the Institute of
Biology of the University of Campinas in partial
fulfillment of the requirements for the degree of
Doctor, in Genetics and Molecular Biology, in
the area of Animal Genetics and Evolution.*

ESTE EXEMPLAR CORRESPONDE À VERSÃO
FINAL DA TESE DEFENDIDA PELA ALUNA
NATACHA AZUSSA MIGITA E ORIENTADA PELO
DR. JOSÉ ANDRES YUNES E CO-ORIENTADA
PELA DRA. KATLIN BRAUER MASSIRER.

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

M588m Migita, Natacha Azussa, 1988-
Mutações genéticas na leucemia linfóide aguda pediátrica do subgrupo 'B-other' / Natacha Azussa Migita. – Campinas, SP : [s.n.], 2022.

Orientador: José Andres Yunes.

Coorientador: Katlin Brauer Massirer.

Tese (doutorado) – Universidade Estadual de Campinas, Instituto de Biologia.

1. Leucemia linfóide aguda. 2. Mutação. 3. Aberrações cromossômicas. 4. Leucemia de células B. 5. Diagnóstico clínico. 6. Sequenciamento de nucleotídeos em larga escala. I. Yunes, José Andres, 1967-. II. Massirer, Katlin Brauer, 1975-. III. Universidade Estadual de Campinas. Instituto de Biologia. IV. Título.

Informações para Biblioteca Digital

Título em outro idioma: Genetic mutations in pediatric B-other acute lymphoblastic leukemia

Palavras-chave em inglês:

Acute lymphoblastic leukemia

Mutation

Chromosome aberrations

Leukemia, B-cell

Clinical diagnosis

High-throughput nucleotide sequencing

Área de concentração: Genética Animal e Evolução

Titulação: Doutora em Genética e Biologia Molecular

Banca examinadora:

José Andres Yunes [Orientador]

Carlos Alberto Scrideli

Letícia Fröhlich Archangelo

Fernando Moreira Simabuco

Juliana Godoy Assumpção

Data de defesa: 20-01-2022

Programa de Pós-Graduação: Genética e Biologia Molecular

Identificação e informações acadêmicas do(a) aluno(a)

- ORCID do autor: <https://orcid.org/0000-0002-1326-3841>

- Currículo Lattes do autor: <http://lattes.cnpq.br/8675765940952782>

Campinas, 20 de janeiro de 2022

COMISSÃO EXAMINADORA

Dr. José Andrés Yunes (Presidente da Banca)

Dr. Carlos Alberto Scrideli

Dra. Letícia Frohlich Archangelo

Dr. Fernando Moreira Simabuco

Dra. Juliana Godoy Assumpcao

Os membros da Comissão Examinadora acima assinaram a Ata de Defesa, que se encontra no processo de vida acadêmica da aluna.

Dedicatória

Dedico este trabalho à todas as crianças tratadas no Centro Infantil Boldrini, na esperança de que o conhecimento escrito nessa tese contribua de alguma forma para melhorar o diagnóstico, tratamento e qualidade de vida dos pacientes, trazendo conforto e esperança para seus familiares...

À minha família...

Aos meus irmãos Nagisa, Tsukassa e Yutaka, pela paciência, dedicação e apoio concedido em todas as etapas da minha vida.

Ao meu pai Mitsutomo e minha batyan Teruyo (in memorian)...

*“Be the things, you loved most about
the people who are gone”*

Agradecimentos

Agradeço primeiramente à minha família e amigos pelo amor, carinho e apoio. Agradeço por estarem comigo nos momentos mais difíceis da minha vida. Obrigada, amo vocês!

Agradeço ao meu orientador, Dr. José Andrés Yunes, por ter me dado a oportunidade de realizar este trabalho e permitir colaborar em vários outros. Sou muito grata pela confiança e paciência em ensinar, pelas discussões científicas e bate-papos. Agradeço pela liberdade de crescimento profissional e pessoal, “permitir-se deixar...”, aqui tive o suporte que me deu mais segurança e coragem de acreditar em mim e continuar na pesquisa.

Agradeço à Dra Silvia Brandalise por permitir desenvolver esse projeto no Centro Infantil Boldrini e disponibilizar toda a infraestrutura necessária para a sua conclusão.

À minha co-orientadora, Dra. Katlin Massirer. Sou muito grata pela sua dedicação e preocupação comigo dentro e fora da vida acadêmica.

Agradeço à todos os membros do Laboratório de Biologia Molecular, são pessoas incríveis e sempre dispostos a ajudar; Juliana, Yanca, Samara, Diego, Zé Ricardo, Leo Artico, Marina, Leo Pissinato, Guilherme Navarro, Márcio, Victor, Gabriella, Daniella, Ana Clara, Gustavo Lopes e Bruna Blue. Todos me deram um apoio enorme nesses meses que tentei ficar isolada para finalizar as análises e escrita da minha tese. Obrigada pela compreensão!

Agradeço à todos os amigos e colegas do Centro de Pesquisa Boldrini pelo carinho, conversas e companhia nos momentos de lazer; Bruna Lima, Priscila, Letícia, Aryanne, Giovanna, Livia, Juan, Gustavo Seguchi, Lisandra, Dr. Pedro, Dr. João Meidanis, Aninha, Mariana, Juliana Ruas, Mayara, Felipe, Tássia, Jeniffer, Natalia Paiva, Guilherme, Prof. Fabio Gozzo, Haroldo, Dani Luccon, Maria Eugênia, Mônica, Carol, Patrícia Jotta, Nathy Cury, Bruninha, Luisa, Maraysa, Ana Mosca, Francisco e Luis. Agradeço também aos demais colegas que direta ou indiretamente contribuíram para que este trabalho fosse realizado. Em especial, agradeço a Patrícia Jotta, Natália Paiva e Victor Vasconcelos pela colaboração direta neste trabalho. Agradeço ao Zé Ricardo, Leo Artico e Juliana Ronchi pelas discussões científicas e brainstorming que permitiram transformar ideias em ação e deu mais foco no meu trabalho. Agradeço a Lilian, Daniella Lucon e a Maria Eugênia pela ajuda com as análises de citogenéticas.

À todos os funcionários e colegas do Hospital Boldrini por facilitar o nosso trabalho do dia-a-dia, pelo espaço de trabalho compartilhado durante os anos iniciais do meu doutorado e conversas animadas na copinha; às meninas que sempre nos ajudam a manter a limpeza do nosso prédio.

Um agradecimento especial ao Prof. Dr. Gene Yeo por ter me recebido durante o meu estágio em San Diego na UCSD (Yeo Lab, Sanford Consortium for Regenerative Medicine), ter me dado a oportunidade de discutir ciência com grandes pesquisadores e desenvolver o meu trabalho. Agradeço aos colegas de laboratório com quem tive o prazer de conviver e celebrar vários aniversários na cafeteria que tocava bossa nova. Agradeço aos amigos que eu fiz no Instituto, inclusive amigos brasileiros do Laboratório do Prof. Dr. Alysson Muotri. Em especial, agradeço ao Stefan, Jaclyn, Elliot e Anthony pela ajuda com os experimentos.

Agradeço a minha família filipina, Alice e Rollie, obrigada por terem me recebido tão bem em San Diego e me levado para conhecer vários lugares. Obrigada pelas festas de família, pelas conversas, por tentarem me ensinar tagalo. Miss u guys!

Agradeço a Bia e ao Di por terem ido me visitar no ano que eu passei fora. A visita de vocês diminuiu muito a tristeza e a saudade que eu tinha. Agradeço aos meus amigos de infância e de longa data de Capão Bonito que tenho um enorme carinho e até hoje não deixamos de marcar aquela reunião e papear sobre tudo. Agradeço à todos os ex-moradores da extinta Rep. Stalker, amigos que eu tive o prazer de dividir a casa e parte da minha vida em Campinas. Obrigada pelas risadas do dia-a-dia, pelos jogos de tabuleiro e pelas jantas incríveis. Agradeço ao Cássio e ao Leo pelas nossas noites de descontração, comida boa e conversas. Vocês fazem os meus dias (ops, noites) mais felizes.

Agradeço aos membros da banca examinadora por terem aceito participar deste importante momento da minha vida e contribuir para a finalização deste trabalho.

E por fim, agradeço à CAPES (Coordenação de Aperfeiçoamento de Pessoal de Nível Superior) pelo apoio financeiro durante os anos de doutorado, além de conceder a bolsa de estudos PDSE-CAPES (88881.190380/2018-01) para desenvolver o meu projeto no exterior.

"O presente trabalho foi realizado com apoio da Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Código de Financiamento 001".

"You may not always have a comfortable life. And you will not always be able to solve all the world's problems all at once. But don't ever underestimate the impact you can have, because history has shown us that courage can be contagious, and hope can take on a life of its own." — Michelle Obama

RESUMO

A Leucemia Linfóide Aguda (LLA) é o tipo de câncer mais comum na infância, correspondendo a cerca de 25% de todos os cânceres pediátricos e mais de 80% de todas as leucemias que ocorrem até os 15 anos. Nas últimas décadas, avanços no tratamento e cuidados de suporte aumentaram as taxas de cura para cerca de 80-90% nos países desenvolvidos. O mesmo não se deu em países em desenvolvimento, onde a falta de acesso às análises genéticas impossibilita a correta alocação dos pacientes nos diferentes grupos de risco. A LLA-B corresponde a cerca de 85% dos casos de LLA pediátrica e é categorizada em seis subtipos clássicos: Hiperdiploide, Hipodiploide, *ETV6-RUNX1*, *TCF3-PBX1*, *BCR-ABL1* e rearranjos de *KMT2A*. Recentemente, o sequenciamento NGS revelou novas alterações genéticas e subtipos moleculares de LLA-B entre pacientes classificados como LLA *B-other*, um grupo heterogêneo de LLA sem as alterações cromossômicas clássicas. Neste trabalho investigamos o espectro de mutações genéticas no subgrupo LLA *B-other*, buscando definir novos subgrupos da LLA e encontrar alterações “acionáveis” ou com impacto na evolução clínica dos pacientes. Entre 1999 e 2017, um total de 562 pacientes LLA-B, foram inscritos nos protocolos GBTLI (LLA-1999 e LLA-2009) no Centro Boldrini. Destes, 482 foram analisados em relação às anormalidades cromossômicas clássicas e 144 (~30%) foram classificados como LLA *B-other*, objeto deste estudo. Transcritos de fusão, mutações SNV e níveis de expressão gênica foram analisados por sequenciamento de RNA direcionado com um kit comercial (*TruSight PanCancer*, *Illumina*). Alterações de número de cópias (CNA) foram analisadas por *digitalMLPA* capaz de detectar simultaneamente 56 genes importantes na leucemia (*D007-ALL probemix*, *MRC-Holland*). Os dados de RNA-seq foram analisados com diversas ferramentas de bioinformática (*Illumina-Apps* e *open-source codes*) e *Coffalyser.Net* (*MRC-Holland*) para o *digitalMLPA*. Setenta por cento dos 144 casos de LLA *B-other* analisados foram classificados dentro dos novos subtipos moleculares: *DUX4-cluster* (7,26%), *iAMP21* (1,45%), *ZNF384-rearranged* (1,86%), *MEF2D-rearranged* (0,82%), *NUTM1-rearranged* (0,82%), *PAX5-driven* (3,73%); fusões *ABL-class* (1,45%), *JAK2-fusion* (0,62%), *IGH-EPOR* (0,2%) e *CRLF2-high"only"* (1,66%); 9,96% dos casos não puderam ser classificados e foram denominados *B-other "rest"*. Análise de mutações SNV e *indel* revelou uma alta frequência de mutações em *NRAS* (23%), *MDC1* (22%), *KRAS* (14%), *FLT3* (8%), *PAX5* (8%), *STAT6* (8%), *SETD2* (7%), *MYC* (6%), *IKZF1* (6%) e *KMT2D* (7%). Mutações na via RAS (particularmente, *KRAS* e *FLT3*) mostraram-se associadas com *B-other "rest"*. Além disso, observou-se que mutações nas vias RAS e JAK/STAT são mutuamente exclusivas. Em relação aos CNAs, confirmamos o valor prognóstico dos marcadores utilizados pelo grupo *UKALL* (Reino Unido), que inclui análise de deleções de *EBF1*, *IKZF1*, *PAX5*, *CDKN2A/B*, *ETV6*, *BTG1*, *RB1* e *PAR1*. Mais importante, revelamos que deleções envolvendo *TP53* (del17p) e *CASP8AP2* estão associados a pior prognóstico (5-anos LFS de 29±0,17% vs 85±0,04%, $p<0,001$) e

deverão ser investigadas em um número maior de pacientes. Acreditamos que métodos NGS ainda são relativamente caros, mas o desenvolvimento de um painel de RNA-seq, incluindo menor número de genes, os relevantes para a LLA, pode tornar esta técnica mais acessível.

ABSTRACT

Acute lymphoblastic leukemia (ALL) is the most common childhood malignancy, accounting for 25% of pediatric cancers and more than 80% of pediatric ALL. Over the last decades, advances in the treatment and supportive care, have dramatically increased its 5-year survival rate to about 80-90% in developed countries. Despite this improvement, wide disparities remain across countries. Notably, one of the critical issues in clinical practice of developing countries is the lack of accessibility to genomic screening for genetic abnormalities, which in modern treatment protocols represents an important setting to identify patients at risk for treatment adjustment. B-cell precursor ALL (BCP-ALL) accounts for 85% of pediatric ALL and has been classically categorized into six molecular subtypes: Hyperdiploidy, Hypodiploidy, *ETV6-RUNX1*, *TCF3-PBX1*, *BCR-ABL1*, *KMT2A*-rearrangements. Recently, several genetic studies using high-throughput sequencing have revealed novel genetic alterations and molecular subtypes of BCP-ALL among patients belonging to the so called “B-other” group of ALL, *i.e.* heterogeneous group of ALLs lacking the classical chromosomal alterations. Thus, we sought to investigate the spectrum of mutations in B-other ALL and find possible druggable lesions and prognostic markers. Between 1999 and 2017, a total of 562 BCP-ALL patients were enrolled in the GBTLI protocols (ALL-1999 and ALL-2009) at Centro Boldrini. Of these, 482 were screened with respect to the classical chromosomal abnormalities and 144 (~30%) were ascribed to the B-other group. Patient samples were analyzed for fusion transcripts, single nucleotide variations (SNV) and gene expression by targeted RNA sequencing using a commercial kit interrogating 1,385 genes (*TruSight PanCancer*, *Illumina*) and for recurrent copy number alterations (CNA) by *digitalMLPA* targeting 56 key leukemia-related genes and regions (*D007-ALL probemix*, *MRC-Holland*). Data was analyzed using dedicated bioinformatics tools (*Illumina-Apps* and *open-source codes*) and *Coffalyser.Net* (*MRC-Holland*), respectively. Seventy percent of the 144 B-other ALL cases analyzed could be ascribed to one of the novel molecular subtypes: DUX4-cluster (7.26%), iAMP21 (1.45%), *ZNF384*-rearranged (1.86%), *MEF2D*-rearranged (0.82%), *NUTM1*-rearranged (0.82%), *PAX5*-driven (3.73%); *ABL*-class fusion (1.45%), *JAK2*-fusion (0.62%), *IGH-EPOR* fusion (0.2%), and *CRLF2*-high‘only’ (1.66%); 9.96% of cases could not be classified and were named B-other ‘rest’. The analysis of SNV and small indel revealed high frequency mutations in *NRAS* (23%), *MDC1* (22%), *KRAS* (14%), *FLT3* (8%), *PAX5* (8%), *STAT6* (8%), *SETD2* (7%), *MYC* (6%), *IKZF1* (6%) and *KMT2D* (7%) in our B-other cohort. RAS pathway mutations (particularly, *KRAS* and *FLT3*) were associated with B-other ‘rest’. Besides, RAS and JAK/STAT pathway mutations were mutually exclusive. Regarding CNAs, we confirmed the prognostic value of the UKALL-CNA classifier using *EBF1*, *IKZF1*, *PAX5*, *CDKN2A/B*, *ETV6*, *BTG1*, *RB1*, and *PAR1* deletions. More importantly, we revealed for the first time a strong prognostic impact of deletions involving *TP53* (del17p) and *CASP8AP2* (5-year LFS of 29±0.17% vs

84±0.04%, $p<0.001$) and it should be investigated in larger number of patients. We believe that NGS methods are still relatively expensive, but the development of a smaller RNA-seq panel, including only the ALL relevant genes, is possible and could make this technique more accessible in the near future.

LISTA DE FIGURAS E TABELAS

INTRODUÇÃO

Figura 1. Resumo esquemático das bases genéticas e moleculares do câncer.	18
Figura 2. Anormalidades genéticas e frequência estimada das principais alterações observadas na Leucemia Linfóide Aguda (LLA) pediátrica.	21
Figura 3. Novos marcadores preditivos de risco em LLA-B.	25
Tabela 1. Fusões gênicas identificadas em LLA BCR-ABL1-like e possíveis inibidores.	27

CAPÍTULO I: MANUSCRITO 1

Table 1. Clinical characteristics of 144 pediatric B-other patients analyzed at Boldrini Children Center with GBLTI-1999 and GBLTI-2009 protocols.	37
Figure 1. Circos plot showing the overview of gene fusions discovered by Targeted RNA-sequencing and schematically organized into chromosome position.	45
Figure 2. Unsupervised hierarchical clustering analyses (HCA) identified specific subgroups of patients with shared gene expression patterns.	46
Figure 3. Frequency and demography associations within B-other subgroups.	47
Supplemental Table S1: Demographic, clinical, and genetic abnormalities distribution of all pediatric patients diagnosed at Boldrini Center according to Protocols GBLTI-1999 and GBLTI-2009.	53
Supplemental Figure S1. Cohort distribution and workflow.	54
Supplemental Figure S2. <i>AJ2E</i> -class fusion description.	57
Supplemental Figure S3. Validation of known and novel fusion genes by RT-PCR.	58
Supplemental Figure S4. <i>ZNF384</i> , <i>MEF2D</i> and <i>NUTM1</i> rearrangements description.	59
Supplemental Figure S5. <i>iAMP21</i> subgroup description.	60
Supplemental Figure S6. <i>DUX4</i> -cluster subgroup description.	61
Supplemental Figure S7. <i>CRLF2</i> expression and underlying genomic alterations observed in our cohort.	62
Supplemental Figure S8. <i>GATA3</i> polymorphism analysis.	64
Supplemental Table S2: List of primers sequence used for RT-PCR validation and <i>GATA3</i> risk allele.	65
Supplemental Table S3: List of Gene-sets used for HCA.	66
Supplemental Table S4: Statistic results for demographic correlation analysis.	68

CAPÍTULO II: PERFIL DAS MUTAÇÕES SNV E SMALL INDEL NA LLA B-OTHER

Figure 1. Carga mutacional da nossa coorte de 144 pacientes LLA B-other analisada ("Boldrini") em comparação a outras coortes do banco de dados TCGA (<i>The Cancer Atlas Genome</i>).	72
Figure 2. Classificação das variantes (SNV e small indels) encontradas em 144 pacientes dos subgrupos moleculares de LLA B-other.	74
Figure 3. Frequência e espectro de mutações observados entre os subgrupos moleculares de LLA B-other.	75
Figure 4. Análise de interações e genes significativamente mutados entre os subgrupos moleculares de LLA B-other.	77
Figure 5. Porcentagem de mutações em genes de vias importantes relacionadas ao câncer.	78
Figure 6. Análise integrativa mostrando os genes frequentemente mutados em importantes vias biológicas da leucemia.	79
Figure 7. Resumo esquemático com os subgrupos moleculares de LLA B-other, dados demográficos e os principais genes mutados nas vias de JAK/STAT e RAS.	81

Tabela 1. Análise do valor prognóstico das mutações SNV e <i>small indel</i> em principais vias alteradas na LLA-B (Sobrevida livre de Leucemia calculada em 5 anos).	82
Tabela 2. Análise de sobrevida para os diferentes subgrupos moleculares de LLA B- <i>other</i> tratadas no Centro Infantil Boldrini.	83

CAPÍTULO III: MANUSCRITO 2

Figure 1. Prognostic value of copy number alterations (CNAs) in B- <i>other</i> ALL.	94
Table 1. Outcome analysis for different characteristics in B- <i>other</i> ALL censored in 5-year.	95
Supplementary Table S1. Demography and clinical characteristics of pediatric B- <i>other</i> ALL patients registered at Boldrini's Center according to Protocols GBLTI-1999 and 2009.	101
Supplementary Figure S1. Schematic CNAs detected by digital MLPA in B- <i>other</i> ALL.	103
Supplementary Figure S2. Kaplan-Meier survival curves of pediatric B- <i>other</i> ALL patients with different profiled CNAs deletions.	104
Supplementary Figure S3. Identification of structurally complex rearrangements at chromosome 21q22 resulting in deletions of <i>RUNX1</i> and <i>ETS2</i> .	105

CONCLUSÃO, PERSPECTIVAS E CONSIDERAÇÕES GERAIS

Figura 1. Representação esquemática das diversas mutações genéticas identificadas no subgrupo molecular LLA B- <i>other</i> e vias de sinalização comumente ativadas na leucemia infantil.	108
Figura 2. Fluxograma explicando as etapas de análise adotada para a caracterização genética dos pacientes pediátricos de LLA B- <i>other</i> .	109

APÊNDICE: MECANISMOS DE RESISTÊNCIA NA LLA-B

Figure 1. Representação esquemática dos mecanismos moleculares envolvidos na resistência às drogas quimioterápicas.	119
Figure 2. Etapas importantes da regulação pós-transcricional por proteínas de ligação a RNA (<i>RNA-Binding Proteins</i> , RBPs).	120
Tabela 1. Proteínas de Ligação a RNAs (<i>RNA-Binding Proteins</i> , RBPs) envolvidas na aquisição de resistência a drogas.	121
Figure 3. Pontos chave da tecnologia de CRISPR/Cas9 e design utilizado para a construção das bibliotecas de RBP CRISPR/Cas9 <i>screening</i> .	122
Figure 4. Caracterização e validação das linhagens celulares BCR-ABL positivas, com diferentes sensibilidades para o imatinibe, utilizadas como modelo experimental no estudo.	124
Figure 5. Proteínas de ligação a RNA (<i>RNA-Binding Proteins</i> , RBPs) vulneráveis a seleção negativa com imatinibe em células BCR-ABL1 (K562-Resistentes e K562-Sensíveis), identificadas pela técnica de RBP CRISPR/Cas9 <i>screening</i> .	127
Tabela 2. Compilado da literatura buscando informações sobre os selecionados RBPs e associação com câncer.	129

SUMÁRIO

APRESENTAÇÃO	15
1 INTRODUÇÃO	16
1.1- Câncer é uma doença genética.....	16
1.2- Leucemia Linfóide Aguda (LLA) pediátrica.....	20
1.3- Desafios na pesquisa clínica da LLA-B pediátrica.....	23
1.4- BCR-ABL1-like (ou Ph-like): Novo subgrupo de LLA-B de alto risco?.....	26
1.5- LLA “B-other”: LLA sem subtipo molecular definido.....	28
2 JUSTIFICATIVA	28
3 OBJETIVOS	29
3.1- Objetivo geral.....	29
3.2- Objetivos específicos.....	29
4 SUJEITOS DA PESQUISA E ASPECTOS ÉTICOS	30
5 CAPÍTULO I: MANUSCRITO 1	31
6 CAPÍTULO II: MUTAÇÕES SNV E <i>SMALL INDEL</i> NA LLA B-OTHER	70
6.1- Análise de Bioinformática.....	70
6.2- Análise das mutações SNV e <i>small indel</i> nos pacientes LLA B-other.....	73
6.3- Associações entre alterações genéticas e dados dos pacientes LLA B-other.....	80
6.4- Discussão das principais mutações SNV e <i>small indel</i> encontradas.....	83
7 CAPÍTULO III: MANUSCRITO 2	88
8 CONCLUSÃO	106
9 PERSPECTIVAS E CONSIDERAÇÕES FINAIS	107
10 METODOLOGIA GERAL	110
11 REFERÊNCIAS	114
12 APÊNDICE	118
1 APÊNDICE: BREVE INTRODUÇÃO - MECANISMOS DE RESISTÊNCIA NA LLA-B...	119
1.1- Resistência a drogas e Proteínas de Ligação a RNA (RBPs).....	119
1.2- <i>Screening</i> genômico funcional: ‘RBP-focused CRISPR/Cas9 screening’.....	121
1.3- O papel das RBPs no mecanismo de resistência ao inibidor de tirosina quinase, imatinibe, em LLA-B portadoras da translocação <i>BCR-ABL1</i>	123
2 APÊNDICE: RESULTADOS.....	125
2.1- Resultados parciais.....	125
3 APÊNDICE: CONCLUSÕES PARCIAIS E PERSPECTIVAS.....	126
4 APÊNDICE: REFERÊNCIAS BIBLIOGRÁFICAS.....	130
13 ANEXOS	132
13.1-Anexo I: Parecer do Comitê de Ética (CEP).....	132
13.2-Anexo II: Declaração sobre Direitos Autorais.....	139

APRESENTAÇÃO DA TESE

Esta tese de doutorado, apresentada ao Programa de Pós-Graduação em Genética e Biologia Molecular, UNICAMP-Campinas, SP, está estruturada na forma de dois manuscritos científicos, resultados complementares e um apêndice. Assim, a tese é apresentada em treze sessões: 1) Introdução; 2) Justificativa; 3) Objetivos; 4) Sujeitos da Pesquisa; 5) *Manuscrito 1*; 6) Resultados complementares ao manuscrito 1- Mutações somáticas; 7) *Manuscrito 2*; 8) Conclusões; 9) Perspectivas e Considerações Finais; 10) Metodologia Geral; 11) Anexos; 12) Referências Bibliográficas; 13) *Apêndice*-Resultados obtidos no estágio no exterior.

Em resumo, essa tese de doutorado buscou estudar as alterações genéticas das leucemias linfóide aguda pediátrica do subgrupo “*B-other*”. Foram caracterizadas as mutações genéticas e investigadas as associações entre os subgrupos moleculares com os dados clínicos e prognóstico dos pacientes tratados no Centro Infantil Boldrini. Resultados apresentados nas sessões 5), 6) e 7). Além do estudo desses biomarcadores e subgrupos moleculares de LLA B-other (incluídos nesse grupo os casos de BCR-ABL1-like), um objetivo secundário da tese foi o de investigar o mecanismo de resistência ao imatinibe na leucemia BCR-ABL1+. Os resultados preliminares desse estudo estão apresentados na sessão “13) Apêndice”, na qual buscamos estudar o papel das Proteínas de ligação a RNA (*RNA-Binding Proteins*, RBPs) no mecanismo de resistência ao imatinibe em células de leucemia BCR-ABL1+ através da mais avançada técnica de “*CRISPR Cas9 screening*”. Resultados apresentado na sessão 13).

1 INTRODUÇÃO

1.1 Câncer é uma doença genética

Câncer é o termo comumente utilizado para descrever um enorme espectro de doenças caracterizadas por crescimento celular anormal e descontrolado, e causado principalmente por alterações genéticas e epigenéticas. Os cânceres são distintamente classificados do ponto de vista etiológico e morfológico, apresentando particularidades genéticas que resultam em progressão variável da doença, os quais dificultam um tratamento clínico universal.^{1,2} Segundo dados publicados pela Organização Mundial da Saúde (OMS), o câncer é a segunda maior causa de mortalidade no mundo, responsável por quase 10 milhões de mortes em 2020 (OMS, 2020).³ Dentre estes, aproximadamente 70% das mortes por câncer ocorrem em países subdesenvolvidos.⁴

O câncer é uma doença genética - ou seja, o desenvolvimento de células malignas envolve várias lesões ou modificações genéticas, que combinados controlam e dirigem a transformação de uma célula normal até o estágio de câncer. Esse processo complexo envolve um conjunto de características que podem ser descritas como as “*hallmarks*” do câncer, ou seja, mutações em vias que levam a (1) sinalização de proliferação celular permanente; (2) potencial replicativo ilimitado; (3) escape dos processos inibitórios de proliferação; (4) evasão a mecanismos de morte celular programada ou apoptose; (5) evasão do sistema imune; (6) promoção de processos inflamatórios; (7) indução da angiogênese; (8) ativação de metástase e invasão tecidual; (9) reprogramação do metabolismo energético; e (10) instabilidade genômica e mutação (Figura 1A). As “*hallmarks*” do câncer foi um marco para os estudos sobre a biologia do câncer. Inicialmente apresentada pelos pesquisadores Hanahan D. e Weinberg RA. como seis características comuns aos tumores, e após em 2011, atualizada para dez características, hoje, em 2022, com o advento das avançadas tecnologias genômicas, Hanahan D. apresenta e discute um novo conjunto de características importantes para o desenvolvimento e progressão do câncer (i.e., “*unlocking phenotypic plasticity*”, “*nonmutational epigenetic reprogramming*”, “*polymorphic microbiomes*” e “*senescent cells*”) a fim de melhorar o atual conceito das “*hallmarks*”.⁵

É importante lembrar que algumas mutações genéticas que promovem o câncer podem ser herdadas dos genitores se essas alterações estiverem presentes nas células germinativas (*i.e.*, óvulos e espermatozóides), os quais representam apenas 5-10% de todos os cânceres. A maioria dos cânceres são causados por mutações adquiridas durante a vida, chamadas de mutações somáticas. Essas alterações genéticas que ocorrem após a concepção podem ser resultado de erros que ocorrem quando as células se dividem ou da influência de fatores ambientais, tais como a exposição a substâncias químicas e cancerígenas que danificam o DNA, e radiação, como os raios ultravioleta do sol.^{1,2}

A célula cancerígena dificilmente apresenta um único gene defeituoso, pelo contrário, ele é uma somatória de mutações, e que muitas vezes não contribuem igualmente para a progressão da doença. Existem cerca de 25.000 genes em cada célula humana, mas, somente quando ocorre acúmulo de mutações em certos genes-chave é que o câncer se desenvolve. Esses genes-chave podem ser agrupados em quatro grandes classes: genes promotores de crescimento (proto-oncogenes), genes inibidores do crescimento (supressores de tumor), genes que regulam a morte celular programada e genes de reparo do DNA.⁵ De maneira geral, estima-se que 3 a 12 mutações cumulativas em alguns dos 2.000 genes potencialmente oncogênicos seriam necessários para desencadear a doença.⁶ Além disso, há dois tipos de mutações: (1) mutações “*driver*”, que conferem vantagem de crescimento para as células portadoras, são causalmente implicadas no desenvolvimento do câncer e, portanto, são selecionadas positivamente; (2) mutações “*passenger*”, que possivelmente decorrem do processo de instabilidade genômica do câncer, mas não conferem vantagem de crescimento, logo, são neutras.⁷ Na Figura 1 são mostrados esquematicamente alguns dos mecanismos através dos quais os proto-oncogenes são ativados nos cânceres e o processo de expansão clonal.

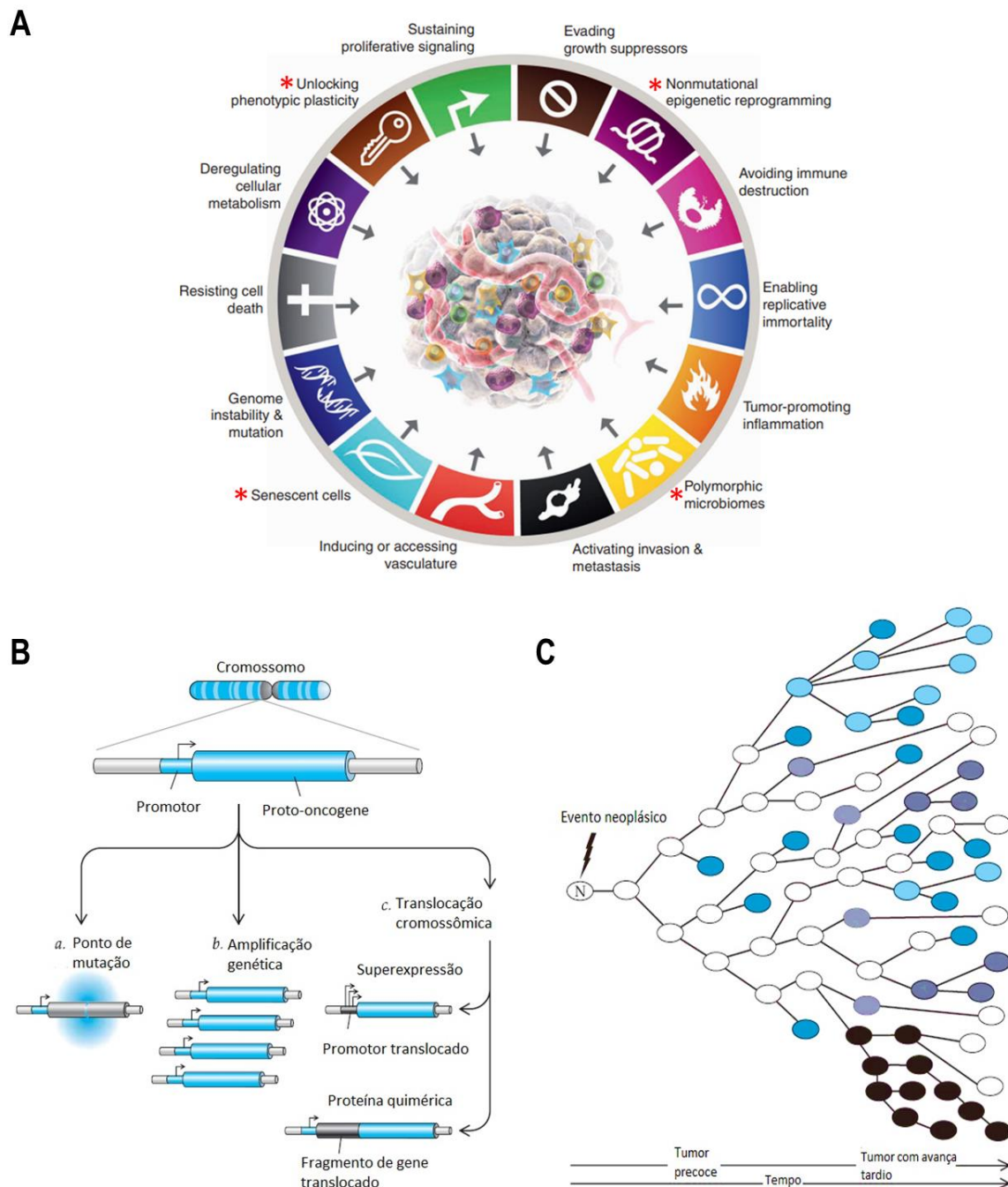


Figura 1. Resumo esquemático das bases genéticas e moleculares do câncer. (A) A mais recente publicação com as “hallmarks” do câncer. Em destaque (asterisco), Hanahan D (2022) propõe a adição de novas marcas envolvendo “desbloqueio de plasticidade fenotípica”, “reprogramação epigenética não mutacional”, “microbiomas polimórficos” e “células senescentes”, como características importantes para o desenvolvimento e progressão tumoral. (B) Mecanismos de ativação de proto-oncogenes. Os genes podem ser ativados no câncer como resultado de (a) mutações de ponto que alteram a sequência de aminoácidos; (b) ampliações do gene resultando em níveis mais elevados de expressão gênica; ou (c) translocações cromossômicas que levam à justaposição de parte de dois genes e/ou de um gene com um promotor constitutivamente ativado, aumentando a expressão, ou que produzam uma proteína quimérica nova, derivada de fragmentos de genes normalmente presentes em cromossomos diferentes. (C) Expansão clonal e progressão de um fenótipo maligno. Após o início de um evento genético neoplásico e limitante de velocidade, a proliferação de células neoplásicas é estimulada por vantagens seletivas e mutações adicionais que conferem uma vantagem complementar, e, consequentemente, são selecionadas durante a expansão do clone maligno. A divisão celular constante e a perda da estabilidade genômica facilitam a ocorrência de mutações. As cores indicam os descendentes clonais de células individuais.

O avanço no conhecimento da biologia do câncer possibilitou uma melhor compreensão da genética da doença, suas particularidades e variabilidades. A multidisciplinaridade de conhecimentos proveniente das ciências básicas – aliado ao conhecimento clínico e avanços de técnicas de diagnóstico – permitiu um grande avanço da medicina de precisão, com impacto positivo no diagnóstico, tratamento, manejo, prevenção, sobrevida e qualidade de vida dos pacientes com câncer. Nesse sentido, sabe-se que os cânceres infantis possuem características próprias e bem diferentes em relação ao câncer em adultos. Predominantemente, o câncer pediátrico afeta células embrionárias, constituídas em sua maioria de células indiferenciadas. Os cânceres mais comuns são as leucemias, tumores do sistema nervoso central, linfomas e tumores sólidos como o neuroblastoma, sarcomas e o tumor de Wilms. A maioria dos cânceres infantis tem relativamente poucas alterações genéticas e muitas vezes não se enquadram para utilizar drogas alvo-específico para tratamentos já desenvolvidos e aprovados em cânceres de adultos. Também, é notável a heterogeneidade nos tipos de alterações genéticas que impulsionam o crescimento de tumores pediátricos, estes incluem alterações no número de cópias (CNAs), fusões gênicas, rearranjos complexos do DNA (*chromoplexy*) e alterações cromossômicas.⁸ Vários estudos moleculares e iniciativas (“*Consortium*”) de grupos na área de tumores pediátricos estão sendo realizados com o objetivo de desvendar as particularidades genéticas e fornecer outras opções de tratamento para as crianças.⁹⁻¹² No Brasil, o câncer infantil já representa a primeira causa de morte (8% do total) por doença entre crianças e adolescentes de 1 a 19 anos. Segundo o Instituto Nacional de Câncer (INCA), estima-se que para cada ano do triênio 2020/2022, sejam diagnosticados 8.460 novos casos de câncer infanto-juvenis (4.310 em meninos e 4.150 em meninas).¹³ Logo, a busca por novas terapias alvo-específicos e o desenvolvimento de protocolos de tratamento que sejam menos tóxicos e causem menos efeitos adversos (agudos e tardios) nas crianças são um dos desafios na pesquisa do câncer infantil.

1.2 Leucemia Linfóide Aguda (LLA) pediátrica

Leucemia é um termo geral dado a um grupo de cânceres de células sanguíneas. É uma doença heterogênea, caracterizada por uma disfunção das células tronco da medula óssea levando a uma proliferação clonal e descontrolada de células precursoras de origem mieloide ou linfóide. Nesse sentido, a leucemia pode ser subdividida em quatro tipos principais que variam substancialmente em sua origem celular e comportamento clínico: Leucemia Mieloide Aguda (LMA); Leucemia Linfóide Aguda (LLA); Leucemia Mieloide Crônica (LMC) e Leucemia Linfóide Crônica (LLC). Tanto adultos quanto crianças podem desenvolver leucemia, mas certos tipos são mais comuns em diferentes faixas etárias.^{14,15}

A Leucemia Linfóide Aguda (LLA) é o tipo de câncer mais comum na infância, correspondendo a cerca de 25% de todos os casos de cânceres nessa faixa etária e a 80% de todas as leucemias que ocorrem até os 15 anos de idade.^{13,16} Tradicionalmente, a LLA é classificada em LLA-B precursora (quando as células afetadas são da linhagem de células B, e que corresponde a 85% dos casos) e em LLA-T (da linhagem de células T, e que corresponde a aprox.15% dos casos).¹⁷⁻¹⁹

De um modo geral, as LLAs estão amplamente associadas a anormalidades cromossômicas, tanto numéricas como estruturais. Assim, além das classificações por tipo celular, a LLA-B e a LLA-T são subdivididas de acordo com anormalidades genéticas, incluindo aneuploidias e translocações cromossômicas.²⁰ O conhecimento destas características é importante para a adequação do tratamento e alocação desses pacientes em grupos de risco, pois o prognóstico difere consideravelmente entre seus subtipos— além de variar também em função das características clínico-biológicas de cada paciente, como idade, contagem leucocitária ao diagnóstico e resposta inicial ao tratamento (*i.e.*, doença residual mínima, DRM).^{21,22}

A Figura 2 apresenta a frequência dos subtipos de LLA descritos na literatura internacional, assim como os dados de índices de sobrevida obtidos com o protocolo do Grupo Brasileiro de Tratamento das Leucemias na Infância (GBTLI LLA 1993-1999)²³.

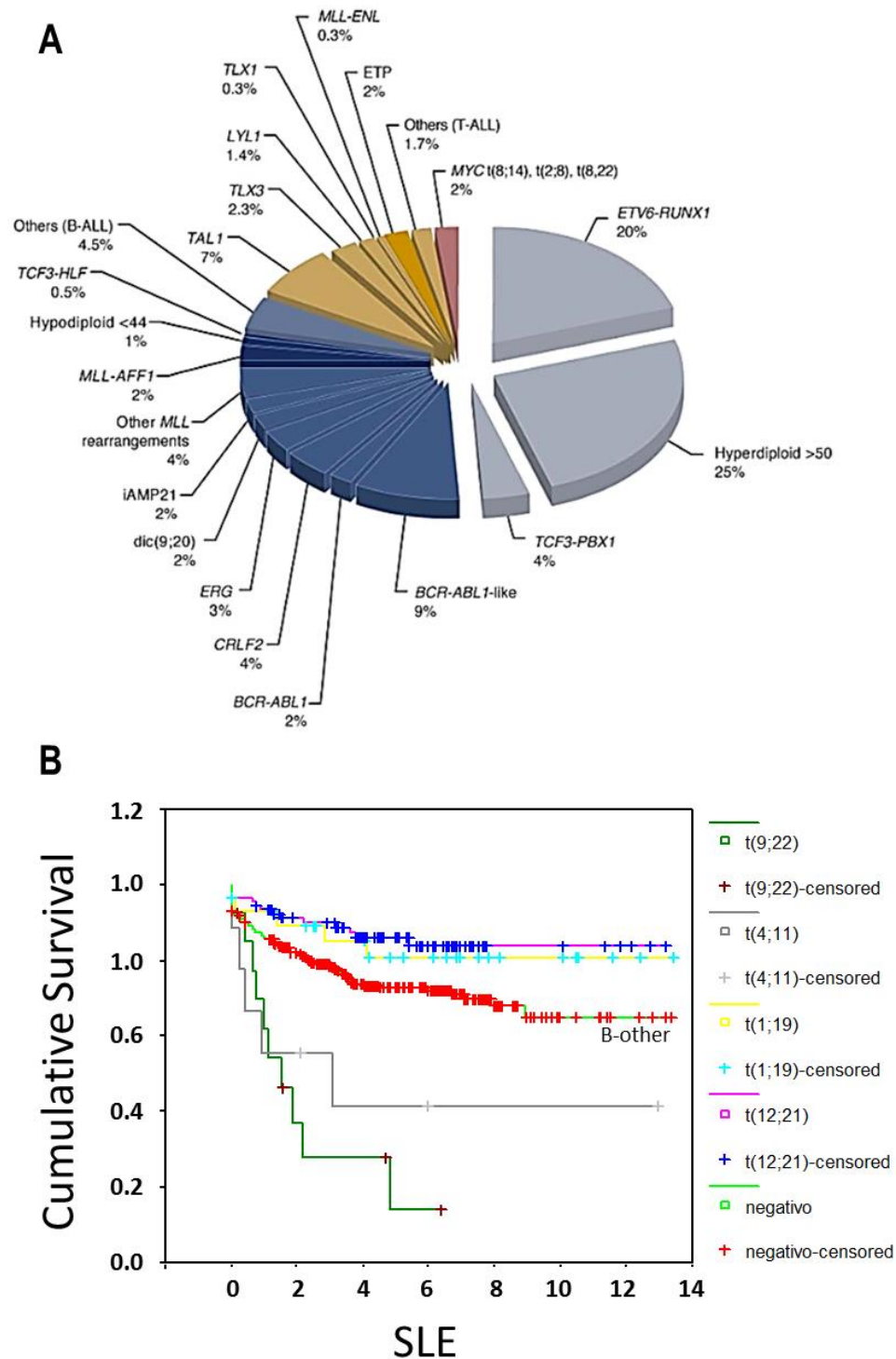


Figura 2. Anormalidades genéticas e frequência estimada das principais alterações observadas na Leucemia Linfóide Aguda (LLA) pediátrica. (A) Frequência estimada dos principais subtipos de LLA pediátrica identificados por imunofenótipo ou presença de alterações de cariótipo ou translocações cromossômicas. O grupo “Others” compreende os casos sem nenhuma alteração genética definida. Alterações genéticas comumente associadas à LLA-B estão em azul e aquelas evidenciadas exclusivamente em casos de LLA-T estão indicadas em amarelo. As porções destacadas em azul escuro (LLA-B) e amarelo escuro (LLA-T) indicam os subtipos geralmente associados com piores prognósticos. Adaptado de Pui, CH et al, (2012)²⁴. (B) Sobrevida livre de eventos (SLE) para os diferentes subtipos de LLA no protocolo de tratamento do grupo brasileiro (GBTLI LLA-1993-1999)²³. Na ocasião, o grupo brasileiro não diferenciava os novos subgrupos moleculares, tais como BCR-ABL1-like, iAMP21, CRLF2 e pacientes com alterações em ERG, os quais ficaram incorporados ao grupo “B-other”. Sobrevida Livre de Eventos considerados: morte e recaída. As curvas foram traçadas pelo método *Kaplan-Meier* e comparadas pelo teste *Log Rank*.

Usualmente, o tratamento da LLA é baseado na quimioterapia com múltiplas drogas, os quais incluem três fases principais (1) fase de indução da remissão, que dura aproximadamente 1 mês; (2) seguido de uma fase de consolidação da remissão, que dura entre 6 e 9 meses e; (3) uma fase final de manutenção da remissão, que pode durar por 1 a 2 anos, dependendo do protocolo de tratamento. Além destas 3 fases, os protocolos terapêuticos incluem o tratamento direcionado ao combate e/ou profilaxia da leucemia no sistema nervoso central com uso de radioterapia e/ou quimioterapia intratecal. A dose, intervalo e combinação de quimioterápicos usados varia de acordo com o subtipo de LLA (GBLTI-2009, dados não publicados)²³.

Nesse sentido, o resultado do tratamento da LLA melhorou muito em relação às décadas de 60 e 70 e hoje superam a taxa de 85% de cura nos países desenvolvidos.²⁵ Uma característica importante da eficácia dos protocolos atuais de tratamento da LLA pediátrica é a alocação dos pacientes em diferentes grupos de risco, combinação de drogas e intensidade de tratamento.^{26,27} Por exemplo, no caso da leucemia portadora da translocação t(9;22), também conhecido como cromossomo *Philadélphia*, e que origina a proteína de fusão *BCR-ABL1*, tem sido paradigmática. A incorporação de imatinibe, um inibidor de tirosina quinase (TKI), no tratamento desse subgrupo de alto risco melhorou além das expectativas a sobrevida das crianças para mais de 80%, quando historicamente era de apenas 36%.²⁸ Entretanto, apesar da significativa melhora no índice de cura da LLA e do tratamento para os casos de alto risco, ainda, ~20% dos pacientes pediátricos (e mais de 50% dos pacientes adultos) sofrem recaída da doença.^{29,30} Se a recaída for tardia (após 36 meses do diagnóstico) as chances de cura chegam a 50%, mas se for muito precoce (até 18 meses do diagnóstico) as chances são inferiores a 30%.³¹ Nesses casos, aumentar a intensidade do tratamento não parece ser a solução. A ocorrência de recaídas no grupo de alto risco, ou seja, nos casos que já recebem intensidade máxima de tratamento quimioterápico, indica que para vencer os casos de LLA resistente será necessário a incorporação de novas drogas e/ou modalidades terapêuticas direcionadas.

1.3 Pesquisa clínica da LLA-B pediátrica

Nas últimas décadas, vários grupos têm direcionado esforços para à identificação das alterações genéticas subjacentes à LLA-B e sua associação com a resposta clínica, de modo a auxiliar na alocação dos pacientes nos diferentes grupos de risco e tratamento. Diversas anormalidades genéticas recorrentes foram relatadas na LLA-B, algumas das quais, como rearranjos do gene *KMT2A (MLL)* em crianças lactentes ou a translocação cromossômica $t(9;22)(BCR-ABL)$; foram descritas o papel dessas alterações primárias na gênese da doença e ao fenótipo de maior resistência *in vitro* à quimioterapia,^{32,33} e baixa resposta ao tratamento³⁴. Nos atuais protocolos de tratamento da LLA pediátrica, pacientes com LLA portadora da $t(9;22)(BCR-ABL1)$ ou rearranjos do *KMT2A (MLL)* em crianças com menos de 1 ano, já são alocados em grupos diferentes dos demais pacientes. No caso das crianças com a $t(9;22)(BCR-ABL1)$, além dos quimioterápicos convencionais, faz-se uso do imatinibe, droga alvo-específica para a proteína quimérica (GBLTI LLA-2009)²³. Além disso, hoje já temos o conhecimento de que existem anormalidades genéticas, como a translocação cromossômica $t(12;21)(ETV6-RUNX1)$ que caracterizam casos de LLA-B com maior quimiosensibilidade e resposta favorável ao tratamento^{35,36} mesmo após recaída^{37,38}, o qual possibilitaria reduzir as doses quimioterápicas, evitando a toxicidade e efeitos adversos nesses pacientes.

Como mencionado anteriormente, aneuploidias e translocações cromossômicas são anormalidades primárias típicas da LLA, enquanto anormalidades secundárias são geralmente alterações do número de cópias (CNA) (*i.e.*, micro-deleções e amplificações) e mutações pontuais.³⁹ De maneira geral, anormalidades primárias estão presentes em todas as células que compõem o clone leucêmico e definem as principais características da leucemia. Por outro lado, as anormalidades secundárias estão presentes em apenas subconjuntos de células leucêmicas, dando origem a uma composição subclonal da doença. Na LLA, observa-se uma forte correlação entre a anormalidade cromossômica primária e a espectro de mutações secundárias ou cooperantes em cada subtipo. Por exemplo, deleções nos genes *ETV6*, *CDKN2A/B* e *PAX5* são observadas com uma frequência de 54%, 22% e 22% no subtipo com a translocação $t(12;21)(ETV6-RUNX1)$, respectivamente. Enquanto que deleções nos genes *IKZF1* (64%), *CDKN2A/B* (48%) e *PAX5* (45%) são frequentes nas leucemias com $t(9;22)(BCR-ABL1)$.⁴⁰ Existem subtipos sem alterações

primárias evidentes, porém com mutações em genes importantes para a via da leucemogênese, como observado em 9% das LLA-T, nos quais mutações focais (*i.e.*, restritas a uma região estreita do éxon) no gene *IL7R* levam a ganho de função, e promovem a transformação e crescimento tumoral.⁴¹

O advento de novas metodologias para análises genômicas têm possibilitado identificar uma série de novas alterações moleculares recorrentes e mutações cooperativas na LLA-B que podem ser usadas em combinação como biomarcadores de prognóstico. Por exemplo, no trabalho de Hamadeh L et al, 2019⁴², eles descrevem e validam um método robusto de classificação dos pacientes em grupos de risco, que leva em conta as alterações primárias e o perfil de alteração do número de cópias (CNAs) de oito genes frequentemente deletados em LLA-B, o qual eles denominam “*UKALL-CNA classifier*” (Figura 3). Observou-se que, principalmente, entre os grupos de risco intermediário, definidos pela citogenética convencional, haviam pacientes que poderiam ser realocados em grupos de risco baixo e outros em risco alto de recaída graças às análises de CNAs. Nesse sentido, esta tese visa preencher uma lacuna para os casos de pacientes sem subtipo genético definido, ou seja, aqueles que não apresentam nenhuma alteração genética que pode ser identificada por exames padrão de cariotipagem, FISH (*Fluorescent In Situ Hybridization*), ou RT-PCR, e que são agrupados e classificados como LLA “B-other” de risco intermediário.

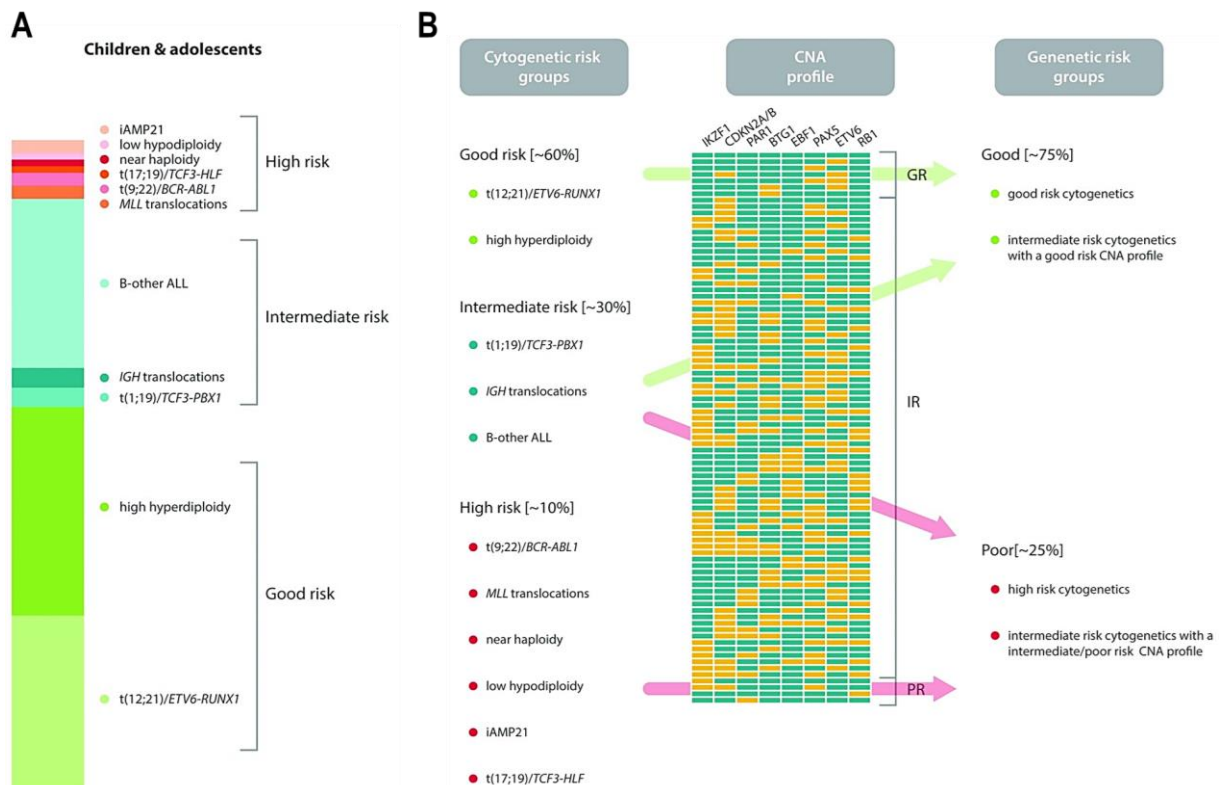


Figura 3. Novos biomarcadores preditivos de risco em LLA-B. (A) Frequência estimada das principais anormalidades cromossômicas primárias e a sua relevância prognóstica em crianças e adolescentes com Leucemia Linfóide Aguda B precursora (LLA-B). (B) Proposta de uma nova classificação de risco para a LLA-B com base na análise integrada dos grupos de risco definidos pelas anormalidades cromossômicas primárias e concomitante análise do perfil de alteração do número de cópias (CNAs) de oito genes comumente deletados na LLA-B. (CNA *profile*: amarelo-deletado e verde-não deletado). Prognóstico “corrigido” e pacientes com baixo e alto risco de recaída podem ser realocados em grupos de risco. Figuras obtidas do trabalho de Moorman, AV, (2016)⁴³.

A mais recente publicação da OMS (2016)⁴⁴ em relação à classificação das neoplasias hematológicas evidencia a necessidade de uma revisão dos métodos diagnósticos e classificação dos subtipos de LLA. A edição de 2016 não se trata de uma “nova” classificação, mas sim, uma revisão da edição anterior (2008), e agrega aspectos clínicos, morfológicos, imunofenotípicos e genético-moleculares de relevância para tratamento e prognóstico dos pacientes. Historicamente, houve vários avanços na identificação de novos biomarcadores gerais ou associados a subgrupos específicos de LLA, tais biomarcadores foram descobertos, em grande parte, por análises de sequenciamento de última geração (*Nex-Generation Sequencing*, NGS) ou análise da expressão gênica global. A “nova” classificação da OMS ressalta que as alterações genético-moleculares não apenas caracterizam entidades específicas e/ou provisórias de LLA, mas são essenciais para identificar produtos gênicos e importantes vias de sinalização como alvos terapêuticos.

Em suma, os critérios contemporâneos de classificação das LLA discutidos pela OMS, levam em conta dados clínicos, citomorfológicos e genético-moleculares para o diagnóstico e estratificação dos pacientes em grupos de risco, com a finalidade de tratamento ajustado ao risco de recidiva. Além disso, preconiza-se monitorar e avaliar a progressão da doença (*follow-up*) por meio da pesquisa de doença residual mínima (DRM). Em relação ao aspecto genético-molecular, é importante ressaltar que nem todas as leucemias estão associadas com anormalidades cromossômicas (citogenéticas) detectáveis pelo cariótipo e/ou por FISH (exemplo, LLA B-*other*).

1.4 BCR-ABL1-like (ou Ph-like): Novo subgrupo de LLA-B de alto risco?

Em 2009, Mullighan *et al*⁴⁵ e Den Boer *et al*⁴⁶ realizaram dois trabalhos independentes e caracterizaram um subgrupo molecular de LLA-B de prognóstico desfavorável, denominado BCR-ABL1-like ou Ph-like. Esse subgrupo foi identificado com base no seu perfil de expressão gênica, que é muito similar ao perfil de expressão de células de pacientes LLA-B t(9;22)(*BCR-ABL1*), apesar de não apresentarem essa translocação cromossômica. Os casos de BCR-ABL1-like compreendem aproximadamente 9% das crianças com LLA-B (Figura 2) e assim como os pacientes com a t(9;22)(*BCR-ABL1*), estão associados a maior risco de recaída, independentemente de idade, contagem leucocitária, citogenética e níveis de doença residual mínima.^{45,46}

O padrão de expressão gênica semelhante a BCR-ABL1-positivo sugere a presença de alterações genéticas capazes de substituir a atividade do gene de fusão *BCR-ABL1*. De fato, extensivas análises genéticas em 154 casos de BCR-ABL1-like evidenciaram a presença de mutações ou fusões gênicas ativadoras de tirosina quinases e dos receptores de citocinas em 91% desses casos. Cerca de 12,6% dos pacientes BCR-ABL1-like apresentavam rearranjos envolvendo os genes de quinases “*ABL-class*”, tais como o *ABL1*, *ABL2*, *CSF1R* e *PDGFRB*; e outros casos apresentavam rearranjos envolvendo *EPOR* (3,9%), *JAK2* (7,4%) e *CRLF2* (49,7%). Aproximadamente, metade dos casos de BCR-ABL1-like apresentavam rearranjos cromossômicos envolvendo o gene receptor de citocina *CRLF2* (e.g., translocação IGH-CRLF2 e uma deleção focal resultando na junção P2RY8-CRLF2) que ocasionam a superexpressão do mesmo.^{47,48} E entre os casos com rearranjo de *CRLF2*, aproximadamente um terço apresentavam mutações concomitantes em gene *JAK2*,

resultando na ativação da via JAK-STAT.⁴⁹ Também observaram-se mutações em *IL7R*, *FLT3*, *SH2B3*, *JAK1*, *JAK3* e *TYK2*, entre outras.⁵⁰ O subgrupo BCR-ABL1-like apresenta elevada frequência de micro-deleções no gene *IKZF1*, o qual codifica um fator de transcrição de IKAROS (*early lymphoid transcription factor IKAROS*)^{45,46}.

Digno de nota, para as fusões gênicas envolvendo os genes *ABL1*, *ABL2*, *CSF1R*, *PDGFRB*, *JAK2* e *EPOR*, já foi demonstrado que as LLAs portadoras dessas alterações são sensíveis a inibidores de tirosina quinases (TKIs) já aprovados e em uso clínico.⁵⁰ Na Tabela 1 estão destacadas as fusões identificadas em BCR-ABL1-like e os TKIs supostamente eficazes contra esses rearranjos. Chama a atenção a variedade de genes parceiros das quinases “*ABL-class*”.

Tabela 1. Fusões gênicas identificadas em LLA BCR-ABL1-like e possíveis inibidores. Tabela adaptada do trabalho de Roberts K.G., et al (2014)⁵⁰, na qual foram incluídos alguns novos rearranjos gênicos em *PDGFRB (Ishibashi, Takeshi et al., 2016)⁵¹ e em **JAK2 (Kawamura, M., et al., 2015)⁵².

Gene (quinase)	Inibidor (TKIs)	Nº de Pacientes /casos	Nº Genes parceiros de fusão	5' Genes
<i>ABL1</i>	Dasatinib	14	6	<i>ETV6</i> , <i>NUP214</i> , <i>RCSD1</i> , <i>RANBP2</i> , <i>SNX2</i> , <i>ZMIZ1</i>
<i>ABL2</i>	Dasatinib	7	3	<i>PAG1</i> , <i>RCSD1</i> , <i>ZC3HAV1</i>
<i>CSF1R</i>	Dasatinib	4	1	<i>SSBP2</i>
<i>PDGFRB</i>	Dasatinib	11	4	<i>EBF1</i> , <i>SSBP2</i> , <i>TNIP1</i> , <i>ZEB2</i> , <i>ATF7IP*</i>
<i>CRLF2</i>	inibidor de JAK2	30	2	<i>IGH</i> , <i>P2RY8</i>
<i>JAK2</i>	inibidor de JAK2	19	10	<i>ATF7IP</i> , <i>BCR</i> , <i>EBF1</i> , <i>ETV6</i> , <i>PAX5</i> , <i>PPFIBP1</i> , <i>SSBP2</i> , <i>STRN3</i> , <i>TERF2</i> , <i>TPR</i> , <i>SPAG9**</i>
<i>EPOR</i>	inibidor de JAK2	9	2	<i>IGH</i> , <i>IGK</i>
<i>DGKH</i>	Desconhecido	1	1	<i>ZFAND3</i>
<i>IL2RB</i>	Inibidor de JAK1 e JAK3	1	1	<i>MYH9</i>
<i>NTRK3</i>	Crizotinib	1	1	<i>ETV6</i>
<i>PTK2B</i>	inibidor de FAK	1	2	<i>KDM6A</i> , <i>STAG2</i>
<i>TSLP</i>	inibidor de JAK2	1	1	<i>IQGAP2</i>
<i>TYK2</i>	inibidor de TYK2	1	1	<i>MYB</i>

Os inibidores de tirosina quinase listados são conhecidos ou supostos para serem ativos contra os respectivos rearranjos de quinase, de acordo com modelos experimentais.

1.5 LLA “B-other”: LLA sem subtipo molecular definido

Como mencionado anteriormente, o subgrupo LLA “B-other” – é um grupo heterogêneo de LLAs que são agrupadas por não apresentarem as aneuploidias ou translocações cromossômicas conhecidas (*i.e.*, $t(9;22)(BCR-ABL1)$, o rearranjo de *KMT2A* (*MLL*) em 11q23, $t(1;19)(TCF3(E2A)-PBX1)$, $t(17;19)(TCF3-HLF)$, $t(12;21)(ETV6-RUNX1)$, hiperdiploidia (HeH) e hipodiploidia, e representam em torno de 20-30% das LLAs-B. Logo, destrinchar o perfil genético das LLA B-other e avaliar a evolução clínica desses pacientes tem sido um dos focos principais das pesquisas em LLA nos últimos dez anos⁵³⁻⁵⁸.

2 JUSTIFICATIVA

Nas últimas décadas, diversos grupos internacionais de oncologia pediátrica têm demonstrado a importância de análises genômicas na medicina de precisão. Grande número de mutações recorrentes foram descobertas e poderiam servir como potenciais biomarcadores para uso no diagnóstico ou prognóstico, além da possibilidade de serem alvos para novas drogas. O nosso estudo tem como objetivo caracterizar as alterações genéticas do subgrupo LLA B-other, correlacionando-as com a clínica dos pacientes.

Identificar alterações genéticas nos casos de LLA B-other pode auxiliar na alocação dos pacientes em grupos de risco, bem como no desenho de futuros estudos clínicos com quimioterápicos direcionados contra as proteínas ou vias afetadas por essas mutações. As anormalidades genéticas “driver” são marcadores prognósticos mais confiáveis, provavelmente devido ao fato de definirem as principais características do clone leucêmico. Logo, o foco da maioria dos algoritmos de classificação das LLAs é baseado nas anormalidades cromossômicas/genéticas “driver”. Dentre estas, é mais urgente a identificação dos casos com alterações acionáveis, que poderiam ajudar a definir tratamentos específicos e já disponíveis, como por exemplo, nos casos com fusões envolvendo os genes “ABL-class”, os quais, esses pacientes se beneficiariam de tratamento com o imatinibe e o dasatinib,⁵⁰ drogas já aprovadas pela ANVISA.

3 OBJETIVOS

3.1 Objetivo geral

O objetivo desta tese foi estudar as alterações genéticas das LLA-B do subgrupo B-*other*, correlacionando os achados com os dados clínicos dos pacientes.

3.2 Objetivos específicos

Capítulo I: Manuscrito 1

- Estabelecer método diagnóstico de *Next-Generation Sequencing* (NGS) baseado em painel de “*Targeted RNA-sequencing*” para identificação de mutações SNV e fusões gênicas recorrentes ou novas em LLA do subgrupo B-*other*.
- Padronizar análises de bioinformática para detectar mutações e/ou ocorrência de transcritos quiméricos, bem como caracterizar subgrupos com mesmo perfil de expressão gênica.
- Analisar a frequência dos subgrupos moleculares encontrados em pacientes LLA B-*other* do Centro Infantil Boldrini e correlacioná-los com os dados demográficos.

Capítulo II: Perfil das mutações SNV e small indel na LLA B-*other*

- Descrever o perfil de mutações somáticas pontuais e pequenas deleções/inserções (i.e., *single-nucleotide variant* - SNV e *small indel*), e vias de sinalização enriquecidas dentro dos subgrupos de LLA B-*other*.
- Correlacionar as mutações SNV e *small indel* e as vias de sinalização enriquecidas com os dados demográficos dos pacientes LLA B-*other*.

Capítulo III: Manuscrito 2

- Caracterizar as mutações de alteração do número de cópias (*copy-number alterations*, CNAs) através do método *Digital Multiplex Ligation-dependent Probe Amplification* (dMLPA) em pacientes do subgrupo LLA B-*other* do Centro Infantil Boldrini.
- Analisar o valor preditivo das mutações de CNAs no prognóstico de pacientes LLA B-*other* do Centro Infantil Boldrini.

4 SUJEITOS DA PESQUISA E ASPECTOS ÉTICOS

Para este estudo foram utilizadas amostras de DNA, RNA e dados de expressão gênica provenientes de um total de 266 pacientes registrados no Centro Infantil Boldrini e tratados seguindo os Protocolos GBTLI LLA-1999 e GBTLI LLA-2009. Tratando-se de um estudo retrospectivo, no qual o material utilizado consistiu de material genético excedente, extraído previamente para fins diagnósticos, os pacientes não passaram por nenhum tipo de desconforto ou alteração no tratamento e/ou acompanhamento pós-terapia decorrente do estudo. Considerando que todos os pacientes já haviam assinado Termo de Consentimento Livre e Esclarecido (TCLE) para participação no estudo clínico, foi solicitado ao Comitê de Ética em Pesquisa (CEP) do Centro Infantil Boldrini (CAAE 51991015.0.0000.5376) a dispensa de assinatura de um novo TCLE ou Termos de Assentimento, pedidos estes que foram aprovados (Parecer: 1.383.710).

Este projeto foi adicionado como adendo ao projeto *“Método para identificar a Leucemia Linfóide Aguda (LLA) pediátrica do subgrupo BCR-ABL1-like”* cadastrado na Plataforma Brasil (CAAE 51991015.0.0000.5376) em 14.12.2015 e aprovado pelo CEP/CONEP em 20.06.2018. Todos os documentos são apresentados no item “Documentos Anexos”.

5 CAPÍTULO I: MANUSCRITO 1

Genetics of pediatric B-other Acute Lymphoblastic Leukemia

Natacha Azussa Migita^{1,2}, Patricia Yoshiokka Jotta¹, Natália Nascimento Paiva¹, Victor Sande Vasconcelos^{1,2}, Amilcar Cardoso de Azevedo¹, Silvia Regina Brandalise¹, Katlin Brauer Massirer³, and José Andrés Yunes^{1,4}

¹ Centro Infantil Boldrini, Campinas, SP, Brazil.

² Graduate Program in Genetics and Molecular Biology, IB, University State of Campinas, Campinas, SP, Brazil.

³ Center for Medicinal Chemistry (CQMED), Center for Molecular Biology and Genetic Engineering (CBMEG), University of Campinas (UNICAMP), Campinas, SP, Brazil.

⁴ Genetics Department, Faculty of Medical Sciences, University of Campinas, Campinas, SP, Brazil.

Contact information for correspondence: Dr. José Andrés Yunes, Laboratório de Biologia Molecular, Centro de Pesquisa Boldrini, Rua Marcia Mendes 619, Campinas–SP 13083-884, Brazil; Phone: (+55) 19 37879096; e-mail: andres@boldrini.org.br

Acknowledgments: The authors would like to thank Lilian Giroto Zambaldi and Daniele Ribeiro Lucon for the excellent technical help with cytogenetics analysis. We would like to thank Guilherme Giusti Navarro and Gabriel Lopes Centoducatte for helping with bioinformatics data management. NAM, VSV and NNP received scholarship from Brazilian funding agencies (CAPES and CNPq). JAY received a Productivity fellowship from the Brazilian National Counsel of Technological and Scientific Development (CNPq, process 305896/2013-0 and 301596/2017-4). This work was supported by grants to JAY from Brazilian Health's Ministry Program, PRONON (Project DRM SIPAR 25000.057709/2015-01), CNPq (process 429811/2016-0) and Instituto Ronald McDonald (project IRM 2016109). KBM received financial support through FAPESP (Grants 14/21704-9 and 14/50897-0).

Keywords: B-other Acute Lymphoblastic Leukemia; Gene Fusions; Gene Expression Profiling; Targeted-RNA sequencing.

Abstract

Important clinically and biological subtypes of leukemia has been described within 20% to 30% of pediatric B-cell precursor acute lymphoblastic leukemia (BCP-ALL) until recently described as “B-other”. Well-established diagnostic centers often use large-scale genomic analysis to identify systematically the molecular features in B-other patients, further clarifying their mutational spectrum and identifying new oncogenic subtypes. In this study, we describe the genetic alterations of a cohort of 144 pediatric B-other diagnosed samples by performing a Targeted RNA-sequencing (*Trusight RNA PanCancer, Illumina*) and using available bioinformatics analysis. We could refine the classification of groups based on recurrent non-overlapping rearrangements and in most cases, with distinct gene expression profile. Our cohort of 144 B-other cases represented approximately 32% and 28% of all pediatric BCP-ALL enrolled in the protocols GBTLI-1999 and -2009, respectively. Approximately 70% of the cases analyzed could be classified in one of the new molecular subgroups: *DUX4*-cluster (7.26%), *iAMP21* (1.45%), *ZNF384*-rearranged (1.86%), *MEF2D*-rearranged (0.82%), *NUTM1*-rearranged (0.82%), *PAX5*-driven (3.73%; including *PAX5* fusions, intragenic amplifications and *PAX5* p.P80R mutations); *AJ2E*-class fusion (including *ABL*-class, *JAK2*-class, *EPOR*-class fusions) and *CRLF2*-high ‘only’ subgroups that represents 3.9% of cases with JAK/STAT alterations (typically linked to BCR-ABL1-like or Ph-like) and; 9.96% of cases that were not classified into any of these subtypes were named B-other ‘rest’. Three out of 11 patients had *CRLF2*-high expression (mainly *P2RY8-CRLF2*) co-occurring with one *iAMP21* and two *PAX5*-driven mutations. Analysis of unsupervised hierarchical clustering, based on published gene-sets and data processing by t-SNE algorithm reveals many subgroups with specific expression signatures within B-other, especially *DUX4*-cluster. Thus, our study shows Brazilian population-based frequencies of known and novel BCP-ALL subtypes, with recurrent genetic aberrations and their associations with demographic data. This study strengthens and widens the current knowledge of B-other ALL and provides an objective basis for optimization of commercially available panels as the best option for clinical practice, through which risk stratification can be improved and actionable and druggable aberrations timewise discovered.

Introduction

Acute lymphoblastic leukemia (ALL) is the most common malignancy of childhood, accounting for 25% of pediatric cancers.¹ Over the last several decades, advances in the treatment and supportive care, have dramatically increased its 5-year survival rate to about 80-90% in developed countries.²⁻⁴ This remarkable achievement is attributed to more precise stratification into risk subentities, which is based on clinical and genetic abnormalities features, as well as early treatment responses assessed by the level of minimal residual disease (MRD).^{5,6} In this context, the key to risk stratification includes the identification of specific genetic aberrations involved in leukemogenesis, mainly chromosomal rearrangements, aneuploidy and cooperating mutations.^{7,8} A number of well-known subtypes of B-cell precursor acute lymphoblastic leukemia (BCP-ALL) can be classified by standard genetic analyses, notably t(9;22) *BCR-ABL1* (Ph-positive), the rearrangement of *KMT2A* (*MLL*) at 11q23, t(1;19) *TCF3* (*E2A*)-*PBX1*, t(17;19) *TCF3-HLF*, t(5;14) *IL3-IGH*, t(12;21) *ETV6-RUNX1*, high hyperdiploidy (HeH), and hypodiploidy.^{9,10}

Until some years ago, approximately 20% to 30% of pediatric BCP-ALL were called "B-other", because they did not have any of those characteristic genetic aberrations at diagnosis, and the underlying driver events were unknown. Extensive genetic studies of "B-other" cases have revealed groups of common, nonoverlapping aberrations that are strong independent predictive factors of outcome and separated genetic abnormalities subtypes, such as *DUX4*-cluster, *iAMP21*, *ZNF384*-rearranged, *MEF2D*-rearranged, *NUTM1*-rearranged, *PAX5*-driven, Ph-like (or *BCR-ABL1*-like) and others specific somatic mutations.¹¹⁻¹³ Those new discoveries were enabled by advanced analytical technologies, such as Next Generation Sequencing (NGS) methods. RNA-sequencing analysis have stratified a variety of subgroups and contributed to understand the molecular basis of ALL and its relationship with prognostic markers to achieve precision therapy.¹⁴ However, sophisticated interplay of known and novel gene fusions, sequence mutations, and gene expression profile, especially in Brazilian patients needs to be further addressed to refine the current classification of BCP-ALL and to reveal the incidence of various subtypes and genetic aberrations. So far, there is no data about our population and unfortunately, we need more efforts to translate these new genetic findings into diagnostic algorithms to improve therapeutic strategies in low-middle income countries.

Here, we present a genetic study where we evaluated signatures for complex abnormalities and gene expression characterization of 144 pediatric ‘B-other’ classified cases. We focused on defining new prognostic subgroups, in particular actionable lesions to be used for treatment management. For this purpose, we performed a Targeted RNA-sequencing analysis using a commercial panel to assess the feasibility and the benefits of introducing NGS technologies into routine diagnostics. We have also aimed to determine population frequencies of those novel ALL subtypes derived from ‘B-other’ and to explore their associations with demographic and clinical data. A number of recurrent gene fusions and gene expression patterns were identified. Besides validating the known and new subtypes (e.g., *iAMP21*, *ABL*-class fusions, *JAK2*-class fusions, *EPOR*-class fusions, *ZNF384*-rearranged, *MEF2D*-rearranged, *NUTM1*-rearranged) defined by chromosomal abnormalities, we could integrate genomic analysis to refine our classification based on gene expression profile and somatic point mutations, to distinguished the *DUX4*-cluster, *PAX5*-driven (diverse *PAX5* alterations, including rearrangements, intragenic amplifications or *PAX5* hotspot mutations as p.P80R; in our study, 13 cases of *PAX5*^{alt} and 5 cases of *PAX5* p.P80R) and *CRLF2*-high ‘only’ subgroups. Also, a number of cases showed a “-like” profile that share similar gene expression features but lack consistent biomarkers. Moreover, still 33.4% (48/144) of B-other cases was classified as ‘rest’ because the underlying genetic driver events could not be clearly revealed in this study. Patients with *ABL*-, *JAK2*- and *EPOR*-class fusions and *CRLF2* overexpression were significantly associated with high white blood cells (WBC) count at diagnosis ($\geq 50 \times 10^9/L$). Unpredicted association with gender was significantly found, we observed a high frequency of girls in B-other ‘rest’ and a tendency of boys in *PAX5*-driven subgroup, with no clear explanation for that. In summary, this study provides the comprehensive analysis of a selected cohort of pediatric B-other ALL patients, contributing to an enlargement view of the frequency and type of genetic alterations that can be biomarkers for risk stratification as well as targets for precision medicine.

Methods

Patients and study design

In total, 482 B-cell precursor acute lymphoblastic leukemia (BCP-ALL) patients were enrolled at Boldrini Center. This study included 266 retrospective BCP-ALL samples from children (≤ 18 years) registered into the Brazilian GBTLI ("Brazilian Childhood Cooperative Group")¹⁵ ALL-1999 and 2009 protocols at Boldrini Center. Leukemia diagnosis was established based on morphology, immunophenotype, cytogenetics, or RT-PCR as part of clinical routine diagnostics. B-other samples included belong to consecutive patients and was selected according to the negative results from molecular biology analysis and/or availability of the biological material. The study was approved by the institutional ethical committee (CAAE 51991015.0.0000.5376). All patients, parents or guardians provided informed consent for sample collection and research in accord with the Declaration of Helsinki. Clinical characteristics of the B-other and classical subtype cases are summarized in Table 1 and Supplemental Table S1.

Library preparation and Targeted RNA-sequencing

Sequenced cohort is distributed in 144 "B-other" samples (i.e., normal cytogenetics or non-classical cytogenetic abnormalities) and 40 randomly selected samples with high hyperdiploidy (HeH, > 50 chromosomes or DNA index > 1.16), *TCF3 (E2A)-PBX1*, *ETV6-RUNX1 (TEL-AML1)*, *BCR-ABL1* and *KMT2A (MLL)*-rearrangements, as classical subtype controls (Supplementary Figure S1). RNA-sequencing library was constructed using the commercial panel-based kit *TruSight RNA Pan-Cancer (Illumina)* according to the manufacturer's instruction. Libraries were selectively enriched for 1385 cancer-related genes and can detect fusions between genes targeted by the panel but also fusions of targeted genes with non-target genes. Basically, cDNA was generated from 50ng of cleaved RNA fragments, end-repaired and ligated to sequencing adapters (*Illumina*), followed by addition of sequence indexes. Small fragments of below 100 bp and unligated adapters were removed using the AMPure® beads (*Agencourt Bioscience*). Sequencing libraries were then hybridized with probes designed for the coding regions of the expressed cancer-associated genes and captured with magnetic beads to create the final stranded library. The final libraries were quantified using KAPA Biosystems qPCR kit (*KAPA Biosystems*) and, after adding PhiX control, the enriched libraries were diluted into 14-18pM and then sequenced on the MiSeq® instrument using MiSeq® Reagent Kit v3 (150 cycles). We used batches of eight samples *per run* to obtain approximately 3 million reads per sample (2x76 cycles). A detailed list of targeted regions can be obtained from Illumina database (https://support.illumina.com/sequencing/sequencing_kits/trusight-rna-pan-cancer-panel/downloads.html).

Sequencing Data Analysis

An integrative bioinformatics analysis was performed to explore the results and was divided into three operations: i) Fusion detection; ii) Important SNV and small indel analysis and; iii) Gene expression analysis and clustering. For that, a wide number of commercial, open-source, and laboratory-developed tools were used for each of these steps and default parameters were used for most of them. Further details on sequencing and bioinformatics analysis are provided in Supplementary Data.

Routine diagnostic and Molecular validation tests

FISH, RT-PCR, and Immunophenotyping by flow cytometry was used as primary detection approaches. The validation of putative fusion events was performed by routine procedures according to laboratory standard guidelines and with validated reagents. Some fusions and genetic abnormalities were orthogonally validated by RT-PCR (Supplementary Figure S3 and primers are listed in Table S2) and/or by FISH.

Statistical analysis

Associations between genetic abnormalities and demographic data were analyzed by using two-sided Fisher's exact test with Monte Carlo simulations (based on 10000 sampled tables with starting seed 2000000), and $p \leq 0.05$ was considered statistically significant. Other analyses were also performed using IBM® SPSS® Statistics, WinSTAT and Prism software, version 7.0 (GraphPad Software) and/or R software (www.r-project.org).

Results and Discussion

Clinical characteristics of B-other ALL group at diagnosis

A total of 144 patients from GBTLI-1999 (n=86) and GBTLI-2009 (n=58) protocols were classified as B-other and analyzed in this study. Statistical differences were not found for demographic parameters (including sex, age and initial WBC count) between them. The median age at diagnosis was 7.37 years and males were predominant (55.6%). Our B-other cohort comprised around 86% standard risk (SR) and 14% high-risk (HR) patients at diagnosis by NCI risk classification. The GBTLI-1999 and GBTLI-2009 considered other parameters as criteria to risk at the end of induction. After re-classifying those who had poor response at the end of induction, we observed an increase in the frequency of patients at risk, approximately 30%, revealing the need for further improved treatment strategy (Table 1).

Table 1. Clinical characteristics of 144 pediatric B-other patients analyzed at Boldrini Children's Center with GBTLI-1999 and GBTLI-2009 protocols.

	All B-other	GBTLI-1999	GBTLI-2009	Statistical (§p value)
Total of B-other	144	86	58	
Age at diagnosis (years old)				
Median	7.37	6.93*	7.95	
Range	0.28-17.85 (average 8.27)	0.28-17.85 (average 7.84)	1.41-17.66 (average 8.90)	
Gender				0.496
Male	80 (55.6%)	50 (58.1%)	30 (51.7%)	
Female	64 (44.4%)	36 (41.9%)	28 (48.3%)	
WBC count at diagnosis ($\times 10^9/L$)				
Median	21.34	28.65	14.03	
Range	1.3-459.0	1.4-459.0	1.3-240.0	
WBC count at risk				0.1025
< 50 $\times 10^9/L$	97	53 (61.6%)	44 (75.9%)	
$\geq 50 \times 10^9/L$	47	33 (38.4%)	14 (24.1%)	
NCI risk at diagnosis				0.3398
Standard risk	124 (86.1%)	72 (83.7%)	52 (89.6%)	
High risk	20 (13.9%)	14 (16.3%)	6 (10.4%)	
GBTLI** risk group				0.199
Standard risk	100 (69.5%)	56 (65%)	44 (76%)	
High risk	44 (30.5%)	30 (35%)	14 (24%)	

ALL, acute lymphoblastic leukemia; WBC, white blood cell count; *One patient with outlier age (21 year) was registered and treated with GBTLI-1999. **GBTLI risk group, re-classification at the end of induction after poor or unsatisfactory response to the treatment as follow; GBTLI-1999 protocol considered patients to be at SR if they met all the following criteria: age ≥ 1.00 and < 10 years; white blood cells (WBC) count at diagnosis $< 50 \times 10^9/L$; a rapid early response day 7 (WBC $< 50 \times 10^9/L$), day 14 (no peripheral blasts and $< 25\%$ blasts in bone marrow) and day 28 ($< 5\%$ blasts in BM) of induction therapy. The prognostic value of WBC count at day 7 was validated on the previous GBTLI ALL-93 protocol (Brandalise SR, 1996).¹⁶ Mediastinal mass, central nervous system (CNS) or other extramedullary leukemic involvement, as well as immunophenotype and cytogenetics, were not considered for risk classification. Patients with mature B-cell ALL were excluded. Bone marrow (BM) specimens were collected on days 14 and 28 of induction. For the GBTLI-2009, risk classification considered minimum residual disease after induction with chemotherapy, involvement in central nervous system and testis. § Fisher test was used to calculate cohort p-value between protocols. The results from this analysis shows that there are no significant differences, so all patients were analyzed together.

Recurrent fusion genes identified in B-other ALL by Targeted RNA-seq

The gene fusions were analyzed by four different bioinformatics pipelines, such as TopHat-Fusion, Manta, DRAGEN and Cicero (detailed in Supplemental Methods). Next, all predicted fusions were manually curated based on quality score, detection by more than one method and knowledge-based of expected true fusions. Exploring our cohort of 144 B-other pediatric samples, *i.e.*, annotated as negative for fusion genes by the standard techniques, we validated 85 samples as negative for reliable fusion genes. In the remaining 59 investigated samples (41%), we identified recurrent as well as novel events with prognostic significance, a subset of which were validated by orthogonal methods (Figure 1 and Supplemental Results).

Our cohort included 11 fusions involving *ABL*-class genes (*i.e.*, *ABL1* and *PDGFRB*) and *JAK/STAT*-class genes (*i.e.*, *JAK2* and *EPOR*), and we grouped them as “AJ2E-class fusion”. Commonly missed by many popular fusion detection tools, we found one *IGH-EPOR* fusion by CICERO only, and later confirmed by overexpression of *EPOR* gene (Supplemental Figure S2). Of note, those alterations are included as part of the provisional risk subgroup called Ph-like (or *BCR-ABL1*-like)^{17,18}, which were recently described in the B-ALL classification by the World Health Organization (WHO)¹⁰. Ph-like are characterized by a specific gene expression profile similar to that of Ph-positive ALL and lack of single founding genomic alteration. Indeed, this subtype are defined by multiple genetic aberrations, including over 30 rearrangements involving kinase, cytokine, and cytokine-receptor genes. These alterations can be classified into discrete subgroups, based on the signaling pathway they affect. Thus, in our study, common abnormalities as *ABL*-class fusions, *JAK2*-class fusions and *EPOR* fusions were considered as a single molecular subgroup, which we call AJ2E-class fusion, and *CRLF2* alterations (*IGH-CRLF2* and *P2RY8-CRLF2*) which result in *CRLF2* overexpression were considered as other entity (*CRLF2*-high subgroup).

In addition, this study reports several *PAX5* fusions, including *PAX5-ZNF521*, *PAX5-AUTS2*, *PAX5-CBFA2T3*, *PAX5-CBFA2T2*, *PAX5-ELN*, *PAX5-ZBED3*, *PAX5-ETV6*, *PAX5-ZCCHC7* and *PAX5-MLLT3*, with the last two cases carrying also a *P2RY8-CRLF2* rearrangement. These alterations are classified into discrete subgroups of *PAX5*-driven, called *PAX5alt*, including all *PAX5* rearrangements and, sequence mutations and intragenic amplification (*i.e.*, *PAX5*^{AMP}, which we found in 4 patients). The second subgroup within *PAX5*-driven subtype, called *PAX5* P80R was found in 5 cases of our cohort, its described to exhibit distinct

gene expression profile and is characterized by the p.Pro80Arg mutation that leads to loss of *PAX5* activity. *PAX5* mutations has been observed in 2% of B-ALL and can be altered by different abnormalities, such as deletions, gene fusions, and point mutations.¹⁹ In our cohort *PAX5*-driven represents 3.7% of BCP-ALL.

We report the identification of recurrent rearrangements involving the genes *ZNF384* and *MEF2D* in approximately 6.3% (9/144) and 2.8% (4/144) within B-other cases, respectively. The identification of recurrent rearrangements involving the genes *ZNF384* and *MEF2D* are complementary to studies described by Hirabayashi, S et al., (2017)²⁰, Gu, Z. et al, (2016)²¹, and reviewed by Lilljebjörn, H. and Fioretos, T., (2017)¹⁴, where different *ZNF384* (nine 5'-fusion partners) and *MEF2D* (six 3'-fusion partners) rearrangements have been reported. In addition to those previously reported *ZNF384*-related and *MEF2D*-related partner genes, we identified novel partners not yet described in the literature, such as *NCOA3* and *SPI1* to *ZNF384* and *PYGO2* to *MEF2D* (gene fusion depicted in Supplemental Figure S4A-B and D-E). Patients harboring *ZNF384*-related and *MEF2D*-related fusion genes have been considered a minor subgroup with prognostic relevancy and able to management therapy.^{22,23} Noteworthy, in our cohort, we could identify 2.8% (4/144) rearrangements involving *NUTM1* with *CUX1*, *KAT6A* and *ZNF618* as partner genes (Supplemental Figure S4-C). *NUTM1*-rearranged cases have been described as a rare but recurrent event in infants (2/4 cases in our cohort), particularly among those who lack *KMT2A*-rearrangements, with eight 5'-fusion partners described elsewhere.^{24,25} Our data point that *NUTM1* is normally not expressed in leukemic cells, although shows high expression only in patients with the fusion (Supplemental Figure S4-F.).

Intrachromosomal amplification of chromosome 21 (iAMP21) is an entity recognized and included by WHO as risk subgroup.¹⁰ It's relatively rare (~2% of cases) and represents a complex structural change within chromosome 21, including amplification of a region including the *RUNX1*, *ETS2* and *ERG* genes.^{26,27} In our cohort, we confirmed 7 patients (1.45% of 482 BCP-ALL cases) by FISH, showing 3 or more copies of *RUNX1* gene, clustered on a single abnormal chromosome 21. We could also validate iAMP21 subtype using MLPA kit (*P327 iAMP21-ERG*, MRC Rolland) with the density of probes along the long arm of chromosome 21 showing a high peak ratio (≥ 1.8) (Supplemental Figure S5). In addition, one sample also harbor *P2RY8-CRLF2* fusion, described before as being present in around 20% of iAMP21.²⁸

We detected four B-other patients that harbored *ETV6*-rearranged alterations (i.e., *ETV6-BBS9*, *ETV6-MANSC1*, *ETV6-INO80D* and *ETV6-LINC01252*). Rare or sporadic fusion genes were found in eight patients, and some fusion could be confirmed by RT-PCR (Supplemental Figure S3). Furthermore, fusion detection by Targeted RNA-seq was highly consistent with results obtained from cytogenetic and RT-PCR analyses for classical genetic markers as *BCR-ABL1*, *ETV6-RUNX1*, *TCF3-PBX1* and rearrangements involving *KMT2A* (i.e., *AFF1*(chr4), *MLLT1* (chr19) and *MLLT3* (chr9)). In our analysis, some patients harboring those primary alterations showed additional gene fusions, as *TFG-GPR128* and *IKZF1-ZPBP* with *BCR-ABL1* and *PAX5-ZCCHC7* with *ETV6-RUNX1* alteration. In contrast, the *TCF3-PBX1* and *KMT2A*-rearranged cases were not seen any. Despite the low number of positive controls analyzed, this implies that the Targeted RNA-seq analysis provided a robust overview of the entire fusion gene landscape of ALL. Overall, comparison of these controls results revealed consistency in n=38/40 cases, albeit with two undetected *ETV6-RUNX1* fusion in RT-PCR analyses (data not shown).

Distinct Gene Expression Groups and Associations with Genetic Abnormalities

In addition to identifying recurrent fusion genes, Targeted RNA-seq simultaneously measures the expression of all captured genes within each sample. Thus, to comprehensively identify BCP-ALL subtypes, we attempted to classify subgroups based on their gene expression signatures by Hierarchical Clustering Analysis (HCA) and using a gene-sets listed in our panel which is representative of subtype-specific gene expression profile from published studies (Figure 2A-E. and the list of gene-sets used are in the Supplemental data). Using 27 genes out of 187 genes published by Yasuda T. et al, 2016²⁹, we successfully separated 35 out of 85 cases as negative for fusions as *DUX4*-cluster subgroup with a distinct gene expression profile (Figure 2A-“Re-classification”). In our cohort, *DUX4*-cluster account for 7.3% of BCP-ALL. For most patients, we could not verify clearly the fusion transcripts supporting the presence of *DUX4* rearrangement, but in accordance with published studies, *DUX4*-cluster patients had significantly higher *DUX4* expression compared to remaining ALLs (Supplementary Figure 6B-C). *DUX4*-cluster subgroup is enriched in *ERG* deregulation, i.e., aberrant *ERG* expression, expression of *ERGalt* (an alternative isoform) and often associated with *ERG* deletion,³⁰ which we

also observed for our samples (Supplementary Figure S6). One *DUX4*-cluster patient also showed an *ETV6*-rearrangement (Figure 3A).

To verify co-clustering of Ph-like and *ETV6*-*RUNX1*-like subtypes with respective classical genetic groups, we amended the cohort with four *BCR-ABL1*-positive and twenty-four *ETV6-RUNX1*-positive ALLs, and we also added other classical subtypes, such as *TCF3-PBX1* (n=3), *KMT2A*-rearrangements (n=6) and HeH (n=3) cases. Unlike *DUX4*-cluster and *ETV6-RUNX1*-positive/-like ALL, Ph-like did not consistently co-cluster together in any unsupervised HCA, confirming that the gene expression profile of Ph-like ALL is not well defined. Moreover, using the array expression data previously performed in a large cohort (205 patients, which includes some samples from our 144 B-other) and with more classical subtypes, we also noticed the non-specific pattern of Ph-like expression which overlaps with many well-defined genetic abnormalities (e.g., *iAMP21*, *DUX4*-cluster, *PAX5*-driven and *ZNF384*-rearranged) (Figure 2 – “Ph-like PAM”), also noticed by Reshmi et al, 2017³¹. As a consensus gene expression profile to define this subgroup has failed to emerge, many independent international studies have been using different panels to identify these cases. Given the non-specificity of Ph-like signature, we decided to not stratify them as a subgroup within B-other.

Among recently reported abnormalities in ALL, a subset of BCP-ALL patients has been characterized by overexpression of the *CRLF2* (cytokine receptor-like factor 2) gene, which is involved in controlling cell proliferation via JAK/STAT pathway.³¹ The distribution of *CRLF2* gene expression levels in the total cohort were analyzed by ranking the expression values, and a sigmoid curve was observed with “outliers” expression (Supplemental Figure S7A.). Eleven patients out of 184 presented *CRLF2* expression 3 times higher than the median expression (values in log 2 TPM+1). Moreover, 6 out of 11 resulted to be positive for the *P2RY8-CRLF2* fusion by RT-PCR and 4/11 has *IGH-CRLF2* translocation by FISH analysis and only one was negative for *P2RY8-CRLF2* (unfortunately with no available material for FISH analysis) (Supplemental Figure S7B-C.). *CRLF2*-high patients were found to co-occurs with *JAK2* mutations (n=6) and *CRLF2* mutation (p.F232C, n=3), and these genetic lesions can collaborate to induce ligand-independent activation of downstream signal transduction pathways.

In addition, gene expression from Targeted RNA-seq was mapped by using unsupervised clustering with t-distributed stochastic neighbor embedding (tSNE)

analysis.³² Overall, *DUX4*-cluster and *ETV6-RUNX1* displayed a very distinct gene expression profile, as expected (Figure 2F.). Hyperdiploidy, *PAX5*-driven and groups defined by rearrangements as *KMT2A*-rearranged, *TCF3-PBX1*, *ZNF384*-rearranged, *MEF2D*-rearranged and *NUTM1*-rearranged showed a peculiar gene expression profile. Unexpectedly, the Ph-positive cases and Ph-like alterations (AJ2E-class fusions and *CRLF2*-high subgroups) showed the most undefined gene expression profiles. B-other ‘rest’ cases who lack the founding rearrangements were clustered with similar gene expression profiles to established subtypes. Such cases commonly harbored alternate chromosomal rearrangements—for example alternate *ETV6* fusions (*ETV6-BBS9* and *ETV6-INO80D*), which was found in two cases and can be classified as *ETV6-RUNX1*-like. In addition, it suggested that cases with a modal chromosome number 47–50 and similar gene expression profiles to those in high hyperdiploid ALL (characterized by a non-random gain of chromosomes X, 4, 6, 10, 14, 17, 18 and 21) can be defined as low hyperdiploid (in our cohort nine B-other ‘rest’ cases with at least one chromosome gain were clustered together with high hyperdiploid ALL). For the others, we did not find any common genetic lesion that can explain the co-clustering with the established ALL subtypes.

Combining and using different gene-sets already published, we could verify a clustering tendency of subtypes with well-defined genetic abnormalities. These data suggest that the global transcription changes in tumor cells vary markedly depending on the genetic lesion that underlies the initiation of the leukemic process. Although, we could not use all genes listed in the published gene-sets, the segregate expression signatures showed us the future utility of our panel to define known and novel subtypes within B-other. Thus, from the 144 patients diagnosed as B-other, seven subgroups were genetically depicted in this work: *DUX4*-cluster, *iAMP21*, *ZNF384*-rearranged, *MEF2D*-rearranged, *NUTM1*-rearranged, AJ2E-class fusions and *CRLF2*-high ‘only’ cases (Figure 3C.).

Association of *GATA3* SNPs with the selected B-other subgroup

Through a series of independent genome-wide association studies (GWAS) in different ethnic populations, several risk loci for ALL susceptibility have been identified (e.g., *ARID5B*, *IKZF1*, *CEBPE*, *PIP4K2A*, *CDKN2A* and *GATA3*), and associated with distinct ALL subtypes.^{33,34} Notably, two inherited genetic variation in *GATA3*, rs3824662 and rs3781093, are strongly associated with treatment outcomes and

particularly relevant in those of Hispanic ancestry.³⁵ It's known that GATA3 as a novel susceptibility locus is especially impacted by subtypes with somatic lesions underlying cytokines and kinase-activating alterations as in Ph-like.³⁶ In our replication analysis, risk alleles at both GATA3 SNPs were consistently overrepresented in *CRLF2*-high'only' subgroup compared to the frequencies calculated by Perez-Andreu V et al, 2013³⁴. The C risk allele at rs3781093 and the A risk allele at rs3824662 were overrepresented in *CRLF2*-high'only', followed by *MEF2D*-rearranged and AJ2E-class fusion. Interestingly, B-other 'rest' has the comparative frequency of risk allele as samples non-ALL. Although many studies report the risk allele at rs3824662 with significant effect in the disease, here we did not find a strong association with outcome, and modest decrease in survival percentage were only seen for the rs3781093 risk allele (but with no statistical significance) (Figure 3A and Supplemental Figure S8).

Demographic features and distribution of B-other

After demonstrating that B-other were characterized by multiple new targets of rearrangement with distinct transcriptional signatures, we have undertaken a discriminative approach to identify demographic features particularly associated with those groups. After testing for gender biases, we observed that girls were significantly more enriched in B-other 'rest' subgroup ($p=0.049$), whereas boys showed a tendency in *PAX5*-driven subgroup (Fisher exact test, $p=0.043$; Monte Carlo, $p=0.071$) (Figure 3B.). While currently we cannot fully understand the mechanisms that underlies gender bias, this observation sheds new light on possible gender differences in the stages of leukemogenesis. AJ2E-class and *CRLF2*-high (commonly grouped as Ph-like) were significantly associated with higher white blood cell (WBC) count at diagnosis, $p<0.007$ and $p<0.042$, respectively. Moreover, AJ2E-class fusion cases were associated with NCI high risk classification ($p<0.038$). On the other hand, B-other 'rest' were associated with standard risk group in GBLTI classification ($p=0.007$). In total, 8 out of 144 B-other ALLs had CALLA negative immunophenotype (CALLA/CD10, cutoff 20%) and those were significantly enriched in *ZNF384*-rearranged ($p<10^{-7}$). No other strongly correlation were observed among the other subtypes (Supplemental Table S4.).

Conclusion

The major contribution of our study was the detection of genetic alterations in samples from children with B-other ALL by Targeted RNA-sequencing analysis. The typically repertoire used in clinical diagnostic to profile genetic alterations of BCP-ALL include: immunophenotyping, karyotyping, FISH, or RT-PCR. Those methods restrict the number of genetic loci that can simultaneously be analyzed and are limited to the classical ALL subtypes. Performing deep analysis using Targeted RNA-seq revealed the complex genetic composition of our B-other ALL cohort. Our results provided a substantial number of novel genetic abnormalities and revealed additional subgroups with recurrent and non-overlapping rearrangements defining distinct gene expression profiles. Thus, our data, contributes to new knowledge about the frequency of druggable and actionable risk groups. One of the limitations in our study was the regionally restricted group of patients, which may contribute to discrepancies with other studies; therefore, collaborative efforts with larger sample sizes are needed. In addition, copy number alterations (CNAs) and SNVs mutations play important roles in leukemogenesis and even in prognostic risk, therefore, in the future, more comprehensive and integrative approaches can be warranted to dissect the underlying complexity of ALL. The integration of the NGS panel in clinical practice could complement the standard diagnostic techniques and allow substantial improvements in both patient risk stratification and precision medicine. Noteworthy, NGS methods are still relatively expensive, but has the potential to be at least partially optimized, which may reduce costs. To this end, we aim in the future develop an RNA-seq custom panel to targeted only leukemia-related genes, and perhaps, make this technique more accessible to other diagnostic centers.

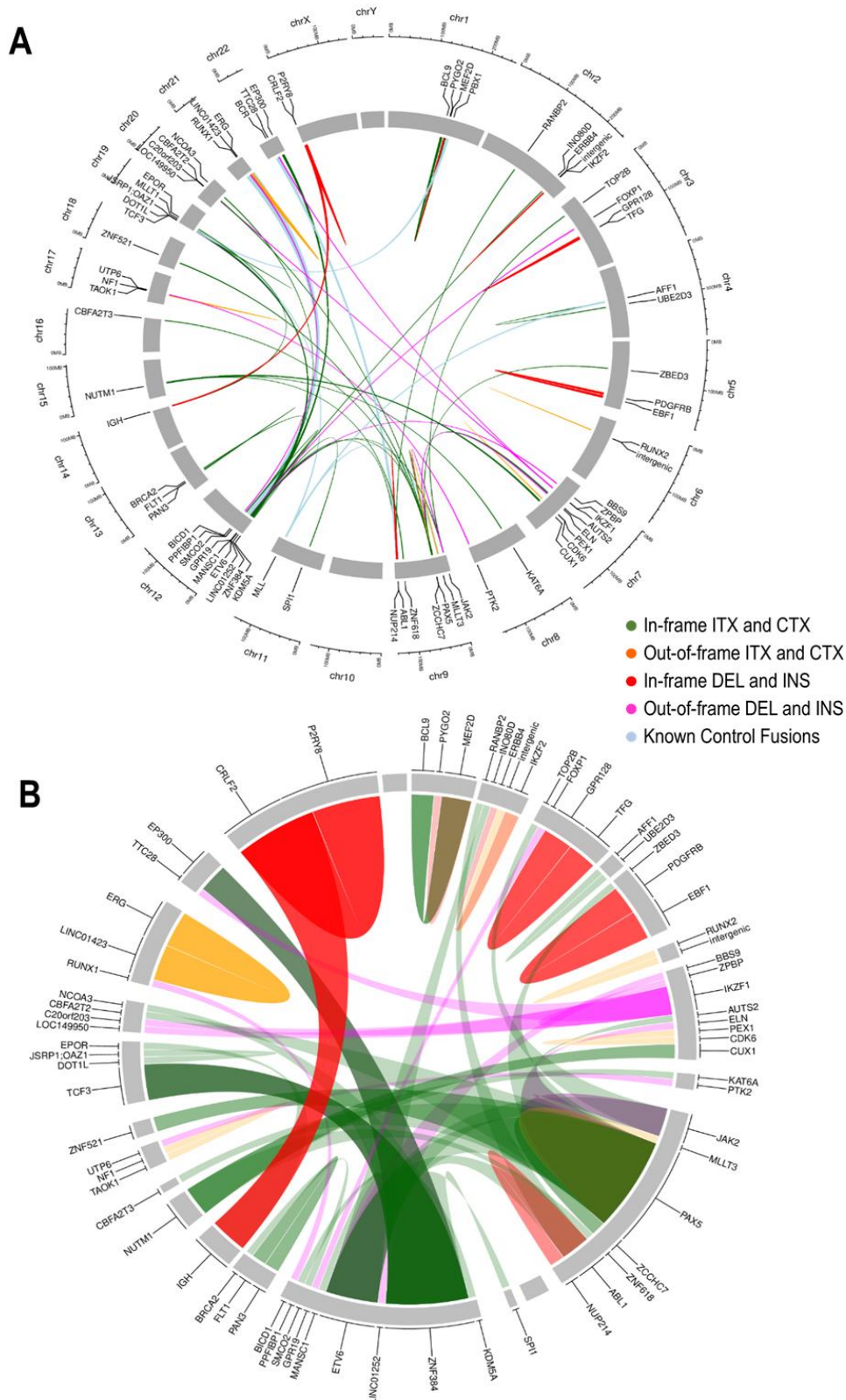
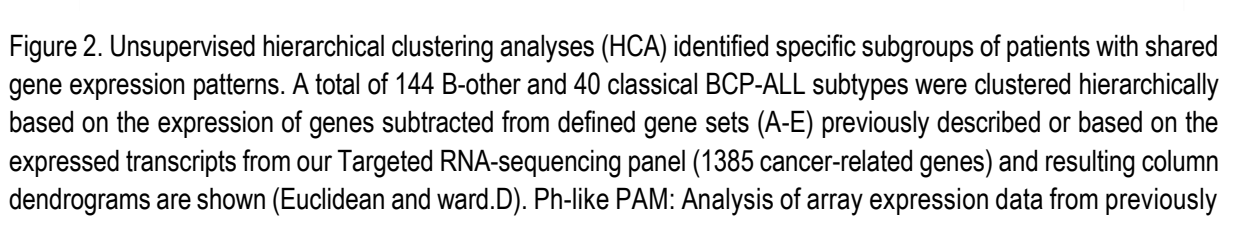
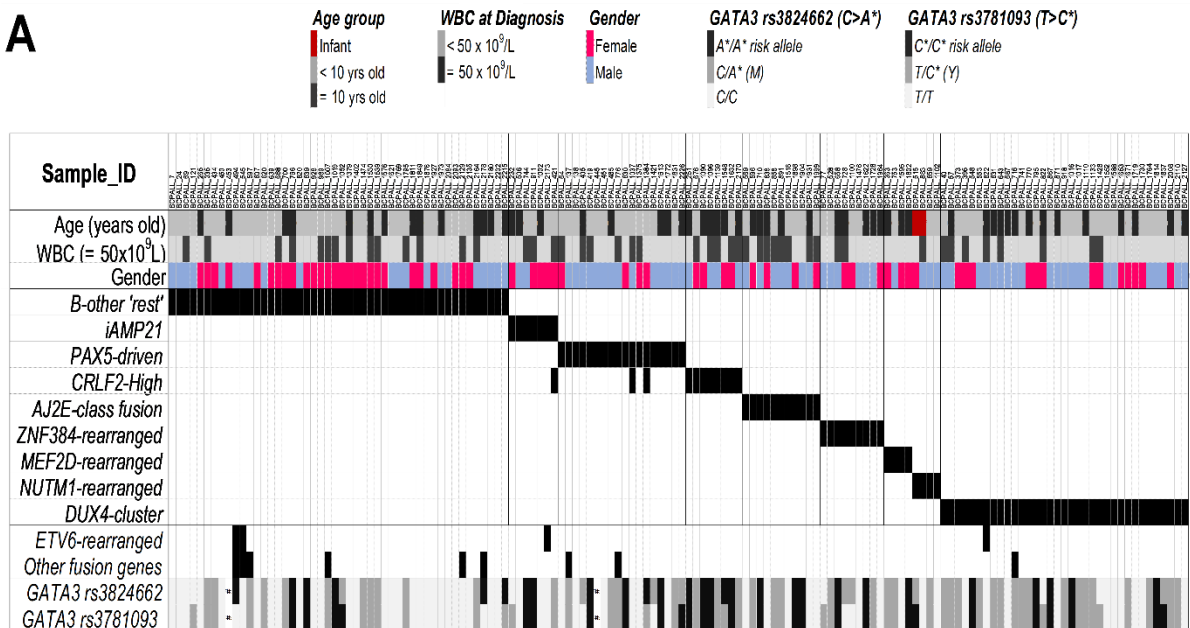
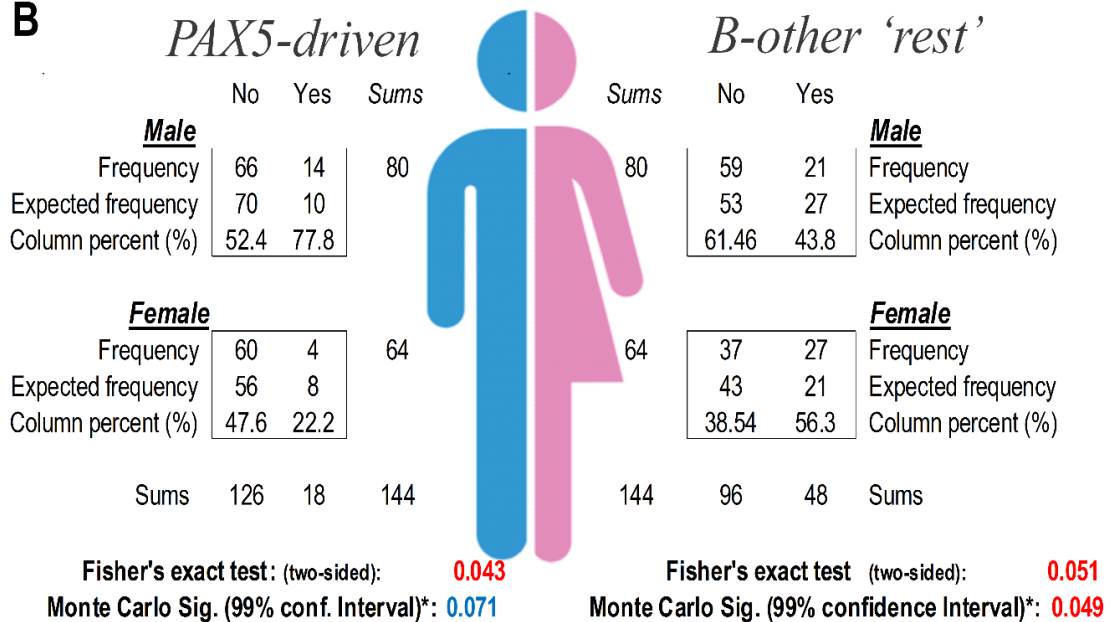


Figure 1. Circos plot showing the overview of gene fusions discovered by Targeted RNA-sequencing and schematically organized into chromosome position. The line links the two partner genes in a fusion and the colored legend shows the rearrangement type. **A.** Genes are arranged according to their genomic position in the chromosome (clockwise from chromosome 1-22 followed by X and Y). All well-known gene fusions are shown as light blue ribbons. **B.** Zooming fusion genes in to show in more details the rearrangements found specifically in B-other classified cases. Ribbon widths are proportional to the frequency of a specific gene in a fusion event.



Mutações genéticas na Leucemia Linfóide Aguda pediátrica do subgrupo “B-other”

studies were used to perform a Prediction Analysis for Microarrays (PAM). Gene sets are described in Supplementary Table S3. Genetic annotation is split in three lanes: 1st lane shows initial classification with classical subtypes, 2nd lane shows our finds with recurrent genetic aberrations, and 3rd lane shows a discriminative attempt to define Ph-like by PAM using array expression data. F. Gene expression signatures of 144 B-other cases and 40 well-known subtypes shown in a two-dimensional tSNE plot. Each dot represents a sample colored by subgroup. The 731 most variably expressed genes were selected and processed by the tSNE algorithm with a perplexity score of 30. t-SNE, t-distributed Stochastic Neighbor Embedding.

A**B**

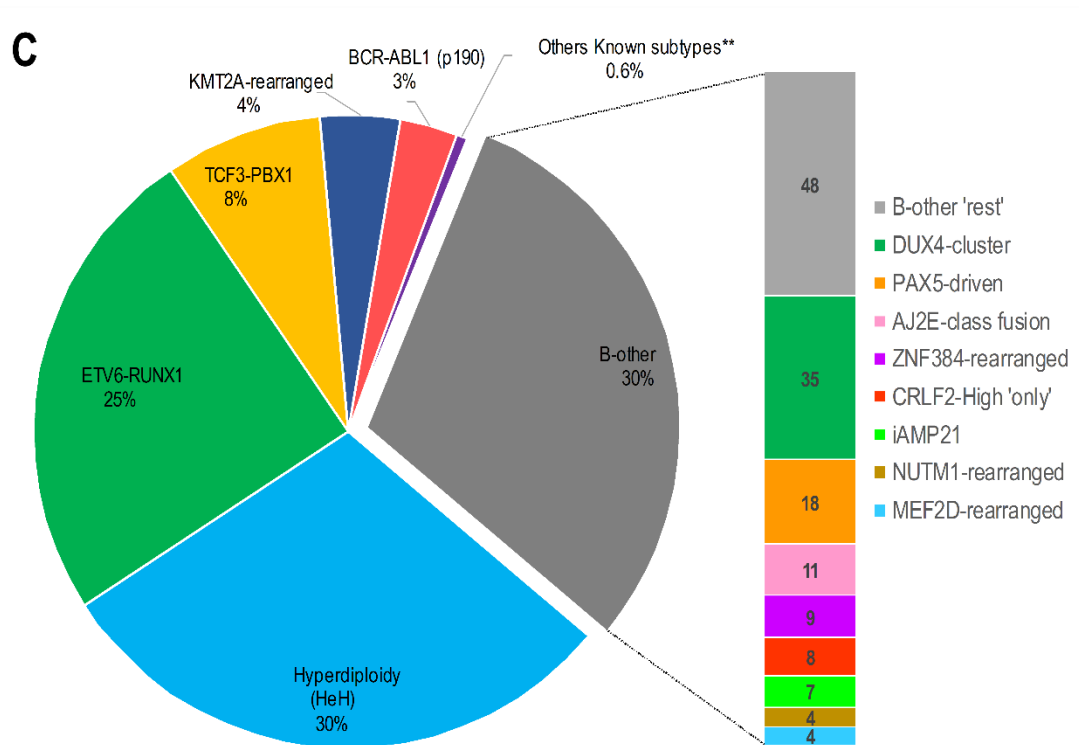


Figure 3. Frequency and demography associations within B-other subgroups. A. Graphical overview of molecular subgroups, demographic and clinical characteristics of all 144 B-other acute lymphoblastic leukemia (ALL) cases. Subgroups are clustered based on recurrent genetic abnormalities defined by gene fusions and gene-expression profile detected by Targeted RNA-sequencing. Categories as age (yrs, years), WBC; initial white blood cell count, and gender are in the top. Each column represents a single patient. B. Gender biases observed in B-other 'rest' and *PAX5*-driven. C. Frequency and distribution of ALL subtypes among 482 BCP-ALL patients consecutively diagnosed in the GBLTI-1999 and GBLTI-2009 at Boldrini's Center. The pie chart depicts the percentages of each classical subtypes and the genetic characterization and frequency of all B-other cases analyzed in this study. In addition to subtypes defined by gene rearrangements, *PAX5*-driven includes intrachromosomal amplifications and *PAX5* p.P80R mutation. Aberrations which are not well-defined with genetic mutation (e.g., Ph-like and ETV6-RUNX1-like) are not shown. Three patients with *CRLF2*-high expression were found in co-occurrence with *PAX5*-driven and *iAMP21*, but the remaining 8 patients were studied as *CRLF2*-high 'only' classification.

AUTHOR CONTRIBUTIONS

Contribution: N.A.M. and J.A.Y. conceived, designed the study, performed the data interpretation and wrote the manuscript; N.A.M. and P.Y.J performed all Targeted RNA-sequencing experiments; N.A.M performed the validation assays by RT-PCR; P.Y.J., G.L.C. and N.P.N. contributed with GATA3 polymorphisms analysis by Sanger sequencing; N.A.M., V.S.V. and K.B.M. contributed with bioinformatics analysis; A.C.A and S.R.B. collected the samples and patient's clinical data; and N.A.M. and J.A.Y. performed statistical analysis. All authors revised the manuscript.

Conflict-of-interest disclosure: The authors declare no competing financial interests.

References

- ¹ Ward E, DeSantis C, Robbins A, Kohler B, Jemal A. Childhood and adolescent cancer statistics, 2014. *CA Cancer J Clin*. 2014 Mar-Apr;64(2):83-103. doi: 10.3322/caac.21219. Epub 2014 Jan 31.
- ² Schrappe M, Reiter A, Ludwig WD, Harbott J, Zimmermann M, Hiddemann W, Niemeyer C, Henze G, Feldges A, Zintl F, Kornhuber B, Ritter J, Welte K, Gadner H, Riehm H. Improved outcome in childhood acute lymphoblastic leukemia despite reduced use of anthracyclines and cranial radiotherapy: results of trial ALL-BFM 90. German-Austrian-Swiss ALL-BFM Study Group. *Blood*. 2000 Jun 1;95(11):3310-22.
- ³ Moorman AV, Ensor HM, Richards SM, Chilton L, Schwab C, Kinsey SE, Vora A, Mitchell CD, Harrison CJ. Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: results from the UK Medical Research Council ALL97/99 randomised trial. *Lancet Oncol*. 2010 May;11(5):429-38. doi: 10.1016/S1470-2045(10)70066-8. Epub 2010 Apr 19. Erratum in: *Lancet Oncol*. 2010 Jun;11(6):516.
- ⁴ Hunger SP, Lu X, Devidas M, Camitta BM, Gaynon PS, Winick NJ, Reaman GH, Carroll WL. Improved survival for children and adolescents with acute lymphoblastic leukemia between 1990 and 2005: a report from the children's oncology group. *J Clin Oncol*. 2012 May 10;30(14):1663-9. doi: 10.1200/JCO.2011.37.8018. Epub 2012 Mar 12.
- ⁵ Michael J, Borowitz M, Meenakshi Devidas, Stephen P. Hunger, W. Paul Bowman, Andrew J. Carroll, William L. Carroll, Stephen Linda, Paul L. Martin, D. Jeanette Pullen, David Viswanatha, Cheryl L. Willman, Naomi Winick, Bruce M. Camitta; for the Children's Oncology Group13, Clinical significance of minimal residual disease in childhood acute lymphoblastic leukemia and its relationship to other prognostic factors: a Children's Oncology Group study. *Blood* 2008; 111 (12): 5477–5485. doi: <https://doi.org/10.1182/blood-2008-01-132837>
- ⁶ Schrappe M, Bleckmann K, Zimmermann M, Biondi A, Möricke A, Locatelli F, Cario G, Rizzari C, Attarbaschi A, Valsecchi MG, Bartram CR, Barisone E, Niggli F, Niemeyer C, Testi AM, Mann G, Ziino O, Schäfer B, Panzer-Grümayer R, Beier R, Parasole R, Göhring G, Ludwig WD, Casale F, Schlegel PG, Basso G, Conter V. Reduced-Intensity Delayed Intensification in Standard-Risk Pediatric Acute Lymphoblastic Leukemia Defined by Undetectable Minimal Residual Disease: Results of an International Randomized Trial (AIEOP-BFM ALL 2000). *J Clin Oncol*. 2018 Jan 20;36(3):244-253. doi: 10.1200/JCO.2017.74.4946. Epub 2017 Nov 17.
- ⁷ Roberts KG, et al. (2012) Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer Cell* 22:153–166.
- ⁸ Inaba H, Mullighan CG. Pediatric acute lymphoblastic leukemia. *Haematologica*. 2020 Nov 1;105(11):2524-2539. doi: 10.3324/haematol.2020.247031.
- ⁹ Hunger SP, Mullighan CG. Acute Lymphoblastic Leukemia in Children. *N Engl J Med*. 2015 Oct 15;373(16):1541-52. doi: 10.1056/NEJMra1400972.
- ¹⁰ Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M, Vardiman JW. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood*. 2016 May 19;127(20):2391-405. doi: 10.1182/blood-2016-03-643544. Epub 2016 Apr 11.
- ¹¹ Liu YF, Wang BY, Zhang WN, Huang JY, Li BS, Zhang M, Jiang L, Li JF, Wang MJ, Dai YJ, Zhang ZG, Wang Q, Kong J, Chen B, Zhu YM, Weng XQ, Shen ZX, Li JM, Wang J, Yan XJ, Li Y, Liang YM, Liu L, Chen XQ, Zhang WG, Yan JS, Hu JD, Shen SH, Chen J, Gu LJ, Pei D, Li Y, Wu G, Zhou X, Ren RB, Cheng C, Yang JJ, Wang KK, Wang SY, Zhang J, Mi JQ, Pui CH, Tang JY, Chen Z, Chen SJ. Genomic Profiling of Adult and Pediatric B-cell Acute Lymphoblastic Leukemia. *EBioMedicine*. 2016 Jun;8:173-183. doi: 10.1016/j.ebiom.2016.04.038. Epub 2016 May 13.
- ¹² Gu Z, Churchman ML, Roberts KG, Moore I, Zhou X, Nakitandwe J, Hagiwara K, Pelletier S, Gingras S, Berns H, Payne-Turner D, Hill A, Iacobucci I, Shi L, Pounds S, Cheng C, Pei D, Qu C, Newman S, Devidas M, Dai Y, Reshmi SC, Gastier-Foster J, Raetz EA, Borowitz MJ, Wood BL, Carroll WL, Zweidler-McKay PA, Rabin KR, Mattano LA, Maloney KW, Rambaldi A, Spinelli O, Radich JP, Minden MD, Rowe JM, Luger S, Litzow MR, Tallman MS, Racevskis J, Zhang Y, Bhatia R, Kohlschmidt J, Mrózek K, Bloomfield CD, Stock W, Kornblau S, Kantarjian HM, Konopleva M, Evans WE, Jeha S, Pui CH, Yang J, Paietta E, Downing JR, Relling MV, Zhang J, Loh ML, Hunger SP, Mullighan CG. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat Genet*. 2019 Feb;51(2):296-307. doi: 10.1038/s41588-018-0315-5. Epub 2019 Jan 14.
- ¹³ Zaliouva M, Stuchly J, Winkowska L, Musilova A, Fiser K, Slamova M, Starkova J, Vaskova M, Hrusak O, Sramkova L, Stary J, Zuna J, Trka J. Genomic landscape of pediatric B-other acute lymphoblastic leukemia in a consecutive European cohort. *Haematologica*. 2019 Jul;104(7):1396-1406. doi: 10.3324/haematol.2018.204974. Epub 2019 Jan 10.
- ¹⁴ Lilljebjörn H, Fioretos T. New oncogenic subtypes in pediatric B-cell precursor acute lymphoblastic leukemia. *Blood*. 2017 Sep 21;130(12):1395-1401. doi: 10.1182/blood-2017-05-742643. Epub 2017 Aug 4..
- ¹⁵ Sociedade Brasileira de Oncologia Pediátrica. Protocolo brasileiro de tratamento da leucemia linfóide aguda na infância GBTLLA-2009. São Paulo: Campinas; 2011.
- ¹⁶ Brandalise SR: Prognostic value of day 8 peripheral blood response for children with acute lymphocytic leukemia, in Zander AR, (ed): *Gene Technology: Stem Cell and Leukemia Research*. Berlin, Springer-Verlag, 1996, pp 421-428.
- ¹⁷ Den Boer, M.L.; van Slegtenhorst, M.; De Menezes, R.X.; Cheok, M.H.; Buijs-Gladdines, J.G.; Peters, S.T.; Van Zutven, L.J.; Beverloo, H.B.; Van der Spek, P.J.; Escherich, G.; et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: A genome-wide classification study. *Lancet Oncol*. 2009, 10, 125–134.

- ¹⁸ Mullighan, C.G.; Su, X.; Zhang, J.; Radtke, I.; Phillips, L.A.; Miller, C.B.; Ma, J.; Liu, W.; Cheng, C.; Schulman, B.A.; et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N. Engl. J. Med.* 2009, 360, 470–480.
- ¹⁹ Li JF, Dai YT, Lilljebjörn H, Shen SH, Cui BW, Bai L, Liu YF, Qian MX, Kubota Y, Kiyoi H, Matsumura I, Miyazaki Y, Olsson L, Tan AM, Ariffin H, Chen J, Takita J, Yasuda T, Mano H, Johansson B, Yang JJ, Yeoh AE, Hayakawa F, Chen Z, Pui CH, Fioretos T, Chen SJ, Huang JY. Transcriptional landscape of B cell precursor acute lymphoblastic leukemia based on an international study of 1,223 cases. *Proc Natl Acad Sci U S A.* 2018 Dec 11;115(50):E11711-E11720. doi: 10.1073/pnas.1814397115. Epub 2018 Nov 28.
- ²⁰ Hirabayashi S, Ohki K, Nakabayashi K, Ichikawa H, Momozawa Y, Okamura K, Yaguchi A, Terada K, Saito Y, Yoshimi A, Ogata-Kawata H, Sakamoto H, Kato M, Fujimura J, Hino M, Kinoshita A, Kakuda H, Kurosawa H, Kato K, Kajiwara R, Moriwaki K, Morimoto T, Nakamura K, Noguchi Y, Osumi T, Sakashita K, Takita J, Yuza Y, Matsuda K, Yoshida T, Matsumoto K, Hata K, Kubo M, Matsubara Y, Fukushima T, Koh K, Manabe A, Ohara A, Kiyokawa N; Tokyo Children's Cancer Study Group (TCCSG). ZNF384-related fusion genes define a subgroup of childhood B-cell precursor acute lymphoblastic leukemia with a characteristic immunotype. *Haematologica.* 2017 Jan;102(1):118-129. doi: 10.3324/haematol.2016.151035. Epub 2016 Sep 15.
- ²¹ Gu, Z. *et al.* Genomic analyses identify recurrent *MEF2D* fusions in acute lymphoblastic leukaemia. *Nat. Commun.* 7, 13331 doi: 10.1038/ncomms13331 (2016).
- ²² Hirabayashi S, Butler ER, Ohki K, Kiyokawa N, Bergmann AK, Möricke A, Boer JM, Cavé H, Cazzaniga G, Yeoh AEJ, Sanada M, Imamura T, Inaba H, Mullighan C, Loh ML, Norén-Nyström U, Pastorczak A, Shih LY, Zaliouva M, Pui CH, Haas OA, Harrison CJ, Moorman AV, Manabe A. Clinical characteristics and outcomes of B-ALL with ZNF384 rearrangements: a retrospective analysis by the Ponte di Legno Childhood ALL Working Group. *Leukemia.* 2021 Nov;35(11):3272-3277. doi: 10.1038/s41375-021-01199-0. Epub 2021 Mar 10.
- ²³ Ohki K, Kiyokawa N, Saito Y, Hirabayashi S, Nakabayashi K, Ichikawa H, Momozawa Y, Okamura K, Yoshimi A, Ogata-Kawata H, Sakamoto H, Kato M, Fukushima K, Hasegawa D, Fukushima H, Imai M, Kajiwara R, Koike T, Komori I, Matsui A, Mori M, Moriwaki K, Noguchi Y, Park MJ, Ueda T, Yamamoto S, Matsuda K, Yoshida T, Matsumoto K, Hata K, Kubo M, Matsubara Y, Takahashi H, Fukushima T, Hayashi Y, Koh K, Manabe A, Ohara A; Tokyo Children's Cancer Study Group (TCCSG). Clinical and molecular characteristics of *MEF2D* fusion-positive B-cell precursor acute lymphoblastic leukemia in childhood, including a novel translocation resulting in *MEF2D-HNRNP1* gene fusion. *Haematologica.* 2019 Jan;104(1):128-137. doi: 10.3324/haematol.2017.186320. Epub 2018 Aug 31.
- ²⁴ Hormann FM, Hoogkamer AQ, Beverloo HB, Boeree A, Dingjan I, Wattel MM, et al. NUTM1 is a recurrent fusion gene partner in B-cell precursor acute lymphoblastic leukemia associated with increased expression of genes on chromosome band 10p12.31-12.2. *Haematologica.* 2019;104(Oct):e455–9.
- ²⁵ Boer, J.M., Valsecchi, M.G., Hormann, F.M. et al. Favorable outcome of NUTM1-rearranged infant and pediatric B cell precursor acute lymphoblastic leukemia in a collaborative international study. *Leukemia* 35, 2978–2982 (2021). <https://doi.org/10.1038/s41375-021-01333-y>
- ²⁶ Harrison CJ. Blood spotlight on iAMP21 acute lymphoblastic leukemia (ALL), a high-risk pediatric disease. *Blood* 2015; 125(9): 1383–1386
- ²⁷ Tsuchiya KD, Davis B, Gardner RA. Is intrachromosomal amplification of chromosome 21 (iAMP21) always intrachromosomal? *Cancer Genet.* 2017 Dec;218-219:10-14. doi: 10.1016/j.cancergen.2017.08.005. Epub 2017 Aug 31.
- ²⁸ Ryan SL, Matheson E, Grossmann V, et al. The role of the RAS pathway in iAMP21-ALL. *Leukemia* 2016; 30:1824–1831.; Russell LJ, Capasso M, Vater I, et al. Deregulated expression of cytokine receptor gene, CRLF2, is involved in lymphoid transformation in B-cell precursor acute lymphoblastic leukemia. *Blood* 2009;114:2688–2698.
- ²⁹ Yasuda, T.; Tsuzuki, S.; Kawazu, M.; Hayakawa, F.; Kojima, S.; Ueno, T.; Imoto, N.; Kohsaka, S.; Kunita, A.; Doi, K.; et al. Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults. *Nat. Genet.* 2016, 48, 569–574.
- ³⁰ Zhang, J.; McCastlain, K.; Yoshihara, H.; Xu, B.; Chang, Y.; Churchman, M.L.; Wu, G.; Li, Y.; Wei, L.; Iacobucci, I.; et al. Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nat. Genet.* 2016, 48, 1481–1489.
- ³¹ Reshmi, S.C.; Harvey, R.C.; Roberts, K.G.; Stonerock, E.; Smith, A.; Jenkins, H.; Chen, I.M.; Valentine, M.; Liu, Y.; Li, Y.; et al. Targetable kinase gene fusions in high-risk B-ALL: A study from the Children's Oncology Group. *Blood* 2017, 129, 3352–3361.
- ³² John CR, Watson D, Russ D, et al. M3C: Monte Carlo reference-based consensus clustering. *Sci Rep.* 2020;10(1):1816. Published 2020 Feb 4.
- ³³ Trevino LR, Yang W, French D, Hunger SP, Carroll WL, Devidas M, Willman C, Neale G, Downing J, Raimondi SC, Pui CH, Evans WE, Relling MV. Germline genomic variants associated with childhood acute lymphoblastic leukemia. *Nat Genet.* 2009;41:1001–5.
- ³⁴ Perez-Andreu V, Roberts KG, Harvey RC, Yang W, Cheng C, Pei D, Xu H, Gastier-Foster J, E S, Lim JY, Chen IM, Fan Y, Devidas M, et al. Inherited GATA3 variants are associated with Ph-like childhood acute lymphoblastic leukemia and risk of relapse. *Nat Genet.* 2013;45:1494–8.
- ³⁵ Raca G, Abdel-Azim H, Yue F, Broach J, Payne JL, Reeves ME, Gowda C, Schramm J, Desai D, Dovat E, Hu T, Berg AS, Bhojwani D, Payne KJ, Dovat S. Increased incidence of IKZF1 deletions and IGH-CRLF2 translocations in B-ALL of Hispanic/Latino children-a novel health disparity. *Leukemia.* 2021 Aug;35(8):2399-2402. doi: 10.1038/s41375-021-01133-4. Epub 2021 Feb 2.

-
- ³⁶ Perez-Andreu V, Roberts KG, Xu H, Smith C, Zhang H, Yang W, Harvey RC, Payne-Turner D, Devidas M, Cheng IM, Carroll WL, Heerema NA, Carroll AJ, et al. A genome-wide association study of susceptibility to acute lymphoblastic leukemia in adolescents and young adults. *Blood*. 2015;125:680–6.

Supplemental DATA for:
Genetics of pediatric B-other Acute Lymphoblastic Leukemia

CONTENT:

SUPPLEMENTAL METHODS

SUPPLEMENTAL FIGURES

Figure S1. Cohort distribution and workflow.

Figure S2. AJ2E-class fusion description.

Figure S3. Validation of known and novel fusion genes by RT-PCR.

Figure S4. ZNF384, MEF2D and NUTM1 rearrangements description.

Figure S5. iAMP21 subgroup description.

Figure S6. DUX4-cluster subgroup description.

Figure S7. CRLF2 expression and underlying genomic alterations observed in our cohort.

Figure S8. GATA3 polymorphism analysis.

SUPPLEMENTAL TABLES

Table S1: Demographic, clinical, and genetic abnormalities distribution of all pediatric patients diagnosed at Boldrini Center according to Protocols GBLTI-1999 and GBLTI-2009.

Table S2: List of primers sequence used for RT-PCR validation.

Table S3: List of Gene-sets used for HCA.

Table S4: Statistic results for demographic correlation analysis.

SUPPLEMENTAL REFERENCES

Supplemental Methods

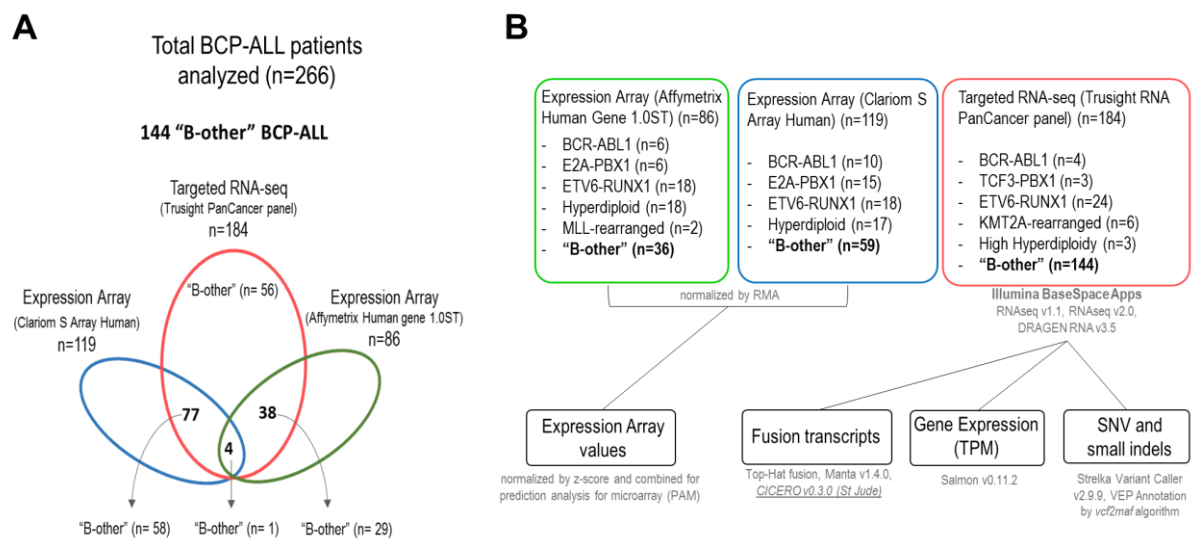
Demographic, clinical, and genetic abnormalities distribution

A total of 272 and 210 pediatric patients (≤ 18 years) with BCP-ALL were diagnosed and treated at Boldrini Children's Center (Brazil) with GBTLI-1999 and GBTLI-2009 trials, respectively. The cohort analyzed by Targeted RNA-sequencing include 144 "B-other" (*i.e.*, normal cytogenetics or non-classical cytogenetic abnormalities) and 40 randomly selected samples, such as high hyperdiploidy ($n=3$), *TCF3-PBX1* ($n=3$), *ETV6-RUNX1* ($n=24$), *BCR-ABL1* ($n=4$) and *KMT2A (MLL)*-rearrangements ($n=6$), as classical subtype controls. The clinical and cytogenetic characteristics of all patients are detailed in Table S1 and clinical data from the B-other cohort are summarized in Table 1. In total, from 482 BCP-ALL patients diagnosed at Boldrini Center, 144 (30%) were B-other.

Table S1. Demographic, clinical, and genetic abnormalities distribution of all pediatric patients diagnosed at Boldrini Center according to Protocols GBTLI-1999 and GBTLI-2009¹.

	GBTLI-1999	GBTLI-2009	Statistical (^{\$} <i>p</i> value)
	No. of patients (%)	No. of patients (%)	
Total	397	269	
Not Included or Not available (NA)*	55 (13.8%)	25 (9.3%)	0.1175
T-ALL	58 (14.6%)	30 (11.1%)	0.297
Down Syndrome (DS) ALL	12 (3.0%)	4 (1.5%)	0.303
BCP-ALL non-DS	272 (68.5%)	210 (78.1%)	0.3027
BCP-ALL (routinely diagnostic)			
Age at diagnosis (years old)			
Median	4.2**	5.04	
Range	0-17.85	0.48-17.66	
Gender			
Male	151 (55.5%)	107 (51%)	0.3575
Female	121 (44.5%)	103 (49%)	
Genetic abnormalities			
Hyperdiploidy (HeH)	67 (24.6%)	76 (36.2%)	0.0453
<i>ETV6-RUNX1</i>	79 (29%)	40 (19%)	0.0493
<i>TCF3 (E2A)-PBX1</i>	22 (8.1%)	17 (8.1%)	>0.999
<i>KMT2A (MLL)</i> -rearranged	9 (3.3%)	11 (5.2%)	0.3617
<i>BCR-ABL1</i>	8 (2.9%)	6 (2.9%)	>0.999
Others Known subtypes***	1 (0.4%)	2 (0.9%)	0.5835
B-other	86 (31.6%)	58 (27.6%)	0.503

ALL, acute lymphoblastic leukemia; WBC, white blood cell; High-Hyperdiploid (HeH, ≥ 51 chromosomes); Patients with Down Syndrome were not included in our analysis. *NA: Samples was not included due to lack of molecular biology diagnosis and/or not material available for analysis. **One patient with an outlier age (≥ 21 years old) was registered and treated with GBTLI-1999. ***Others known subtypes include Hypodiploidy and *t(8;14)*. \$Fisher test was used to calculate cohort *p*-value between protocols. The results from this analysis shows that there are no significant differences in the majority of variables. However, the frequency of HeH and *ETV6-RUNX1* were significantly different between protocols.



Supplementary Figure S1. Cohort distribution and workflow. A total of 266 patients with diagnosed BCP-ALL were enrolled in genomic studies at our center. Two hundred and five were previously analyzed by Array expression (where 77 samples analyzed with “Clariom S” overlapped with Targeted RNA-seq and 58 of them were B-other cases; and 38 samples analyzed with “Affymetrix Hugene” were also analyzed with Targeted RNA-seq and 29/38 were B-other cases; and 4 cases BCP-ALL cases were analyzed with both expression arrays and Targeted RNA-seq, with one B-other case). Thus, 184 out of 266 underwent Targeted RNA-sequencing, of whom 144 were B-other cases. A, Venn diagram showing overlap among samples analyzed by Targeted RNA-sequencing and Array expression in this study. B, Analytical workflow depicting cohorts (classical subtypes and B-other) and genomic assays. Output files of genomic data were analyzed by several pipelines followed by manual curation and controls.

Samples preparation

Mononuclear cells were purified from the diagnostic bone marrow or peripheral blood samples by density gradient centrifugation and were cryopreserved. Cells for karyotyping and FISH analysis were stored in Carnoy’s fixative. Genomic DNA and RNA was extracted from diagnostic leukemia samples using the *Illustra Mini Spin kit* (GE Healthcare Life Sciences, Chicago, IL, USA). Nucleic acids was quantified using *Qubit Fluorometer* (Thermo Fisher Scientific, Waltham, MA, USA), and quality was assessed by agarose gel electrophoresis. For sequencing, RNA integrity was checked by chip electrophoresis using the *RNA 6000 Nano Kit* on the Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA) and all samples input had a threshold of $DV_{200} > 50\%$.

Targeted RNA-sequencing Data Processing

Base calling of data from Miseq systems was performed on the instrument as the resulting *.bcl* files were filtered, de-multiplexed, and converted to paired FastQ format. The first part of data analysis was performed by using the dedicated *Illumina BaseSpace* apps RNA-seq Alignment version 1.1, RNA-seq Alignment version 2.0, and DRAGEN RNA Pipeline version 3.5 by read mapping on GRCh37/hg19 reference genome with standard settings

(<https://basespace.illumina.com/apps>). For fusion detection, in parallel, the same algorithm performed by St. Jude Hospital called CICERO² was also used. The source code is available from <https://github.com/stjude/Cicero> and the predicted gene fusions can then be imported to a visualization tool for manual curation and inspection (<https://proteinpaint.stjude.org/FusionEditor/>). Reported rearrangements were manually reviewed and were visualized in Integrative Genomics Viewer (IGV)³ to keep the reliable ones and non-correlated fusion were excluded as likely false positives. The following filter strategy was used for selection of fusion candidates: all fusions reported as 'PASS' by RNA Alignment v1.1 and v2.2, DRAGEN RNA v3.5 and CICERO were considered, unless: (1) Only one software annotated the event as 'PASS'; (2) fusion was caused by misalignment of reads to a pseudogene. In addition, fusions involving *IGH*, *ERG* and *CRLF2*, only detected by either DRAGEN RNA or CICERO were included.

The selected variants (SNVs and small indels) and gene expression data from RNA-seq Alignment version 2.0 (quantification of reference genes and transcripts by TPM method), using Strelka Variant caller and Salmon algorithms, respectively, was used.

Gene expression by microarray and analysis

Array expression was assessed using either *Affymetrix Human Gene 1.0ST array* (Affymetrix, Santa Clara, CA) and *Clariom™ array* (ThermoFisher, Waltham, MA), according to the manufacturer's protocol. Raw signals intensities were analyzed with 'affy' package⁴ and normalized by Robust Multi-array average (RMA) available at Bioconductor/R package⁵. This process consists of the following steps: raw reading files ".cel", background correction of the gross values of the fluorescence intensity by whole array adjustment and normalization by quantile, then Log2 transformed values were used for subsequent clustering analysis.

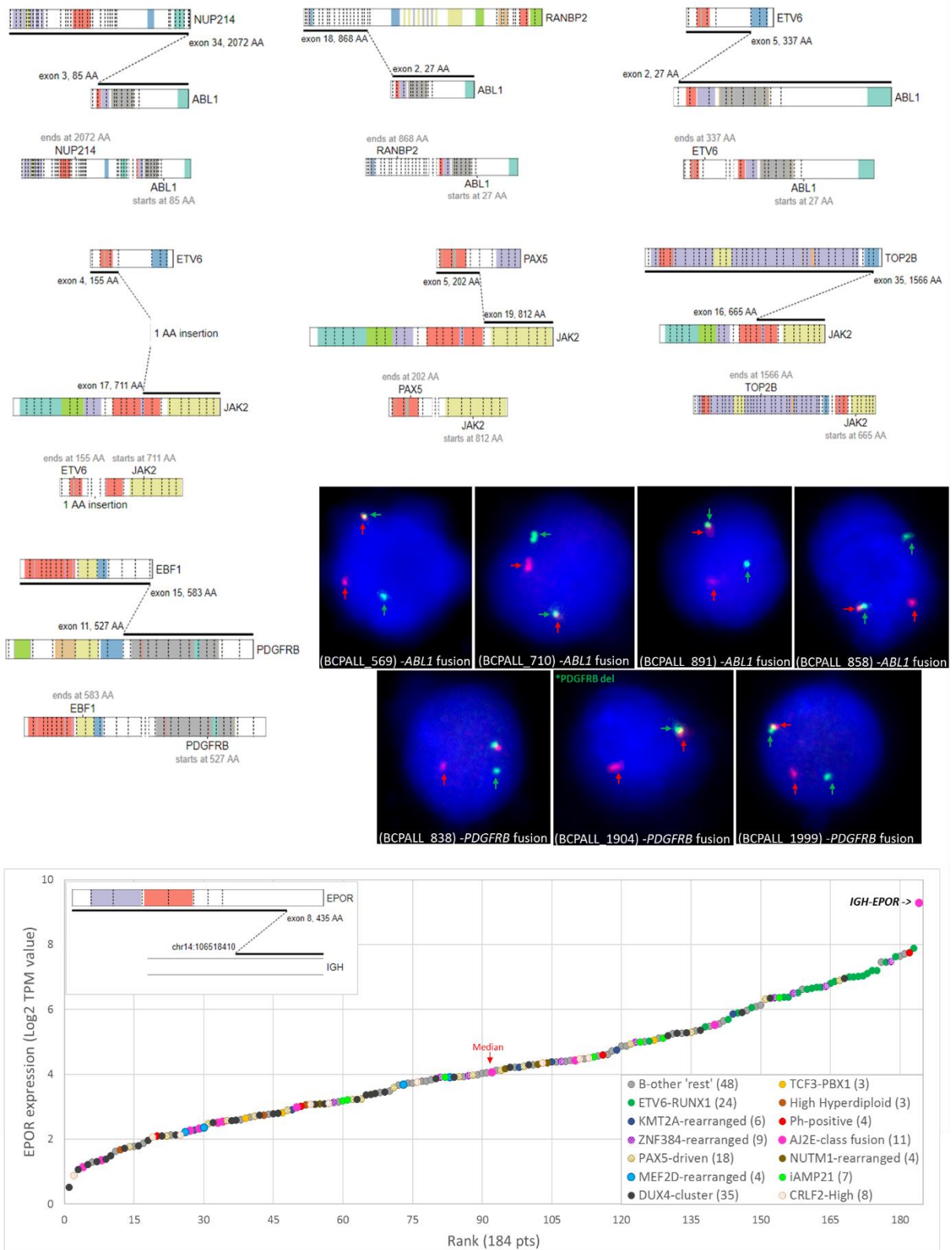
Hierarchical Clustering Analysis (HCA)

Individual gene expression (TPM) values were log2 transformed and then, they were subjected to Euclidean Hierarchical Clustering Analysis (HCA) with Ward's method. Different gene sets already published and previously used for clustering analysis by gene expression were matched with our gene list, and then also used for HCA. The list of genes used for each attempt to achieve gene expression profile-based subtypes are detailed in Table S3. For in-house subgroup clusterization we performed t-distributed stochastic neighbor embedding (tSNE) analysis. Basically, the TPM gene expression were log2 transformed, and expression values was filtered by *varFilter* function from *genefilter* R package (v.1.74.0). It was settled with cutoff of 0.5 and IQR as base function, resulting in 731 most variables genes. The resulting variables values present a normal distribution and were used for post analysis. The t-SNE plot was generated by *tsne* function from *M3C* R package (v1.14.0) using perplexity score of 30.

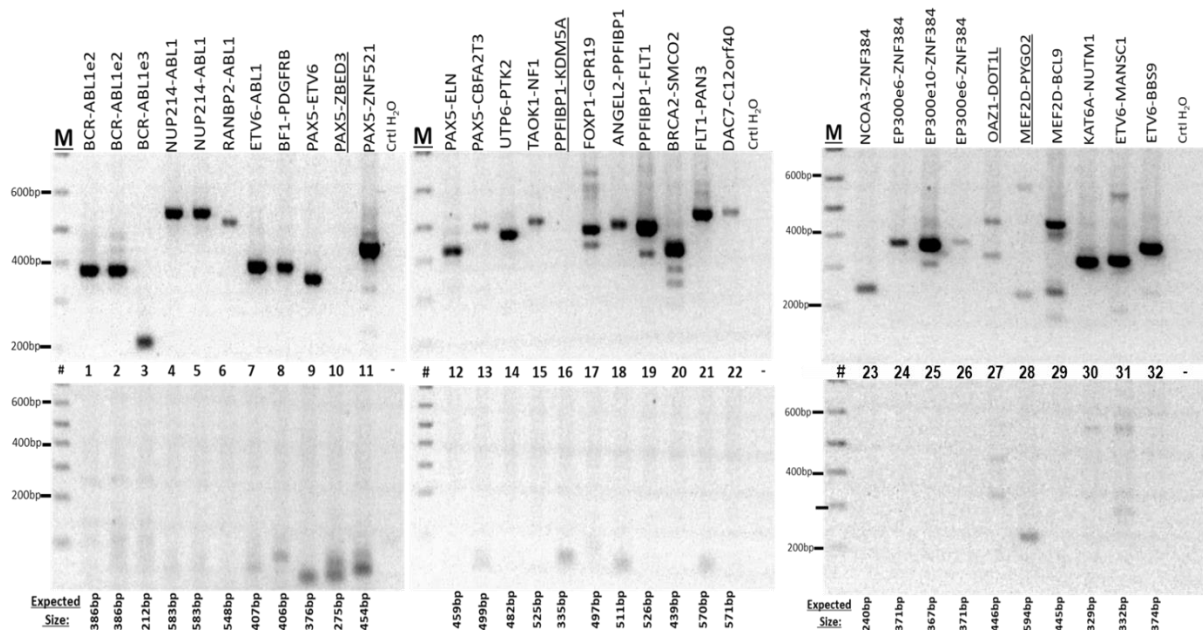
GATA3 rs3824662 and rs3781093 risk polymorphisms

Genotyping for risk polymorphisms in *GATA3*, rs3824662 and rs3781093, was done for 142 (out of 144 B-other samples, two DNA samples was not available). Risk allele was identified in fragments amplified by PCR and Nested-PCR reactions of the *GATA3* gene, analyzed by Sanger sequencing using Big Dye Terminator v3.1 (Applied Biosystems) and aligned to the gene reference. The rs3781093 polymorphism was amplified using the primers forward 5'-TTCCTGTGCTCTGTTCTT-3' and reverse 5'-GGCTCAGGATAACAATG-3, and the product was sequenced. The rs3824662 polymorphism was amplified using the primers forward 5'-TATCACCTCCCCACCA-3' and reverse 5'-GGAAAGCCCCAGATCAA-3'; and then, a second amplification was carried out in Nested-PCR reaction using the forward primers 5'-GTAAAACGACGGCCAGTGATCATGTCAGGCTGGGAGGT-3' (universal sequence M13, underlined) and reverse 5'-TAATACGACTCACTATAGGGGGGTTACGT' (universal sequence T7, underlined). The rs3781093 polymorphisms were sequenced using 5pmol of the forward and/or reverse primers, and rs3824662 polymorphisms were sequenced using 5pmol of the universal M13 and/or T7 primers.

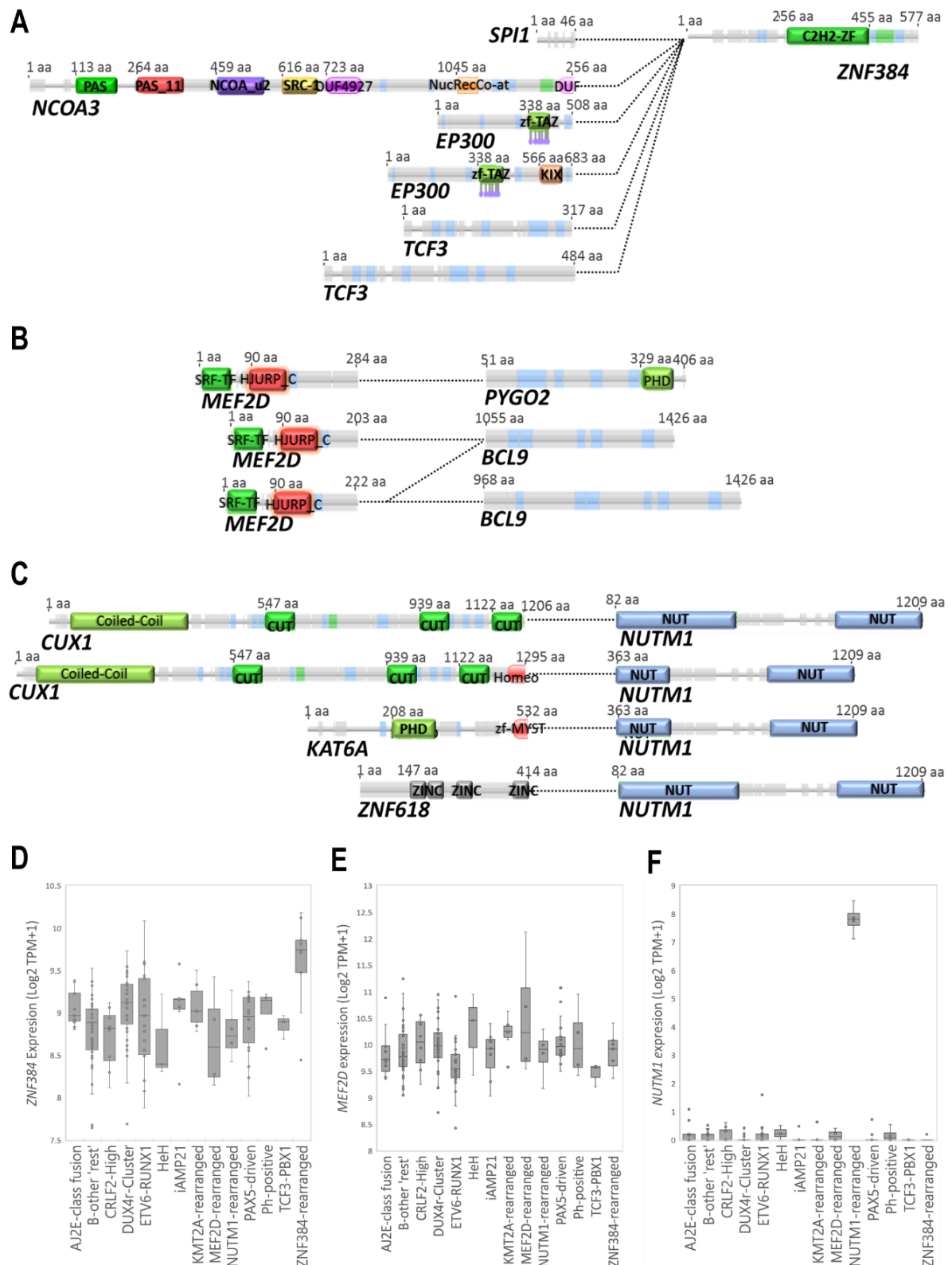
Supplemental Results



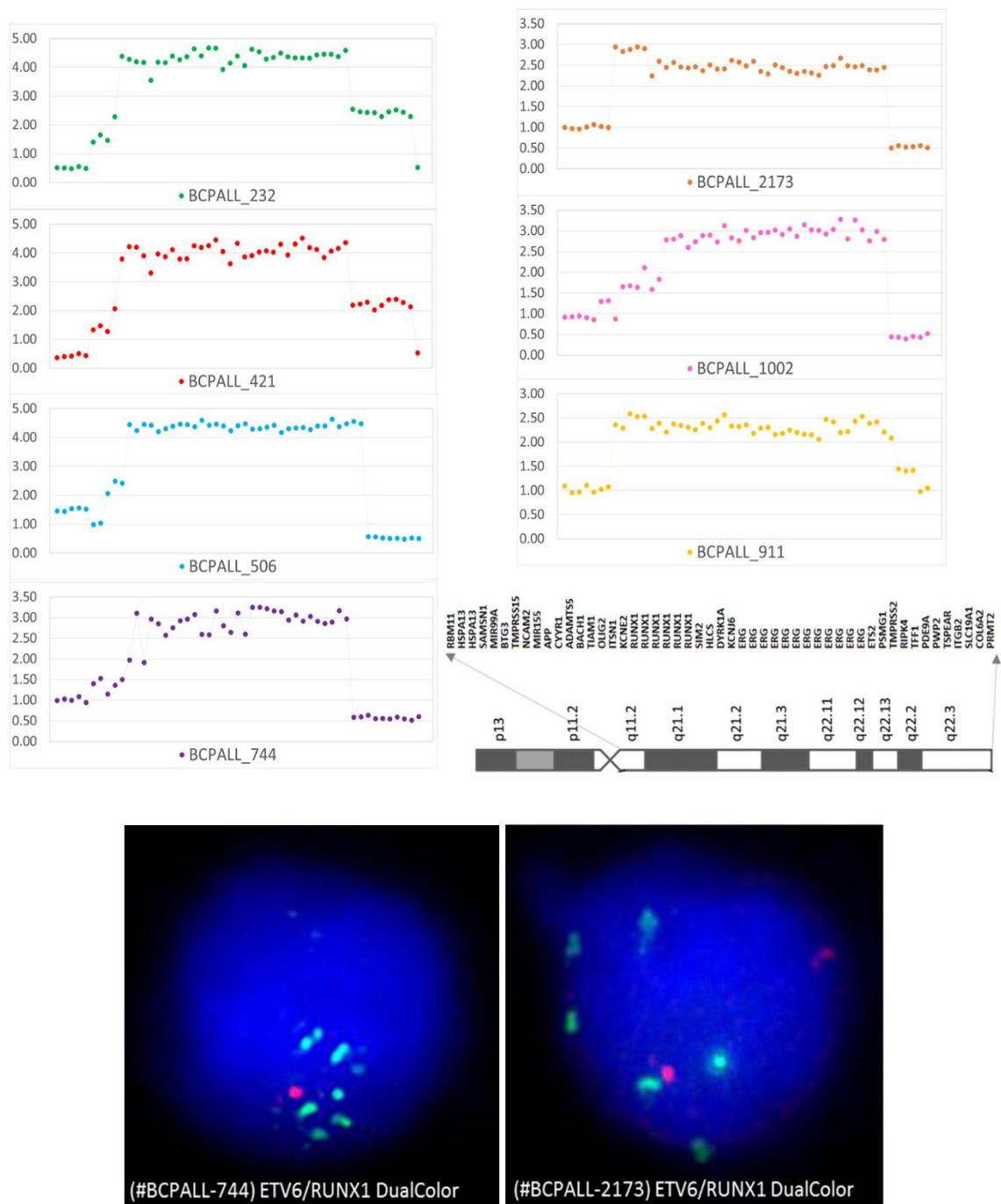
Supplemental Figure S2. AJ2E-class fusion. A. FusionEditor visualization of ABL-class, JAK/STAT-class and EPOR-class fusion genes. Breakpoints are structure represented and dot line indicate the link with partner gene; colored blocks represent protein functional domains. B. Orthogonal validation by FISH for selected samples with material available. C. Overexpression of EPOR gene in only one patient with the IGH-EPOR fusion gene.



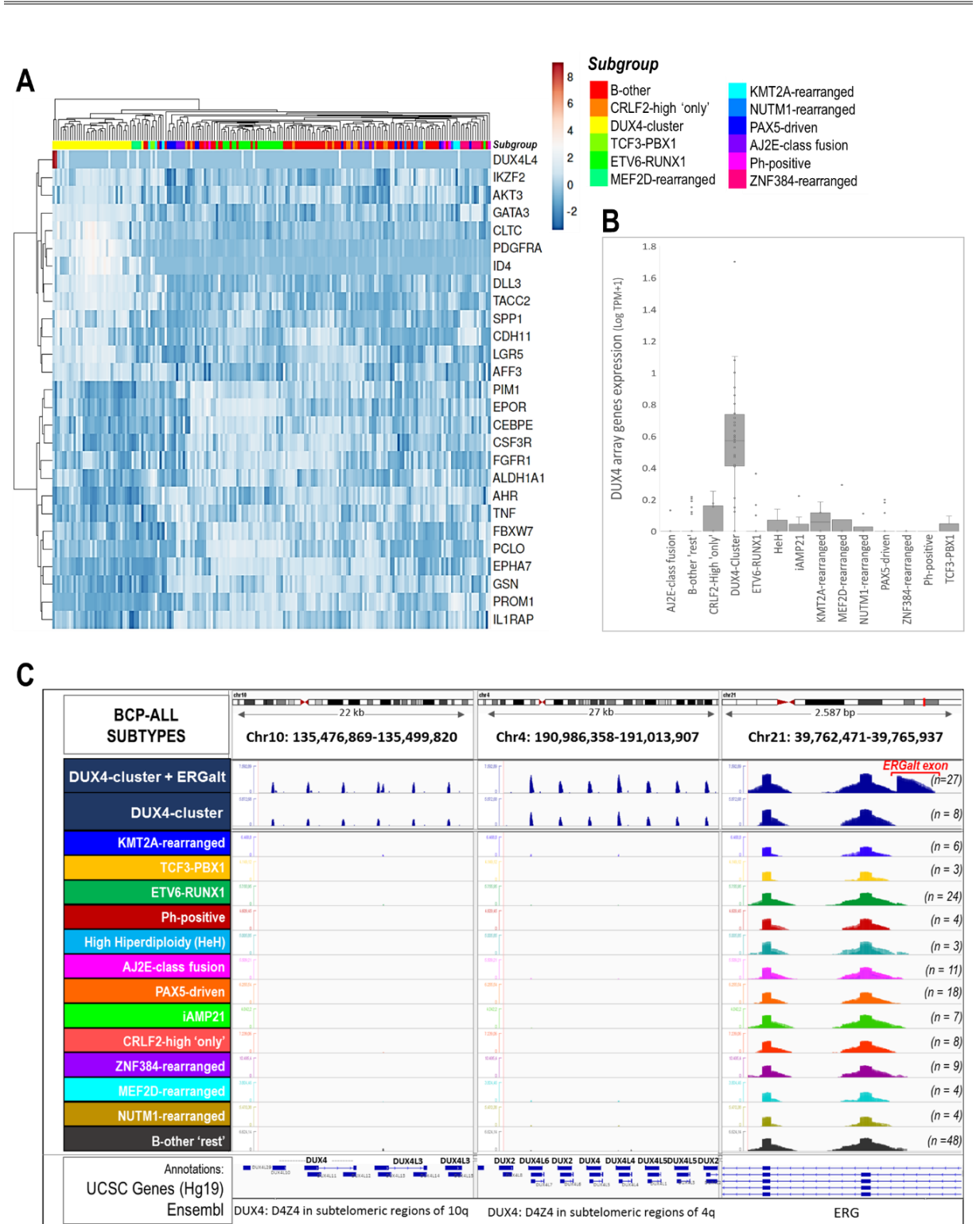
Supplemental Figure S3. Validation of known and novel fusion genes by RT-PCR. Classical rearrangements, recurrent fusions (e.g., ABL-class, PAX5-, ZNF384-, MEF2D and ETV6-rearrangements) and rare gene fusions such as TAOK1-NF1, UTP6-PTK2, FOXP1-GPR19, ANGEL2-PPFIBP1, PPFIBP1-FLT1, BRCA2-SMCO2, FLT1-PAN3, HDAC7-C12orf40, were identified and validated by RT-PCR. Although some fusions were detected on sequencing, we were not able to validate it by RT-PCR (underline) on our first try. Upper panel shows the amplification of the fragments (fusion genes) containing the breakpoint region. Lower panel shows control reaction with remission samples. Ctrl H₂O - negative control. M- Ladder 1kb (ThermoFisher).



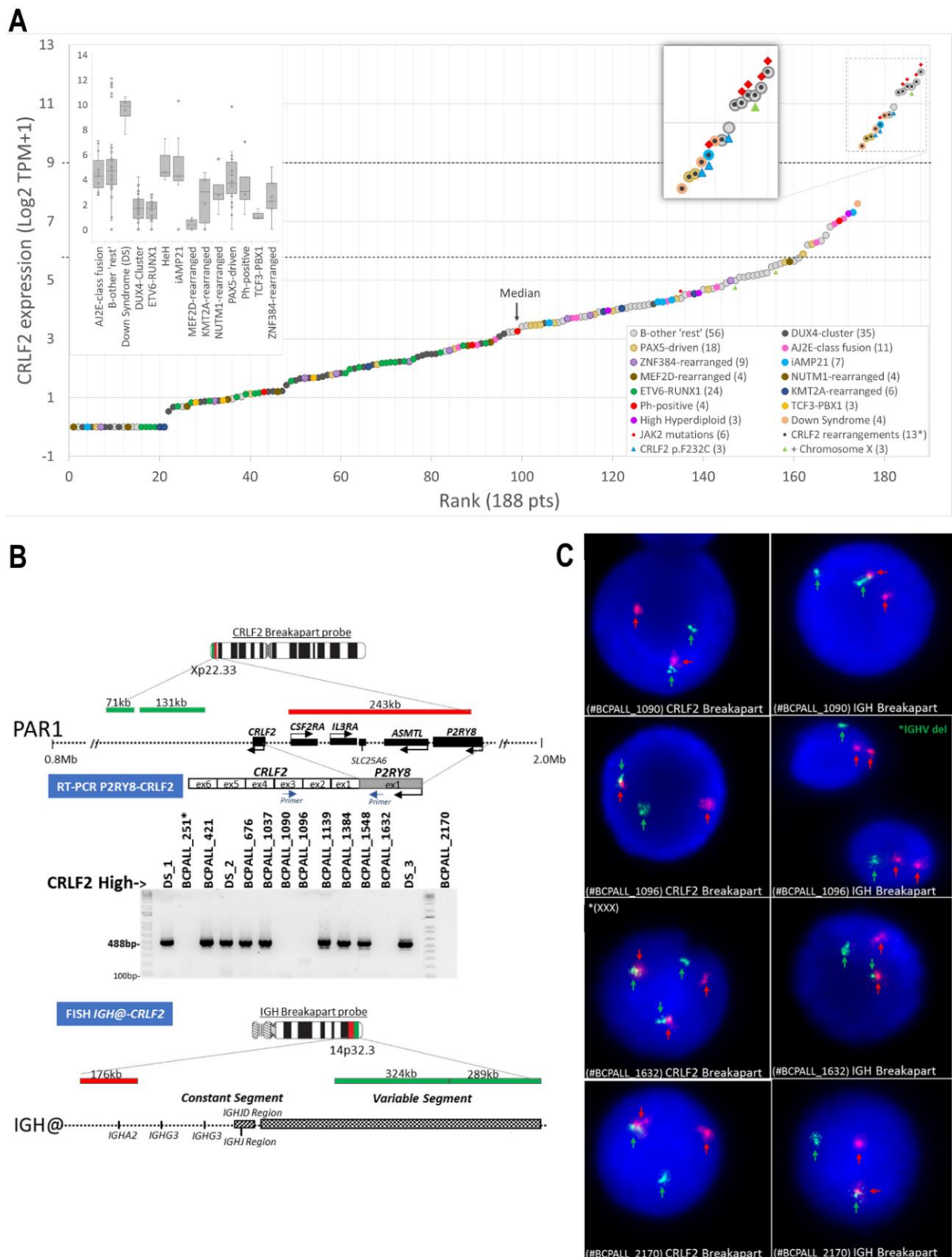
Supplemental Figure S4. Schematic representation of ZNF384, MEF2D and NUTM1 rearrangements found in our cohort. Respectively in A., B. and C., breakpoints are structure represented and dot line indicate the fusion region with the partner gene; colored blocks represent protein functional domains. D. and F. ZNF384 and NUTM1 genes are highly expressed in leukemic lymphoblasts containing the fusion gene compared to other subtypes. E. The same could not be seen for MEF2D-rearranged cases.



Supplemental Figure S5. Recently identified chromosomal aberrations in BCP-ALL and included as provisional risk subtype, iAMP21. A. Seven out of 144 B-other patients were classified as iAMP21 and had high MLPA peak ratios (≥ 1.8) along the long arm of chromosome 21, while the majority of patients with < 5 RUNX1 presented low MLPA peak ratios (< 1.8 , data not shown). The iAMP21-positive cases shows a “step ladder-like” amplification pattern, as schematized. B. We re-evaluated samples to search for extra copies of the RUNX1 gene by fluorescence in situ hybridization (FISH) as internationally used to define an iAMP21. Metaphase FISH demonstrates multiple RUNX1 signals (green) on the chromosome 21.

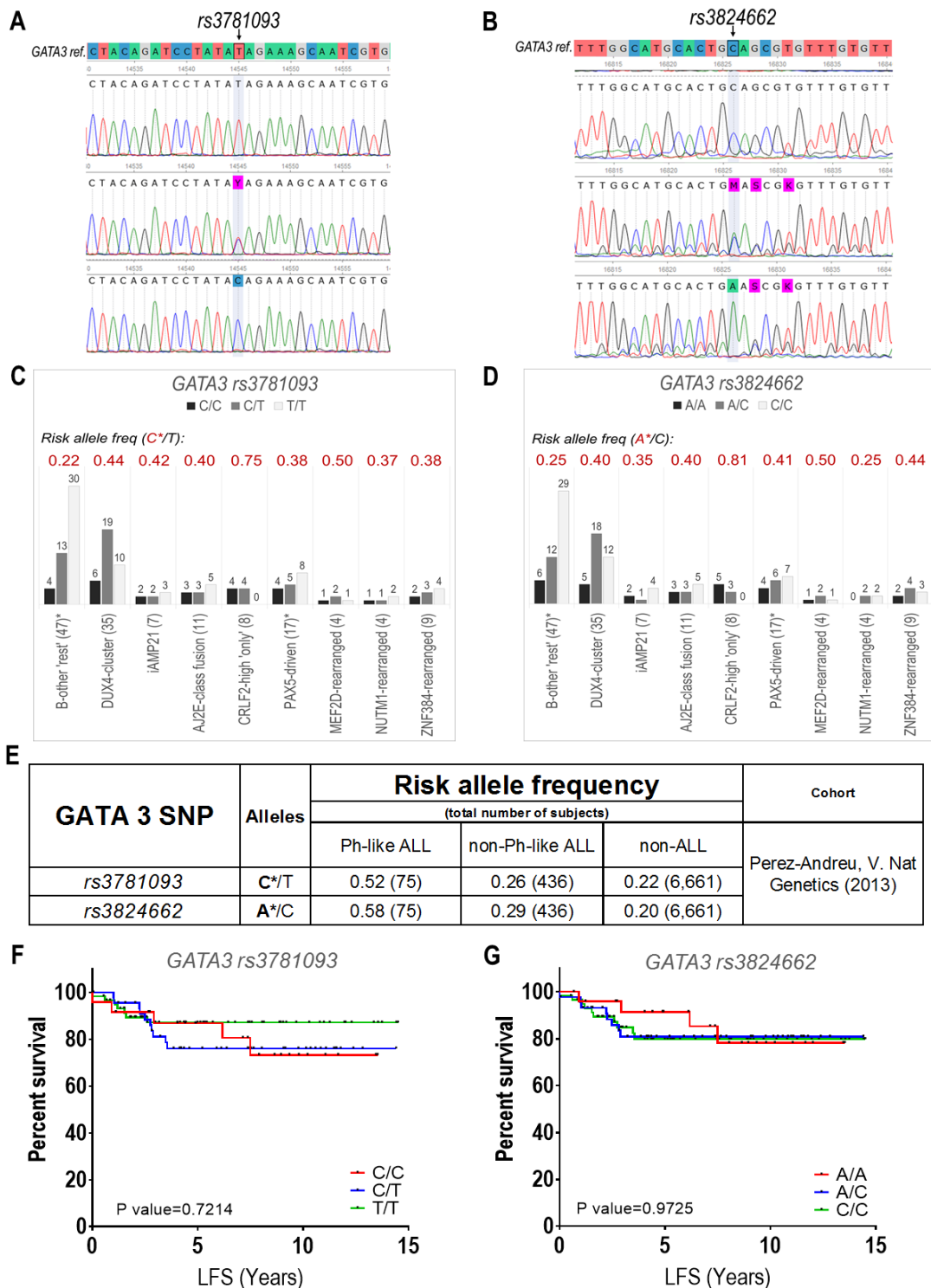


Supplemental Figure S6. DUX4-cluster subgroup. A. Unsupervised Hierarchical Clustering with 27 out of 187 genes published by Yasuda T. et al (Nat Genet., 2016) used to separate this subgroup. B. DUX4 expression is shown in box plots of Log2 TPM+1 values from Targeted RNA-seq analysis with BCP-ALL cases grouped by subtype along the x axis. C. Mapping and verification of DUX4-cluster subgroup features. The plots shows read depth coverage of selected three regions, i.e., DUX4 array expression in both chr4 and chr10 telomere regions, and ERG gene on chr21. DUX4 gene is exclusively expressed in DUX4-cluster subgroup, and is associated with ERG deregulation ('ERGalt' exon, cases with expression of exon 6 alt). The location of the ERGalt exon is marked in red. Screenshot of IGV browser.



Supplemental Figure S7. CRLF2 expression and underlying genomic alterations observed in this subgroup. A. Rank-based of normalized expression (Log2 TPM+1) of CRLF2 its showed as sigmoid curve and proved to be effective to define CRLF2-rearrangements subtype. y-axis represents CRLF2 expression; x-axis, represents 188 BCP-ALL patients ordered from lowest to highest CRLF2 expression levels. Dashed box highlights samples with CRLF2 expression levels more than 3-fold higher than the median level and represent samples with CRLF2-associated mutations (e.g., P2RY8-CRLF2, IGH-CRLF2, CRLF2 p.R683G and JAK2 mutations) as detailed.

Supplemental Figure S7 [Cont.] B. Orthogonal validation of the 11 B-other ALL cases with *CRLF2* rearrangement (two samples has also PAX5-driven and one has iAMP21 alteration). Schematic representation of the genomic location from PAR1 region and IGH, with respective probes used for FISH analyses. P2RY8-CRLF2 rearrangement were validated in 6 cases with RT-PCR assays (primers in supplemental Table S2.), and we used samples from Down Syndrome BCP-ALL patients that frequently shows P2RY8-CRLF2 fusion. C. FISH analysis for *IGH-CRLF2* were confirmed in 4 cases negative for P2RY8-CRLF2. There was no material for sample #BCPALL_251. Co-localization of the IGH probe (green) and CRLF2 probe (red) generating a yellowish fusion signal (arrows).



Supplemental Figure S8. GATA3 SNPs (*rs3781093* and *rs3824662*) risk allele frequencies within B-other subgroups. A. and B. GATA3 SNPs, *rs3781093* and *rs3824662*, respectively, were genotyped in the Boldrini B-other ALL cohort by Sanger sequencing. C. and D. Frequency of risk allele within B-other molecular subgroups. E. Risk allele frequency calculated by Perez-Andreu, V, et al, 2013 in the Children's Oncology Group (COG) AALL0232 cohort group. Analysis of association of SNP genotype and Ph-like ALL was evaluated by comparing allele frequency between Ph-like ALL and non-ALL, and also between Ph-like ALL and non-Ph-like ALL, after adjusting for genetic ancestry. F. and G. The leukemia-free survival (LFS) was compared by genotyped *rs3781093* and *rs3824662*, respectively, in 142 (2/144 was excluded due not DNA availability) patients analyzed in our cohort, with no significant p-value for both.

Table S2. Forward and Reverse primer sequences used for the validated fusions transcripts by RT-PCR and GAT1A3 polymorphisms.

Fusion; Duplication	Primer Forward	Forward Sequence (5' ->3')	Primer Reverse	Reverse Sequence (5' ->3')	~Amplen(bp)
NUP214-ABL1 (chr9-chr9)	NUP214e31F	GGCTTTGGATCCACAGCTACCT	ABL1e3R2	TGTAGTTGCTTGGGACCCAGCCTTG	~583
RANBP2-ABL1 (chr2-chr9)	RANBP2e16F2	TGTTCTTTGGAAATGCAGATTCA	ABL3-R	CCATTTTGGTTTGGGCTTCACACC	~548
ETV6-ABL1 (chr12-chr9)	ETV6e5F3	CGGCACTCCGTGGATTTTCAACAGT	ABL3-R	CCATTTTGGTTTGGGCTTCACACC	~407
EBF1-PDGFRB (chr5-chr5)	EBF1e14F2	CACGACCATGAACGGATACGGCTCT	PDGFRBe13R2	TTTCATCGTGCCCTGAGAATGGCTC	~406
PAX5-ETV6 (chr9-chr12)	PAX5e3F1	CAAGCTGGGGTAAATTGGAGG	ETV6e3R	TTTGGTCAGCAGCAGGAGAG	~376
PAX5- ZNF521 (chr9-chr18)	PAX5e4F1	ACAATGACACCGTGCCTAGCGTCAG	ZNF521e4R	TTGACACGGGTATGGAAGCC	~554
PAX5-ELN (chr9-chr7)	PAX5e6F2	ATGCGGGGAGACTTGTTCAC	ELNe3R	GGAAAGTAACTCGGGGAA	~499
PAX5-CBFA2T3 (chr9-chr16)	PAX5e6F2	ATGCGGGGAGACTTGTTCAC	CBFA13e2R	GATGTGTGTGGGCTGAAG	~459
ETV6-BBS9 (chr12-chr7)	ETV6e1F	GGGGAGAGGAAAGGAAAGTGG	BBS9e20R	TGCCGCTTCAGAAATAGCAA	~374
ETV6-MANSC1 (chr12-chr12)	ETV6e1F	GGGGAGAGGAAAGGAAAGTGG	MANSC1e4R	TGTGATGATGGGCTAGGGGA	~332
EP300-ZNF384 (chr22-chr12)	EP300e5F	TCCCCCTCAAAAATGCTGCT	ZNF384e3-4R#	CATTGTGTTCTCGATCTGACTGA	~371
EP300-ZNF384 (chr22-chr12)	EP300e9F	TGCTCGGAAAGTTGAAGGGG	ZNF384e3-4R#	CATTGTGTTCTCGATCTGACTGA	~367
NCOA3-ZNF384 (chr20-chr2)	NCOA3e20F	AATGTGACTGCTTCCCCAG	ZNF384e3-4R#	CATTGTGTTCTCGATCTGACTGA	~240
MEF2D-PYG02 (chr1-chr1)	MEF2De4F	CTGCTGGAGGCAAGTACCG	PYG02e3R	GGATGCAACGAGTGAATCCA	~594
MEF2D-PYG02 (chr1-chr1)	MEF2De5F	CAGCAGCCAGCATACAGAG	PYG02e3_Rv	GGCTCCGAAATCATCTTCAA	~411
MEF2D-BCL9 (chr1-chr1)	MEF2De4F	CTGCTGGAGGCAAGTACCG	BCL9e9R	GTGGGTTCTGGGAAAGGTT	~445
KAT6A-NUTM1 (chr8-chr15)	KAT6Ae8F	GACAATCAGGATGGCTGGGA	NUTM1e5R	AGTTTCAAAAGGGGTTGCAGGA	~329
TAOK1-NF1 (chr17-chr17)	TAOK1e7F	TGCTGACTTTGGCTCTGCTT	NF1e37R	CTGCGGAGAGCTTGCTGTA	~525
UTP6-PTK2 (chr18-chr8)	UTP6e14F	CGAGTCAAGAGCCCTGACA	PTK2e11R	GGTCAAGCATATTCTCCGA	~482
FOXp1-GPR19 (chr3-chr12)	FOXp1e11F	CCACAGGCAACATCACAGC	GPR19e2R	GGTCAAGATAGCCCTTGAT	~497
BRCA2-SMCO2 (chr13-chr12)	BRCA2	TGACAGATTCTAAACTGCCAAGTC	SMCO2e2R	AGCAAACTCTGCATTGCACC	~439
FLT1-PAN3 (chr13-chr13)	FLT1e13F	AGGGCTCTGTGGAAAGTTGAG	PAN3e3R	GGGCTGAATATGGCTTGCA	~570
P2RY8-CRLF2 (chrX-chrX)	P2RY8e1F	TTGCAAGTTTGTGGACAGATGGAA	CRLF2e3R	GTCTAGGAGGACCCCGAAGTGTGA	~488
BCR-ABL1 (chr22-chr9), p190	BCRe1F1	GTGCCATTAAGCGGCACCGGCACT	ABL1e3R2	TGTAAGTTGCTTGGGACCCAGCCTTG	~386
BCR-ABL1 (chr22-chr9), p190	BCRe1A	GACTGCAGCTCCAAATGAGAAC	ABL1e3B	GTTTGGGCTTCACACCAATTC	~347, 413, 521
ETV6-RUNX1 (chr12-chr21)	TEL-A bis	TGCACCTCTGATCTGTAAC	AML1-B tris	AACGCCTCGCTCATCTTGC	~259, 298
MLL-AFF4 (chr11-4)	MLL-A	CCGCCTCAGCACCTAC	A4-B	TGTACAGAGCTGAAGGTCG	~184, 353, 427
TCF3-PBX1 (chr19-chr1)	E2A-A	CACCAAGCTCATGCACAAC	PBX-B	TCGCAAGAGATTCAACAG	~373, 400, 504
GATA 3 rs3781093	rs3781093-F	TTCTGTGCTCTGTTCTT	rs3781093-R	GGCTCAGGATTAACAATG	596
GATA 3 rs3824662	rs3824662-F	TATCACCCTCCCCACCA	rs3824662-R	GGAAAGCCCCAGATCAA	805

Supplementary Tables S2 and S3.

*A detailed list of targeted regions can be obtained from Illumina (https://support.illumina.com/s equencing/sequencing_kits/tru sight-rna-pan-cancer-panel/downloads.html).

F. 731 most variable genes from t-SNE algorithm processing.

t-SNE algorithm (731 genes) ^F

RAD52	PRDM1	CYLD	ACSG61	GAB1	TPO	H1-3	ACCA	C11orf1	CNA2	ANGPT1	COL6A3	PTK2	MAF	PCLO	NTRK1	AC002986.1
HECW1	CAAMK2B	APLP2	FZD3	FBXW7	SOS1	H2BC11	PPARG	THBS1	SETD7	PRKCA	CIP2A	CXCL8	AURKB	PRG2	ATL1	BC02122-PA8P1
BAAP2L1	STYK1	CAD	EVA1	SH3D19	MSH6	HIST1H3B	SERPINF1	NDUFNF1	SERP2	SORBS2	TAFA2	RAO3	PER1	TEAD1	TAFA2	ITGB3
ETV1	HDAC7	BCORL1	NDRG1	ZBTB16	BC19	CDKNA1A	TEF2	NDL	PKR1	FLCN	CTLA4	ROBO1	MAGED1	MITE	DL1	RP11-133K1.2
IFRD1	AHR	MECOM	TRPS1	LPXN	CD58	RUNX2	PCNA	ITPKA	SSBP2	EPH8	WDR3	WNT10B	GAT42	SHTN1	RCSD1	ITGB3
PROM1	CASP8	AURKA	TUSC3	DTX4	MYCL	IRF1	ATP8A2	SMAD6	ARRGAP26	FZD7	CCDC5A	ZFPW2	HMG2P46	FNBP1	ZNF521	RP11-48M22.4
GAS7	NGFR	GNAT1	PPP1R13L	CCND1	AKT3	BMP4	MTUS2	PAB1	ICFBP3	PSD3	MAZD1	CDK1	EGR3	EPOR	RNR3	AC012309.1
MATK	CTSA	PPP1R13B	AMH	POU2AF1	ID3	IL1B	MACROD1	CYP11B1	ICFBP3	GNAAQ	IL15	FOS	CITFA	TC16	RASGEF1A	AC004223.3
CD79B	ERBB3	ICAM1	LY1	PP1BP1	MPL	PAX8	MYH11	TACC2	CONB3	BACH1	EBF1	MGMT	GAS1	SRRY4	MAGEE1	AC008738.5
TEAD3	FGFR2	FLT3LG	CCNE1	BCL7A	FASLG	CD70	BTG1	NCOM4	GPC3	TIAM1	EBF1	MGMT	GAS1	FANCA	SNHG5	AC008738.5
JARID2	EM1	FCGBP	CD79A	GLI1	RPS6K41	FOSB	DUSP26	TEF1	TACC1	BUB1B	ACS16	CHCHD7	PRF1	FANCM	MABF	AP007697
CDKL5	MA3	DL3	JAK3	WNT5B	CENPE	ID1	CHL1	AT1C	NEIB	KIT	ACS16	CHCHD7	PRF1	FANCM	MABF	
DLEC1	FGFR3	ESR1	KMT2B	ALDH2	RAC2	KDM5C	SORT1	STAT4	CDKN2B	KIT	ACS16	CHCHD7	PRF1	FANCM	MABF	
MARKKIP2	INPP5A	CDKN1B	WNT5B	ALDH2	RAC2	KDM5C	SORT1	STAT4	CDKN2B	KIT	ACS16	CHCHD7	PRF1	FANCM	MABF	
TEMN1	BC13	COL9A3	PK3CG	RASAL1	MB8	PRKCG	NOTCH2	ITGAV	GN	ERG	COL1A2	PTCRA	P2R8	NBR1	RAUBP17	
ANLN	TGFB3	EZR	CAM1	IFNG	SGK1	HSP22	MYCN	PRG2	GN	ERG	COL1A2	PTCRA	P2R8	NBR1	RAUBP17	
ICN	NEED4	TGFB2	MET	PTPN6	SP1	BCL11B	LY2RA	LEF1	ASTN2	CREBL1	FZD6	GPR34	EXT1	DPO	CRU2	
BRO41	CLTCL1	MAR3K1	LENG	CER8L	CCND2	HIP1	PGFFA	EGF	NOTCH1	ASTN2	CREBL1	FZD6	GPR34	EXT1	DPO	
ZWIND1	FTL3	TLL2	GRB10	SOD2	CNTRL	GNM1	ET51	ETV6	AD3	DUSP2	ALDH1A1	METTL18	CSF1R	APOD	ANKRD28	
RGPD5	RPS6K42	JAK2	NOD1	CLK1	NR43	RAQ2	DTX1	ETV6	AD3	DUSP2	ALDH1A1	METTL18	CSF1R	APOD	ANKRD28	
IGF1	PRKACA	ABLIM1	EPHB6	BACH2	CSF3R	SPECC1	CO36	PRICKLE1	MK67	WASF2	TSHR	ID4	RGPD6	IL1RAP	BORCS8-MEF2B	
WNTFR1	TRFC	CEP170B	SERPINE1	PHF1	TGF33	POT1	EPH47	LG85	ADM	AUTS2	REEP3	CEBPB	FAHCF	PAG5	NUTM2D	
HGF	SMC1A	GADD45B	RP3	CCND3	YFEL5	SNO	ITGAV	TG12	NCAM1	SLC45A3	REI	RHOD	LHP16	SRGAP3	AKR1C3	
ADGRA2	FCGR2B	SH3BP1	EZH2	CCND3	YFEL5	SNO	ITGAV	TG12	NCAM1	SLC45A3	REI	RHOD	LHP16	SRGAP3	AKR1C3	
CP51	TP63	XBPI	AHR	CCND3	YFEL5	SNO	ITGAV	TG12	NCAM1	SLC45A3	REI	RHOD	LHP16	SRGAP3	AKR1C3	
BIRC3	PTGS2	PDGFB	GLI3	VEGFA	CD274	DLL4	AH1	IGF1R	CHK1	BTG2	WEE1	OLR1	CSF1	H2AC11	CEBPD	
RT1EL1-TNFRSF6B	ST6GAL1	NIN	LHX2	TTK	TGFB1	PMREB	TEC	ARRDC4	HMGA2	ACE	PRKCB	TNFRSF10D	MAML2	SCN8A	RP11-297K8.2	
FAS	NOTCH3	HIF1A	DNM1	GHR	EGR1	CDKN2D	MDM2	PML	TRAP	TRAP	NAB2	PC	SS18L1	ARHGFE12	AC093818.2	
CD44	TEAD2	TCL1A	KANK1	LOX	SOC2	BCL2L2	WNT10A	FANCI	DOCK1	KALRN	SMAD3	CMKLRT1	FOXO4	FLNA	ERCC6	
IKZF2	GTSE1	TRIP1	GAT3	RASGEF2	ITK	TNFRSF10B	LMO7	FOXO1	FOXO1	BR3K1	TPM4	FUT1	SOCS3	SRC	RP4-78K5.1	
MAP2K3	PAG1	NFKBIA	BMRF1A	RAD50	NR3C1	ACR2A	SHC2	CDKN1C	EDNRB	KSR1	RECQL4	KLK2	POLDA	AC010801.2	TNF	
FLT4	TOP2B	NFATC2	FGFR	NR3C1	ACR2A	SHC2	CDKN1C	EDNRB	KSR1	RECQL4	KLK2	POLDA	AC010801.2	TNF		
CDH1	NFAB2	NFATC2	FGFR	NR3C1	ACR2A	SHC2	CDKN1C	EDNRB	KSR1	RECQL4	KLK2	POLDA	AC010801.2	TNF		
PHF23	FGFR1	PLCB4	NEU1L1	LFR	PLCB4	NEU1L1	LFR	PLCB4	NEU1L1	LFR	PLCB4	NEU1L1	LFR	PLCB4	NEU1L1	
TNC	ITGA8	EZF1	DKK1	PDGFRB	NHBA	SXN9	BRP1	ERBB2	CENPU	WDR59	GN4	ETV4	WNT1	TNFA2	AL158212.2	
EPH43	MAP2	CACNA1F	ZMIZ1	KAT2B	KIAP1549	AFDN	BN1	MYO1F	DST	LRP5	ETV4	WNT1	TNFA2	AL158212.2		
IOF1	HOXA9	SYP	RNF43	KAT2B	KIAP1549	AFDN	BN1	MYO1F	DST	LRP5	ETV4	WNT1	TNFA2	AL158212.2		
GOFC	TP73	ELF4	CCL2	HES1	EGR2	LAM6	KLF4	SETBP1	PLP3	FEV1	RHOH	SOX11	SOC51	DNM3	GAS5	
TNFRSF17	RUNX1T1	GATA1	ABCC3	IOCG	CDKN2C	TFRA3	EMRP2	IKBAE	HEX	VCAM1	LR7	FZD8	MYBL1	TFDP1	CERPA	
SXN29	CC	GATA1	ABCC3	IOCG	CDKN2C	TFRA3	EMRP2	IKBAE	HEX	VCAM1	LR7	FZD8	MYBL1	TFDP1	CERPA	
LTBP1	NDG80	ARRGFE7	HLF	SPTBN1	STIL	TFRA3	MYC	EM14	AF3	CDBA	ATF3	IRS1	JUN	IRS2	PEG3	
ELN	MEF2C	FGF9	MAP2K6	STAT1	ACVR1C	TRAF3	MYC	EM14	AF3	CDBA	ATF3	IRS1	JUN	IRS2	PEG3	
HERPUD1	CD14B	FLT1	WSN1	IGFBP2	SDC4	RARA	TFAP2A	ACR3	JAZF1	SLC66A3	AR	MLF1	TAL2	HIST1H4I	HOXA10	
RAD51	SLC7A5	ALDOC	LIR1	PLCG1	LGALS3	IF4	PAF2A	ACR3	JAZF1	SLC66A3	AR	MLF1	TAL2	HIST1H4I	HOXA10	
CYFP2	TRAF5	IL2IR	INP4B	FLT2	BCAS4	CCTB8	YAP1	LPP	AK5	SGP22	CRADD	CD28	BTIA	HIST1H2AL	SDHD	

Supplementary Table S4.

^a Based on 10000 sampled tables with starting seed 2000000.

	Valid Cases N=144	Fisher's test	Monte Carlo Sig.	
			99% Confidence Interval	
		Approx. Sig.	Sig. ^a	Lower Bound Upper Bound
B-other 'rest'	GBLTI Risk classification (Standard vs High Risk)	.005	.007	.005 .009
	NCI risk classification (Standard vs High Risk)	.486	.609	.596 .621
	Age (<10 or ≥10 yrs)	.115	.147	.138 .156
	Gender (Male vs Female)	.044	.049	.043 .054
	WBC count (≥ 50x10 ⁹ L)	.530	.579	.567 .592
	CALLA (negative or low <20% CD10)	.198	.273	.261 .284
DUX4- cluster	GBLTI Risk classification (Standard vs High Risk)	.256	.301	.289 .313
	NCI risk classification (Standard vs High Risk)	.723	.784	.773 .795
	Age (<10 or ≥10 yrs)	.801	.845	.836 .855
	Gender (Male vs Female)	.318	.340	.328 .352
	WBC count (≥ 50x10 ⁹ L)	.156	.217	.206 .228
	CALLA (negative or low <20% CD10)	.962	1.000	1.000 1.000
iAMP21	GBLTI Risk classification (Standard vs High Risk)	.072	.103	.095 .111
	NCI risk classification (Standard vs High Risk)	.290	.596	.583 .609
	Age (<10 or ≥10 yrs)	.795	1.000	1.000 1.000
	Gender (Male vs Female)	.141	.246	.235 .257
	WBC count (≥ 50x10 ⁹ L)	.554	.685	.673 .697
	CALLA (negative or low <20% CD10)	.511	1.000	1.000 1.000
PAX5- driven	GBLTI Risk classification (Standard vs High Risk)	.171	.274	.263 .286
	NCI risk classification (Standard vs High Risk)	.306	.474	.461 .487
	Age (<10 or ≥10 yrs)	.331	.444	.431 .457
	Gender (Male vs Female)	.043	.071	.064 .077
	WBC count (≥ 50x10 ⁹ L)	.946	1.000	1.000 1.000
	CALLA (negative or low <20% CD10)	.271	.598	.586 .611
AJ2E-class fusion	GBLTI Risk classification (Standard vs High Risk)	.108	.174	.164 .184
	NCI risk classification (Standard vs High Risk)	.018	.038	.033 .043
	Age (<10 or ≥10 yrs)	.606	.748	.737 .759
	Gender (Male vs Female)	.575	.754	.742 .765
	WBC count (≥ 50x10 ⁹ L)	.003	.007	.004 .009
	CALLA (negative or low <20% CD10)	.403	.640	.627 .652
CRLF2- High	GBLTI Risk classification (Standard vs High Risk)	.108	.174	.164 .184
	NCI risk classification (Standard vs High Risk)	.151	.163	.153 .172
	Age (<10 or ≥10 yrs)	.245	.339	.327 .351
	Gender (Male vs Female)	.483	.537	.524 .550
	WBC count (≥ 50x10 ⁹ L)	.023	.042	.036 .047
	CALLA (negative or low <20% CD10)	.594	1.000	1.000 1.000
ZNF384- rearranged	GBLTI Risk classification (Standard vs High Risk)	.575	.721	.710 .733
	NCI risk classification (Standard vs High Risk)	.409	.610	.597 .622
	Age (<10 or ≥10 yrs)	.268	.308	.296 .320
	Gender (Male vs Female)	.488	.732	.721 .744
	WBC count (≥ 50x10 ⁹ L)	.491	.717	.705 .729
	CALLA (negative or low <20% CD10)	.000	.000	.000 .001
MEF2D- rearranged	GBLTI Risk classification (Standard vs High Risk)	.807	1.000	1.000 1.000
	NCI risk classification (Standard vs High Risk)	.429	.650	.638 .663
	Age (<10 or ≥10 yrs)	.124	.300	.288 .312
	Gender (Male vs Female)	.212	.322	.310 .334
	WBC count (≥ 50x10 ⁹ L)	.158	.309	.297 .321
	CALLA (negative or low <20% CD10)	.623	1.000	1.000 1.000
NUTM1- rearranged	GBLTI Risk classification (Standard vs High Risk)	.392	.588	.575 .600
	NCI risk classification (Standard vs High Risk)	.429	.650	.638 .663
	Age (<10 or ≥10 yrs)	.111	.163	.154 .173
	Gender (Male vs Female)	.427	.632	.619 .644
	WBC count (≥ 50x10 ⁹ L)	.741	1.000	1.000 1.000
	CALLA (negative or low <20% CD10)	.623	1.000	1.000 1.000

Softwares information:

RNAseq v1.1 (Illumina App)

Isis (Analysis Software) ----- version 2.6.25.18
 Read mapping using the STAR aligner ----- version STAR_2.5.0b
 FPKM estimation using Cufflinks 2 ----- version 2.2.1
 Assembly of novel transcripts with Cufflinks 2 ----- version 2.2.1
 Variant calling (SNVs and small indels) with the Isaac Variant caller ----- version 2.3.13.31
 Fusion calling with Manta ----- version 1.4.0
 Optional: Fusion calling with TopHat-Fusion

RNAseq v2.0 (Illumina App)

Isis (Analysis Software) ----- version 3.19.1.12+master
 SAMtools ----- version 0.1.20
 Read mapping using the STAR aligner ----- version STAR_2.6.1a
 Quantification of reference genes and transcripts using Salmon ----- version 0.11.2
 Variant calling (SNVs and small indels) using the Strelka Variant caller ----- version 2.9.9
 Fusion calling with Manta ----- version 1.4.0

DRAGEN RNA Pipeline v3.5 (Illumina App)

Illumina Bio-IT Platform that provides ultra-rapid secondary analysis of sequencing data using field-

CICERO v0.3.0 (St Jude Hospital)

Source Code GitHub Repository. 2020; <https://github.com/stjude/Cicero>.

Tian, L., Li, Y., Edmonson, M.N. et al. CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. *Genome Biol*

Supplemental References

- ¹ Sociedade Brasileira de Oncologia Pediátrica. Protocolo de Tratamento da Leucemia Linfoblástica Aguda na Infância-GBTLI-LLA-1999, Campinas; 2000
- ² Tian L, Li Y, Edmonson MN, Zhou X, Newman S, McLeod C, Thrasher A, Liu Y, Tang B, Rusch MC, Easton J, Ma J, Davis E, Trull A, Michael JR, Szlachta K, Mullighan C, Baker SJ, Downing JR, Ellison DW, Zhang J. CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. *Genome Biol*. 2020 May 28;21(1):126. doi: 10.1186/s13059-020-02043-x.
- ³ Robinson JT, Thorvaldsdottir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. *Nature biotechnology*. 2011 Jan;29(1):24-6.
- ⁴ Gautier L, Cope L, Bolstad BM, Irizarry RA (2004). "affy-analysis of Affymetrix GeneChip data at the probe level." *Bioinformatics*, 20(3), 307–315. ISSN 1367-4803
- ⁵ Gentleman RC, Carey VJ, Bates DM, Bolstad B, Dettling M, Dudoit S, Ellis B, Gautier L, Ge Y, Gentry J, Hornik K, Hothorn T, Huber W, Iacus S, Irizarry R, Leisch F, Li C, Maechler M, Rossini AJ, Sawitzki G, Smyth G, Tierney L, Yang JY, Zhang J. Bioconductor: open software development for computational biology and bioinformatics. *Genome Biol*. 2004;5(10):R80. doi: 10.1186/gb-2004-5-10-r80. Epub 2004 Sep 15.

6 CAPÍTULO II: PERFIL DAS MUTAÇÕES SNV E *SMALL INDEL* NA LLA B-OTHER

Para identificar mutações gênicas do tipo pontuais e pequenas deleções/inserções (SNV e *small indel*), realizamos a análise com o painel comercialmente disponível de RNA-seq direcionado (*TruSight RNA Pan-Cancer*, *Illumina*) e sequenciamento na plataforma *MiSeq* (*Illumina*) de 144 pacientes LLA B-*other*. O painel contém sondas biotiniladas que tem como alvo uma lista de 1.385 genes (21.043 regiões exônicas-alvo) relacionados ao câncer.

Vale lembrar que os casos de LLA e os dados de sequenciamento com o kit *TruSight RNA Pan-Cancer* são os mesmos do Capítulo Primeiro. No capítulo primeiro apresentamos as fusões gênicas e a análise do nível de expressão gênica, que serviram para a discriminação dos diferentes tipos de LLA dentro do subgrupo B-*other*. Neste capítulo apresentamos os resultados de mutações SNV e *small indel*, além de analisar vias de sinalização enriquecidas e detectar possíveis assinaturas mutacionais dentro dos diferentes tipos de LLA pertencentes a B-*other*.

Conforme apresentado no Capítulo Primeiro foram analisados 144 casos de LLA B-*other*, divididos em: B-*other* ‘rest’ (n=48), *DUX4*-cluster (n=35), *PAX5*-driven (n=18), *AJ2E-class fusion* (n=11, composta por pacientes com fusões envolvendo os genes *ABL-class*, fusões de *JAK2* e *EPOR*), *CRLF2*-high ‘only’ (n=8; e consequente expressão elevada de *CRLF2*, os quais 2 pacientes em *PAX5*-driven e 1 paciente em *iAMP21* também possuem o rearranjo *P2RY8-CRLF2* e expressão elevada de *CRLF2*), *ZNF384*-rearranged (n=9), *iAMP21* (n=7), *MEF2D*-rearranged (n=4) e *NUTM1*-rearranged (n=4). Muitos desses casos também foram classificados quanto ao perfil de expressão do subgrupo BCR-ABL1-like (Ph-like), pois dispúnhamos de dados de expressão gênica com microarranjos de DNA Affymetrix. Porém, esses dados não são apresentados, pois em nossas análises, foi mais informativa a discriminação dos casos com “*AJ2E-class fusion*” e “*CRLF2*-high” do que a discriminação de Ph-like, que é um grupo mais amplo, heterogêneo e de classificação mais duvidosa (pois depende de dados de expressão gênica).

6.1 Análise de Bioinformática

O grande volume de dados gerados pelo NGS requer a utilização de ferramentas computacionais elaboradas, a fim de que apenas as poucas variantes

genéticas de interesse sejam identificadas em meio a centenas ou milhares de variantes não relacionadas ao fenótipo patogênico (*i.e.*, variantes de polimorfismos).

As mutações de *SNV* e *small indel* foram detectados com o algoritmo *Strelka2*⁵⁹, seguindo os parâmetros padrão. Anotados com *Ensembl Variant Effect Predictor (VEP)*⁶⁰ para registrar as frequências populacionais, o tipo de mutação, prever o impacto funcional e anexar os códigos de variantes já existentes, com base no banco de dados dbSNP (NCBI) e COSMIC (*Catalogue of Somatic Mutations in Cancer*). Os arquivos finais *Variant Call Format (.vcf)* contendo informações de todas as variantes somáticas anotadas foram convertidos em arquivos *Mutation Annotation Format (.maf)* através do algoritmo *vcf2maf*⁶¹.

Além dos filtros utilizados pelos algoritmos acima mencionados, filtros adicionais foram aplicados visando excluir os polimorfismos e selecionar apenas mutações somáticas das células leucêmicas. Utilizando as anotações compiladas pelo VEP, os critérios para a filtragem de priorização foram: (1) exclusão de variantes com uma frequência populacional (*Minor Allele Frequency, MAF*) maior do que 1% (*gnomAD exomes combined population*); (2) exclusão de variantes com impacto funcional classificadas como 'LOW' e 'MODIFIER' de acordo com a anotação do *Ensembl*; (3) inclusão apenas de variantes exônicas e; (4) exclusão de variantes não patogênicas, ou seja, "*tolerated*" e "*benign*", de acordo com análises do SIFT v5.2 e PolyPhen-2 v2.2. Após a aplicação de todos esses filtros, observamos que a carga mutacional da nossa coorte ficou na média estimada para tumores hematológicos quando comparados com o banco de tumores do TCGA (*The Cancer Atlas Genome*) (Figura 1).

Os arquivos *.maf* filtrados foram utilizados para as subseqüentes análises com o pacote *Maftools*⁶², que permite analisar de uma forma integrativa o conjunto de mutações encontradas nos subgrupos moleculares de LLA B-*other*; tais como indicar possíveis mutações *driver*, vias enriquecidas, análises de associação com os dados clínicos, análises estatísticas e, além disso, pode gerar os resultados graficamente.

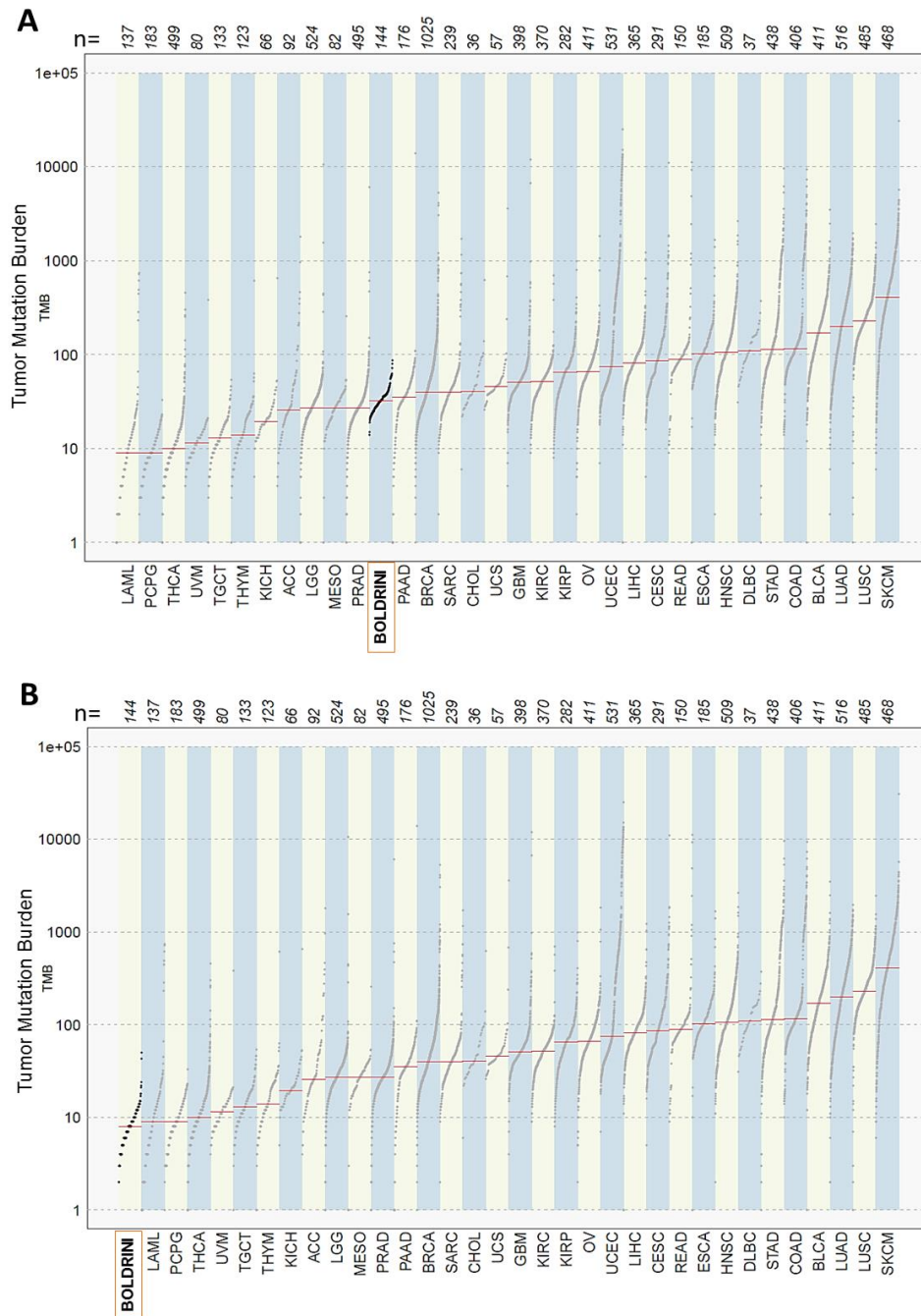


Figura 1. Carga mutacional da nossa coorte de 144 pacientes LLA B-other analisada (“Boldrini”) em comparação a outras coortes do banco de dados TCGA (*The Cancer Atlas Genome*). A. Antes dos filtros adicionais e B. Após a aplicação dos filtros (vide texto). A taxa de mutação ficou semelhante a esperada carga mutacional para cânceres hematológicos como o LAML, e considerando o sequenciamento direcionado para apenas 1.385 genes relacionados ao câncer (maior cobertura vertical no sequenciamento em relação a exomas completos), no extremo oposto observamos cânceres com alta taxa de mutação, tais como câncer de pulmão (LUSC) e melanoma (SKCM). Siglas: LAML- Acute Myeloid Leukemia, ACC-Adrenocortical carcinoma, BLCA-Bladder Urothelial Carcinoma, LGG-Brain Lower Grade Glioma, BRCA-Breast invasive carcinoma, CESC-Cervical squamous cell carcinoma and endocervical adenocarcinoma, CHOL-Cholangiocarcinoma, LCML-Chronic Myelogenous Leukemia, COAD-Colon adenocarcinoma, CNTL-Controls, ESCA-Esophageal carcinoma, FPPP-FFPE Pilot Phase II, GBM-Glioblastoma multiforme, HNSC-Head and Neck squamous cell carcinoma, KICH-Kidney Chromophobe, KIRC-Kidney renal clear cell carcinoma, KIRP-Kidney renal papillary cell carcinoma, LIHC-Liver hepatocellular carcinoma, LUAD-Lung adenocarcinoma, LUSC-Lung squamous cell carcinoma, DLBC-Lymphoid Neoplasm Diffuse Large B-cell Lymphoma, MESO-Mesothelioma, MISC-Miscellaneous, OV-Ovarian serous cystadenocarcinoma, PAAD-Pancreatic adenocarcinoma, PCPG-Pheochromocytoma and Paraganglioma, PRAD-Prostate adenocarcinoma, READ-Rectum adenocarcinoma, SARC-Sarcoma, SKCM-Skin Cutaneous Melanoma, STAD-Stomach adenocarcinoma, TGCT-Testicular Germ Cell Tumors, THYM-Thymoma, THCA-Thyroid carcinoma, UCS-Uterine Carcinosarcoma, UCEC-Uterine Corpus Endometrial Carcinoma, UVM-Uveal Melanoma.

6.2 Análise das mutações SNV e *small indel* nos pacientes LLA B-*other*

Após a aplicação dos filtros descritos acima, foram observadas um total de 1150 mutações SNV e *small indel* (i.e., 18 deleções *frameshift*, 20 inserções *frameshift*, 12 deleções *in-frame*, 13 inserções *in-frame*, 1009 mutações *missense*, 69 mutações *nonsense* e 1 *translation start site*) distribuídos em 539 genes. Como mencionado anteriormente, os cânceres infantis têm uma taxa de mutação baixa e em sua maioria, surgem em poucos genes. Além disso, pelo fato de estarmos usando um painel direcionado de captura, a cobertura horizontal é muito maior em comparação com sequenciamentos de exomas completos realizados pelo TCGA. Aqui, observamos uma média de 7 mutações somáticas por paciente. A maioria das variantes observadas foram mutações pontuais do tipo *missense*, representando 88% do total de variantes. A troca de nucleotídeo mais comumente foi C>T (33%), seguido por C>A (28%) e T>C (19%), caracterizando uma predominância de transições sobre transversões (Figura 2).

Os dez genes do painel mais frequentemente mutados foram *NRAS* (23%), *MDC1* (22%), *KRAS* (14%), *FLT3* (8%), *PAX5* (8%), *STAT6* (8%), *SETD2* (7%), *MYC* (6%), *IKZF1* (6%) e *KMT2D* (7%) (Figura 2E). Na Figura 3, podem ser observados em mais detalhes os top 42 genes mutados (com pelo menos 3% de frequência, ponto de corte considerado para visualização gráfica) em relação aos subgrupos moleculares de LLA B-*other*.

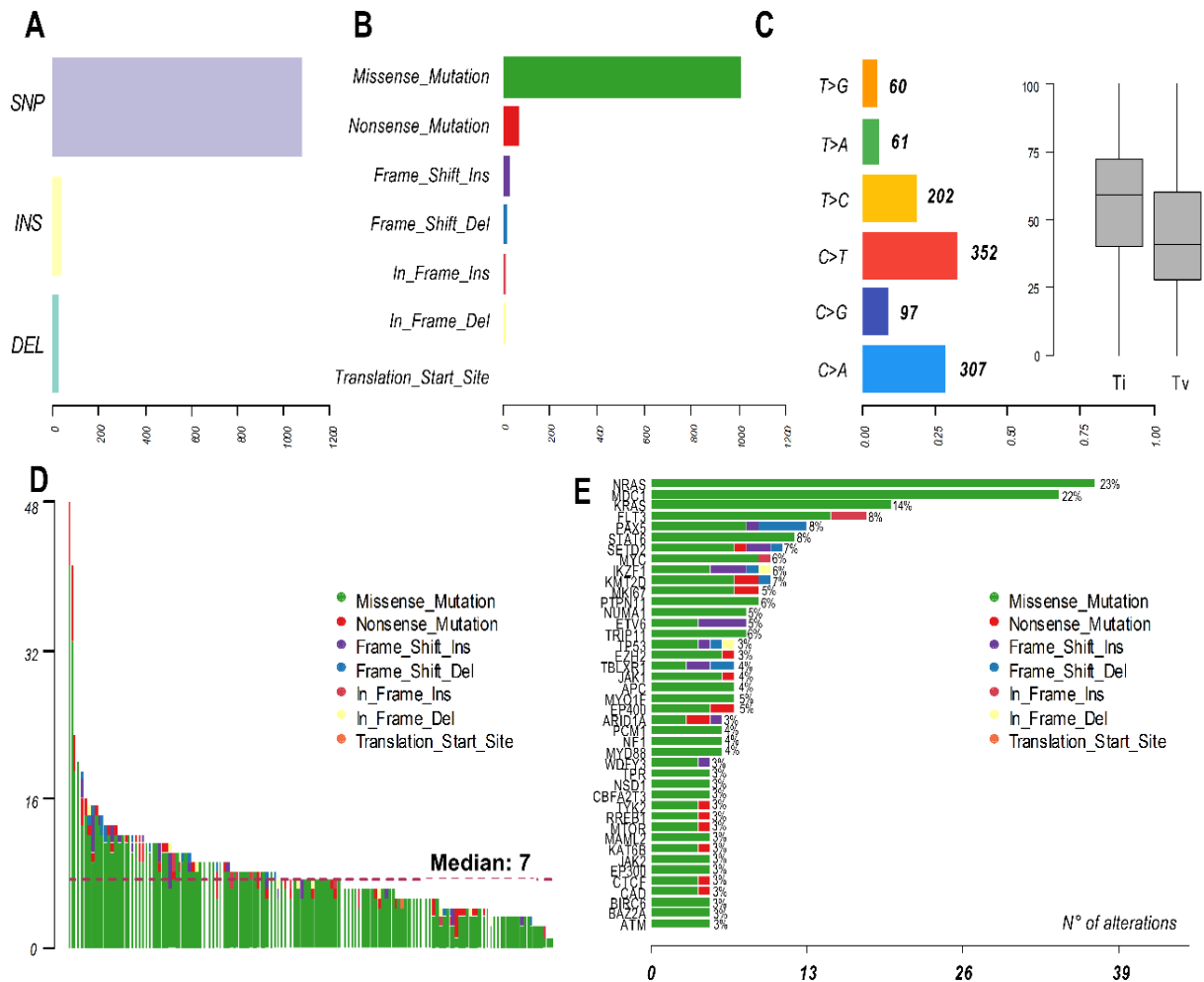


Figura 2. Classificação das variantes (SNVs e small indels) encontradas em 144 pacientes dos subgrupos moleculares de LLA B-other. A. Tipo da mutação; B. Classificação das mutações; C. Distribuição e troca de nucleotídeos mais frequentes; D. Número de mutações por amostra; E. Lista dos top 42 genes mais frequentemente mutados.

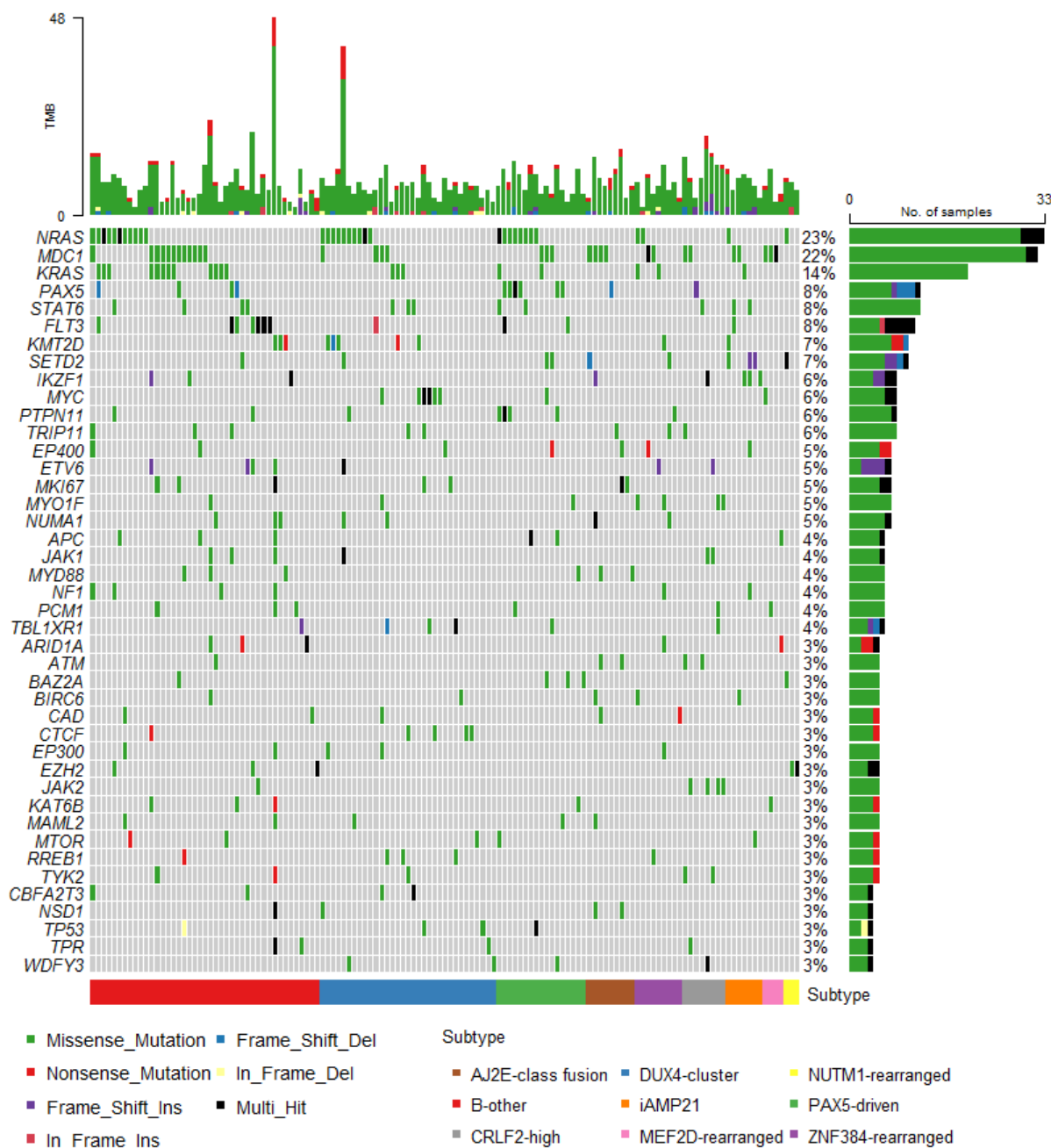


Figura 3. Frequência e espectro de mutações observados entre os subgrupos moleculares de LLA B-other. Gráfico de *oncoplot* exibindo o perfil de mutações encontradas entre os subgrupos. Os top 42 genes mutados estão ordenados de acordo com a sua frequência e tipo de mutação, e foram observados em 133 (92.36%) dos 144 pacientes analisados e agrupados pelos subtipos, como indicado pela anotação (barras coloridas – B-other ‘rest’ em vermelho).

Para identificar os genes ou conjunto de genes que estão alterados de forma mútua ou de forma exclusiva, realizamos uma análise de correlação e observamos que *FLT3/PAX5*, *JAK1/MKI67*, *EP300/KMT2D*, *EP300/CAD*, *EP300/MAML2*, *MAML2/NSD1* co-ocorrem de forma significativa ($p < 0.01$); e de forma

interessante *MDC1/NRAS* são mutuamente exclusivos e estão frequentemente mutados nos pacientes analisados (Figura 4A).

Em relação aos genes significativamente enriquecidos entre os subgrupos moleculares de LLA B-*other*, observamos 18 genes mutados ($fdr < 0.05$ e $*fdr < 0.01$) em grupos específicos. Digno de atenção, mutações em *JAK2* e *JAK1*, assim como mutações em genes menos associados à LLA como *ATM* (*ATM Serine/Threonine Kinase*) e *MYO1F* (*Myosin IF*) foram enriquecidos no subgrupo *CRLF2*-high ‘only’. Mutações em *TYK2* (*Tyrosine Kinase 2*), outro membro da família de proteínas Jak, também foi observado no subgrupo *CRLF2*-high ‘only’, como esperado. Mutações em *PAX5*, *PTPN11* e *NRAS* estão associados com o subgrupo *PAX5*-driven, além do gene *BAZ2A* (*Bromodomain Adjacent to Zinc finger domain 2A*), não descrito anteriormente. Mutações em *MYC* no grupo *DUX4*-cluster já foi descrito, porém o enriquecimento de mutações no gene *CTCF* (*CCCTC-Binding Factor*) é novo. Mutações SNV e small indel nos genes *IKZF1* e *SETD2* (*SET Domain Containing 2, Histone Lysine Methyltransferase*) em *iAMP21* não foram descritas anteriormente, apenas deleções. Mutações no gene *EZH2* (*Enhancer of Zeste 2 Polycomb Repressive Complex 2 Subunit*) em *NUTM1*-rearranged; *MDC1* (*Mediator Of DNA Damage Checkpoint 1*) em *MEF2D*-rearranged e; *NSD1* (*Nuclear Receptor Binding SET Domain Protein 1*) e *ATM* em *AJ2E*-class fusion não foram descritas anteriormente. Chama a nossa atenção o enriquecimento de mutações em *KRAS* (A146P, A146T, K117N, Q61P, G13D) e *FLT3* (R845G, Y842H, R595I e uma mutação de inserção in frame R595_E596insAPWDGSDPG) no subgrupo molecular B-*other* ‘rest’ (i.e., pacientes LLA-B para os quais não foram observadas alterações genéticas recorrentes e descritas na literatura que pudessem ser usadas para agrupá-los em subgrupo específico) (Figura 4B).

Utilizamos o algoritmo *OncodriveCLUST*⁶³, que analisa as informações contidas no banco de dados “*Catalogue of Somatic Mutations in Cancer (COSMIC)*”⁶⁴ e *Cancer Gene Census (CGC)*⁶⁵ para identificar possíveis genes *driver* que carregam mutações agrupadas em regiões específicas (*hotspots*) da sequência de aminoácidos. Essa abordagem computacional é voltada para identificar mutações de ganho de função, os quais geralmente são positivamente selecionadas durante a tumorigênese. Na figura 4C, observamos o enriquecimento de genes e o número de mutações observadas em torno de *hotspots* (colchetes). Muitos genes relacionados à leucemia (e.g., *CRLF2* (F232C), *AFF4* (G829R), *IL7R* (éxon 6), *KRAS*, *PAX5*, *PTPN11*, *FLT3* e

JAK2 (R683G) foram significativamente enriquecidos em regiões ativadoras ($\text{fdr} < 0.05$). Outros genes pouco descritos na leucemia, tais como *KDM1A* (D390G), *MAPK1* (G245V) e *SEPT2* (L329H) também foram enriquecidos. É importante mencionar que existem outros algoritmos mais estabelecidos, como o *MutSigCV*, que analisa se a frequência de mutações em tal gene é maior do que o esperado em uma coorte de amostras tumorais, o qual poderia complementar nossas análises.

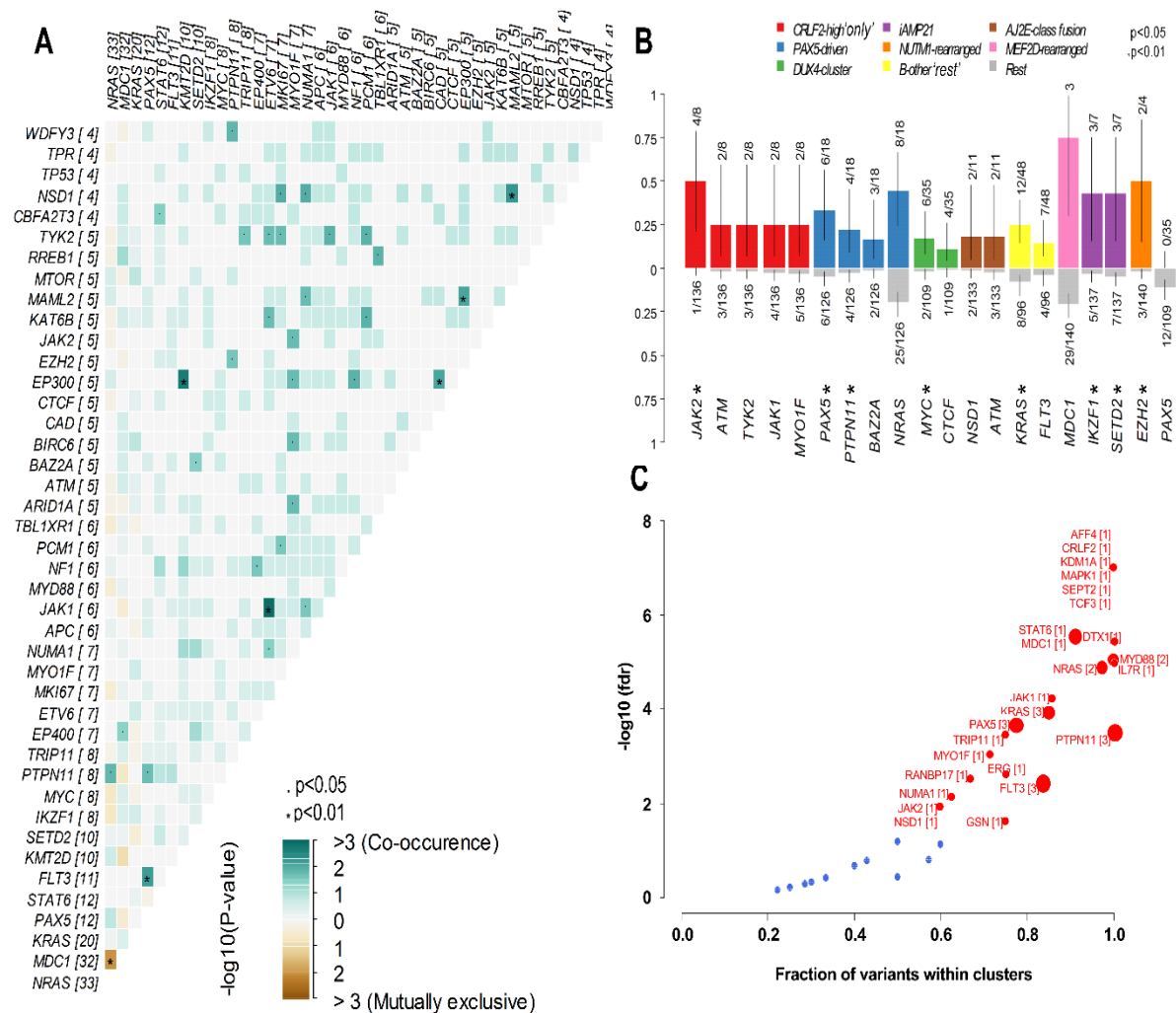


Figura 4. Análise de interações e genes significativamente mutados entre os subgrupos moleculares de LLA B-other. A. Genes mutuamente exclusivos ou que co-ocorrem entre os top 42 genes frequentemente alterados na coorte estudada. B. Análise de associação entre os genes mutados e os subgrupos moleculares (*Fisher exact test*, *significant associations*, $p\text{-value} < 0.05$ and $*p < 0.01$). Considerando genes mutados em no mínimo 3 pacientes analisados. (*Performs fishers test on 2x2 contingency table for WT/Mutants in group of interest vs rest of the sample*). C. Genes driver identificados pelo algoritmo *OncodriveCLUST* ($\text{fdr} < 0.05$).

Além disso, a análise integrativa para identificar mutações recorrentes dentro de uma mesma via biológica mostrou um enriquecimento de alterações em genes da via dos receptores tirosina quinase-Ras (RTK-RAS pathway), 83/144 dos pacientes B-other e 24/85 dos genes dentro da via (Figura 5). Genes frequentemente mutados e selecionados de acordo com o envolvimento em vias importantes da leucemogênese como “*B-cell development/Transcriptional, Cell Cycle/TP53 pathway, Epigenetic/Chromatin modifiers, JAK/STAT, Wnt, PI3K/mTOR, NOTCH and Others*” (lista de genes e vias descritos na literatura)^{66,67} foram compilados e representados na Figura 6. Observa-se que mutações em vias específicas estão enriquecidos nos subgrupos de B-other.

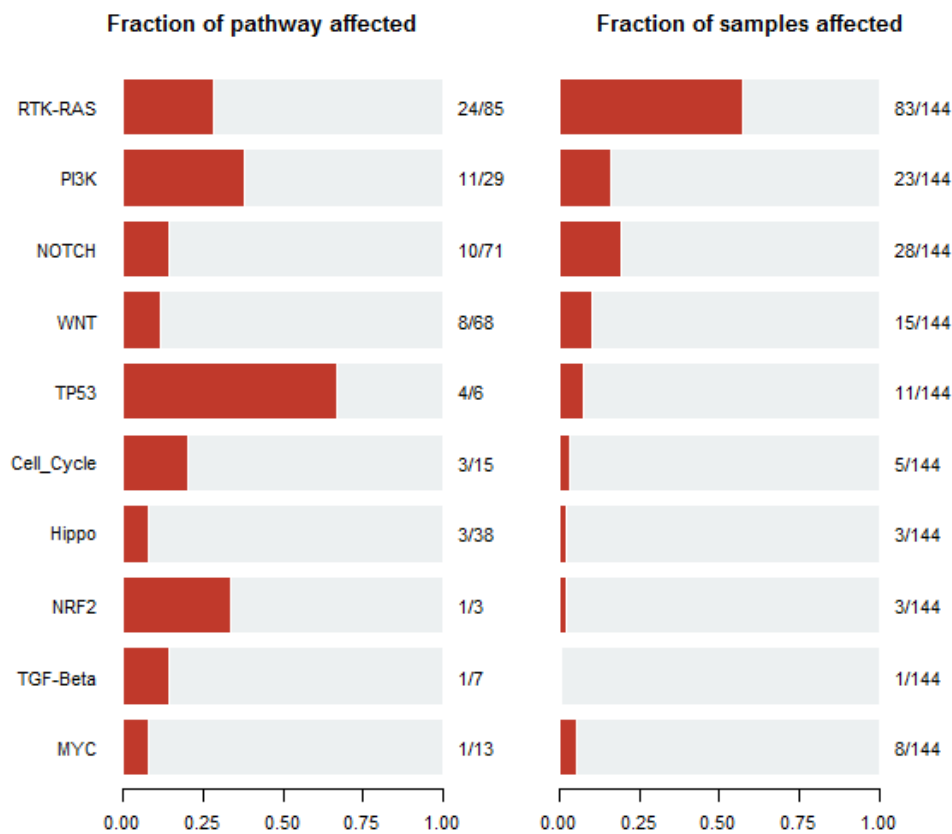
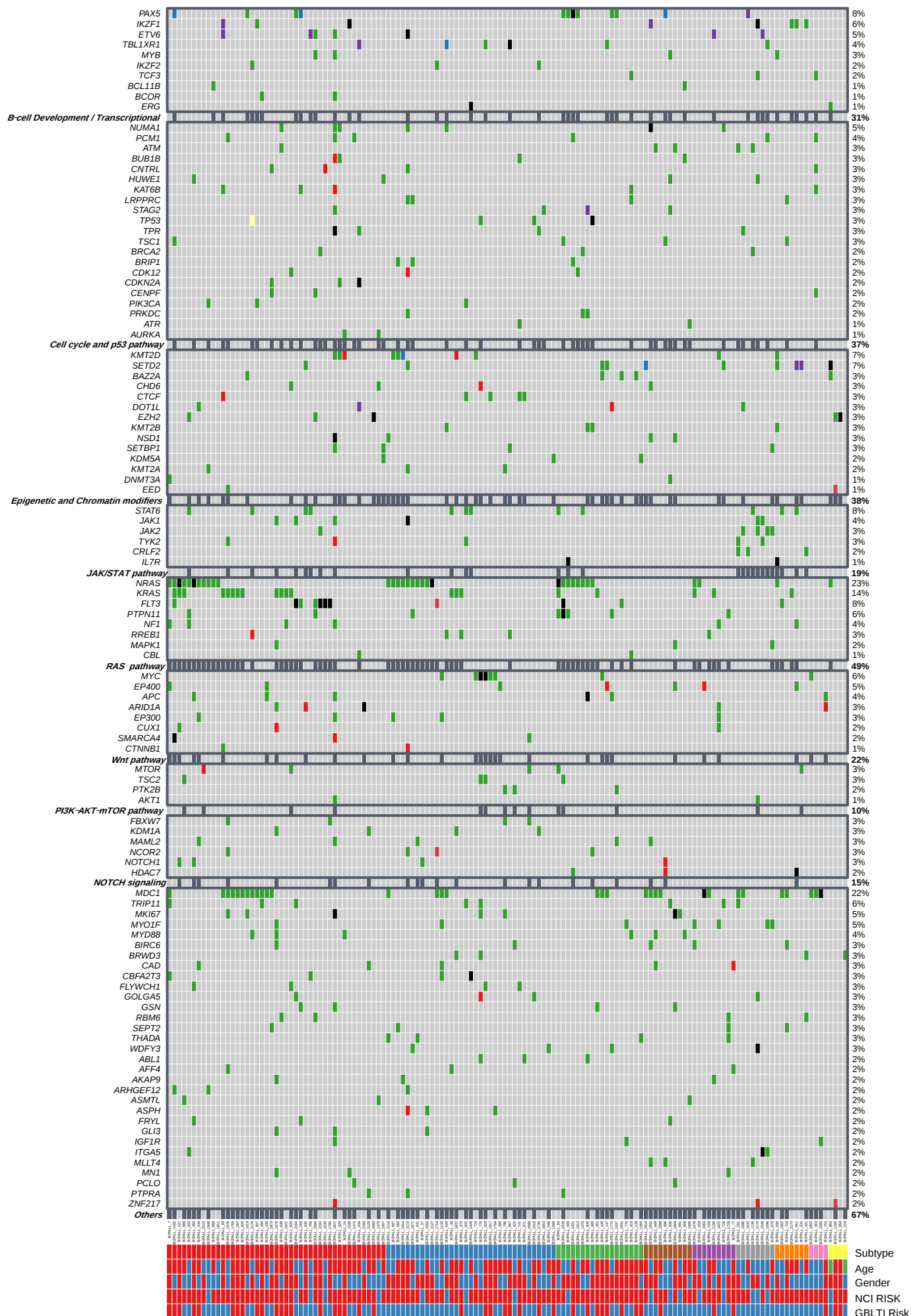


Figura 5. Porcentagem de mutações em genes de vias importantes relacionadas ao câncer. Fração de genes mutados para cada via e fração de pacientes com mutações em genes mutados para essa via.

Figura 6 [Abaixo]. Análise integrativa mostrando os genes frequentemente mutados em importantes vias biológicas da leucemia (compilados da literatura). Gráfico de Oncoplot mostrando em detalhes os genes e as vias enriquecidas entre os subgrupos moleculares de LLA B-other. Dados clínicos como idade, gênero e risco estão apresentados abaixo (barras coloridas – B-other ‘rest’ em vermelho).

Figura 6.



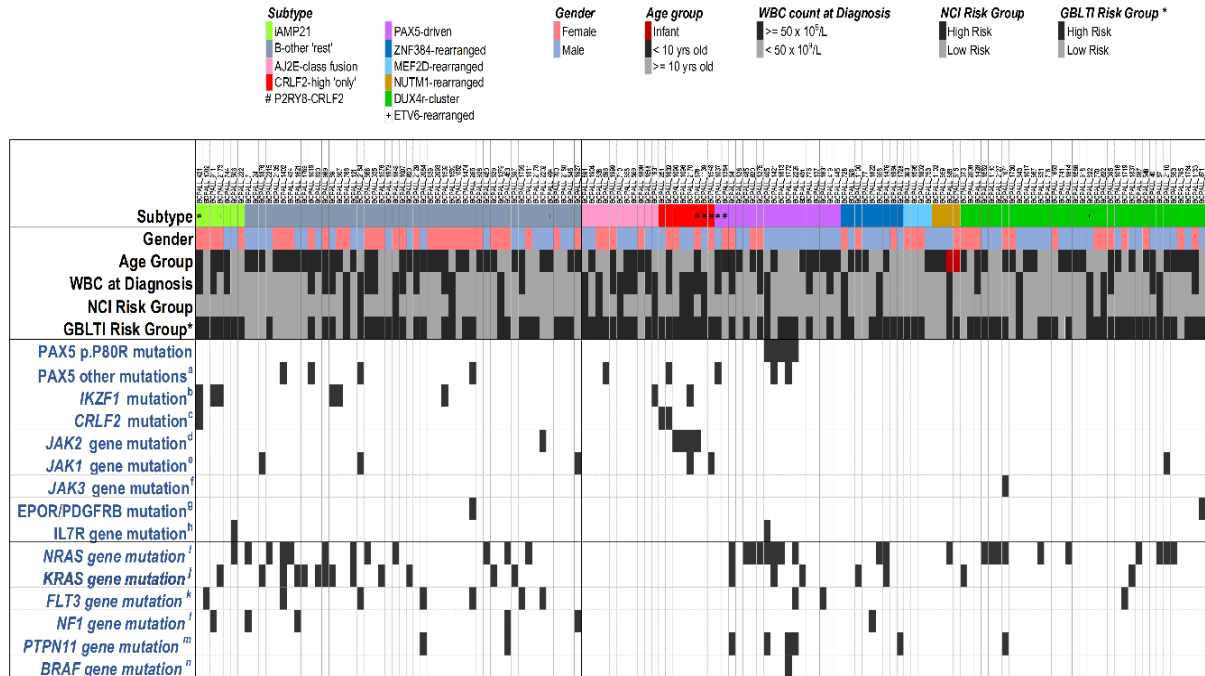
Além de confirmar várias mutações pontuais bem conhecidas na leucemia, tais como *PAX5* p.P80R (n=5), *JAK2* p.R683G (n=3), *CRLF2* p.F232C (n=3) e *IKZF1* p.N159Y (n=1); genes frequentemente mutados em mais de uma região da sequência de aminoácidos também foram observadas, tais como *NRAS*, *KRAS*, *FLT3*, *KMT2D*, *PTPN11*, *TP53*, *NF1*, *EZH2* entre outras. Um número significativo de mutações em genes pouco descritos, mas com possíveis funções na leucemia também foram encontradas (i.e., *MDC1*, *STAT6*, *SETD2*, *MKI67*, *NUMA1*, *ETV6*, *TRIP11*, *TBLXR1*, *ARID1A* entre outras).

6.3 Associações entre alterações genéticas e dados biológicos/clínicos dos pacientes LLA B-other

Em geral, com as análises acima discutidas, observou-se que mutações em genes das principais vias da leucemia estão enriquecidas em subgrupos moleculares específicos, indicando um possível papel dessas alterações na progressão da doença e/ou à um possível fenótipo clínico. Por exemplo, mutações em genes da via Ras foram correlacionadas positivamente com o subgrupo molecular B-other 'rest' (p-value=0.02). De maneira interessante, 22 pacientes com mutações em genes *driver* que ativam vias de JAK/STAT (i.e., fusões *AJ2E*-class e alta expressão de *CRLF2*) não possuem mutações em vias de RAS (mutuamente exclusivos, p-value<0.0001).

Análise de correlação das vias enriquecidas com os dados demográficos desses pacientes foram realizadas e as associações que merecem destaque são; (1) B-other 'rest' apresentou um enriquecimento de meninas (p-value: 0.05) e mutações em genes da via RAS (p-value: 0.02). (2) Fato intrigante é que entre todos os subgrupos de B-other, os meninos estão positivamente associados com mutações em RAS (p-value=0.04). Não temos uma clara explicação para essa associação significativa com o gênero. (3) Observamos também que os genes da via RAS cujas mutações mais se correlacionam significativamente com o subgrupo B-other 'rest' é o gene *KRAS* p=0.01 e *FLT3* p=0.04; (4) Mutações em *MYC* estão enriquecidas no subgrupo *DUX4*-cluster (p-value=0.05); (5) *AJ2E*-class fusion e *CRLF2*-high 'only' se correlacionam significativamente com a classificação de alto risco pelo protocolo GBLTI (p-value=0.02) e alta contagem de leucócitos (WBC) ao diagnóstico (p-value<0.001) e quase metade dos pacientes com alterações em *CRLF2* têm mutações

em JAK2. Como reportado na literatura, essas lesões genéticas mostraram-se associadas ao subtipo de alto risco “Ph-like” (Figura 7).



*GBTLI risk: *Patients were considered to be 'low risk' if they met all the following criteria: age ≥ 1.00 and up to 10 years; WBCs at diagnosis lower than 50,000/ μL ; a rapid early response day 7 (WBC < 5,000/ μL), day 14 (no peripheral blasts and < 25% blasts in BM) and day 28 (< 5% blasts in BM) of induction therapy. Mediastinal mass, CNS, or other extramedullary leukemic involvement, as well as, immunophenotype and cytogenetics, were not considered for risk classification. ^a PAX5 mutations as p.S352Nfs*52, p.M335Wfs*68, p.V208Cfs*70, p.P34T, p.V26Gfs*49, p.V26G and p.V26Ffs*3. ^b IKZF1 mutation as p.G158S, p.D22Gfs*6, p.V57Gfs*6, p.R97Kfs*22, p.A226Pfs*4, p.L160_L161del, p.H163D, p.L192P and p.N159Y. ^c CRLF2 mutation as p.F232C. ^d JAK2 mutation as p.R683G, p.A658S and p.L925P. ^e JAK1 mutation as p.E668D, p.E667* and p.A634D. ^f JAK3 mutation as p.E161K. ^g EPOR and PDGFRB mutations as p.S433A, p.R922H, respectively. ^h IL7R mutation as p.L241S, p.L242Cfs*4, p.L242*, p.L242F and p.T244Nfs*26. ⁱ NRAS mutation as p.A146T, p.Q61H, p.Q61K, p.T58I, p.G13D, p.G13C, p.G12D, p.G12A, p.G12V, p.G12S, p.G12C and p.G12R. ^j KRAS mutation as p.A146V, p.A146P, p.A146T, p.K117N, p.Q61P, p.T58I, p.D33E, p.G13D, p.G12V, p.G12D and p.G12S. ^k FLT3 mutation as p.L939_T940insVGV, p.I867T, p.R845G, p.Y842H, p.N841K, p.N841Y, p.D835_I836insPA, p.D835G, p.D835Y, p.R834Q, p.N676K, p.R595_E596insAPWDGSDPG, p.R595I, p.F594C, p.F594L, p.D593E and p.Y591D. ^l NF1 mutation as p.R156L, p.C383R, p.G1602W, p.R1769Q, p.R1846Q and p.K2628N. ^m PTPN11 mutation as p.D61V, p.A72S, p.E76G, p.S502P, p.S502L, p.T507K, p.Q510L and p.Q510H. ⁿ BRAF mutation as p.D594A.

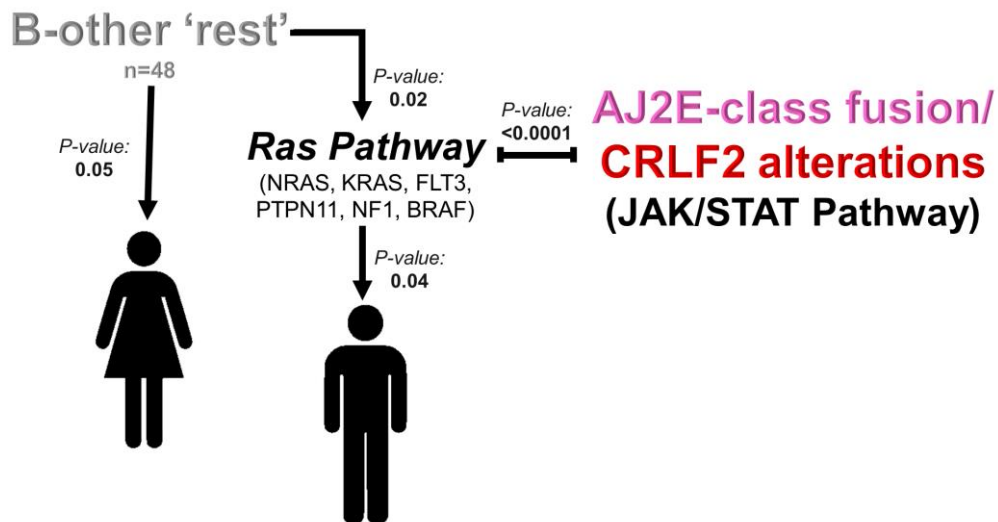


Figure 7. Resumo esquemático dos subgrupos moleculares de LLA B-other, dados demográficos e os principais genes mutados nas vias de JAK/STAT e RAS observados no nosso estudo. Associação das alterações genéticas com os subgrupos moleculares e fatores clínicos foram analisados pelo teste exato de Fisher (Fisher's exact test, two-sided).

Para analisar o valor prognóstico dessas alterações na evolução clínica dos pacientes, avaliamos a sobrevida livre de leucemia (*Leukemia Free Survival*, LFS) para as principais vias alteradas e para os diferentes subgrupos moleculares de LLA B-*other*. A LFS foi calculada a partir da data de remissão completa até a data do último acompanhamento ou o primeiro evento (recaída ou morte relacionada a leucemia). A doença refratária foi considerada um evento no tempo zero. Foram excluídos da análise pacientes que foram a óbito por motivos não relacionados a doença (sepse). Desta forma, foram incluídos 132/144 pacientes diagnosticados com LLA B-*other* para análise de desfecho clínico. A mediana de seguimento do estudo foi de 5 anos. A LFS estimada da coorte total analisada foi de 83±4% (132/144).

Em relação ao valor prognóstico das principais vias alteradas na LLA B-*other*, observamos uma associação entre mutações em genes da via JAK/STAT e um pior prognóstico de forma significativa (5-y LFS, 72±10% versus 85±4%; p=0,004). Apesar da via RAS não ter apresentado um valor prognóstico ruim, o gene KRAS apresentou uma tendência de fator preditivo de risco (5-y LFS, 64±12% versus 86±3%; p=0,06).

Tabela 1. Análise do valor prognóstico das mutações SNV e *small indel* nas principais vias alteradas da LLA-B (frequência analisada para as mutações encontradas em 140 (97,22%) das 132/144 LLA B-*other* analisadas.

	# de amostras com alterações na via	Sobrevida livre de leucemia (Leukemia free Syrvrival, LFS)	Estatística (§p-valor)
		5-anos ± STDE	
Desenvolvimento de células B e Transcricional	39	0.80 ± 0.07 vs 0.84 ± 0.04	0.264
Ciclo celular e p53	49	0.78 ± 0.06 vs 0.86 ± 0.04	0.402
Epigenética e Modificadores de cromatina	51	0.90 ± 0.04 vs 0.77 ± 0.05	0.212
Via JAK/STAT	25	0.72 ± 0.10 vs 0.85 ± 0.04	0.004
Via de RAS	63	0.88 ± 0.04 vs 0.77 ± 0.06	0.225
gene KRAS	18	0.64 ± 0.12 vs 0.86 ± 0.03	0.066
gene FLT3	9	0.89 ± 0.10 vs 0.83 ± 0.04	0.800
Via de Wnt	29	0.91 ± 0.06 vs 0.80 ± 0.04	0.201
Via de PI3K-AKT-mTOR	12	0.69 ± 0.15 vs 0.84 ± 0.04	0.279
Sinalização de NOTCH	18	0.82 ± 0.04 vs 0.86 ± 0.10	0.851
Outras	87	0.78 ± 0.05 vs 0.92 ± 0.04	0.215

Em relação ao valor prognóstico dos subgrupos moleculares de LLA B-*other*, nenhum grupo apresentou um pior desfecho clínico significativo, porém observamos que iAMP21, ZNF384-rearranged, MEF2D-rearranged, NUTM1-

rearranged, *AJ2E*-class fusion e *CRLF2*-high 'only' tiveram uma estimativa de sobrevida igual ou abaixo de 75% (Tabela 2).

Tabela 2. Análise de sobrevida para os diferentes subgrupos moleculares de LLA B-*other* tratadas no Centro Infantil Boldrini. Probabilidade estimada de sobrevida livre de leucemia (*Leukemia-Free Survival*, LFS) no tempo de 5 anos.

	LLA B-other N (%)	Analisado N (%)	Sobrevida Livre de Leucemia (Leukemia-free survival, LFS)			Estatística (§p-valor)
			5-anos ± STDE			
			TOTAL 144	TOTAL 132	0.83 ± 0.03	
Subgrupos						
B-other 'rest'	48 (33.3%)	45 (34.1%)	0.84 ± 0.06	vs	0.82 ± 0.04	0.739
DUX4-cluster	35 (24.3%)	29 (22%)	0.89 ± 0.06	vs	0.81 ± 0.04	0.182
PAX5-driven	18 (12.5%)	18 (13.6%)	0.88 ± 0.08	vs	0.82 ± 0.04	0.409
iAMP21	7 (4.8%)	6 (4.5%)	0.75 ± 0.22	vs	0.83 ± 0.04	0.286
ZNF384-rearranged	9 (6.3%)	9 (6.8%)	0.76 ± 0.15	vs	0.84 ± 0.04	0.794
MEF2D-rearranged	4 (2.8%)	4 (3%)	0.75 ± 0.22	vs	0.83 ± 0.04	0.659
NUTM1-rearranged	4 (2.8%)	4 (3%)	0.75 ± 0.22	vs	0.83 ± 0.04	0.751
AJ2E-class fusion	11 (7.6%)	9 (6.8%)	0.71 ± 0.18	vs	0.84 ± 0.03	0.173
CRLF2-high 'only'	8 (5.6%)	8 (6.1%)	0.75 ± 0.15	vs	0.84 ± 0.04	0.103
CRLF2-high*	11 (7.6%)	11 (8.3%)	0.82 ± 0.12	vs	0.83 ± 0.04	0.349
alterações em JAK/STAT**	22 (15.3%)	20 (15.2%)	0.77 ± 0.10	vs	0.84 ± 0.04	0.113

*Três de 11 pacientes com expressão elevada de CRLF2 (principalmente P2RY8-CRLF2 e IGH-CRLF2), foram classificados como um iAMP21 e duas como PAX5-driven.

**Pacientes com alterações CRLF2-high e fusões AJ2E-class foram analisados conjuntamente, uma vez que ambos ativam a via JAK/STAT.

6.4 Discussão das principais mutações *SNV* e *small indel* encontradas

Neste capítulo, buscamos aprofundar nossas análises para identificar mutações recorrentes nos diferentes subtipos de LLA B-*other*. Avaliamos o perfil de alterações somáticas *SNV* e *small indel* em nove subgrupos moleculares de LLA B-*other* e observamos assinaturas mutacionais em grupos específicos. O espectro de mutações que ocorre em determinado subgrupo de LLA B-*other* pode afetar vias oncogênicas específicas, e conduzir a doença para diferentes fenótipos clínicos. Embora muitas das mutações *SNV* e *small indel* frequentemente observados na LLA B-*other* tenham sido já descritas, nossa análise integrativa conseguiu identificar novas mutações, novas associações e vias enriquecidas entre os subgrupos moleculares.

DUX4-cluster é um subgrupo já descrito na literatura internacional, compreendendo cerca de 4-5% das LLA-B. É caracterizado pela expressão de *DUX4* (*Double Homeobox 4*) e desregulação do gene *ERG* (*ETS Transcription Factor ERG*), resultando na expressão de uma isoforma alternativa de *ERG*, *ERGalt*, e na frequente deleção do gene *ERG*.⁶⁸ Assim como descrito anteriormente por Zhang J et al, 2016⁶⁸, aqui também observamos um enriquecimento em mutações no fator de transcrição *MYC*, os quais não foram observados nos outros subgrupos moleculares. De maneira

pioneira, observamos um enriquecimento de mutações no gene *CTCF* em leucemias LLA-B do subgrupo DUX4-cluster. *CTCF* é um regulador transcricional com vários domínios conservados de zinc finger (ZF) e está associado as alterações epigenéticas que ocorrem nas leucemias. Sabe-se que o gene *DUX4* está localizado na região subtelomérica do cromossomo 4 e 10, conhecida como “*D4Z4 repeat array*”. A sua expressão é regulada por vários fatores, incluindo a modificação epigenética, a modificadores de cromatina e a presença de *enhancers* miogênicos (por isso também está relacionado a distrofias musculares, como o *Facioscapulohumeral dystrophy*, FSHD). E nesse sentido, revisando a literatura, encontramos a associação entre *DUX4* e *CTCF* na distrofia FSHD, no qual observou-se que *CTCF* pode regular a transcrição por criar sítios acessíveis ou inacessíveis na cromatina e no caso de *DUX4*, age reprimindo a sua transcrição.⁶⁹

O subgrupo molecular *PAX5-driven* foi um termo criado pelo grupo do professor Charles Mullighan (St. Jude) e descrito no trabalho de Gu Z et al, 2019⁷⁰. *PAX5-driven* engloba dois subtipos com perfis de expressão gênica distintos, mas que geneticamente são caracterizados pela frequente alteração no fator de transcrição, *PAX5*. Tais alterações envolvem rearranjos cromossômicos de *PAX5* com diversos parceiros de fusão; amplificação intragênica (*PAX5*^{AMP}- entre os éxons 2-5) e mutações SNV, os quais são agrupados como *PAX5alt*, e o segundo subtipo *PAX5 P80R* é caracterizado pela mutação p.Pro80Arg. No nosso trabalho encontramos 5 casos p.P80R e 2 casos com outras variantes mutagênicas (p.P34T e deleção p.V26Fs*3). Além disso, é descrito que mutações em genes da via RAS são frequentes nesse subgrupo, principalmente *NRAS* e *PTPN11*, o qual também observamos na nossa coorte.

Em relação ao subgrupo *CRLF2-high ‘only’*, nossos achados corroboram com os dados internacionais, *i.e.* aproximadamente um terço apresentam mutações concomitantes em *JAK2*, ou mutações em outros genes importantes para a ativação da via JAK/STAT, tais como *JAK1* e *TYK2*.^{49,50} Já entre os casos *AJ2E-class fusion*, encontramos mutações recorrentes de forma significativa em *NSD1* (p.C1021F) e *ATM* (p.L240P e p.G2765V) do que nos outros subgrupos, e até onde sabemos, nenhuma associação anterior desses genes com a LLA pediátrica foi relatada. Mutações recorrentes em genes pouco conhecidos e/ou que estão enriquecidos em subgrupos constituídos por poucos pacientes deve ser analisada com cuidado, por conta de possível viés amostral. Nesta tese procuramos controlar eventuais artefatos

de análise estatística de associação utilizando a correção de Monte Carlo quando da realização de teste de Fisher.

O gene *MDC1* é um dos mais frequentemente mutados na nossa coorte e está enriquecido no subgrupo *MEF2D*-rearranged (n=3/4). Ele codifica uma proteína que interage com importantes complexos (e.g., *ATM*, *CHECK1* e *CHEK2*) que medeiam a parada do ciclo celular em resposta a danos no DNA.⁷¹ Mutações em *MDC1* foram descritos em tumores sólidos de adulto, mas nenhuma literatura em leucemia pediátrica foi encontrada, o que é surpreendente dada a alta frequência *MDC1* em nosso estudo.

O grupo *NUTM1*-rearranged está enriquecido com mutações no gene *EZH2* (n=2/4) em relação aos outros. *EZH2* é muito conhecida e estudada na hematopoiese normal e câncer. Esse gene codifica uma proteína do complexo repressivo 2 (*PRC2*, do *polycomb group*, *PcG*) envolvida na regulação epigenética. Possui atividade de histona metiltransferase ao metilar a lisina (K)-27 da histona H3 (H3K27), modificando a conformação da cromatina e promovendo repressão gênica. A proteína Ezh2 é necessária na medula óssea para progressão de células pro-B em células pre-B e células B imaturas, com possível função nas fases iniciais da diferenciação das células B.⁷² Sua inativação causada por mutação ou baixa expressão contribui para a patogênese e está associado a um mau prognóstico.

Dois importantes genes da via “*B-cell Development and Transcriptional*” estavam entre os nossos *tops* genes frequentemente mutados e merecem destaque pela relação com o subtipo de LLA t(12;21)*ETV6-RUNX1*, considerado de baixo risco. *ETV6* é um gene que codifica um repressor transcricional e desempenha um papel crítico na hematopoiese. Um total de 8 mutações em *ETV6* foram detectadas em 7 casos, 4 das quais eram mutações de “*Frame_Shift_Ins*”. Sabe-se que mutações de perda de função do alelo *ETV6* são um evento secundário frequente em LLA do subtipo t(12;21)*ETV6-RUNX1*.⁷³ Assim, é interessante entender se as alterações de *ETV6* independente do rearranjo primário contribui para o desenvolvimento da leucemia. O gene *TBL1XR1* (*TBL1X Receptor 1*) recruta o complexo ubiquitina/proteassoma 19S e ativa a transcrição através de receptores nucleares. Embora deleções focais de *TBL1XR1* tenham sido descritas na LLA pediátrica do subtipo t(12;21)*ETV6-RUNX1*,⁷⁴ o nosso trabalho é o primeiro a encontrar variantes mutacionais do tipo de *SNP* (p.F487L, p.F487I, p.W376C), DEL (p.E393Ffs*22 e

p.H139Yfs*4) e INS (p.S461Vfs*17 e p.D315Qfs*25) no gene *TBL1XR1* e em pacientes LLA B-*other*.

Os genes da via RAS estão frequentemente alterados no câncer. Mutações em RAS são comuns nas leucemias pediátricas de LLA-B, principalmente nos genes *NRAS* e *KRAS*.⁷⁵ Entretanto, não há estudos prévios que tenham analisado a frequência e o prognóstico dessas mutações nas LLA B-*other*. Nós observamos alterações significativas na via de RAS, principalmente de *KRAS* e *FLT3* no subgrupo B-*other* ‘rest’, indicando um papel central desses genes na progressão tumoral. Além disso, foi importante notar que os casos *AJ2E*-class fusion e *CRLF2*-high ‘only’, os quais ativam principalmente as vias JAK/STAT e possuem similaridades com os casos de BCR-ABL1-positivo (e muitas vezes são agrupados como BCR-ABL1-like), tiveram uma correlação negativa com mutações em RAS, sugerindo mecanismos tumorais divergentes. De fato, no artigo recente de Chan LN et al, 2020⁷⁶, foi demonstrado que o processo de leucemogênese chega a ser interrompido caso haja mutações concomitantes em JAK/STAT e RAS/ERK. O motivo disto, segundo se pensa, é uma característica própria das células B, decorrente de um processo sequencial de ativação dessas vias em duas etapas, a primeira visando a expansão celular e a segunda a parada da proliferação e rearranjo da cadeia leve dos genes de imunoglobulina visando a produção de um receptor funcional de células B, *i.e.* um anticorpo monoclonal. Resumidamente, o programa de maturação dos precursores de célula B, no estágio de pro-B, sobrevivem e proliferam graças à sinalização JAK/STAT oriunda dos receptores de citocina (e.g., *IL7R*, *CRLF2*, *EPOR*, *PDGFRA/B*), com destaque para o *IL7R*. Com a expressão de um pre-receptor de células B (pre-BCR) funcional, ocorre a ativação de *BLNK* e da via RAS/ERK que suprimem a transcrição da via JAK/STAT e levam ao rearranjo da cadeia leve para a obtenção de um receptor *BCR* funcional. Lesões oncogênicas nas LLA-B frequentemente mimetizam essas sinalizações, ou por meio da ativação de JAK/STAT no estágio pro-B ou, mutações em genes que ativam a via de sinalização RAS/ERK (e.g., *NRAS*, *KRAS*, *PTPN11*, *NF1* e *BRAF*), mimetizando a sinalização pre-B. O estudo de Chan LN et al, 2020⁷⁶, mostrou que, embora as mutações ativadoras de JAK/STAT e RAS/ERK sejam encontradas com frequência nas LLA com imunofenótipo correspondentes aos estágios pro-B e pre-B, elas ocorrem de forma mutuamente exclusiva. Juntas, elas ocorrem em apenas cerca de 3% dos casos. Além disso, os pesquisadores mostraram que quando uma dessas vias oncogênicas é ativada, as células suprimem ativamente

a outra via. A reativação genética ou farmacológica da via divergente que fora suprimida na doença, reverte a transformação maligna, o que pode abrir caminho para novas abordagens terapêuticas. Corroborando com esse estudo, observamos que a mesma convergência para uma única via ‘driver’ principal, JAK/STAT ou RAS/ERK também se dá nas células de pacientes LLA B-*other*.

Em relação as mutações de importância clínica, observamos que as alterações genéticas envolvendo a via JAK/STAT apresentaram um significativo valor preditivo de risco ($p=0.004$). Essa via está comumente alterada nos subgrupos classificados como sendo de alto risco, *i.e.*, Ph-like. Digno de nota, importantes mutações (fusões gênicas tais como *ABL*-class e mutações em *JAK2*) que afetam essa via estão sendo alvo de estudos clínicos com inibidores específicos, tais como imatinibe e ruxolitinib⁷⁷. Em nosso estudo, não encontramos uma diferença significativa de sobrevida entre os subgrupos moleculares de LLA B-*other*, porém *iAMP21*, *ZNF384*-rearranged, *MEF2D*-rearranged, *NUTM1*-rearranged, *AJ2E*-class fusion e *CRLF2*-high ‘only’ tiveram uma estimativa de sobrevida desfavorável (5-y LFS $\leq 75\%$) em comparação com os outros grupos. Nesse sentido, já é de conhecimento que alguns destes são classificados como grupos de alto risco (*iAMP21*, *AJ2E*-class, *CRLF2*-high ‘only’ e *MEF2D*-rearranged)⁷⁸, porém *NUTM1*-rearranged é descrito como um grupo de baixo risco. De fato, dentre os quatro casos com a fusão envolvendo *NUTM1*, uma criança (<1 ano) apresentou LLA refratária e as outras três não apresentaram nenhum evento até a data do último censo. Assim, podemos considerar que talvez, o pequeno número de casos *NUTM1*-rearranged observado, e apenas um evento (“*outlier*”) ocorrido, tenha diminuído a estimativa de sobrevida desse grupo. Os pacientes com *ZNF384*-rearranged possuem características clínicas que dependem das propriedades funcionais do parceiro de fusão, por exemplo, *TCF3*-*ZNF384* apresenta uma resposta clínica desfavorável quando comparada com a fusão *EP300*-*ZNF384*⁷⁹. Na nossa coorte foram observados quatro casos envolvendo *TCF3* e três envolvendo *EP300*, além de um caso envolvendo *SPI1* e *NCOA3*; sendo os dois últimos, novos parceiros de fusão descritas pela primeira vez em nosso estudo.

7 CAPÍTULO III: MANUSCRITO 2

BRIEF REPORT

Loss of 17p/*TP53* and *CASP8AP2* at diagnosis is associated with poor outcome in pediatric B-other Acute Lymphoblastic Leukemia

Natacha Azussa Migita^{1,2}, Patricia Yoshiokka Jotta¹, Natália Nascimento Paiva¹, Amilcar Cardoso de Azevedo¹, Silvia Regina Brandalise¹, and José Andrés Yunes^{1,3}

¹*Centro Infantil Boldrini, Campinas, SP, Brazil.*

²*Graduate Program in Genetics and Molecular Biology, IB, University State of Campinas, Campinas, SP, Brazil.*

³*Genetics Department, Faculty of Medical Sciences, University of Campinas, Campinas, SP, Brazil.*

Correspondence: Dr. José Andrés Yunes, Laboratório de Biologia Molecular, Centro de Pesquisa Boldrini, Rua Marcia Mendes 619, Campinas–SP 13083-884, Brazil; Phone: (+55) 19 37879096; e-mail: andres@boldrini.org.br

Acknowledgments: The authors thank all participants of the GBTLI-1999 and 2009 trials. NAM and NNP received scholarship from Brazilian funding agencies (CAPES). JAY received a Productivity fellowship from the Brazilian National Counsel of Technological and Scientific Development (CNPq, process 305896/2013-0 and 301596/2017-4). This work was supported by grants to JAY from Brazilian Health's Ministry Program, PRONON (Project DRM SIPAR 25000.057709/2015-01), CNPq (process 429811/2016-0) and Instituto Ronald McDonald (project IRM 2016109).

Conflict-of-interest disclosure: The authors declare no competing financial interests.

Keywords: B-other Acute Lymphoblastic Leukemia; digital MLPA; CNAs; Prognostic markers.

Key Points:

- 17p/*TP53* and *CASP8AP2* loss at diagnosis were associated with an inferior outcome in B-other ALL.
- The UKALL-CNA classifier is valid for B-other ALLs.
- Microdeletions at 21q22 resulting in *RUNX1* and *ETS2* loss at diagnosis were associated with inferior outcome in B-other ALL.

Abstract

Current treatment protocols for B-cell precursor acute lymphoblastic leukemia (BCP-ALL) are based on risk adapted therapy, that ensures appropriate therapy for patients at high risk of relapse while avoiding overtreatment of low risk ones. Genetic abnormalities provide important diagnostic and prognostic information in BCP-ALL and are increasingly used to assign patients to risk groups, hence playing an important role in the improvement of survival rates. Unfortunately, despite this progress, there is still a percentage of patients, commonly grouped as “B-other”, who do not have any of the recurrent genetic alterations and, therefore, cannot be stratified according to risk. There is also great clinical heterogeneity within B-other, which is made up of different molecular subgroups (*DUX4*-cluster, *PAX5*-driven, *ZNF384*-, *MEF2D*-, *NUTM1*-rearranged and Ph-like cases). Thus, to identify important biomarkers predicting risk of relapse in children with B-other ALL, outcome (refractory disease or relapse) of 136 children retrospectively treated with GBTLI protocols (1999 and 2009 trials) was evaluated with regard to copy-number alterations (CNAs) analysis of key leukemia-related genes by NGS-based digital MLPA. We could identify important CNA markers in B-other cases that can be used either individually or in combination to help estimate a patient’s risk of relapse and guide therapy. The presence of deletions in apoptotic genes (*TP53* and *CASP8AP2*) were the strongest prognostic markers of poor outcome at diagnosis in children with B-other ALL.

Introduction

Pediatric B-cell precursor acute lymphoblastic leukemia (BCP-ALL) is characterized by recurrent cytogenetic and molecular aberrations, including chromosomal rearrangements, aneuploidy, gene deletions and amplifications. Certain chromosomal abnormalities define subgroups of ALL with distinctive clinical features. For example, the presence of High Hyperdiploidy or *t*(12;21)(*ETV6-RUNX1*) are associated with good prognosis whereas *KMT2A*-rearrangements, *t*(17;19)(*TCF3-HLF*), hypodiploidy and *t*(9;22)(*BCR-ABL1*) are biomarkers of poorer outcome.¹⁻⁵ Such abnormalities are used in risk stratification of patients. Recently, two large studies, the International Berlin-Frankfurt-Münster (AIEOP-BFM) and the United Kingdom ALL (UKALL) Groups have evaluated the prognostic value of copy-number alterations (CNAs) in genes commonly deleted in BCP-ALL. The AIEOP-BFM study group has recently shown that *IKZF1*_{plus}⁶, defined by a combination of *IKZF1* deletions and other CNA alteration (deletions of *CDKN2A/B*, *PAX5*, or *PAR1* in the absence of *ERG* deletion), confers a significantly higher risk of relapse when associated with positive levels of minimal residual disease (MRD). The UKALL study group defined three CNA risk groups (good risk, GR; intermediate risk, IR; and poor risk, PR) which can be defined based on *EBF1*, *IKZF1*, *PAX5*, *CDKN2A/B*, *ETV6*, *BTG1*, *RB1*, and *PAR1* copy-number status^{7,8}. Even though those new classifiers are based on prevalent CNAs in ALL, it is conceivable that additional CNAs could provide relevant information. In this study, we aimed to evaluate the frequency and clinical relevance of CNA markers, as identified by digital Multiplex Ligation-dependent Probe Amplification (dMLPA) analysis, in 136 pediatric ALL of the “B-other” group. We report the negative prognostic impact of *TP53* loss (del17p or *TP53* deletion) and *CASP8AP2* deletion in a relatively small but significant subset of newly diagnosed B-other ALL patients.

Study design

Diagnostics, risk group assignment, and treatment were performed according to the GBTLI protocol. Between 1999 and 2009, 482 patients with BCP-ALL (aged ≤ 18 years) were registered at the Boldrini Children Center, of which 144 (~30%) were classified as B-other, as had no classical cytogenetic abnormalities. Of this, 136 cases had DNA for subsequent analysis. This study was undertaken with approval from our institutional ethics committee (CAAE 51991015.0.0000.5376), and informed consent was obtained from all patients. As described elsewhere (Migita NA et al, manuscript submitted), targeted RNA sequencing allowed us to classify these 136 ALLs into the novel molecular subgroups known as *DUX4*-cluster, *PAX5*-driven, *ABL*-class fusion, *JAK2*-fusion, *IGH-EPOR*, *CRLF2*-high, *ZNF384r*, *iAMP21*, *MEF2Dr* and *NUTM1r*; those who did not belong to any of these subtypes and formed the largest proportion of cases, were classified as B-other 'rest'. To increase statistical power, *ABL*-class fusion, *JAK2*-fusion, and *IGH-EPOR* cases were merged into a single group called *AJ2E*-class fusions (n=11). *CRLF2*-high and *AJ2E*-class fusion patients were analyzed as separate groups and as a single group since both activate the JAK/STAT pathway (Supplemental data).

Sub-microscopic copy number data and ploidy status were generated by using a commercial dMLPA assay (MRC Holland)⁹ which covers up to 1,000 target sequences and simultaneously detect 56 key target genes and regions implicated in leukemia in a single reaction. It is a multiplexed PCR assay that uses probes specific for exons within genes of interest, enabling the relative copy number of each covered exon to be reported as a ratio against other probe(s) within reference regions and then normalized to healthy individual controls. The dMLPA libraries were sequenced on Illumina MiSeq and the *.fastq* files were analyzed with *Coffalyser.Net 8.0 Software* (MRC Holland). Details are provided in supplemental Methods.

Pairwise comparisons of variables for exploratory purposes were performed using Monte Carlo unbiased estimation of Fisher's exact test. The contribution of CNAs for refractory disease and relapse was analyzed in the *MetaboAnalyst 5.0* platform (<https://www.metaboanalyst.ca/>) using the function "Biomarker analysis" sub-function Receiver Operating Characteristic (ROC). Only CNAs with more than two hits were considered. Leukemia-free survival (LFS) was calculated from the date of complete remission to the date of last follow-up or the first event (relapse or disease-related death). Refractory disease was considered an event at time zero. Kaplan-Meier survival curves were compared by the log-rank test. Analysis were performed using *Winstat* (Microsoft), *SPSS Statistics* (IBM) and *Prism v7.0* (GraphPad).

Results and Discussion

The clinical characteristics at diagnosis for the 136 patients (74 boys and 62 girls) are shown in Table S1. The median age was 7.3 years (range: 0–17 years), and their median leukocyte count was 20.9 (range: 1.4–459.0 $\times 10^9/L$). In total, 92% (n=126) of the cases harbored one or more CNA (either deletion or amplification). *ERG* deletions almost exclusively occurred in the *DUX4*-cluster (p<0.001), even though 15/35 of *DUX4*-cluster cases did not show an *ERG* deletion (Supplemental data). Negative associations of deletions in *MLLT3*, *ETV6* and *RUNX1* were observed in the *DUX4*-cluster (mutually

exclusive associations with p-values of 0.023, 0.013, and 0.033, respectively). *IKZF1* deletion were significantly associated with the broader group of JAK/STAT alteration cases, i.e. *AJ2E*-class fusions and *CRLF2*-high patients ($p=0.003$). *RPS14*, *VPREB1*, *EBF1* and *PAX5* deletions were also enriched in JAK/STAT alteration cases, with the last two deleted genes strongly associated with *CRLF2*-high ($p=0.001$ and $p=0.012$, respectively). We found a positive correlation of *CD200/BTLA* deletions with JAK/STAT alteration and iAMP21 cases ($p<0.001$ and $p=0.008$, respectively). Whereas it was negatively correlated with B-other 'rest' ($p=0.02$). Recurrent somatic deletions of *CD200/BTLA* genes have previously been reported in 2.8% of B-other (and 4.8% of BCP-ALL) at the 3q13.2 locus;¹⁰ here, ~9% of B-other cases showed heterozygous deletion of *CD200/BTLA*. In the novel subtype of BCP-ALL proposed by WHO¹¹, iAMP21, we identified frequent deletions in *ETV6* and *NF1*, and confirmed the expected copy number changes in a "step ladder" pattern, with a common sub-telomeric deletion along chromosome 21. *PAX5*-driven was strongly associated with abnormalities at chromosome 9p, as described elsewhere¹², with 3 cases harboring gene deletions in *EZH2* and *EPHA1*. *ZNF384r* showed frequently 17q11.2 microdeletions, which include the *NF1* and *SUZ12* genes. In *MEF2Dr* cases, deletions in *TBL1XR1*, *MLLT3*, *CDKN2A*, *RUNX1* were observed, which could potentially introduce heterogeneity for risk stratification in this recently discovery subgroup. No CNAs (deletions and gains) were significantly associated with *NUTM1r* cases, however the number of cases analyzed were small (Figure 1A).

Regarding treatment outcome, it is known that *IKZF1* deletions are overrepresented in high-risk patients, for example in t(9;22)*BCR-ABL1* positive cases (~70%)¹³, but here *IKZF1* deletions were not related to dismal diagnosis (5-y LFS, $78\pm 8\%$ versus $84\pm 4\%$; $p=0.1$). *IKZF1* deletion were significantly associated with high WBC count at diagnosis ($p=0.04$). The highest frequency of *CD200/BTLA* deletions was seen in the known poor-prognosis genetic groups, iAMP21 and JAK/STAT altered-cases. Moreover, its deletions were positively correlated with *GATA3* risk allele (rs3824662 and rs3781093; $p=0.001$). However, no significant difference was observed in the 5-y LFS survival rate of patients with *CD200/BTLA* loss ($75\pm 22\%$ versus $81\pm 4\%$; $p=0.7$). Noteworthy, important clinical features (WBC count, NCI risk and Infiltration in CNS) were associated with deletions in genes shown to be associated with prognosis as *PAX5*, *BTG1*, *RB1*, *VPREB1*, *PTEN* and *MTAP*. Of note, from six cases with infiltration in CNS at diagnosis, four co-occurred with *MTAP* deletions and one was not analyzed by dMLPA. The *MTAP* gene is located at 9p21, in proximity to tumor suppressor genes *CDKN2A* and *CDKN2B*, and are recurrently reported to be co-deleted with *CDKN2A* in T-ALL.^{14,15} Perhaps, this association involve the whole chromosome 9p region, including an extended deletion from *MTAP* to *CDKN2B* and might need future studies to prove this relation.

We evaluated the prognostic value of previously described CNA-based risk stratifications by AIEOP-BFM and UKALL study groups. The *IKZF1*_{Plus} group comprised 14% (19/136) of our patients and showed a 5-y LFS of $73\pm 12\%$ compared with $84\pm 9\%$ ($p=0.653$) in patients with *IKZF1* deletion who did not fulfill the *IKZF1*_{Plus} definition and $84\pm 4\%$ ($p=0.10$) in patients without *IKZF1* deletion. It is described that *IKZF1*_{Plus} has a strong prognostic impact only in patients still carrying measurable minimal residual disease (MRD) levels. Unfortunately, we did not have MRD levels for all patients. However, no association between *IKZF1*_{Plus} and LFS could be found when Low and High risk patients

were analyzed in separate (Supplemental data). By using the UKALL-CNA classifier, based on *EBF1*, *IKZF1*, *PAX5*, *CDKN2A/B*, *ETV6*, *BTG1*, *RB1*, and *PAR1* copy-number status, we were able to identify 50 good (5-y LFS, 90±5%), 55 intermediate (5-y LFS, 77±5%) and 31 poor (5-y LFS, 79±9%) risk patients (Log rank test for trend $p=0.04$). Of note, *ZNF384r* was associated with UKALL-CNA GR ($p=0.013$), *CRLF2*-high and *PAX5*-driven with UKALL-CNA IR ($p=0.027$ and $p=0.019$, respectively), and *AJ2E*-class fusion with UKALL-CNA PR ($p=0.003$). Thus, the UKALL-CNA classifier could be also considered for risk stratification among B-other ALLs.

To identify potential CNA biomarkers and evaluate their association with resistant disease (refractory disease or relapse), we performed multivariate ROC curve analysis. Genes deleted in only two cases, genes associated with *iAMP21* (a known prognostic factor) and genes in sex chromosomes were filtered out. By this analysis, we observed that 17p/*TP53* and *CASP8AP2* loss at diagnosis were the most significant biomarkers related to refractory disease or relapse. Survival curve analysis confirmed a significantly lower 5-year leukemia-free survival for loss of *TP53* (25±22% versus 84±4%, $p<0.0001$) and loss of *CASP8AP2* (33±27% versus 83±4%, $p=0.003$). *TP53* encodes the tumor-suppressor protein p53, which has numerous cellular activities including regulation of the cell cycle and apoptosis, and promotes DNA repair in response to DNA damage.¹⁶ Deletions of 17p were found in three out of the four patients with *TP53* loss. Since the single patient (#BCPALL776) with a focal loss of *TP53* relapsed, we suspect that it is the *TP53* loss what matters, and not the entire 17p arm. Importantly, no prognostic significance was seen for *TP53* SNV mutations ($p=0.357$), present in 4 patients. Of note, only one of the 4 cases with 17p/*TP53* loss had concomitant *TP53* SNVs, but without event (#BCPALL485). Unfortunately, we did not have the relapse samples of these patients in order to investigate whether the remaining *TP53* allele was lost or mutated during disease progression to relapse. Mutations or loss of the *TP53* gene are described in 5-13% of relapsed ALL, and are independently predictive for nonresponse to treatment and dismal prognosis in a significant number of patients.¹⁷⁻¹⁹ Yet, mutations as well as deletions of the *TP53* gene are rare (<5%) at first presentation of the disease, thus its significance at diagnosis remains unclear.²⁰⁻²³ Here we found that *TP53* losses occur in 2.9% of B-other ALL at diagnosis, being the strongest predictive marker for relapse. In addition, loss of another pro-apoptotic gene, *CASP8AP2*, showed prognostic significance. This is in agreement with previous reports on lower *CASP8AP2* expression as independent prognostic marker of relapse.²⁴ Downregulation of *CASP8AP2* by epigenetic modifications were also described in pediatric leukemia and resulted in poor treatment response.^{25,26} Interestingly, Swaminathan S et al. (2013) reported an independent prognostic value for deletions in *BACH2* gene located at 6q15, in close proximity of *CASP8AP2*.²⁷ *Bach2* mediates the negative selection of pre-B cells that fail to productively rearrange their V_H-DJ_H gene segments, through p53 activation. In patients with pre-B ALL, this checkpoint control is compromised by *BACH2* loss of function mutations (32% deletion found in BCP-ALL). Decreased expression of *BACH2* were observed in matched samples pairs from the time of diagnosis and the subsequent relapse. Notably, loss of *BACH2* were strongly associated with poor outcome among childhood ALL. Of note, even though *BACH2* deletions were accompanied by *CASP8AP2* deletions, the prognostic relevance of *CASP8AP2* was not investigated. In our dMLPA panel there was no probes for *BACH2*,

thus at present we don't know whether *CASP8AP2* deletions are connected to the dismal outcome or just represents a flag of *BACH2* deletions. .

Curiously, microdeletions at 21q22.12 (*RUNX1*) and 21q22.2 (*ETS2*) were associated with poor probability of 5y-LFS ($61\pm15\%$ versus $84\pm4\%$; $p=0.05$ and $54\pm20\%$ versus $84\pm4\%$; $p=0.03$, respectively). A similar microdeletions event was reported in chronic myeloid leukemia, in which cryptic inversions occurs in this region leading to break and fusion of *RUNX1-ETS2*^{28,29}. We could not detect any *RUNX1-ETS2* fusion by targeted RNAseq. *RUNX1* is a crucial hematopoietic transcription factor recurrently mutated (chromosomal rearrangements, somatic and germline point mutations) in leukemia. Most *RUNX1* mutations lead to loss of function or dominant repression of normal *RUNX1*, disrupting its transcriptional regulation.^{30,31} Interestingly, we found that patients who harbor *RUNX1* deletions with normal *ERG* had worse prognosis when compared with patients harboring *ERG* deletion ($p=0.054$), thus, reinforcing the good outcome ligated to loss of *ERG*³²(Supplemental data).

In conclusion, loss of 17p/*TP53* or *CASP8AP2* occur in about 5% of B-other ALL and are associated with lower LFS. The UKALL-CNA classifier is suitable for the stratification of B-other ALL into risk groups. Remarkably, deletion of *RUNX1* in 8.8% of the B-other ALL cases was associated with dismal patient outcome and possible cryptic inversion needs to be investigated.

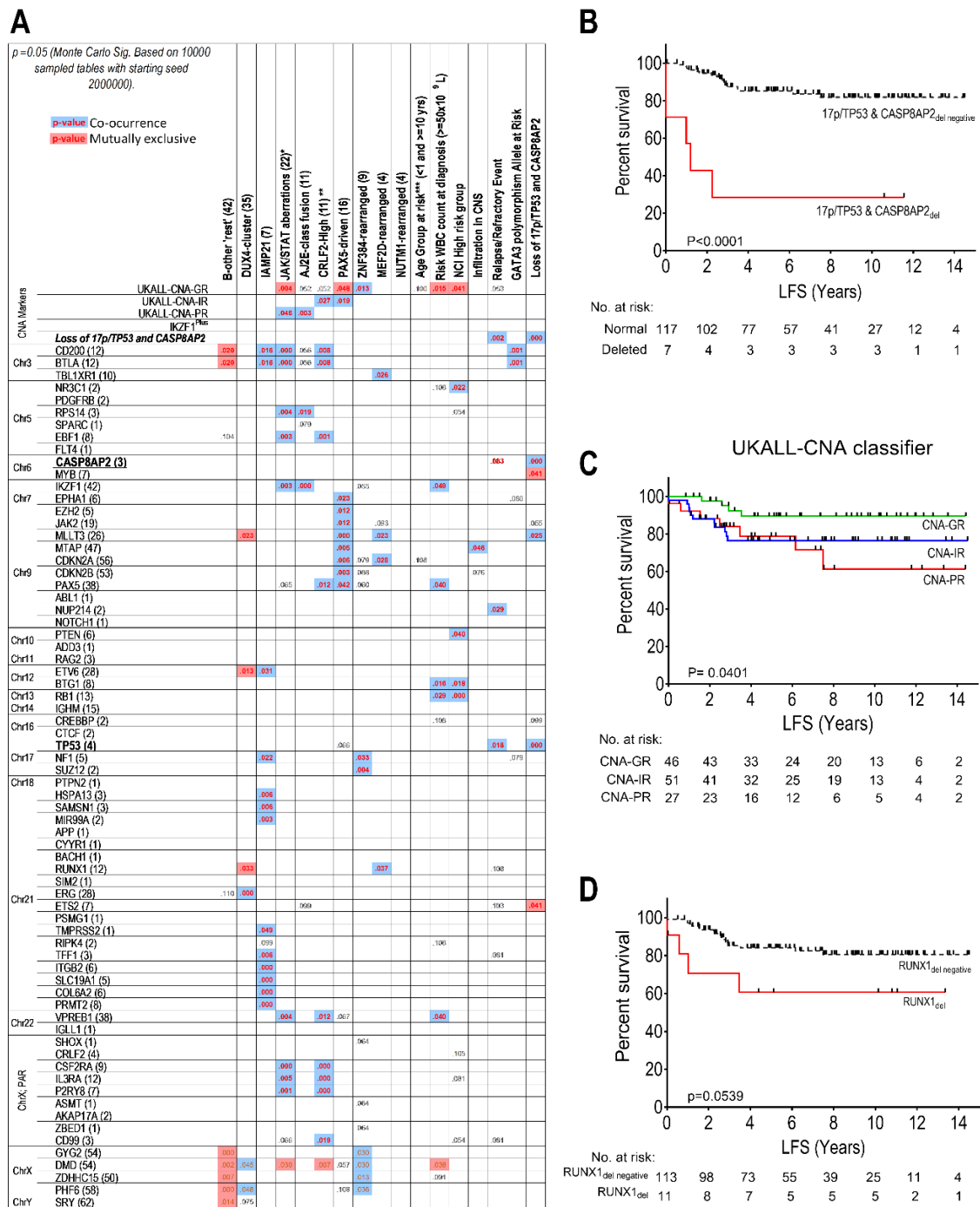


Figure 1. Prognostic value of copy number alterations (CNAs) in B-other ALL. **A.** Association and co-occurrence of copy number aberrations (CNAs) detected by digital MLPA in the different subtypes of pediatric B-other ALL. The proportion of patients per subtype with specific CNAs is shown. Only CNAs with more than two hits were considered for discussion. Color code for correlation figure; red: mutually exclusive; blue: co-occurrence; statistically significant values of $p \leq 0.05$ were considered and figure presents p -value ≤ 0.1 to show tendency for some CNAs. *JAK/STAT alterations include AJ2E-class fusion and CRLF2-high samples. **CRLF2-high include two samples of PAX5-driven and one sample of IAMP21 subgroups. ***One patient with an outlier age (=21 years old) was registered and treated with GBLT1-1999. **B.** Probability of leukemia-free survival (LFS) for patients with 17p/TP53 and CASP8AP2 deletions versus wild type; $p < 0.0001$ by log-rank test. **C.** Probability of leukemia-free survival (LFS) for patients classified as good risk (GR, $n=50$), intermediate risk (IR, $n=55$) and poor risk (PR, $n=31$) profiles based on the UKALL-CNA group; $p=0.04$ by log-rank test for trend. **D.** Probability of leukemia-free survival (LFS) for patients with RUNX1 deletions located at 21q22 versus wild type; $p=0.05$ by log-rank test.

Table 1. Outcome analysis for different characteristics in B-other ALL. Demographic, clinical, genetic and survival analysis censored in 5-year Leukemia-Free Survival (LFS) of patients diagnosed as B-other ALL and treated according to GBLTI protocols. Table shows a descriptive analysis of the clinically relevant and outcome percentage for all 136 B-other patients. §Bold values indicate statistical significance, log-rank test $p \leq 0.05$. *Log rank test for trend. **Important CNAs but with only two deletions thus were not considered for discussion. Abbreviations: LFS, leukemia free survival; WBC, white blood cell; NCI, National Cancer Institute.

	B-other ALL n (%)	Leukemia-free survival (LFS) 5-year $\pm$ STDE		Statistical (§p value)
Total of B-other	136	0.82 $\pm$ 0.04		
Age at diagnosis (years old)				0.791
≥ 1 and < 10 years	82 (60.3%)	0.81 $\pm$ 0.05		
≥ 10 years	54 (39.7%)	0.84 $\pm$ 0.05		
Gender				0.767
Male	74 (54.4%)	0.81 $\pm$ 0.05		
Female	62 (45.6%)	0.83 $\pm$ 0.05		
WBC count at diagnosis ($\times 10^9$/L)				0.004
$< 50 \times 10^9$ /L	91 (66.9%)	0.88 $\pm$ 0.03		
$\geq 50 \times 10^9$ /L	45 (33.1%)	0.69 $\pm$ 0.08		
NCI risk at diagnosis				0.270
Standard risk	116 (85.3%)	0.83 $\pm$ 0.04		
High risk	20 (14.7%)	0.72 $\pm$ 0.11		
Copy Number Alterations				
UKALL-CNA classifier (Logrank test for trend)*				0.040
UKALL-CNA-GR	50 (36.8%)	0.90 $\pm$ 0.05	vs 0.78 $\pm$ 0.05	0.040
UKALL-CNA-IR	55 (40.4%)	0.77 $\pm$ 0.06	vs 0.86 $\pm$ 0.04	0.342
UKALL-CNA-PR	31 (22.8%)	0.79 $\pm$ 0.09	vs 0.83 $\pm$ 0.04	0.183
AIEOP-BFM-IKZF1plus	19 (14%)	0.73 $\pm$ 0.12	vs 0.83 $\pm$ 0.04	0.164
IKZF1 del & No plus	23 (17%)	0.84 $\pm$ 0.09	vs 0.73 $\pm$ 0.12	0.653
IKZF1 del & ERG del	12 (8.8%)	0.80 $\pm$ 0.12	vs 0.77 $\pm$ 0.09	0.780
Loss 17p/TP53 and CASP8AP2				
17p/TP53 del	7 (5.1%)	0.29 $\pm$ 0.17	vs 0.85 $\pm$ 0.04	0.000
TP53 SNV	4 (2.9%)	0.25 $\pm$ 0.22	vs 0.84 $\pm$ 0.04	0.000
CASP8AP2 del	4 (2.9%)	1.00 $\pm$ 0.00	vs 0.81 $\pm$ 0.04	0.357
CDKN2A/2B del	3 (2.2%)	0.33 $\pm$ 0.27	vs 0.83 $\pm$ 0.04	0.003
IKZF1 del	56 (41.2%)	0.75 $\pm$ 0.06	vs 0.87 $\pm$ 0.04	0.258
IKZF1 SNV	42 (30.9%)	0.78 $\pm$ 0.08	vs 0.84 $\pm$ 0.04	0.100
VPREB1 del	8 (5.9%)	0.75 $\pm$ 0.22	vs 0.82 $\pm$ 0.04	0.456
PAX5 del	38 (27.9%)	0.77 $\pm$ 0.08	vs 0.84 $\pm$ 0.04	0.297
ETV6 del	33 (24.3%)	0.78 $\pm$ 0.08	vs 0.83 $\pm$ 0.04	0.433
ERG del	28 (20.6%)	0.69 $\pm$ 0.11	vs 0.85 $\pm$ 0.04	0.183
MLL3 del	28 (20.6%)	0.83 $\pm$ 0.08	vs 0.82 $\pm$ 0.04	0.977
IGHM del	26 (19.1%)	0.72 $\pm$ 0.09	vs 0.85 $\pm$ 0.04	0.231
RB1 del	15 (11.0%)	0.83 $\pm$ 0.11	vs 0.82 $\pm$ 0.04	0.883
PAR1 del	13 (9.6%)	0.71 $\pm$ 0.15	vs 0.83 $\pm$ 0.04	0.316
CD200/BTLA del	13 (9.6%)	0.83 $\pm$ 0.11	vs 0.82 $\pm$ 0.04	0.856
RUNX1 del	12 (8.8%)	0.75 $\pm$ 0.22	vs 0.81 $\pm$ 0.04	0.727
TBL1XR1 del	12 (8.8%)	0.61 $\pm$ 0.15	vs 0.84 $\pm$ 0.04	0.054
BTG1 del	10 (7.4%)	0.76 $\pm$ 0.15	vs 0.82 $\pm$ 0.04	0.812
EBF1 del	8 (5.9%)	1.00 $\pm$ 0.00	vs 0.81 $\pm$ 0.04	0.265
MYB del	8 (5.9%)	0.70 $\pm$ 0.18	vs 0.83 $\pm$ 0.04	0.498
ETS2 del	7 (5.1%)	0.69 $\pm$ 0.19	vs 0.83 $\pm$ 0.04	0.293
PTEN del	7 (5.1%)	0.54 $\pm$ 0.20	vs 0.84 $\pm$ 0.04	0.030
NF1 del	6 (4.4%)	0.75 $\pm$ 0.22	vs 0.82 $\pm$ 0.04	0.719
CREBBP del**	5 (3.7%)	0.33 $\pm$ 0.27	vs 0.83 $\pm$ 0.04	0.078
NR3C1 del**	2 (1.5%)	0.50 $\pm$ 0.35	vs 0.82 $\pm$ 0.04	0.134
	2 (1.5%)	0.50 $\pm$ 0.35	vs 0.83 $\pm$ 0.04	0.097

Authorship

Contribution: N.A.M and J.A.Y designed the study, analyzed the data and wrote the manuscript; N.A.M, P.Y.J and N.P.N performed the digital MLPA experiments; S.R.B and A.C.A performed the collection of the clinical data; and all authors critically reviewed the manuscript.

References

- ¹ Pui CH, Campana D, Evans WE. Childhood acute lymphoblastic leukaemia: current status and future perspectives. *Lancet Oncol.* 2001;2: 597-607.
- ² Moorman AV. The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia. *Blood Rev* 2012; **26**: 123–135. Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 2007;446:758-764
- ³ Moorman AV. The clinical relevance of chromosomal and genomic abnormalities in B-cell precursor acute lymphoblastic leukaemia. *Blood Rev* 2012; **26**: 123–135.
- ⁴ Teachey DT, Hunger SP. Predicting relapse risk in childhood acute lymphoblastic leukaemia. *Br J Haematol* 2013; **162**: 606–620.
- ⁵ Short, NJ, Jabbour, E. Minimal residual disease in acute lymphoblastic leukemia: how to recognize and treat it. *Curr Oncol Rep.* 2017;19(1):6.
- ⁶ Stanulla, M. *et al.* IKZF1plus Defines a New Minimal Residual Disease-Dependent Very-Poor Prognostic Profile in Pediatric B-Cell Precursor Acute Lymphoblastic Leukemia. *J. Clin. Oncol.* **36**, 1240–1249 (2018).
- ⁷ Hamadeh L, Enshaei A, Schwab C, Alonso CN, Attarbaschi A, Barbany G, et al. Validation of the United Kingdom copy-number alteration classifier in 3239 children with B-cell precursor ALL. *Blood Adv.* 2019;3(2):148–57
- ⁸ Moorman, A. V. *et al.* A novel integrated cytogenetic and genomic classification refines risk stratification in pediatric acute lymphoblastic leukemia (ALL). *Blood* **124**, 1434–44 (2014).
- ⁹ Benard-Slagter, A.; Zondervan, I.; de Groot, K.; Ghazavi, F.; Sarhadi, V.; Van Vlierberghe, P.; de Moerloose, B.; Schwab, C.; Vetteranta, K.; Harrison, C.J.; et al. Digital multiplex ligation-dependent probe amplification for detection of key copy number alterations in T- and B-cell lymphoblastic leukemia. *J. Mol. Diagn.* 2017, **19**, 659–672.
- ¹⁰ Ghazavi F, Clappier E, Lammens T, Suciu S, Caye A, Zegrari S, et al. CD200/BTLA deletions in pediatric precursor B-cell acute lymphoblastic leukemia treated according to the EORTC-CLG 58951 protocol. *Haematologica.* 2015;**100**:1311–9.
- ¹¹ Wenzinger C, Williams E, Gru AA. Updates in the Pathology of Precursor Lymphoid Neoplasms in the Revised Fourth Edition of the WHO Classification of Tumors of Hematopoietic and Lymphoid Tissues. *Curr Hematol Malig Rep.* 2018 Aug;13(4):275–288. doi: 10.1007/s11899-018-0456-8.
- ¹² Coysaud E, Struski S, Prade N, Familiades J, Eichner R, Quelen C, Bousquet M, Mugneret F, Talmant P, Pages MP, Lefebvre C, Penther D, Lippert E, Nadal N, Taviaux S, Poppe B, Luquet I, Baranger L, Eclache V, Radford I, Barin C, Mozziconacci MJ, Lafage-Pochitaloff M, Antoine-Poirel H, Charrin C, Perot C, Terre C, Brousset P, Dastugue N, Broccardo C. Wide diversity of PAX5 alterations in B-ALL: a Groupe Francophone de Cytogenetique Hematologique study. *Blood.* 2010 Apr 15;115(15):3089–97. doi: 10.1182/blood-2009-07-234229. Epub 2010 Feb 16. PMID: 20160164.
- ¹³ Martin Stanulla, Hélène Cavé, Anthony V. Moorman; for the International BFM Study Group. **IKZF1** deletions in pediatric acute lymphoblastic leukemia: still a poor prognostic marker?. *Blood* 2020; **135** (4): 252–260.
- ¹⁴ Batova A, Diccianni MB, Nobori T, et al. Frequent deletion in the methylthioadenosine phosphorylase gene in T-cell acute lymphoblastic leukemia: strategies for enzyme-targeted therapy. *Blood.* 1996;88(8):3083-3090.
- ¹⁵ M'soka TJ, Nishioka J, Taga A, et al. Detection of methylthioadenosine phosphorylase (MTAP) and p16 gene deletion in T cell acute lymphoblastic leukemia by real-time quantitative PCR assay. *Leukemia.* 2000;14(5):935-940. doi:10.1038/sj.leu.2401771
- ¹⁶ Bieging KT, Mello SS, Attardi LD. Unravelling mechanisms of p53-mediated tumour suppression. *Nat Rev Cancer.* 2014;14(5):359–370.
- ¹⁷ Sugimoto K, Toyoshima H, Sakai R, Miyagawa K, Hagiwara K, Hirai H et al. Mutations of the p53 gene in lymphoid leukemia. *Blood* 1991; **77**: 1153–1156.
- ¹⁸ Hof J, Krentz S, van Schewick C, Körner G, Shalapour S, Rhein P, Karawajew L, Ludwig WD, Seeger K, Henze G, von Stackelberg A, Hagemeier C, Eckert C, Kirschner-Schwabe R. Mutations and deletions of the TP53 gene predict nonresponse to treatment and poor outcome in first relapse of childhood acute lymphoblastic leukemia. *J Clin Oncol.* 2011 Aug 10;29(23):3185-93. doi: 10.1200/JCO.2011.34.8144. Epub 2011 Jul 11. PMID: 21747090.
- ¹⁹ Waanders E, Gu Z, Dobson SM, et al. Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia. *Blood Cancer Discov.* 2020;1(1):96-111. doi:10.1158/0008-5472.BCD-19-0041
- ²⁰ Hsiao MH, Yu AL, Yeargin J, Ku D, Haas M. Nonhereditary p53 mutations in T-cell acute lymphoblastic leukemia are associated with the relapse phase. *Blood* 1994; **83**: 2922–2930.

-
- ²¹ Fenaux P, Jonveaux P, Quiquandon I, Preudhomme C, Lai JL, Vanrumbeke M et al. Mutations of the p53 gene in B-cell lymphoblastic acute leukemia: a report on 60 cases. *Leukemia* 1992; 6: 42–46.
- ²² Wada M, Bartram CR, Nakamura H, et al: Analysis of p53 mutations in a large series of lymphoid hematologic malignancies of childhood. *Blood* 82:3163-3169, 1993
- ²³ . Moorman AV, Ensor HM, Richards SM, et al: Prognostic effect of chromosomal abnormalities in childhood B-cell precursor acute lymphoblastic leukaemia: Results from the UK Medical Research Council ALL97/99 randomised trial. *Lancet Oncol* 11:429-438, 2010
- ²⁴ Flotho C, Coustan-Smith E, Pei D, et al. Genes contributing to minimal residual disease in childhood acute lymphoblastic leukemia: prognostic significance of CASP8AP2. *Blood* 2006;108:1050–7
- ²⁵ Lee KD, Pai MY, Hsu CC, et al. Targeted Casp8AP2 methylation increases drug resistance in mesenchymal stem cells and cancer cells. *Biochem Biophys Res Commun.* 2012;422(4):578-585. doi:10.1016/j.bbrc.2012.05.029
- ²⁶ Li ZG, Jiao Y, Li WJ, et al. Hypermethylation of two CpG sites upstream of CASP8AP2 promoter influences gene expression and treatment outcome in childhood acute lymphoblastic leukemia. *Leuk Res.* 2013;37(10):1287-1293. doi:10.1016/j.leukres.2013.07.018
- ²⁷ Swaminathan S, Huang C, Geng H, et al. BACH2 mediates negative selection and p53-dependent tumor suppression at the pre-B cell receptor checkpoint. *Nat Med.* 2013;19(8):1014-1022. doi:10.1038/nm.3247
- ²⁸ Gelsi-Boyer V, Trouplin V, Adélaide J, et al. Genome profiling of chronic myelomonocytic leukemia: frequent aberrations of RAS and RUNX1 genes. *BMC Cancer.* 2008;8:299. Published 2008 Oct 16. Doi:10.1186/1471-2407-8-299
- ²⁹ Ochi Y, Yoshida K, Huang YJ, et al. Clonal evolution and clinical implications of genetic abnormalities in blastic transformation of chronic myeloid leukaemia. *Nat Commun.* 2021;12(1):2833. Published 2021 May 14. doi:10.1038/s41467-021-23097-w
- ³⁰ Sood R, Kamikubo Y, Liu P. Role of RUNX1 in hematological malignancies [published correction appears in *Blood*. 2018 Jan 18;131(3):373]. *Blood.* 2017;129(15):2070-2082. doi:10.1182/blood-2016-10-687830
- ³¹ Yang G, Khalaf W, van de Locht L, et al. Transcriptional repression of the Neurofibromatosis-1 tumor suppressor by the t(8;21) fusion protein. *Mol Cell Biol.* 2005;25(14):5869-5879. doi:10.1128/MCB.25.14.5869-5879.2005
- ³² Clappier, E., Auclerc, MF., Rapon, J. et al. An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. *Leukemia* 28, 70–77 (2014). <https://doi.org/10.1038/leu.2013.277>

Supplemental DATA for:

Loss of 17p/TP53 and CASP8AP2 at diagnosis is associated with poor outcome in pediatric B-other Acute Lymphoblastic Leukemia

Natacha Azussa Migita^{1,2}, Patricia Yoshiokka Jotta¹, Natália Nascimento Paiva¹, Amilcar Cardoso de Azevedo¹, Silvia Regina Brandalise¹, and José Andrés Yunes^{1,3}

¹*Centro Infantil Boldrini, Campinas, SP, Brazil.*

²*Graduate Program in Genetics and Molecular Biology, IB, University State of Campinas, Campinas, SP, Brazil.*

³*Genetics Department, Faculty of Medical Sciences, University of Campinas, Campinas, SP, Brazil.*

Contact information for correspondence: Dr. José Andrés Yunes, Laboratório de Biologia Molecular, Centro de Pesquisa Boldrini, Rua Marcia Mendes 619, Campinas–SP 13083-884, Brazil; Phone: (+55) 19 37879096; e-mail: andres@boldrini.org.br

Supplemental Methods

Patients and study design

This study included pediatric ALL patients (≤ 18 years) enrolled in the Brazilian GBTLI (“Brazilian Childhood Cooperative Group”) ALL-1999 and 2009 protocols¹ at the Boldrini Children Center. Leukemia diagnosis was established based on morphology, immunophenotype, cytogenetics, or RT-PCR as part of clinical routine diagnostics. Classical ALL subtypes, such as those with high hyperdiploidy (≥ 51 chromosomes or DNA index >1.16), hypodiploidy, *TCF3-PBX1* (*E2A-PBX1*), *ETV6-RUNX1* (*TEL-AML1*), *BCR-ABL1* and *KMT2A* (*MLL*)-rearrangements were excluded from this study. A total of 144 “B-other” ALLs (*i.e.*, normal cytogenetics or non-recurrent classical chromosomal abnormalities in ALL) were selected, of which only 136 had DNA samples available for the study, being 82 from the GBTLI-99 and 54 from the GBTLI-2009 protocol. Fisher’s exact test was used to compare the group of patients from the two different protocols. The results from this analysis showed that there were no significant differences in terms of clinically relevant variables for the B-other cases that were included in this study (Table S1). The study was approved by the institutional ethical committee (CAAE 51991015.0.0000.5376). Written informed consent was obtained from patients or guardians, and pseudonyms were created for patient data and specimens. Clinical characteristics of the patients are summarized in Table S1.

DNA samples

Mononuclear cells were purified from the diagnostic bone marrow or peripheral blood samples by density gradient centrifugation and were cryopreserved. Cells for

karyotyping and FISH analysis were stored in Carnoy's fixative. Genomic DNA was extracted from diagnostic leukemia samples using the *Illustra Mini Spin kit* (GE Healthcare Life Sciences, Chicago, IL, USA). Nucleic acids was quantified using *Qubit Fluorometer* (Thermo Fisher Scientific, Waltham, MA, USA), and quality was assessed by agarose gel electrophoresis.

Digital Multiplex Ligation-dependent Probe Amplification (dMLPA) for B-other ALL

Copy Number Alteration (CNA) and ploidy status were performed on genomic DNA of 136 out of 144 B-other patients. We use a test version of the D007 probemix (X4, lot0613 and lot0717 and X5 lot0220) for research use only (RUO), which contains up to 600 probes and has recently been developed by MRC-Holland (Amsterdam, the Netherlands)². The novel digitalMLPA can detect simultaneous ALL-associated copy number alterations (CNAs) via Illumina NGS platforms. Experiments were performed according to the manufacturer's instructions. Briefly, 80ng (in a total volume of 4uL) of each sample was mixed with 2uL of a unique barcode solution, followed by denaturation at 98°C for 10 minutes. After denaturation, a mixture of 1.25 µL digitalMLPA probe mix and 1.25 µL digitalMLPA buffer was added to each sample, and reactions were incubated for 16-20 hours at 60°C to ensure hybridization of the probes to the target DNA. The next day, probes were ligated by incubating the reactions with 32 µL of a ligase mastermix containing the ligase-65 enzyme (MRC Holland) and buffers (MRC Holland) at 48°C for 30 minutes, followed by heat inactivation of the Ligase-65 enzyme at 98°C for 5 minutes and an additional incubation at 65°C for 20 minutes to reduce background. PCR amplification of the ligated probes was performed on *Veriti® thermocycler* (Applied Biosystems). After amplification, sample-specific products from different reactions were pooled to generate the final amplicon library, diluted into 50pM and loaded onto an *Illumina MiSeq®* instrument using *MiSeq® Reagent Kit v3* for 116 bp single-read sequencing. Two pooled healthy males and two pooled healthy females, with four samples each pool, were included in the analysis as controls.

After quality assessment of the exported FASTQ files, assignment of the reads to digitalMLPA probes and subsequent data analysis were performed with *Coffalyser.Net 8.0 Software* (MRC Holland). Our runs generated at least 600 single reads for each probe. First, the read number for each probe was normalized by the median read number generated from reference probes hybridizing to usually conservative regions in the same genome (intrasample normalization). The relative read number generated for each probe was then compared with the matching values obtained in all reference samples (intersample normalization). The final probe ratio value, called dosage quotient (DQ), was around 1.0 if the region of interest was unaffected by CNA (i.e., normal copy number =2, for autosomal gene regions), whereas an increased or decreased value indicated the presence and level of gain or loss, respectively. In this study, a DQ intensity ratio between 0.75 and 1.3 was considered to represent a normal copy number, a ratio between 0.25 and 0.75 a monoallelic deletion, and a ratio < 0.25 a biallelic deletion; a ratio between 1.3 and 2.4 one copy gain; and a ratio > 2.0 as two or more copies gains. Although, we should not consider a sample with 100% tumor cell percentage consisting of only one major clone, we do consider the addition of "estimated" values as minor clones in the analysis.

Molecular subgroups of B-other

In our previous study, pediatric B-other ALL samples were analyzed for fusion transcripts, single nucleotide variations (SNV) and gene expression by Targeted RNA sequencing using a commercial kit interrogating 1,385 genes (*TruSight PanCancer*, *Illumina*). All subgroups within B-other were classified based on integrative and comprehensive analysis of results obtained from a Targeted RNA-sequencing methodology. Nine subgroups were considered; *DUX4*-cluster (n=35), *iAMP21* (n=7), *ZNF384*-rearranged (n=9), *MEF2D*-rearranged (n=4), *NUTM1*-rearranged (n=4), *PAX5*-driven (n=16); *ABL*-class fusion (n=7), *JAK2*-class fusion (n=3), *EPOR*-class fusion (n=1), *CRLF2*-high 'only' (n=8) and, a proportion of cases that were not classified into any of these subtypes were named B-other 'rest' (n=42). Three out of 11 patients had *CRLF2*-high expression (mainly *P2RY8-CRLF2* and *IGH-CRLF2*) co-occurring with one *iAMP21* and two *PAX5*-driven mutations.

Statistical analysis

Associations between genetic mutations, demographic data, prognostic factors and response to treatment were analyzed by using two-sided Fisher's exact test with Monte Carlo simulations (based on 10000 sampled tables with starting seed 2000000), and $p \leq 0.05$ was considered statistically significant. Survival analyzes were performed using the Kaplan-Meier method and the difference between curves was calculated using the log-rank test. All statistical procedures were performed with the GraphPad Prism v7 (La Jolla, CA, USA). Follow-up was updated in April 2021. Leukemia-free survival (LFS) was calculated from the date of complete remission to the date of last follow-up or the first event (relapse or death related to disease). Refractory disease was considered an event at time zero. Values of $p \leq 0.05$ were considered statistically significant. Analyses were performed using IBM® SPSS® Statistics, WinSTAT Microsoft® and Prism software version 7.0 (GraphPad Software).

Table S1. Demography and clinical characteristics of pediatric B-other ALL patients registered at Boldrini's Center according to Protocols GBLTI-1999 and -2009.

	All B-other	GBLTI-1999	GBLTI-2009	Statistical ([§] p value)
Total of B-other	136	82	54	
Age at diagnosis (years old)				
Median	7.38	7.05*	7.95	
Range	0.28-17.85 (average 8.36)	0.28-17.85 (average 8.04)	1.41-17.66 (average 8.92)	
Gender				0.482
Male	74 (54.4%)	47 (57.3%)	27 (50%)	
Female	62 (45.6%)	35 (42.7%)	27 (50%)	
WBC count at diagnosis ($\times 10^9/L$)				
Median	20.9	29.45	14.47	
Range	1.4-459.0	1.4-459.0	1.7-240.0	
WBC count at risk				0.093
< 50 $\times 10^9/L$	91	50 (61%)	41 (75.9%)	
$\geq 50 \times 10^9/L$	45	32 (39%)	13 (24.1%)	
NCI risk at diagnosis				0.459
Standard risk	116 (86.1%)	68 (83%)	48 (88.9%)	
High risk	20 (13.9%)	14 (17%)	6 (11.1%)	
GBLTI** risk group				0.253
Standard risk	95 (69.9%)	54 (65.9%)	41 (76%)	
High risk	41 (30.1%)	28 (34.1%)	13 (24%)	

*One patient with an outlier age (=21 years old) was registered and treated with GBLTI-1999. § Fisher test was used to calculate cohort p-value between protocols. The results from this analysis shows that there are no significant differences in the majority of variables. **GBLTI risk group, re-classification at the end of induction after poor or unsatisfactory response to the treatment as follow; GBLTI-1999 protocol considered patients to be at SR if they met all the following criteria: age ≥ 1.00 and < 10 years; white blood cells (WBC) count at diagnosis $< 50 \times 10^9/L$; a rapid early response day 7 (WBC $< 50 \times 10^9/L$), day 14 (no peripheral blasts and $< 25\%$ blasts in bone marrow) and day 28 ($< 5\%$ blasts in BM) of induction therapy. The prognostic value of WBC count at day 7 was validated on the previous GBLTI ALL-93 protocol (Brandalise SR, 1996)³. Mediastinal mass, central nervous system (CNS) or other extramedullary leukemic involvement, as well as immunophenotype and cytogenetics, were not considered for risk classification. Patients with mature B-cell ALL were excluded. Bone marrow (BM) specimens were collected on days 14 and 28 of induction. For the GBLTI-2009, risk classification considered minimum residual disease (MRD) after induction with chemotherapy (day 36), involvement in central nervous system and testis.

Supplemental Results

Supplementary Figure S1. Schematic copy number aberrations (CNA) detected by digital MLPA. Important leukemia-deleted genes are listed and 136 B-other pediatric patients are clustered based on genetic subgroups ('Subtype'), indicated with colored bars.

Supplementary Figure S2. Kaplan-Meier survival curves of pediatric B-other ALL patients with different profiled CNAs deletions. A. LFS according to 17p/*TP53* deletion. B. LFS according to *TP53* SNV mutation. C. LFS according to *CASP8AP2* deletion. D. LFS according to *RUNX1* deletion, associated or not with *ERG* deletion. E-H. LFS according to *IKZF1*plus profile, defined by the AIEOP-BFM study group. The *IKZF1*plus profile is defined by a combination of *IKZF1* deletion and at least one additional deletion of *CDKN2A/B*, *PAX5*, or *PAR1* in the absence of *ERG* deletion.

Supplementary Figure S3. Identification of structurally complex rearrangements at chromosome 21q22 resulting in deletions of *RUNX1* and *ETS2*.

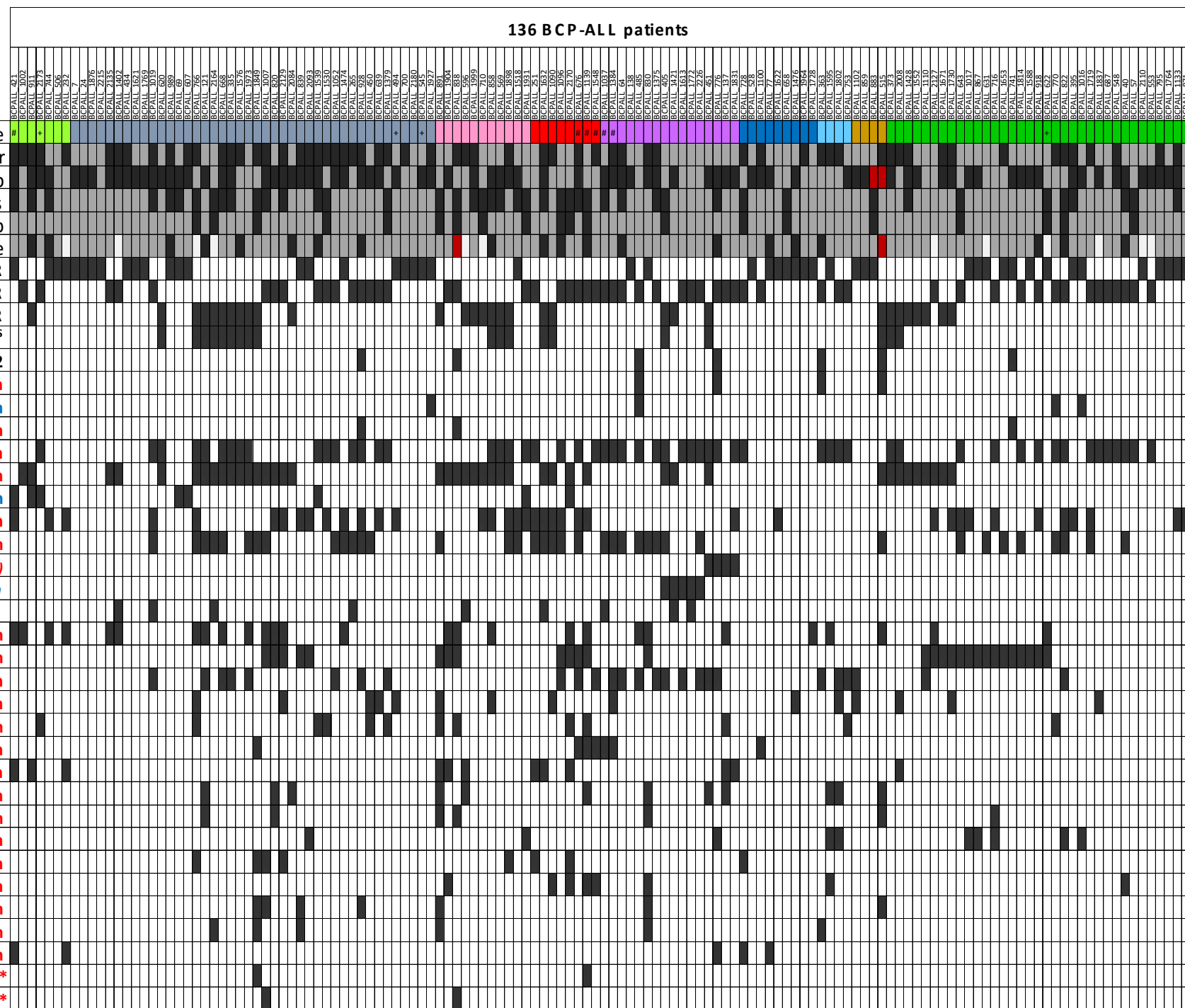
Supplemental References

¹ Sociedade Brasileira de Oncologia Pediátrica. Protocolo brasileiro de tratamento da leucemia linfóide aguda na infância GBTLI LLA-2009. São Paulo: Campinas; 2011.

² Benard-Slagter A, Zondervan I, de Groot K, Ghazavi F, Sarhadi V, Van Vlierberghe P, De Moerloose B, Schwab C, Vettenranta K, Harrison CJ, Knuutila S, Schouten J, Lammens T, Savola S. Digital Multiplex Ligation-Dependent Probe Amplification for Detection of Key Copy Number Alterations in T- and B-Cell Lymphoblastic Leukemia. *J Mol Diagn*. 2017 Sep;19(5):659-672. doi: 10.1016/j.jmoldx.2017.05.004. Epub 2017 Jul 19. PMID: 28736295.

³ Brandalise SR: Prognostic value of day 8 peripheral blood response for children with acute lymphocytic leukemia, in Zander AR, (ed): *Gene Technology: Stem Cell and Leukemia Research*. Berlin, Springer-Verlag, 1996, pp 421-428.

- **red text**: CNAs (dMLPA) result
- **blue text**: SNVs result



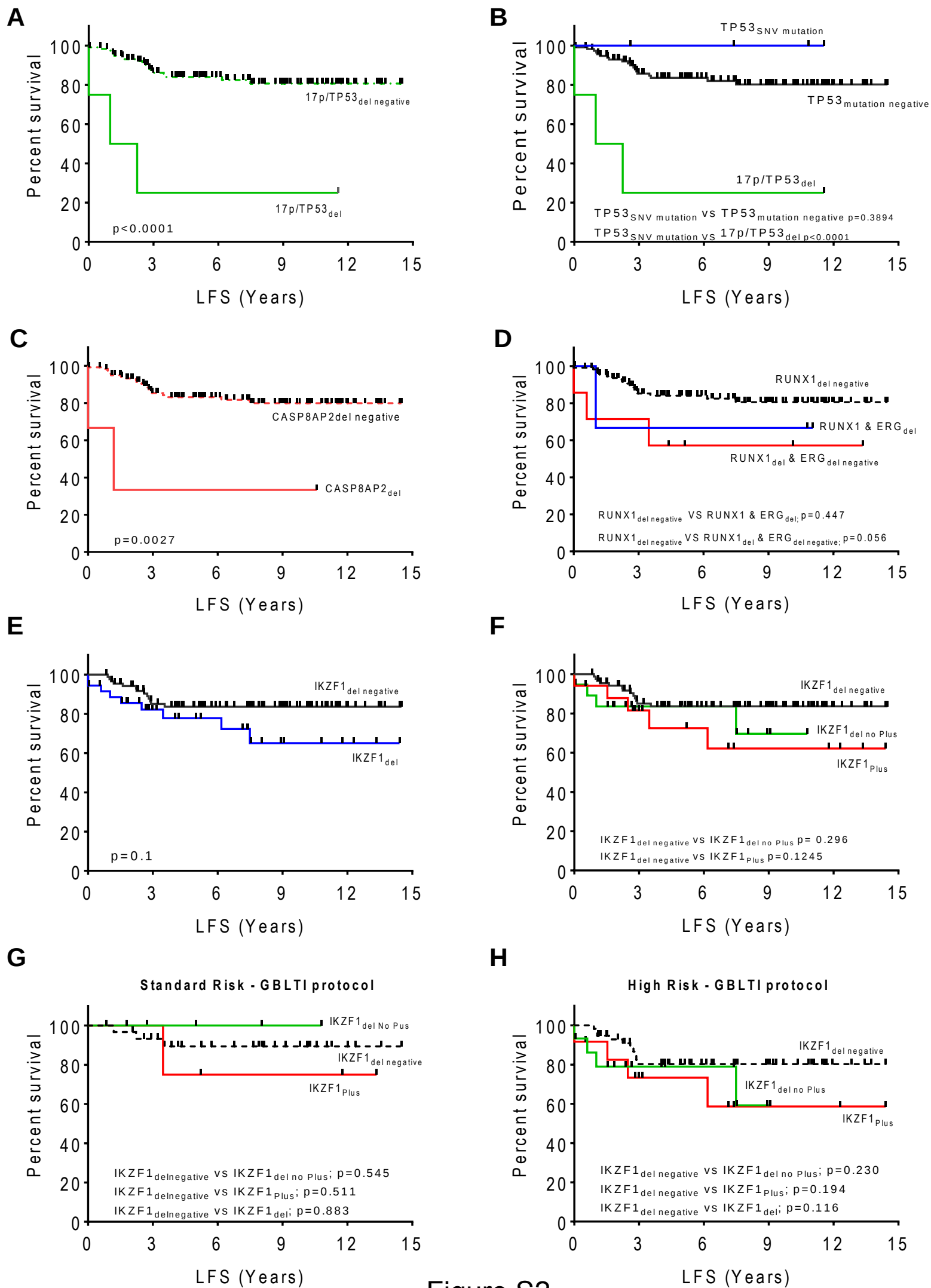
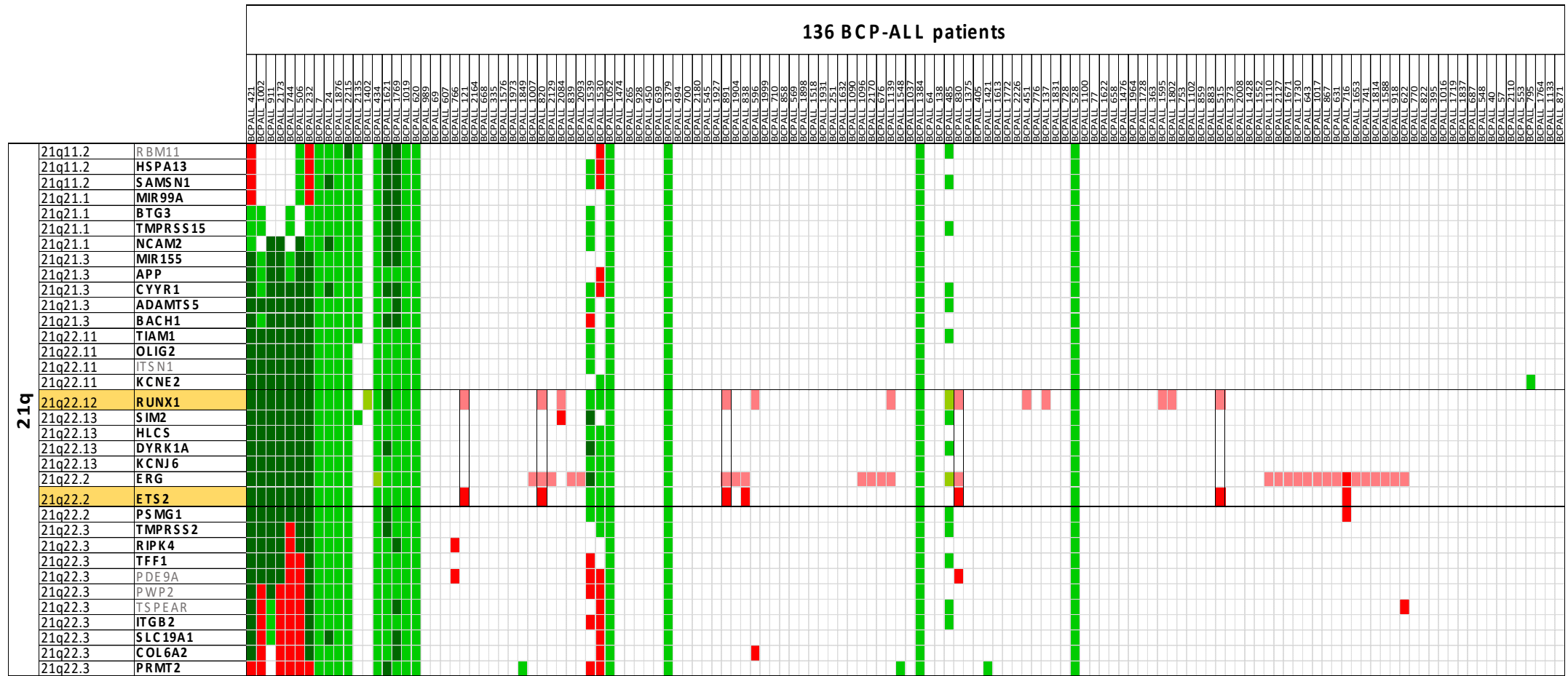


Figure S2.

21q

Focal duplication	Focal deletion
Heterozygous gain	Heterozygous loss
Homozygous gain	Homozygous loss



8 CONCLUSÃO

1. Aproximadamente 70% dos 144 casos de LLA B-other, o qual representa 30% (144/482) dos casos de BCP-ALL registrados no Centro Boldrini, foram analisados e puderam ser classificados em um dos novos subgrupos moleculares: *DUX4*-cluster (7,26%), *iAMP21* (1,45%), *ZNF384*-rearranged (1,86%), *MEF2D*-rearranged (0,82%), *NUTM1*-rearranged (0,82%), *PAX5*-driven (3,73%; incluindo fusões *PAX5*, ampliações intragênicas e mutações de *PAX5* p.P80R); Fusões envolvendo genes *ABL*-class (1,45%), *JAK2*-class (0,62%) e *EPOR*-class (0,2%); *CRLF2*-high 'only' (1,66%) e; 9,96% dos casos que não foram classificados em nenhum desses subtipos foram denominados B-other "rest".
2. A análise de mutações SNV e *small indel* revelou uma alta frequência de mutações em *NRAS* (23%), *MDC1* (22%), *KRAS* (14%), *FLT3* (8%), *PAX5* (8%), *STAT6* (8%), *SETD2* (7%), *MYC* (6%), *IKZF1* (6%) e *KMT2D* (7%).
3. Mutações na via RAS (particularmente, *KRAS* e *FLT3*) mostraram-se associadas com o subgrupo B-other "rest" de forma significativa.
4. Observou-se que mutações em genes das vias de RAS e JAK/STAT são mutuamente exclusivas.
5. Curiosamente, os pacientes classificados como B-other 'rest' estavam enriquecidas de meninas e o subgrupo *PAX5*-driven de meninos.
6. Pacientes com mutações SNV e *small indel* que afetam a via JAK/STAT apresentaram uma pior estimativa de sobrevida livre de leucemia (5-y LFS, 72±10% versus 85±4%; p=0,004).
7. A perda de dois genes ao diagnóstico, *TP53* (del17p ou deleção de TP53) e *CASP8AP2* relacionados à apoptose, estão associadas a um pior prognóstico nas LLAs B-other.
8. A classificação baseada na análise de alteração no número de cópias (CNAs) criada pelo grupo internacional *United Kingdom ALL* (UKALL) é válida para o prognóstico de pacientes do subgrupo LLA B-other.
9. Microdeleções na região 21q22, levando a perda dos genes *RUNX1* e *ETS2*, foram associadas com um desfecho clínico desfavorável nas LLAs B-other.

9 PERSPECTIVAS E CONSIDERAÇÕES FINAIS

Este é o primeiro estudo nacional com o objetivo de analisar em grande escala mutações somáticas em pacientes pediátricos de LLA do subtipo “B-other”. Deciframos o espectro de mutações genéticas que estão presentes nos 144 pacientes pediátricos, até então, classificados de forma conjunta como LLA B-other (Figura 1-Resumo esquemático). Conseguimos subdividi-los em importantes subgrupos moleculares, identificar mutações oncogênicas *driver*/recorrentes e associá-las com as características demográficas e clínicas dos pacientes. O grande número de mutações identificadas em nosso trabalho também inclui mutações que não foram descritas na literatura e, portanto, ainda não estão caracterizadas funcionalmente. Assim, mais estudos serão necessários para investigar seus papéis funcionais na leucemia.

Como perspectivas futuras, podemos listar importantes questões que nos chamam a atenção e precisam ser desenvolvidas e/ou melhor estudadas;

1. Desenvolver um painel menor e personalizado de RNA-seq para capturar apenas genes importantes para o diagnóstico e manejo clínico dos pacientes com leucemia; além de tornar esta técnica acessível a outros centros de diagnóstico (Manuscrito 1);
2. Analisar a relação de subgrupos específicos com o gênero. Checar se existe diferença na expressão de genes localizados nos cromossomos sexuais;
3. Estudar a relação oposta de ativação das vias JAK/STAT e RAS;
4. Investigar através de análises por FISH e/ou sequenciamento Sanger, a possível inversão e fusão entre os genes *RUNX1* e *ETS2*, como sendo o resultado das microdeleções observadas na região 21q22. Verificar se essa alteração ocorre em outros subtipos de LLA-B;
5. Estudar funcionalmente a associação observada entre a deleção de *MTAP* e a infiltração de leucemia no SNC. Para isso, podemos injetar células leucêmicas com a deleção do braço 9p e células leucêmicas contendo apenas a deleção de *CDKN2A* e *CDKN2B*; e verificar se há infiltração no SNC de forma diferenciada (Manuscrito 2);
6. Estudar o efeito da adição do inibidor Nutlin-3, uma molécula antagonista de Mdm2 (regulador negativo de p53), em células leucêmicas com a deleção de 17p/*TP53* em um dos alelos. Verificar se o inibidor resgataria a atividade da proteína p53 e induziria apoptose;

Mutações genéticas na Leucemia Linfóide Aguda pediátrica do subgrupo “B-other”

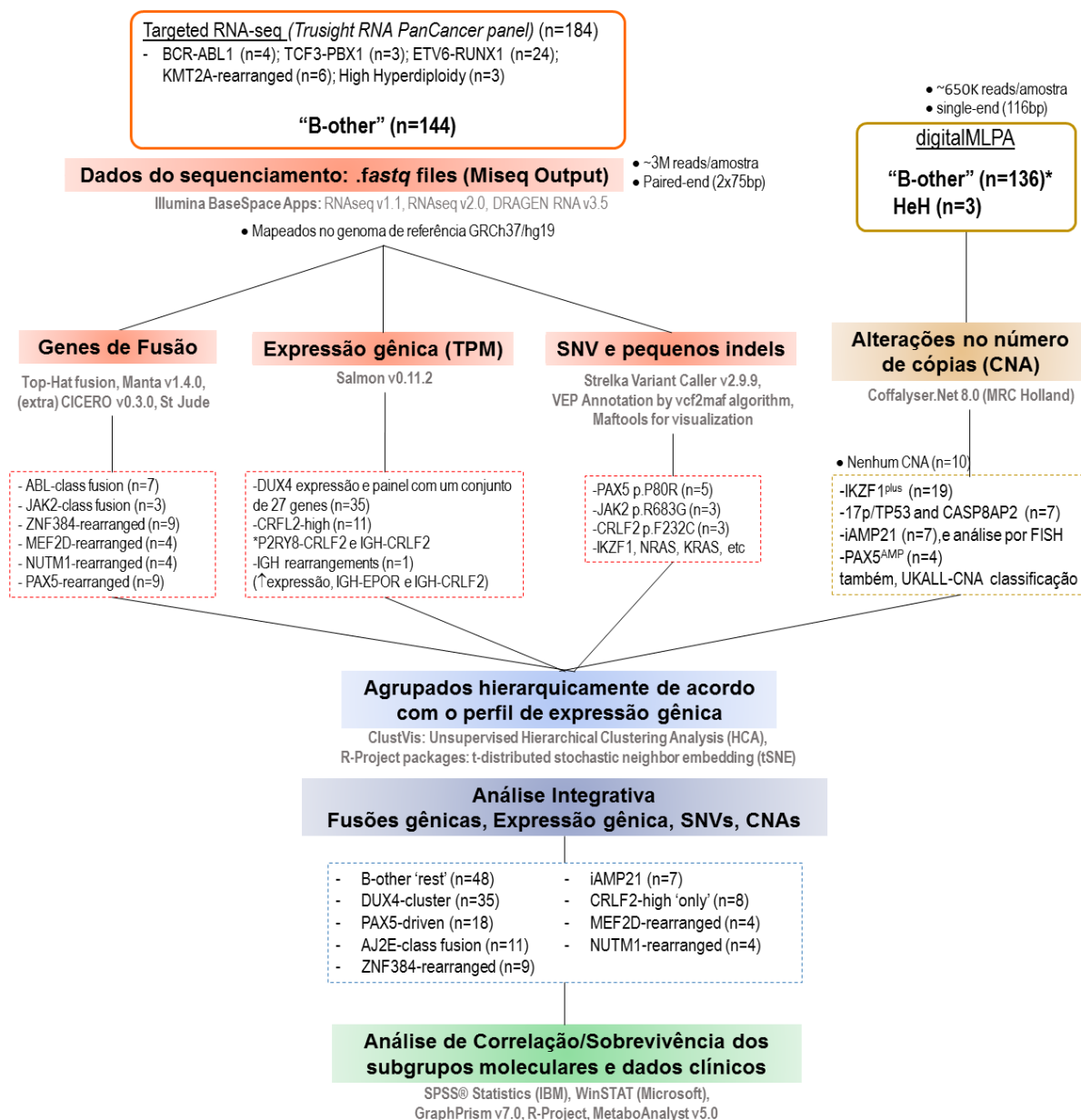


Figura 2. Fluxograma explicando as etapas de análise adotada para a caracterização genética dos pacientes pediátricos de LLA B-other. As caixas retangulares coloridas representam os principais dados gerados pelo sequenciamento (Targeted RNA-seq e dMLPA) e utilizados para as análises subsequentes. Palavras em cinza indicam os *softwares* de análise. Mais informações no corpo do texto. *Oito amostras não foram analisadas pelo dMLPA devido à indisponibilidade de material ou baixa qualidade da amostra de DNA.

10 METODOLOGIA GERAL

Técnicas rotineiras de Biologia Molecular.

As técnicas padrão de Biologia Molecular e Celular, tais como extração de ácidos nucleicos, PCR, RT-PCR, eletroforese, sequenciamento Sanger, etc. estão descritas nos seguintes manuais de laboratório: *Sambrook & Russel, (2001)*⁸⁰ e *Ausubel et al., (1997)*⁸¹. A técnica de imunofenotipagem e citometria de fluxo são descritas no manual *Michael Brown & Carl Wittwer (2000)*⁸². O desenho de *primers* para reações de PCR e/ou sequenciamento foram feitos com o software *PrimePremier*.

Preparação das amostras para o NGS RNA-seq direcionado

A extração e purificação do RNA das amostras (blastos de diagnóstico) foram realizadas em coluna de sílica com o kit *Illustra Mini Spin kit* (GE Healthcare), seguindo as instruções do fabricante. O RNA extraído foi quantificado no *Qubit*, com o kit *RNA BR Assay* (ThermoFisher). A qualidade e a pureza do RNA extraído foram analisadas por gel de agarose e checados no equipamento *Agilent Bioanalyzer 2100* (Agilent Technologies), com a avaliação positiva do RIN e DV₂₀₀>50% para todas as amostras.

Preparação das bibliotecas de cDNA para o NGS RNA-seq direcionado

As bibliotecas de cDNA foram preparadas com 50ng iniciais de RNA total com o kit *TruSight RNA Pan-Cancer Panel* (Illumina), seguindo o protocolo do kit. A biblioteca é formada por regiões selecionadas e enriquecidas em 1.385 genes (21.043 regiões exônicas) relacionados ao câncer. Resumidamente, o protocolo consiste na fragmentação enzimática das moléculas de RNA, síntese da primeira e da segunda fita de cDNA, adenilação do terminal 3', ligação de adaptadores com múltiplos índices à uma das extremidades do cDNA e enriquecimento dos fragmentos de cDNAs via PCR. Na sequência, os fragmentos são hibridizados com as sondas biotiniladas e capturadas com *beads* magnéticas. As bibliotecas enriquecidas foram quantificadas por real-time PCR (*KAPA Biosystems qPCR kit, Wilmington, MA*), diluídas (14-18pM), carregadas nos cartuchos (*MiSeq® Reagent Kit v3, 150 cycles*), e sequenciadas no equipamento *MiSeq* (Illumina). Realizamos a corrida *paired-end* de 8 amostras por *flowcell*, gerando ~3M de *reads* por amostra (2x76 ciclos).

Análises de bioinformática

Análises de bioinformática foram realizadas com várias ferramentas comerciais, *open-source* e/ou desenvolvidas *in loco* pela equipe de bioinformática do laboratório, para de forma integrada explorar os resultados de NGS obtidos. Estas análises podem ser divididas em quatro operações principais: (i) Detecção de fusões gênicas; (ii) análise de expressão gênica e agrupamentos; (iii) análise de SNV e *small indels*; (iv) anotação funcional e enriquecimento de vias. Basicamente, os dados brutos do sequenciamento passam pela trimagem dos adaptadores e exclusão de *reads* de baixa qualidade no próprio equipamento *MiSeq* (algoritmo '*bcl2fastq*'). Os arquivos *.fastq* foram utilizados como *input* para o mapeamento e detecção de variantes e fusões gênicas realizadas com as pipelines *RNA-Seq Alignment v1.1*, *v2.0* e *DRAGEN RNA v3.5*, disponíveis no *BaseSpace Sequence Hub* (Illumina), com os parâmetros padrão de análise. Com o *software STAR v2.5.0b* procedeu-se o mapeamento dos *reads* no genoma humano de referência (*release hg19, Fevereiro de 2009/GRCh37, Gencode gene annotations*), utilizando parâmetros para detecção de *reads* quiméricos. Utilizou-se o *software Salmon* para a contagem dos *reads*, normalização e estimar os valores de TPM dos genes. O *software Cufflinks2*, *Manta*, *STAR* e *CICERO* (mesmo algoritmo utilizado pelo St. Jude Hospital)⁸³ foram utilizados para a detecção de fusões gênicas novas e/ou recorrentes. A chamada de variantes (SNV e *small indel*) foi realizada com o *software Strelka2*. Todas as análises foram inspecionadas e seguiram o guia de boas práticas para identificação e análise de variantes utilizando dados de sequenciamento em larga-escala (*GATK Best practices*),⁸⁴ elaborado pelo *Broad Institute*. Mais detalhes sobre cada etapa de análise de bioinformática são fornecidos nos Materiais e Métodos dos Manuscrito e/ou Capítulo relacionado e um fluxograma é apresentado na Figura 2 – 'Perspectivas e Considerações finais'.

Digital Multiplex Ligation-dependent Probe Amplification (dMLPA)

Para investigar alterações no número de cópias (CNAs) dos pacientes, o DNA genômico de 136 (dos 144 pacientes B-other) foram extraídas. Oito pacientes não tinham material disponível para análise. Utilizamos um kit teste "*probemix D007*" (X4, lot0613 e lot0717 e X5 lot0220), gentilmente cedidos pela empresa *MRC-Holland* (*Amsterdam, the Netherlands*). O kit contém em torno de 600 sondas e possibilitam detectar mais de 56 CNAs importantes para a leucemia através da tecnologia NGS da

Illumina. O protocolo foi realizado de acordo com as instruções do fabricante, e utiliza 80ng (em um volume total de 4uL) de cada amostra identificada por *barcodes*. Ao final, as bibliotecas são agrupadas e diluídas a 50pM e carregados no cartucho *MiSeq Reagent Kit v3 (Illumina, single-end, 116 bp)*. Para normalização das análises, amostras genotipicamente normais de oito homens e oito mulheres foram sequenciados na mesma corrida. Após checagem da qualidade do sequenciamento, os arquivos *.fastq* foram analisados no software *Coffalyser.Net 8.0* (MRC Holland), desenvolvido especificamente para análises de CNAs. Os valores de ‘*probe ratio*’ normalizados são chamados de quociente de dosagem (DQ). Quando o valor de DQ fica em torno de 1,0, a região de interesse possui o número de cópias normais, e um valor aumentado ou diminuído indica ganho ou perda, respectivamente. Neste estudo, valores de DQ entre 0,75 e 1,3 foram considerados cópias normais; valores entre 0,25 e 0,75 deleção monoalélica; valores <0,25 deleção bialélica; valores entre 1,3 e 2,4 ganho de uma cópia; e valores >2,0 ganho de duas ou mais cópias. Mais detalhes sobre cada etapa de análise de bioinformática são fornecidos nos Materiais e Métodos do Manuscrito e/ou Capítulo relacionado.

Identificação de polimorfismos de risco em GATA3

Os polimorfismos de risco em *GATA3*, rs3824662 e rs3781093 foram identificadas a partir do sequenciamento de fragmentos amplificados por reações de PCR e *Nested-PCR* do gene *GATA3*. O polimorfismo rs3781093 foi amplificado utilizando os *primers forward* 5'-TTCCTGTGCTCTGTTCTT-3' e *reverse* 5'-GGCTCAGGATAACAATG-3, e sequenciados a partir do produto obtido na reação de PCR. O polimorfismo rs3824662 foi amplificado utilizando os *primers forward* 5'-TATCACCTCCCCACCA-3' e *reverse* 5'-GGAAAGCCCCAGATCAA-3'; a partir do produto obtido na reação de PCR foi realizada uma segunda amplificação em reação de *Nested-PCR* utilizando os *primers forward* 5'-GTAAACGACGGCCAGTGATCATGTCAGGCTGGGAGGT-3' (sequência universal M13, indicada pela sequência sublinhada) e *reverse* 5'-TAATACGACTCACTATAGGGGGTTTCACTTGGGGGCC-3' (sequência universal T7, indicada pela sequência sublinhada). Os polimorfismos rs3781093 e rs3824662 foram sequenciados utilizando o kit *Big Dye Terminator v3.1* (Applied Biosystems) e 5 pmol dos *primers forward* e/ou *reverse*, e dos *primers* universais M13 e/ou T7,

respectivamente. As sequências obtidas foram comparadas à sequência de referência do gene.

Análises estatísticas

As associações entre as alterações genéticas e outras variáveis categóricas (dados clínicos e subgrupos) foram analisados com os *Softwares SPSS® Statistics (IBM®)*, utilizando o teste exato de Fisher e/ou o método de Monte Carlo, e valores de $p \leq 0,05$ foi considerado estatisticamente significativo. A sobrevida livre de leucemia (*Leukemia-Free Survival*, LFS) foi calculada a partir da data de remissão completa até a data do último acompanhamento ou o primeiro evento (recidiva ou morte decorrente da doença). A doença refratária foi considerada um evento no tempo zero. As curvas de sobrevida de *Kaplan-Meier* foram comparadas pelo teste de log-rank. O último “follow-up” foi atualizado em abril de 2021. As análises também foram realizadas pelo *Prism versão 7.0 (GraphPad Software)* e/ou *WinSTAT (Microsoft)*.

Para selecionar importantes marcadores prognósticos, os dados de CNAs foram analisados na plataforma *MetaboAnalyst 5.0* (<https://www.metaboanalyst.ca/>) usando a função “*Biomarker analysis*” sub-função ROC (*Receiver Operating Characteristic*). Trata-se de uma análise supervisionada, ou seja, dados de prognóstico foram associados as mutações observadas para cada paciente. Paralelamente, para a aferição da existência de correlação entre os dados, foi empregado o coeficiente de correlação de *Spearman* (para dados não paramétricos) no *WinSTAT*. Valores de $p \leq 0,05$ foram considerados significativos.

11 REFERÊNCIAS BIBLIOGRÁFICAS

- ¹ Stratton MR, Campbell PJ, Futreal PA. The cancer genome. *Nature*. 2009 Apr 9;458(7239):719-24. doi: 10.1038/nature07943.
- ² Stratton MR. Exploring the genomes of cancer cells: progress and promise. *Science*. 2011 Mar 25;331(6024):1553-8. doi: 10.1126/science.1204040.
- ³ World Health Organization (WHO). Global Health Estimates 2020: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2019. WHO; 2020. Accessed October 11, 2021. Website: who.int/data/gho/data/themes/mortality-and-global-health-estimates/ghes-leading-causes-of-death
- ⁴ Ferlay J., Ervik M.L.F., Colombet M., Mery L., Piñeros M. Global Cancer Observatory: Cancer Today. International Agency for Research on Cancer; Lyon, France: 2021
- ⁵ Hanahan, Douglas. "Hallmarks of Cancer: New Dimensions." *Cancer discovery* vol. 12,1 (2022): 31-46. doi:10.1158/2159-8290.CD-21-1059
- ⁶ Renan MJ. How many mutations are required for tumorigenesis? Implications from human cancer data. *Mol Carcinog*. 1993;7(3):139-46. doi: 10.1002/mc.2940070303.
- ⁷ Greenman C, Stephens P, Smith R, et al. Patterns of somatic mutation in human cancer genomes. *Nature*. 2007;446(7132):153-158. doi:10.1038/nature05610
- ⁸ Vogelstein B., Papadopoulos N., Velculescu V.E., Zhou S., Diaz L.A., Jr., Kinzler K.W. Cancer genome landscapes. *Science*. 2013;339:1546–1558. doi: 10.1126/science.1235122
- ⁹ Ma X., Liu Y., Liu Y., Alexandrov L.B., Edmonson M.N., Gawad C., Zhou X., Li Y., Rusch M.C., Easton J., et al. Pan-cancer genome and transcriptome analyses of 1,699 paediatric leukaemias and solid tumours. *Nature*. 2018;555:371–376. doi: 10.1038/nature25795.
- ¹⁰ Gröbner S.N., Worst B.C., Weischenfeldt J., Buchhalter I., Kleinheinz K., Rudneva V.A., Johann P.D., Balasubramanian G.P., Segura-Wang M., Brabetz S., et al. The landscape of genomic alterations across childhood cancers. *Nature*. 2018;555:321–327. doi: 10.1038/nature25480.
- ¹¹ Sweet-Cordero E.A., Biegel J.A. The genomic landscape of pediatric cancers: Implications for diagnosis and treatment. *Science*. 2019;363:1170–1175. doi: 10.1126/science.aaw3535.
- ¹² Pan-Cancer Analysis Of Whole Genomes The ICGC/TCGA Pan-Cancer Analysis of Whole Genomes Consortium. *Nature*. 2020;578:82–93. doi: 10.1038/s41586-020-1969-6.
- ¹³ INCA, 2019. Estimativa 2020: incidência de câncer no Brasil. Instituto Nacional de Câncer José Alencar Gomes da Silva. Coordenação de Prevenção e Vigilância, Rio de Janeiro, p. 122.
- ¹⁴ Cobaleda C, Sánchez-García I. B-cell acute lymphoblastic leukaemia: towards understanding its cellular origin. *Bioessays*. 2009;31(6):600-609. doi:10.1002/bies.200800234
- ¹⁵ Misaghian N, Ligresti G, Steelman LS, et al. Targeting the leukemic stem cell: the Holy Grail of leukemia therapy. *Leukemia*. 2009;23(1):25-42. doi:10.1038/leu.2008.246
- ¹⁶ Lassaletta, A. (2004) Leucemias. *Leucemia linfoblástica aguda*. *Pediatr Integral*; VIII(5):435-442.
- ¹⁷ Smith MT, McHale CM, Wiemels JL, et al. Molecular biomarkers for the study of childhood leukemia. *Toxicol Appl Pharmacol*. 2005;206(2):237-245. doi:10.1016/j.taap.2004.11.026
- ¹⁸ Gaux C. Biology of acute lymphoblastic leukemia (ALL): clinical and therapeutic relevance. *Transfus Apher Sci*. 2011;44(2):183-189. doi:10.1016/j.transci.2011.01.009
- ¹⁹ Sinnett D, Labuda D, Krajcinovic M. Challenges identifying genetic determinants of pediatric cancers--the childhood leukemia experience. *Fam Cancer*. 2006;5(1):35-47. doi:10.1007/s10689-005-2574-4
- ²⁰ Pui CH, Robison LL, Look AT. Acute lymphoblastic leukaemia. *Lancet*. 2008;371(9617):1030-1043. doi:10.1016/S0140-6736(08)60457-2
- ²¹ Inaba H, Greaves M, Mullighan CG. Acute lymphoblastic leukaemia. *Lancet*. 2013;381(9881):1943-1955. doi:10.1016/S0140-6736(12)62187-4
- ²² Goldstone AH, Richards SM, Lazarus HM, et al. In adults with standard-risk acute lymphoblastic leukemia, the greatest benefit is achieved from a matched sibling allogeneic transplantation in first complete remission, and an autologous transplantation is less effective than conventional consolidation/maintenance chemotherapy in all patients: final results of the International ALL Trial (MRC UKALL XII/ECOG E2993). *Blood*. 2008;111(4):1827-1833. doi:10.1182/blood-2007-10-116582
- ²³ Sociedade Brasileira de Oncologia Pediátrica. Protocolo brasileiro de tratamento da leucemia linfóide aguda na infância GBTLI LLA-2009. São Paulo: Campinas; 2011.
- ²⁴ Pui CH, Mullighan CG, Evans WE, Relling MV. Pediatric acute lymphoblastic leukemia: where are we going and how do we get there?. *Blood*. 2012;120(6):1165-1174. doi:10.1182/blood-2012-05-378943
- ²⁵ Hunger SP, Mullighan CG. Acute Lymphoblastic Leukemia in Children. *N Engl J Med*. 2015;373(16):1541-1552. doi:10.1056/NEJMra1400972
- ²⁶ Schultz KR, Pullen DJ, Sather HN, et al. Risk- and response-based classification of childhood B-precursor acute lymphoblastic leukemia: a combined analysis of prognostic markers from the Pediatric Oncology Group (POG) and Children's Cancer Group (CCG). *Blood*. 2007;109(3):926-935. doi:10.1182/blood-2006-01-024729

- 27 Bhojwani D, Howard SC, Pui CH. High-risk childhood acute lymphoblastic leukemia. *Clin Lymphoma Myeloma*. 2009;9 Suppl 3(Suppl 3):S222-S230. doi:10.3816/CLM.2009.s.016
- 28 Hunger SP. Tyrosine kinase inhibitor use in pediatric Philadelphia chromosome-positive acute lymphoblastic anemia. *Hematology Am Soc Hematol Educ Program*. 2011;2011:361-365. doi:10.1182/asheducation-2011.1.361
- 29 Möricke A, Zimmermann M, Reiter A, et al. Long-term results of five consecutive trials in childhood acute lymphoblastic leukemia performed by the ALL-BFM study group from 1981 to 2000. *Leukemia*. 2010;24(2):265-284. doi:10.1038/leu.2009.257
- 30 Salzer WL, Devidas M, Carroll WL, et al. Long-term results of the pediatric oncology group studies for childhood acute lymphoblastic leukemia 1984-2001: a report from the children's oncology group. *Leukemia*. 2010;24(2):355-370. doi:10.1038/leu.2009.261
- 31 Bailey LC, Lange BJ, Rheingold SR, Bunin NJ. Bone-marrow relapse in paediatric acute lymphoblastic leukaemia. *Lancet Oncol*. 2008;9(9):873-883. doi:10.1016/S1470-2045(08)70229-8
- 32 Pieters R, Kaspers GJ, van Wering ER, et al. Cellular drug resistance profiles that might explain the prognostic value of immunophenotype and age in childhood acute lymphoblastic leukemia. *Leukemia*. 1993;7(3):392-397.
- 33 Ramakers-van Woerden NL, Pieters R, Hoelzer D, et al. In vitro drug resistance profile of Philadelphia positive acute lymphoblastic leukemia is heterogeneous and related to age: a report of the Dutch and German Leukemia Study Groups. *Med Pediatr Oncol*. 2002;38(6):379-386. doi:10.1002/mpo.10087.
- 34 Aricò M, Valsecchi MG, Camitta B, et al. Outcome of treatment in children with Philadelphia chromosome-positive acute lymphoblastic leukemia. *N Engl J Med*. 2000;342(14):998-1006. doi:10.1056/NEJM200004063421402.
- 35 Ramakers-van Woerden NL, Pieters R, Loonen AH, Hubbek I, van Drunen E, Beberloo HB et al. TEL/AML1 gene fusion is related to in vitro drug sensitivity for L-asparaginase in childhood acute lymphoblastic leukemia. *Blood* 2000; 96: 1094-1099.
- 36 Whitehead VM, Payment C, Cooley L, Lauer SJ, Mahoney DH, Shuster JJ, et al. The association of the TEL-AML1 chromosomal translocation with the accumulation of methotrexate polyglutamates in lymphoblasts and with ploidy in childhood B-progenitor cells acute lymphoblastic leukemia: a Pediatric Oncology Group study. *Leukemia* 2001; 15: 1081-1088.
- 37 Seeger K, Adams HP, Buchwald D, et al. TEL-AML1 fusion transcript in relapsed childhood acute lymphoblastic leukemia. The Berlin-Frankfurt-Münster Study Group. *Blood*. 1998;91(5):1716-1722.
- 38 Gandemer V, Chevret S, Petit A, et al. Excellent prognosis of late relapses of ETV6/RUNX1-positive childhood acute lymphoblastic leukemia: lessons from the FRALLE 93 protocol. *Haematologica*. 2012;97(11):1743-1750. doi:10.3324/haematol.2011.059584
- 39 Anderson K, Lutz C, van Delft FW, et al. Genetic variegation of clonal architecture and propagating cells in leukaemia. *Nature*. 2011;469(7330):356-361. doi:10.1038/nature09650
- 40 Schwab CJ, Chilton L, Morrison H, et al. Genes commonly deleted in childhood B-cell precursor acute lymphoblastic leukemia: association with cytogenetics and clinical features. *Haematologica*. 2013;98(7):1081-1088. doi:10.3324/haematol.2013.085175
- 41 Zenatti, P.P., D. Ribeiro, W. Li, L. Zuurbier, M.C. Silva, M. Paganin, J. Tritapoe, J.A. Hixon, A.B. Silveira, B.A. Cardoso, et al. (2011). Oncogenic IL7R gain-of-function mutations in childhood T-cell acute lymphoblastic leukemia. *Nat. Genet.* 43:932–939. doi:10.1038/ng.924
- 42 Hamadeh L, Enshaie A, Schwab C, et al. Validation of the United Kingdom copy-number alteration classifier in 3239 children with B-cell precursor ALL. *Blood Adv*. 2019;3(2):148-157. doi:10.1182/bloodadvances.2018025718
- 43 Moorman AV. New and emerging prognostic and predictive genetic biomarkers in B-cell precursor acute lymphoblastic leukemia. *Haematologica*. 2016;101(4):407-416. doi:10.3324/haematol.2015.141101
- 44 Daniel A. Arber, Attilio Orazi, Robert Hasserjian, Jürgen Thiele, Michael J. Borowitz, Michelle M. Le Beau, Clara D. Bloomfield, Mario Cazzola, James W. Vardiman; The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood* 2016; 127 (20): 2391–2405. doi: <https://doi.org/10.1182/blood-2016-03-643544>
- 45 Mullighan CG, Su X, Zhang J, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med*. 2009;360(5):470-480. doi:10.1056/NEJMoa0808253
- 46 Den Boer ML, van Slegtenhorst M, De Menezes RX, et al. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncol*. 2009;10(2):125-134. doi:10.1016/S1470-2045(08)70339-5
- 47 Roberts KG, Morin RD, Zhang J, et al. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer Cell*. 2012;22(2):153-166. doi:10.1016/j.ccr.2012.06.005
- 48 Harvey RC, Mullighan CG, Wang X, et al. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood*. 2010;116(23):4874-4884. doi:10.1182/blood-2009-08-239681
- 49 Ensor HM, Schwab C, Russell LJ, et al. Demographic, clinical, and outcome features of children with acute lymphoblastic leukemia and CRLF2 deregulation: results from the MRC ALL97 clinical trial [published correction appears in *Blood*. 2011 May 5;117(18):5009]. *Blood*. 2011;117(7):2129-2136. doi:10.1182/blood-2010-07-297135
- 50 Roberts KG, Li Y, Payne-Turner D, et al. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N Engl J Med*. 2014;371(11):1005-1015. doi:10.1056/NEJMoa1403088

- ⁵¹ Ishibashi T, Yaguchi A, Terada K, et al. Ph-like ALL-related novel fusion kinase ATF7IP-PDGFRB exhibits high sensitivity to tyrosine kinase inhibitors in murine cells. *Exp Hematol*. 2016;44(3):177-88.e5. doi:10.1016/j.exphem.2015.11.009
- ⁵² Kawamura M, Taki T, Kaku H, Ohki K, Hayashi Y. Identification of SPAG9 as a novel JAK2 fusion partner gene in pediatric acute lymphoblastic leukemia with t(9;17)(p24;q21). *Genes Chromosomes Cancer*. 2015;54(7):401-408. doi:10.1002/gcc.22251
- ⁵³ Yeoh EJ, Ross ME, Shurtleff SA, et al. Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell*. 2002;1(2):133-143. doi:10.1016/s1535-6108(02)00032-6
- ⁵⁴ Mullighan CG, Goorha S, Radtke I, et al. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature*. 2007;446(7137):758-764. doi:10.1038/nature05690
- ⁵⁵ Strefford JC, Worley H, Barber K, et al. Genome complexity in acute lymphoblastic leukemia is revealed by array-based comparative genomic hybridization. *Oncogene*. 2007;26(29):4306-4318. doi: 10.1038/sj.onc.1210190
- ⁵⁶ Harvey RC, Mullighan CG, Wang X, et al. Identification of novel cluster groups in pediatric high-risk B-precursor acute lymphoblastic leukemia with gene expression profiling: correlation with genome-wide DNA copy number alterations, clinical characteristics, and outcome. *Blood*. 2010;116(23):4874-4884.
- ⁵⁷ Liu YF, Wang BY, Zhang WN, et al. Genomic Profiling of Adult and Pediatric B-cell Acute Lymphoblastic Leukemia. *EBioMedicine*. 2016;8:173-183. doi:10.1016/j.ebiom.2016.04.038
- ⁵⁸ Li JF, Dai YT, Lilljebjörn H, et al. Transcriptional landscape of B cell precursor acute lymphoblastic leukemia based on an international study of 1,223 cases. *Proc Natl Acad Sci U S A*. 2018;115(50):E11711-E11720. doi:10.1073/pnas.1814397115
- ⁵⁹ Kim S, Scheffler K, Halpern AL, et al. Strelka2: fast and accurate calling of germline and somatic variants. *Nat Methods*. 2018;15(8):591-594. doi:10.1038/s41592-018-0051-x
- ⁶⁰ McLaren W, Gil L, Hunt SE, Riat HS, Ritchie GR, Thormann A, Flicek P, Cunningham F. The Ensembl Variant Effect Predictor. *Genome Biology* Jun 6;17(1):122. (2016). doi: 10.1186/s13059-016-0974-4
- ⁶¹ Cyriac Kandoth. mskcc/vcf2maf: vcf2maf v1.6.19. (2020). doi:10.5281/zenodo.593251
- ⁶² Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28(11):1747-1756. doi:10.1101/gr.239244.118
- ⁶³ Tamborero D, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUST: exploiting the positional clustering of somatic mutations to identify cancer genes. *Bioinformatics*. 2013 Sep 15;29(18):2238-44. doi: 10.1093/bioinformatics/btt395.
- ⁶⁴ Forbes SA, et al. COSMIC (the Catalogue of Somatic Mutations in Cancer): a resource to investigate acquired mutations in human cancer, *Nucleic Acids Res.*, 2010, vol. 38 (pg. D652-D657)
- ⁶⁵ Futreal PA, et al. A census of human cancer genes, *Nat. Rev. Cancer*, 2004, vol. 4 (pg. 177-183)
- ⁶⁶ Waanders E, Gu Z, Dobson SM, Antić Ž, Crawford JC, Ma X, Edmonson MN, Payne-Turner D, van der Vorst M, Jongmans MCJ, McGuire I, Zhou X, Wang J, Shi L, Pounds S, Pei D, Cheng C, Song G, Fan Y, Shao Y, Rusch M, McCastlain K, Yu J, van Boxtel R, Blokzijl F, Iacobucci I, Roberts KG, Wen J, Wu G, Ma J, Easton J, Neale G, Olsen SR, Nichols KE, Pui CH, Zhang J, Evans WE, Relling MV, Yang JJ, Thomas PG, Dick JE, Kuiper RP, Mullighan CG. Mutational landscape and patterns of clonal evolution in relapsed pediatric acute lymphoblastic leukemia. *Blood Cancer Discov*. 2020 Jul;1(1):96-111. doi: 10.1158/0008-5472.BCD-19-0041.
- ⁶⁷ Jinghui Zhang, Charles G. Mullighan, Richard C. Harvey, Gang Wu, Xiang Chen, Michael Edmonson, Kenneth H. Buetow, William L. Carroll, I-Ming Chen, Meenakshi Devidas, Daniela S. Gerhard, Mignon L. Loh, Gregory H. Reaman, Mary V. Relling, Bruce M. Camitta, W. Paul Bowman, Malcolm A. Smith, Cheryl L. Willman, James R. Downing, Stephen P. Hunger; Key pathways are frequently mutated in high-risk childhood acute lymphoblastic leukemia: a report from the Children's Oncology Group. **Blood** 2011; 118 (11): 3080-3087. doi: <https://doi.org/10.1182/blood-2011-03-341412>
- ⁶⁸ Zhang J, McCastlain K, Yoshihara H, et al. Deregulation of DUX4 and ERG in acute lymphoblastic leukemia. *Nat Genet*. 2016;48(12):1481-1489. doi:10.1038/ng.3691
- ⁶⁹ Haynes P, Bomsztyk K, Miller DG. Sporadic DUX4 expression in FSHD myocytes is associated with incomplete repression by the PRC2 complex and gain of H3K9 acetylation on the contracted D4Z4 allele. *Epigenetics Chromatin*. 2018;11(1):47. Published 2018 Aug 20. doi:10.1186/s13072-018-0215-z
- ⁷⁰ Gu Z, Churchman ML, Roberts KG, et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. *Nat Genet*. 2019;51(2):296-307. doi:10.1038/s41588-018-0315-5
- ⁷¹ Beck C, Castaneda-Zegarra S, Huse C, Xing M, Oksenysh V. Mediator of DNA damage checkpoint protein 1 facilitates V(D)J recombination in cells lacking DNA repair factor XLF. *Biomolecules*. 2019;10(1):60. doi:10.3390/biom10010060
- ⁷² Lund, K., Adams, P. & Copland, M. EZH2 in normal and malignant hematopoiesis. *Leukemia* **28**, 44-49 (2014). <https://doi.org/10.1038/leu.2013.288>
- ⁷³ Papaemmanuil E, Rapado I, Li Y, Potter NE, Wedge DC, Tubio J, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nature genetics*. 2014;46(2):116-25.
- ⁷⁴ Parker H, An Q, Barber K, et al. The complex genomic profile of ETV6-RUNX1 positive acute lymphoblastic leukemia highlights a recurrent deletion of TBL1XR1. *Genes Chromosomes Cancer*. 2008;47(12):1118-1125.
- ⁷⁵ Ward AF, Braun BS, Shannon KM. Targeting oncogenic Ras signaling in hematologic malignancies. *Blood* 2012; **120**: 3397-3406.
- ⁷⁶ Chan, L.N.; Murakami, M.A.; Robinson, M.E.; Caeser, R.; Sadras, T.; Lee, J.; Cosgun, K.N.; Kume, K.; Khairnar, V.; Xiao, G.; et al. Signalling input from divergent pathways subverts b cell transformation. *Nature* 2020, **583**, 845-851.

-
- ⁷⁷Roberts KG, Yang YL, Payne-Turner D, Lin W, Files JK, Dickerson K, et al. Oncogenic role and therapeutic targeting of ABL-class and JAK-STAT activating kinase alterations in Ph-like ALL. *Blood Adv.* 2017;1(20):1657–71.
- ⁷⁸Inaba H, Mullighan CG. Pediatric acute lymphoblastic leukemia. *Haematologica.* 2020;105(11):2524-2539. Published 2020 Nov 1. doi:10.3324/haematol.2020.247031
- ⁷⁹Hirabayashi, S., Butler, E.R., Ohki, K. et al. Clinical characteristics and outcomes of B-ALL with ZNF384 rearrangements: a retrospective analysis by the Ponte di Legno Childhood ALL Working Group. *Leukemia* 35, 3272–3277 (2021). <https://doi.org/10.1038/s41375-021-01199-0>
- ⁸⁰Sambrook, J. & Russel, D.W. (2001). *Molecular Cloning: A laboratory Manual*. 3^a ed. NY: Cold Spring Harbor.
- ⁸¹Ausubel, F. M., et al (1997). *Current Protocols in Molecular Biology*. 4^a ed. Cold Spring Harbor Laboratory Press.
- ⁸²Brown M. and Wittwer C. (2000) Flow cytometry: principles and clinical applications in hematology *Clin. Chem.* **46** 1221-9
- ⁸³Tian, L., Li, Y., Edmonson, M.N. et al. CICERO: a versatile method for detecting complex and diverse driver fusions using cancer RNA sequencing data. *Genome Biol* 21, 126 (2020).
- ⁸⁴Van der Auwera, Geraldine A., et al. "From FastQ data to high-confidence variant calls: the genome analysis toolkit best practices pipeline." *Current protocols in bioinformatics* (2013): 11-10.

12 APÊNDICE

Além do estudo dos subtipos moleculares e biomarcadores para LLA B-*other* (incluídos nesse grupo os casos de BCR-ABL1-like), um objetivo secundário da tese foi o de investigar o mecanismo de resistência ao imatinibe na leucemia BCR-ABL1+. Nesta seção apresentamos resultados e perspectivas relacionados aos nossos esforços iniciais e trabalhos sendo realizados para estudar os mecanismos de resistência a quimioterápicos, principal causa de recaída da doença. O fato de termos casos de pacientes de LLA-B com diferentes alterações genéticas, dados genômicos e de resposta ao tratamento, catalogados nos bancos de dados do Centro Infantil Boldrini, possibilita delinear estudos mecanísticos e funcionais de resistência. Uma vez que cerca de metade dos 144 pacientes apresentavam alterações genéticas conhecidas por ativar tirosina quinases, investigamos o possível papel de proteínas de ligação ao RNA (RBPs) na resistência da leucemia ao inibidor da tirosina quinase (TKI) imatinibe, um medicamento comumente usado no tratamento de LLA e leucemia mieloide crônica (CML) com a fusão BCR-ABL1.

Os resultados aqui apresentados fazem parte do estudo intitulado “*Identification of Imatinib-Sensitizing RNA-binding proteins in leukemia cells with a genome-scale CRISPR/Cas9 screening*”; trabalho sendo realizado em colaboração com o laboratório do Professor Gene W Yeo (*Department of Cellular and Molecular Medicine and Institute for Genomic Medicine and UCSD Stem Cell Program, UCSD, San Diego, CA, USA*) e o laboratório da Dra Katlin Brauer Massirer, do Centro de Biologia Molecular e Engenharia Genética (CBMEG-UNICAMP, Campinas, SP). Neste trabalho, utilizamos a mais moderna técnica de CRISPR/Cas9 *screening* com uma biblioteca direcionada a mais de mil RBPs em experimentos de letalidade sintética para identificar quais RBPs deletados sensibilizariam células K562 (BCR-ABL1+) ao imatinibe e identificar, de forma ampla, possíveis biomarcadores alvo de drogas.

1 BREVE INTRODUÇÃO: MECANISMOS DE RESISTÊNCIA NA LLA-B

1.1 Resistência a drogas e Proteínas de Ligação a RNA (RBPs)

A resistência à quimioterapia pode ser tanto intrínseca, *i.e*, fatores que medeiam a resistência já estavam presentes nas células tumorais; ou adquirida, *i.e*, o tumor que era inicialmente sensível, ao longo do tempo de exposição à droga, seleciona clones e desenvolve resistência. Digno de atenção, em alguns pacientes, a exposição prolongada a um único agente quimioterápico pode levar ao desenvolvimento de resistência a múltiplos compostos estruturalmente não relacionados, processo conhecido como resistência cruzada ou resistência a múltiplas drogas (MDR).^{1,2}

Vários fatores que atuam na farmacocinética e no metabolismo celular podem levar a resistência (Figura 1). Esses fatores incluem mecanismos que limitam a quantidade de droga que entra nas células (influxo) ou aumento do efluxo de droga; à inativação direta da droga ou de processos envolvidos na sua metabolização; alterações que modificam o alvo do fármaco; reparação de dano do DNA; influência do microambiente tumoral (*e.g.*, EMT); modificações epigenéticas; e inibição da morte celular por mudanças nas vias de sinalização apoptótica ou expressão de proteínas de sobrevivência celular. Além disso, é preciso considerar que os tumores possuem um grau elevado de heterogeneidade, o que contribuiria com a resistência a drogas resultante da pressão seletiva imposta pelo tratamento³⁻⁷.

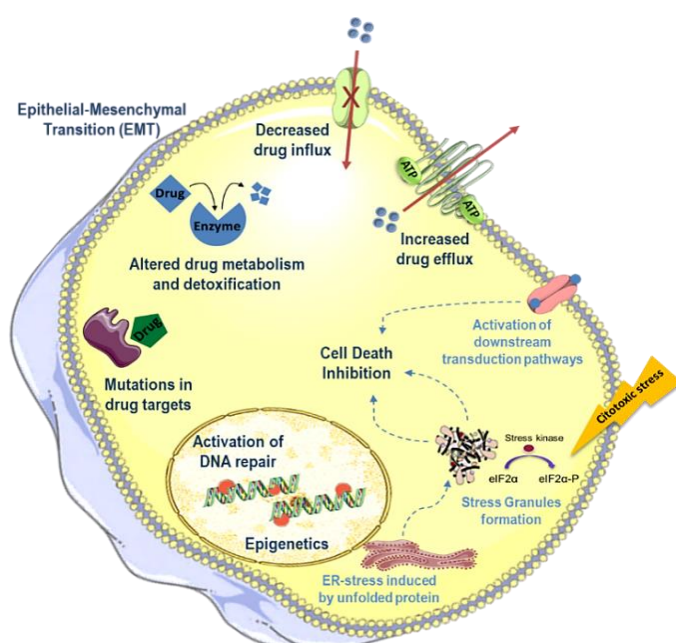


Figura 1. Representação esquemática dos mecanismos moleculares envolvidos na resistência às drogas quimioterápicas. Esses mecanismos podem atuar independentemente ou em conjunto para alterar vias metabólicas e de sinalização nas células tumorais.

A progressão tumoral é em parte impulsionada pela alteração da expressão gênica. Em condições normais, a expressão gênica é regulada em diferentes etapas, desde a transcrição, processamento de RNA, tradução em proteína e até a ativação ou degradação do polipeptídeo formado. Esses controles guiam os processos de diferenciação celular e homeostase celular, bem como, quando desregulados atuam em vias do câncer.^{8,9} Nesse sentido, proteínas de ligação a RNA (RBPs) são um dos principais reguladores pós-transcricionais e atuam em importantes passos da regulação gênica. Ligam-se a vários RNAs celulares controlando o processamento de transcritos, como nas etapas indicadas na Figura 2: *splicing* (1), transporte (2), localização (3), estabilização (4), degradação (5), controle de tradução de transcritos (6) e como guias na regulação gênica por microRNAs (7). Além disso, RBPs estão em grânulos citoplasmáticos (e.g., grânulos de estresse, GS) formados em situações de estresse celular e que desempenham importante função no controle da expressão e destino de mRNAs específicos (8).^{10,11} Logo, RBPs aberrantes podem atuar sob a homeostase celular e regular vias importantes para a progressão tumoral.

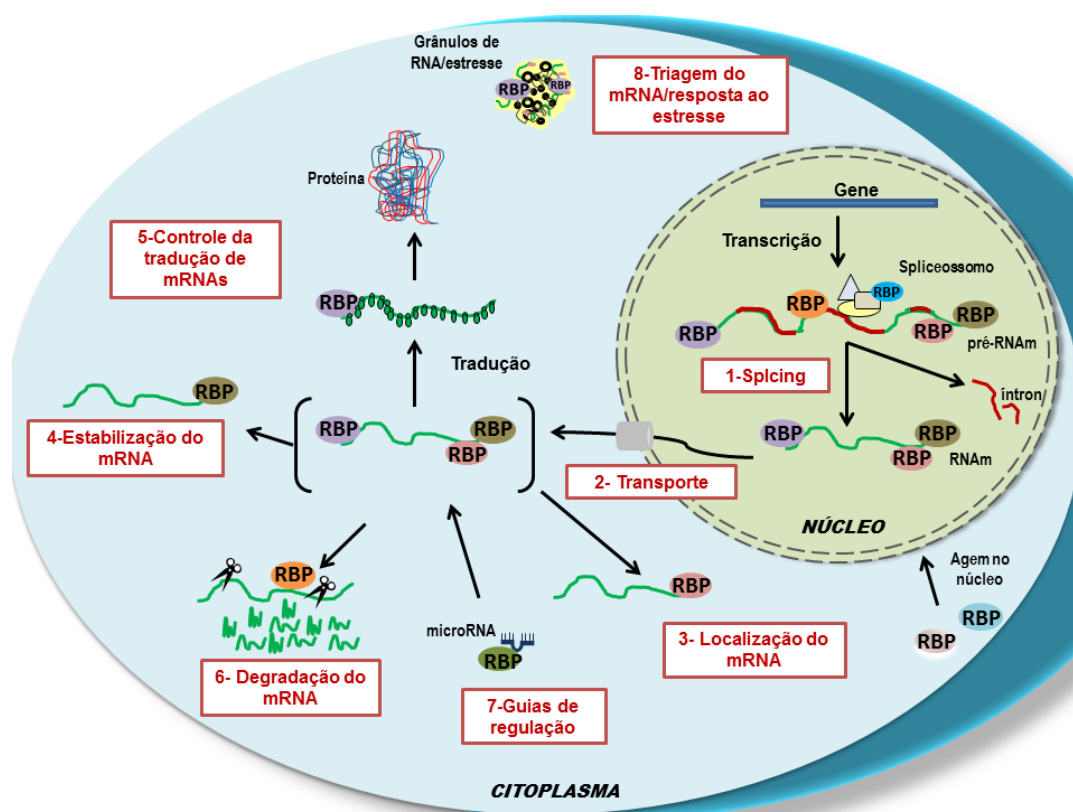


Figura 2 – Etapas importantes da regulação pós-transcricional por proteínas de ligação a RNA (RBPs).

Tabela 1. Proteínas de Ligação a RNAs (*RNA-Binding Proteins*, RBPs) envolvidas na aquisição de resistência a drogas. Tabela retirada do trabalho de Kang et al., 2013.¹²

Regulators	Target genes	Anti-cancer drugs	Cell or tissue types	Function	Effect
HuR	ER	Tamoxifen	Breast cancer cell	mRNA stability ↑	Resistance ↑
HuR	COX-2	Lapatinib	Breast cancer cell	mRNA stability ↑	Invasion ↑ Metastasis ↑
HuR	TGIF	ATO	Hepatocellular carcinoma cell	mRNA stability ↑	Cell death ↓
HuR	dCK	Gemcitabine	Pancreatic cancer	Translation ↑	Sensitivity ↑
HuR	BCL2	Etoposide, Topotecan, Cisplatin	Glioma cell	Translation ↑	Resistance ↑
HuR	TOP2A	Doxorubicin	Hela cell	Translation ↑	Sensitivity ↑
HuR	Several targets	Doxorubicin	Breast cancer cell	Translation ↑	Sensitivity ↑
HuR	Beta-tubulin	Cisplatin	Ovarian cancer cells	Translation ↑	Resistance ↑
RBM3	Unknown	Cisplatin	Epithelial ovarian cancer	Unknown	Sensitivity ↑
RBM5	Unknown	Cisplatin	Non-small cell lung cancer	Unknown	Sensitivity ↑
IMP3	BCRP	Doxorubicin, Mtoxantrone	Breast cancer cell	mRNA stability ↑	Resistance ↑
CUG-BP1	Survivin	Camptothecin	Oesophageal cancer	mRNA stability ↑	Resistance ↑
BRF1	cIAP2	Cisplatin	Head and neck squamous Carcinoma cell lines	mRNA degradation	Sensitivity ↑
HnRNPs H1/H2	TYMP	5'-deoxyfluorouridine	Histiocytic lymphoma cell	Splicing	Resistance ↑
MTDH	Several targets	Mitomycin C, BIBF1120	Endometrial cancer cell line	Stress granule formation?	Resistance ↑

Sabe-se que mutações em RBPs ou alterações na sua regulação e de seus RNAs-alvo levam ao desenvolvimento de doenças¹³, tais como doenças autoimunes¹⁴, oncológicas^{15,16} e neurológicas^{17,18}. Além disso, seu envolvimento na resistência de células tumorais a drogas anti-cancer vem sendo descrito na literatura (Tabela 1). No entanto, ainda são poucos os estudos das RBPs como alvo de drogas, e ainda não se tem provas suficientes que associem o fenótipo de resistência a drogas com RBPs individuais.¹⁹ Deste modo, é necessário mais estudos sobre a participação das RBPs e de seu papel no processamento de RNAs, nos mecanismos de regulação gênica em células resistentes.

1.2 Screening genômico funcional: '*RBP-focused CRISPR/Cas9 screening*'

O sistema CRISPR/Cas9 é baseado em um mecanismo de defesa antiviral encontrado em bactérias, em que a endonuclease Cas9 reconhece as sequências de DNA viral de infecções anteriores e destrói o DNA invasor durante a reinfecção. Atualmente, o sistema CRISPR/Cas9 é utilizado em diversos ensaios de *screening* e ensaios funcionais, e além de possibilitar editar sequências específicas de DNA do genoma, também pode ativar ou desativar a expressão de genes alvo.^{20,21} No nosso estudo, utilizamos a plataforma de *screening* genômico baseado na tecnologia de CRISPR/Cas9 com foco em RBPs, metodologia estabelecida pelo nosso colaborador.

A biblioteca usada já está disponível no banco da Addgene (*RNA-Binding Protein Pooled CRISPR Knockout Library*, from Eugene Yeo, Addgene #141438) e mais informações sobre a biblioteca e metodologia podem ser encontrados na publicação de Wheeler EC et al., 2020.²² Basicamente, a tecnologia “*pooled CRISPR/Cas9 knockout screen targeting RBPs*”, consiste em um processo no qual a endonuclease Cas9 (e.g., *S. pyogenes*) guiada por sgRNA (*single-guide RNA*) encontra uma região específica do DNA e, juntos, formam um complexo ribonucleoproteico. A Cas9 gera uma quebra na fita dupla (*double-strand break*, DSB) no DNA, ativando mecanismos inatos de reparo (i.e., NHEJ e HDR) que são aproveitados para edição de genes (Figura 3A.). A sequência guia de cada sgRNA da biblioteca integrado no genoma da célula hospedeira atua como um “*barcode*” e pode ser utilizado para calcular o número de células editadas pelo sgRNA específico por sequenciamento NGS (Figura 3C.). Esse sistema permite gerar uma biblioteca de células mutantes que podem ser selecionadas positivamente ou negativamente (*screen* fenotípico) e levam a identificação de biomarcadores e alvos terapêuticos.

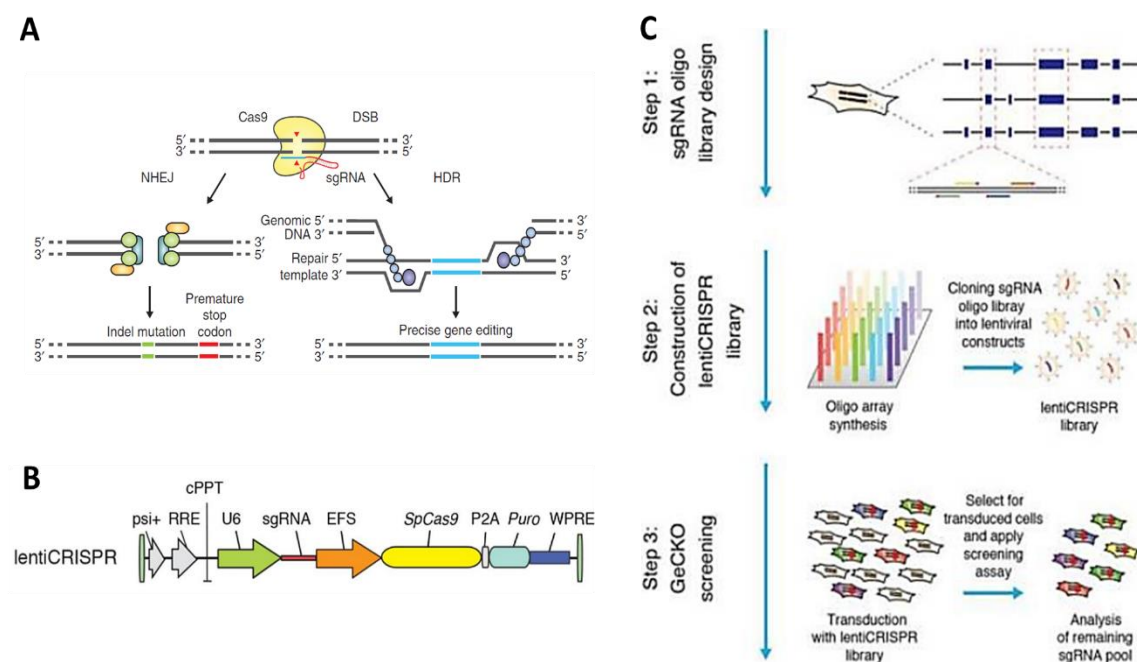


Figura 3. Pontos chave no sistema de CRISPR/Cas9 *screening*. A. Após a Cas9 induzir a quebra da dupla fita de DNA, dois mecanismos de reparo podem ser ativados; a via NHEJ que corrige a quebra através da junção das extremidades e pode resultar na correção errônea, gerando mutações randômicas no sítio de junção (*indels*). Quando essas mutações ocorrem em regiões codificantes, podem levar a uma mudança de leitura (“*frameshift*”) ou gerar códon de parada, resultado na perda de função do gene; alternativamente, uma fita molde pode ser utilizada para reparar o ponto de quebra pela via HDR, a qual permite uma edição sequência específica B. Estrutura do vetor lentiCRISPR usado no *screening*. O plasmídeo lentiCRISPR é um sistema “*all-in-one*”, isto é, um único constructo carrega a sequência de um dos sgRNAs da biblioteca e o gene que codifica a enzima Cas9, além dos agentes seletivos. C. Design da biblioteca Gecko usado para *screens* genômicos e que serviu de modelo para desenvolver o RBP CRISPR/Cas9 *screen* (Shalem, Ophir et al., 2014).²³

1.3 O papel das RBPs no mecanismo de resistência ao inibidor de tirosina quinase, imatinibe, em LLA-B portadora da translocação *BCR-ABL1*.

A proteína quimérica *BCR-ABL1* é resultante da fusão gênica da tirosina-quinase *ABL1* do cromossomo 9 com o gene *BCR* do cromossomo 22 (cromossomo Ph+ ou t(9;22)), e é característica da leucemia mieloide crônica (LMC), mas também é encontrada em 20-30% de casos de LLA de adultos e 5% das LLA infanto-juvenis. Nesses casos, a terapia alvo-específica com o uso de inibidores de tirosina-quinases (TKIs), como o imatinibe, revolucionou o tratamento da doença. O imatinibe é uma droga conhecida por atuar no domínio tirosina-quinase do proto-oncogene *ABL1*, inibindo a sua ativação constitutiva. O imatinibe (1ª geração) e dasatinib (2ª geração) têm sido eficazes no tratamento da leucemia *BCR-ABL1*+, porém ainda há um número significativo de casos que recaem.²⁴ Sabe-se que alguns pacientes não respondem ao tratamento inicial com imatinibe, enquanto outros deixam de responder ao longo da terapia. O principal mecanismo de resistência a essas drogas, descritas até o momento, resulta de alterações da oncoproteína *BCR-ABL1*, tais como a superexpressão de *BCR-ABL1*, amplificação do gene e mutações no domínio de quinase, a qual impede a ligação dos TKIs. Entretanto, fatores independentes de *BCR-ABL1*, tais como a superexpressão de transportadores (e.g., Glicoproteína P) e alterações em vias da sinalização celular também atuam sobre o processo de resistência aos TKIs.

As linhagens celulares K562-R (CRL-3344), K562-S (CRL-3343), AR230-R (CRL-3346) e AR230-S (CRL-3345) são células que possuem a fusão gênica *BCR-ABL1* e são respectivamente, resistentes (R) e sensíveis (S) ao imatinibe (Figura 4). Linhagens celulares usualmente utilizadas como modelos de resistência e estão disponíveis no biobanco da ATCC (atcc.org). As células sensíveis são cultivadas em meio de cultura RPMI-1640 acrescido de 10% de soro fetal bovino, 100UI/mL penicilina, 100pg/mL estreptomicina e incubadas a 37°C e 5%CO₂. As células resistentes são cultivadas com o mesmo meio de cultivo acrescido de 1 µM de imatinibe. Essas linhagens não possuem nenhuma mutação nos domínios da proteína quimérica descritas no contexto linhagens imatinibe-resistentes.³⁴

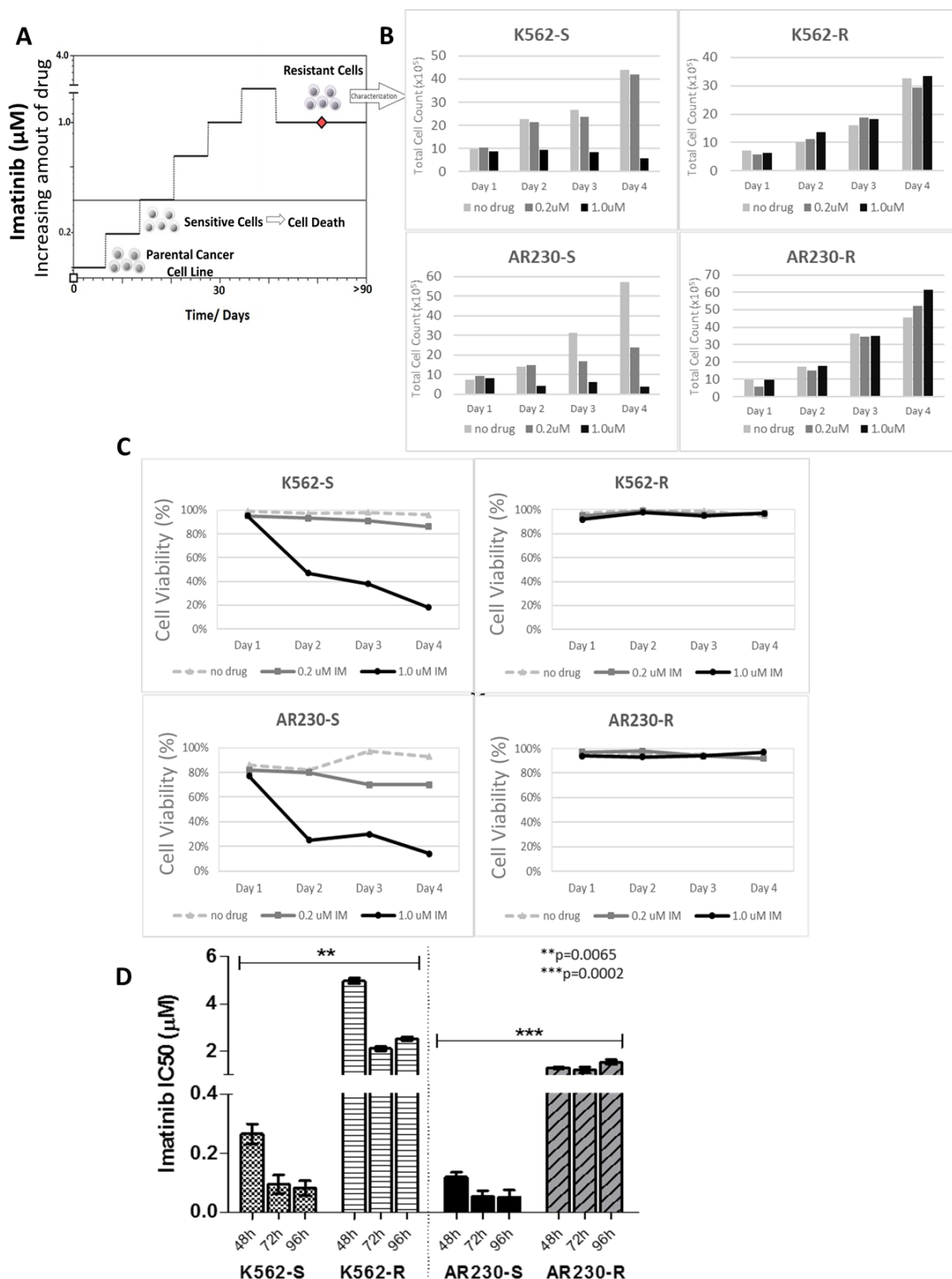


Figura 4. Caracterização e validação das linhagens celulares *BCR-ABL* positivas, com diferentes sensibilidades para o imatinibe, utilizadas como modelo experimental no estudo. A. Processo de seleção *in vitro* de clones resistentes. Células mantidas em culturas são gradualmente expostas a quantidades crescentes de droga. Após longo período de seleção, as linhagens celulares resistentes são mantidas em cultivo com o agente seletor. As linhagens parentais sensíveis são mantidas em cultivo paralelo em meio de cultura normal. B. Proliferação celular avaliada por contagem de células vivas em três concentrações diferentes de imatinibe (0, 0.2 e 1 μM) por 4 dias. C. Viabilidade celular avaliada por contagem de células mortas e vivas com o *trypan blue* nas células sem tratamento e tratadas com duas concentrações diferentes de imatinibe por 4 dias. D. Valores de IC50 calculados para as diferentes linhagens pelo ensaio de MTT (3-[4,5-dimethylthiazol-2-yl]-2,5 diphenyl tetrazolium bromide).

2 RESULTADOS

2.1 Resultados parciais

Para identificar RBPs cuja perda de função tem o potencial de re-sensibilizar células resistentes ao imatinibe, utilizamos uma biblioteca de CRISPR/Cas9 lentiviral contendo 10 sgRNAs para cada 1078 RBPs avaliados, somado a 64 sgRNA contra genes essenciais (controles positivos), e 792 sgRNAs contra regiões não codificantes (controles negativos). Ao contrário das metodologias de “*Genome-wide CRISPR screen*” utilizadas para o genoma inteiro, e que contêm em torno de 3-6 sgRNAs por gene,^{96,99} a nossa biblioteca é menor e contém um maior número de sgRNAs específica para cada RBPs, fornecendo uma maior confiança nos alvos encontrados.

Um total de 20 milhões de células K562-R e K562-S foram transduzidas com a biblioteca de RBP CRISPR/Cas9 em duplicatas biológicas e selecionadas com puromicina 1ug/mL durante 5 dias (período em que as células controle não transduzidas morrem). Células K562-R positivamente transduzida foram separadas em dois grupos, e um grupo foi tratada com 1μM de imatinibe. As células K562-S igualmente transduzidas foram mantidas em paralelo para servir de controle. Amostras de células desses grupos foram coletadas em diferentes tempos, seguido da extração do DNA genômico, preparação das bibliotecas de sequenciamento e análise pelo NGS (Figura 5A.).

As análises computacionais foram realizadas pela colaboradora Jaclyn Einstein (UCSD), com o algoritmo *MAGeCK/MAGeCK-VISPR*. A distribuição de cada sgRNA antes e depois do tratamento foi comparado para determinar o efeito da perda de função de RBPs individuais na sensibilidade ao imatinibe.

Os resultados do sequenciamento NGS mostram que sgRNAs escolhidos randomicamente no *pool* inicial de células transduzidas (dia 0) estão bem representados e em quantidade (número de *reads*) proporcional ao *pool* de sgRNAs presente na biblioteca de plasmídeos usada na preparação dos vetores virais (Figura 5B-gráfico esquerdo). Após o início do tratamento com a droga, podemos observar uma redução significativa na diversidade de sgRNAs presente nas células vivas. Essa redução aumenta com o tempo de tratamento com imatinibe, indicando que há um “*dropout*” de células transduzidas com sgRNAs que tem como alvo genes essenciais para a sobrevivência celular e/ou fenótipo de resistência ao imatinibe (Figura 5B-

gráfico direito). Após realizar uma normalização dos sgRNAs enriquecidos no dia 16, pós início da seleção negativa com imatinibe, entre o grupo resistente tratado e sensível não tratado, em relação ao controle resistente não tratado, identificamos uma lista de 27 RBPs candidatos ($p < 0.05$) que estavam depletados na população de células resistentes e tratados com imatinibe (Figura 5C). Como esperado, os controles positivos e negativos não foram afetados pela seleção com o imatinibe.

Ao analisar as vias celulares associadas com as RBPs selecionadas, observamos que cinco vias importantes se destacam; (1) '*Ribosome, Translation*'; (2) '*Aminoacyl-tRNA biosynthesis*'; (3) '*Proteasome*'; (4) '*mRNA Transport*'; e (5) '*Metabolism of RNA, mRNA splicing*' (Figura 5D.). Entre eles, destacam-se RBPs já descritos na literatura com papel na tumorigênese (Figura 5D-destacado em vermelho).

3 CONCLUSÕES PARCIAIS E PERSPECTIVAS

Em suma, nossos resultados listam 27 RBPs necessários à resistência de células *BCR-ABL+* ao imatinibe. Espera-se que a futura validação funcional e individual desses 27 RBPs aponte novos alvos para mitigar a resistência à terapia com TKI, uma condição que afeta vários pacientes em uso dessas drogas.

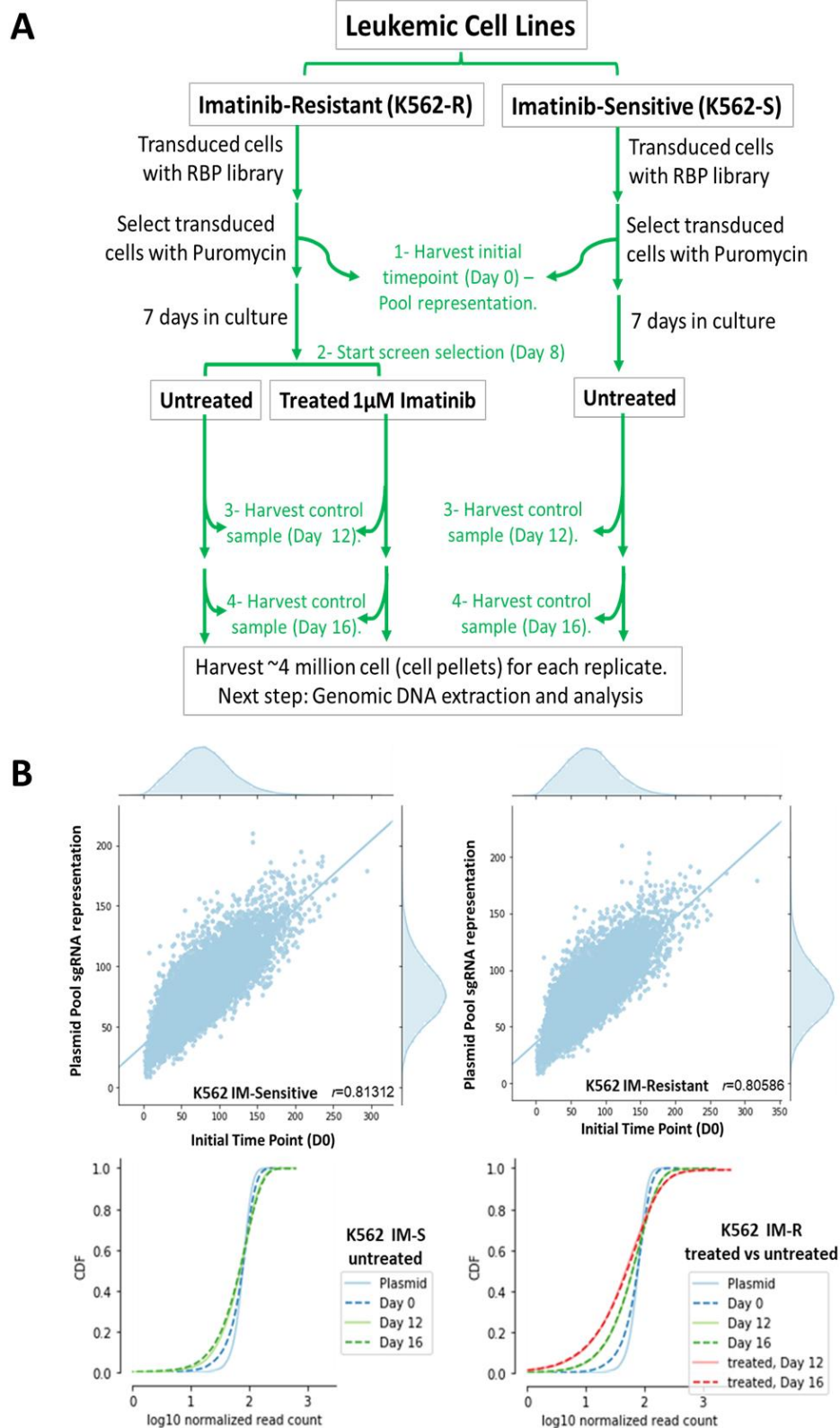


Figura 5. Proteínas de ligação a RNA (RBPs) vulneráveis a seleção negativa com a droga imatinibe em células *BCR-ABL1* (K562-Resistentes e K562-Sensíveis), identificadas pela técnica de RBP CRISPR/Cas9 *screening*. A. Delineamento experimental adotado para o ensaio de *screening*. Cada condição foi realizada em duplicatas independentes. B. Correlação entre a representatividade de sgRNAs no *pool* dos plasmídeos iniciais e ao final da seleção com puromicina (Day 0), gráficos acima. Distribuição cumulativa das ‘read counts’ dos sgRNAs em cada condição, gráficos abaixo. ‘Two-sided Kolmogorov-Smirnov’ teste comparado com o dia 0.

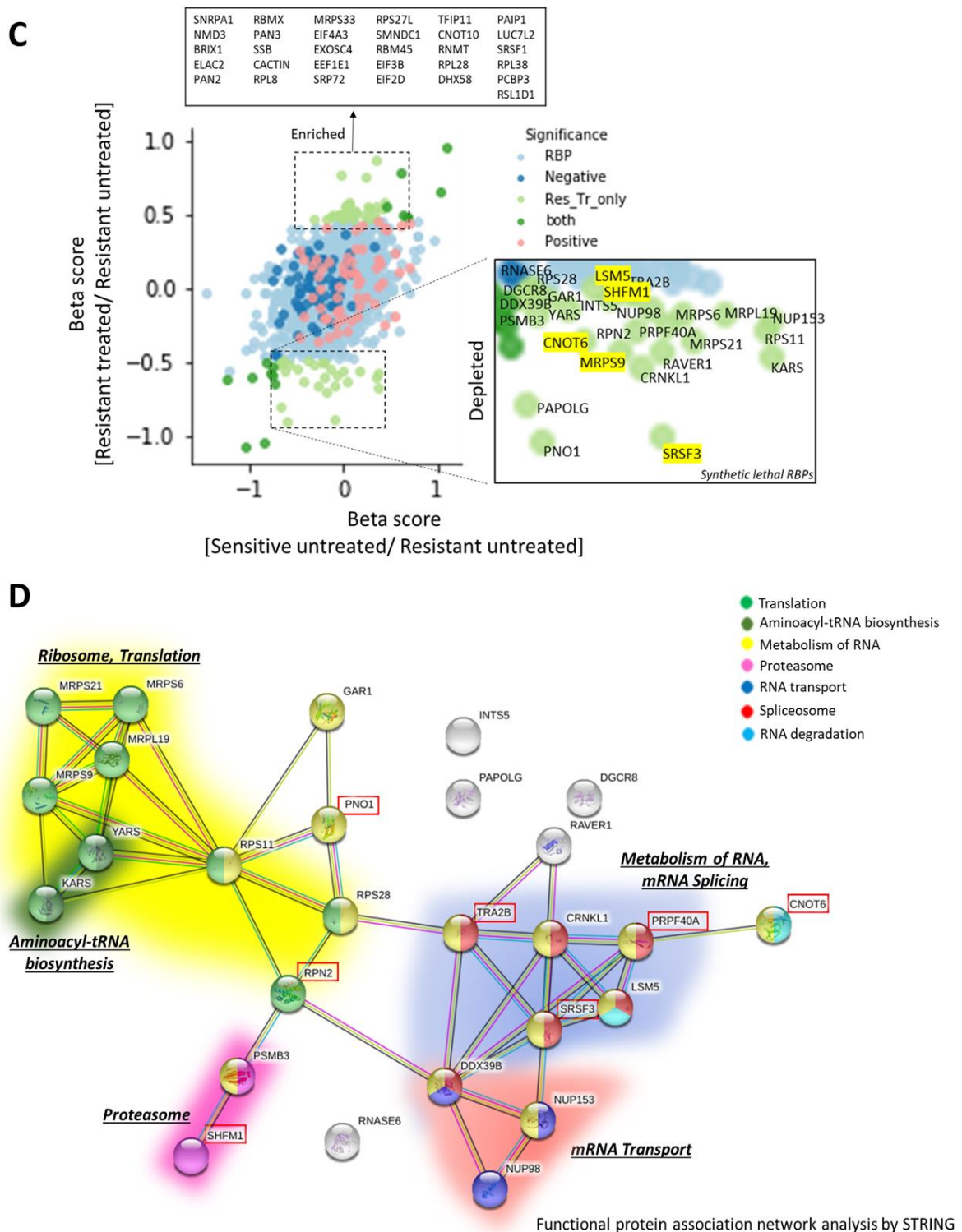


Figura 5 [Cont.] C. Comparação do 'β score' no dia 16 pós início da pressão seletiva com imatinibe. Quadro em destaque lista os possíveis "synthetic lethal" RBPs candidatos e em amarelo são os genes significativamente depletados também no dia 12 pós tratamento com imatinibe. D. Esquema descrevendo as vias biológicas relacionadas com os RBPs candidatos identificados em C. RBPs destacados no quadro vermelho indicam proteínas já descritas com papel na progressão do câncer.

RBP	Nome	Informações funcionais com pilados na literatura	Artigos selecionados para estudo
PN01	Partner Of NOB1 Homolog	Crítico para a biogênese do ribossomo. Regula positivamente a dimetilação de adenosinas do rRNA 18S.	EBF1-Mediated Upregulation of Ribosome Assembly Factor PN01 Contributes to Cancer Progression by Negatively Regulating the p53 Signaling Pathway. [PMID: 30862720]. Ribosome assembly factor PN01 as a potential oncogene involved in tumor growth and progression of colorectal cancer. [PMID: 30862720]
SRSF3	Seime And Arginine Rich Splicing Factor 3	Fator de splicing. Regula IL3 e promove a tumorigênese.	SRSF3-Regulated RNA Alternative Splicing Promotes Glioblastoma Tumorigenicity by Affecting Multiple Cellular Processes. [PMID: 31462429]. Oncogenic splicing factor SRSF3 regulates ILF3 alternative splicing to promote cancer cell proliferation and transformation. [PMID: 30796096]. Downregulation of splicing factor SRSF3 induces p53 β , an alternatively spliced isoform of p53 that promotes cellular senescence. [PMID: 22777358]
RPN2	Ribophorn II	Potencializa a resistência a múltiplas drogas mediada pela glicoproteína P e ABCG2 via ativação da via ERK em câncer gástrico.	Ribophorn II potentiates P-glycoprotein- and ABCG2-mediated multidrug resistance via activating ERK pathway in gastric cancer.[PMID: 30710584] Ribophorn II promotes cell proliferation, migration, and invasion in esophageal cancer cells <i>in vitro</i> and <i>in vivo</i> . [PMID: 30940778]. Downregulation of RPN2 induces apoptosis and inhibits migration and invasion in colon carcinoma.[PMID: 29749494]. MicroRNA-128 targeting RPN2 inhibits cell proliferation and migration through the Akt-p53-cyclin pathway in colorectal cancer cells. [PMID: 30546426]
PRPF40A	Pre-MRNA Processing Factor 40 Homolog A	Desempenha um papel na regulação da morfologia celular e organização do citosqueleto. Pode desempenhar um papel na citocinese. Pode estar envolvido no splicing pré-mRNA.	PRPF40A as a potential diagnostic and prognostic marker is upregulated in pancreatic cancer tissues and cell lines: an integrated bioinformatics data analysis. [PMID: 31303762]. More targets, more pathways and more clues for mutant p53. [PMID: 23817466]
TRA2B	Transformer 2 Beta Homolog	Participa no controle de splicing do pré-mRNA. Pode regular a inclusão ou retirada de exon.	Knocking down the expression of TRA2B inhibits the proliferation and migration of human glioma cells. [PMID: 26298634]. Transformer 2B and miR-204 regulate apoptosis through competitive binding to 3' UTR of BCL2 mRNA. [PMID: 25342468]. Transformer 2B (Tra2B/SFRS10) positively regulates the progression of NSCLC via promoting cell proliferation. [PMID: 24952301]
CNOT6	CCR4-NOT Transcription Complex Subunit 6	Poi (A) nuclease com atividade 3'-5'-RNase. Componente catalítico do complexo CCR4-NOT, que é uma das principais deadenilases de mRNA celulares e está ligado a vários processos celulares, incluindo degradação de mRNA, repressão mediada por miRNA, repressão translacional durante o início da tradução e regulação geral da transcrição.	The Ccr4a (CNOT6) and Ccr4b (CNOT6L) deadenylase subunits of the human Ccr4-Not complex contribute to the prevention of cell death and senescence. [PMID: 21233283]. miR-29c-3p promotes senescence of human mesenchymal stem cells by targeting CNOT6 through p53-p21 and p16-pRB pathways. [PMID: 26792405]
SHFM1 (DSS1)	SEM1 26S Proteasome Complex Subunit	Componente do proteossoma 26S, um complexo multiproteico envolvido na degradação dependente de ATP de proteínas ubiquitinadas.	DSS1 promoter hypomethylation and overexpression predict poor prognosis in melanoma and squamous cell carcinoma patients. [PMID: 27825810]. Upregulated, 7q21-22 amplicon candidate gene SHFM1 confers oncogenic advantage by suppressing p53 function in gastric cancer.[PMID: 25697906]
INTS5	Integrator Complex Subunit 5	Componente do complexo integrador (INT), um complexo envolvido na transcrição U1 e U2 de pequenos RNAs nucleares (snRNA).	The Integrator subunits function in hematopoiesis by modulating Smad/BMP signaling. [PMID: 19605500]

Tabela 2. Compilado da literatura buscando informações sobre os selecionados RBPs e associação com câncer.

4 REFERÊNCIAS BIBLIOGRÁFICAS

- ¹ Zahreddine, H.; Borden, K.L. Mechanisms and insights into drug resistance in cancer. *Front. Pharmacol.* 2013, 4, 28.
- ² Housman G, Byler S, Heerboth S, et al. Drug Resistance in Cancer: An Overview. *Cancers*. 2014;6(3):1769-1792.
- ³ Gorre, M.E.; Mohammed, M.; Ellwood, K.; Hsu, N.; Paquette, R.; Rao, P.N.; Sawyers, C.L. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. *Science* 2001, 293, 876-880
- ⁴ Michael, M.; Doherty, M.M. Tumoral drug metabolism: Overview and its implications for cancer therapy. *J. Clin. Oncol.* 2005, 23, 205-229.
- ⁵ Shang, Y.; Cai, X.; Fan, D. Roles of epithelial-mesenchymal transition in cancer drug resistance. *Curr. Cancer Drug Targets* 2013, 13, 915-929.
- ⁶ Holohan, C.; van Schaeybroeck, S.; Longley, D.B.; Johnston, P.G. Cancer drug resistance: An evolving paradigm. *Nat. Rev.* 2013, 13, 714-726.
- ⁷ Yin, Q.; Shen, J.; Zhang, Z.; Yu, H.; Li, Y. Reversal of multidrug resistance by stimuli-responsive drug delivery systems for therapy of tumor. *Adv. Drug Deliv. Rev.*, 2013, 65 (13-14), 1699-1715.
- ⁸ Wray GA, Hahn MW, Abouheif E, et al. The evolution of transcriptional regulation in eukaryotes. *Mol Biol Evol.* 2003;20(9):1377-1419. doi:10.1093/molbev/msg140
- ⁹ Keene JD. RNA regulons: coordination of post-transcriptional events. *Nat Rev Genet.* 2007;8(7):533-543. doi:10.1038/nrg2111
- ¹⁰ Glisovic T, Bachorik JL, Yong J, Dreyfuss G. RNA-binding proteins and post-transcriptional gene regulation. *FEBS Lett.* 2008;582(14):1977-1986. doi:10.1016/j.febslet.2008.03.004
- ¹¹ Kishore S, Lubner S, Zavolan M. Deciphering the role of RNA-binding proteins in the post-transcriptional control of gene expression. *Brief Funct Genomics.* 2010;9(5-6):391-404. doi:10.1093/bfpg/elq028
- ¹² Kang, H.; Kim, C.; Lee, H.; Kim, W.; Lee, E.K. Post-Transcriptional Controls by Ribonucleoprotein Complexes in the Acquisition of Drug Resistance. *Int. J. Mol. Sci.* 2013, 14, 17204-17220. <https://doi.org/10.3390/ijms140817204>
- ¹³ Lukong KE, Chang KW, Khandjian EW, Richard S. RNA-binding proteins in human genetic disease. *Trends Genet.* 2008;24(8):416-425. doi:10.1016/j.tig.2008.05.004
- ¹⁴ Ivanov P, Anderson P. Post-transcriptional regulatory networks in immunity. *Immunol Rev.* 2013;253(1):253-272. doi:10.1111/immr.12051
- ¹⁵ Sabile AA, Arlt MJ, Muff R, et al. Caprin-1, a novel Cyr61-interacting protein, promotes osteosarcoma tumor growth and lung metastasis in mice. *Biochim Biophys Acta.* 2013;1832(8):1173-1182. doi:10.1016/j.bbdis.2013.03.014
- ¹⁶ Wurth L, Gebauer F. RNA-binding proteins, multifaceted translational regulators in cancer. *Biochim Biophys Acta.* 2015;1849(7):881-886. doi:10.1016/j.bbagr.2014.10.001
- ¹⁷ Ibrahim F, Nakaya T, Mourelatos Z. RNA dysregulation in diseases of motor neurons. *Annu Rev Pathol.* 2012;7:323-352. doi:10.1146/annurev-pathol-011110-130307
- ¹⁸ Kapeli K, Yeo GW. Genome-wide approaches to dissect the roles of RNA binding proteins in translational control: implications for neurological diseases. *Front Neurosci.* 2012;6:144. Published 2012 Oct 2. doi:10.3389/fnins.2012.00144
- ¹⁹ Pereira B, Billaud M, Almeida R. RNA-Binding Proteins in Cancer: Old Players and New Actors. *Trends Cancer.* 2017;3(7):506-528. doi:10.1016/j.trecan.2017.05.003
- ²⁰ Sanjana NE, Shalem O, Zhang F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods.* 2014 Aug;11(8):783-4. doi: 10.1038/nmeth.3047. 10.1038/nmeth.3047
- ²¹ He C, Han S, Chang Y, et al. CRISPR screen in cancer: status quo and future perspectives. *Am J Cancer Res.* 2021;11(4):1031-1050. Published 2021 Apr 15.
- ²² Wheeler EC, Vu AQ, Einstein JM, DiSalvo M, Ahmed N, Van Nostrand EL, Shishkin AA, Jin W, Allbritton NL, Yeo GW. Pooled CRISPR screens with imaging on microarray reveals stress granule-regulatory factors. *Nat Methods.* 2020 May 11. pii: 10.1038/s41592-020-0826-8. doi: 10.1038/s41592-020-0826-8.
- ²³ Shalem O, Sanjana NE, Hartenian E, et al. Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science.* 2014;343(6166):84-87. doi:10.1126/science.1247005
- ²⁴ Hunger SP. Tyrosine kinase inhibitor use in pediatric Philadelphia chromosome-positive acute lymphoblastic anemia. *Hematology Am Soc Hematol Educ Program.* 2011; 2011:361-5.
- ²⁵ Hochhaus A, Kreil S, Corbin AS, La Rosee P, Muller MC, Lahaye T, Hanfstein B, Schoch C, Cross NC, Berger U, GSchaidmeier H, Druker BJ, Hehlmann R. Molecular and chromosomal mechanisms of resistance to *Imatinib* (STI571) therapy. *Leukemia.* 2002; 16:2190-2196.
- ²⁶ Hochhaus A. Cytogenetic and molecular mechanisms of resistance to *Imatinib*. *Semin Hematol.* 2003;40: 69-79.
- ²⁷ Shah NP, Tran C, Lee FY, Chen P, Norris D, Sawyers CL. Overriding *Imatinib* resistance with a novel ABL kinase inhibitor. *Science.* 2004; 305:399-401.
- ²⁸ Weisberg E, Manley PW, Cowan-Jacob SW, Hochhaus A, Griffin JD. Second generation inhibitors of BCR-ABL for the treatment of *Imatinib*-resistant chronic myeloid leukaemia. *Nat Rev Cancer.* 2007; 7:345-356.
- ²⁹ Nimmanapalli R, Bhalla K. Mechanisms of resistance to *Imatinib* mesylate in Bcr-Abl-positive leukemias. *Curr Opin Oncol.* 2002; 14:616-620

-
- ³⁰ Hirayama, C.; Watanabe, H.; Nakashima, R.; Nanbu, T.; Hamada, A.; Kuniyasu, A.; Nakayama, H.; Kawaguchi, T.; Saito, H. Constitutive overexpression of P-glycoprotein, rather than breast cancer resistance protein or organic cation transporter 1, contributes to acquisition of *Imatinib*-resistance in K562 cells. *Pharm. Res.* 2008, 25, 827–835.
- ³¹ Al-Jamal, H.A.; Asmaa, M.J.; Yong, A.C.; Asan, J.M.; Hassan, R.; Johan, M.F. Silencing of suppressor of cytokine signaling-3 due to methylation results in phosphorylation of STAT3 in *Imatinib* resistant BCR-ABL positive chronic myeloid leukemia cells. *Asian Pac. J. Cancer Prev.* 2014, 15, 4555–4561.
- ³² Nambu, T.; Araki, N.; Nakagawa, A.; Kuniyasu, A.; Kawaguchi, T.; Hamada, A.; Saito, H. Contribution of BCR-ABL-independent activation of ERK1/2 to acquired *Imatinib* resistance in K562 chronic myeloid leukemia cells. *Cancer Sci.* 2010, 101, 137–142.
- ³³ Burchert A. Roots of *Imatinib* resistance: a question of self-renewal? *Drug Resist Updat* 2007; 10: 152–161.
- ³⁴ Mahon FX, et al. Selection and characterization of BCR-ABL positive cell lines with differential sensitivity to the tyrosine kinase inhibitor STI571: diverse mechanisms of resistance. *Blood* 96(3): 1070-1079, 2000

13 ANEXOS

13.1 Anexo I: Parecer do Comitê de Ética (CEP).

Campinas, 19/06/2018

Ao CEP-Boldrini

Prezado Sr(a),

Solicito a inclusão de três adendos ao projeto de pesquisa “Método para identificar a Leucemia Linfóide Aguda (LLA) pediátrica do subgrupo BCR-ABL-like” (CAAE 51991015.0.0000.5376).

A logística para identificação das LLA do subgrupo BCR/ABL1-like (ph-like), conforme definida no projeto, consta de duas etapas sequenciais:

- (i) Primeiro faz-se uma etapa de seleção negativa, identificando as LLA que apresentam hiperdiploidia ou alguma das translocações cromossômicas clássicas: t(21;21)(ETV6/RUNX1), t(1;19)(TCF3/PBX1), t(9;22)(BCR/ABL1) ou t(4;11)(MLL/AF4). Casos de LLA com hiperdiploidia ou presença das translocações clássicas pertencem a outros subtipos de LLA e, portanto, não são ph-like.
- (ii) Uma vez excluídos os casos de hiperdiploidia e com translocações clássicas, as demais LLA são analisadas quanto ao nível de expressão de um conjunto de 25 genes marcadores das ph-like.

Após estas duas etapas, há uma terceira cujo objetivo é confirmar se os casos de LLA identificados como ph-like são mesmo ph-like. Para isso, propusemos analisar deleções no gene IKZF1, rearranjos de CRLF2, mutações em Jak1/Jak2 e um polimorfismo no gene GATA3, que são típicos das LLA ph-like.

Adendo 1: qPCR translocações clássicas da LLA

O primeiro adendo é referente à etapa 1 e visa o teste de PCR quantitativo (qPCR) para detecção das translocações cromossômicas clássicas da leucemia linfóide aguda (LLA).

A identificação da hiperdiploidia (etapa 1) é atualmente feita por análise convencional de cariótipo ou análise de índice de DNA por citometria de fluxo.

A identificação das translocações cromossômicas clássicas é feita de forma rotineira há quase 20 anos, pela análise de RT-PCR dos transcritos quiméricos ETV6/RUNX1, TCF3/PBX1, BCR/ABL1, MLL/AF4 de acordo com método de Van Dongen et al. (Leukemia 13:1901–1928; 1999) ou utilizando FISH (principalmente para os casos de rearranjos do MLL).

O Dr Mateus Nobrega Aoki, pesquisador da Fiocruz - Instituto Carlos Chagas (Curitiba, PR), desenvolveu método de detecção dos transcritos quiméricos das LLA por PCR quantitativo. Este novo método pode potencialmente alcançar uma maior especificidade e sensibilidade que o método RT-PCR em uso, com a vantagem de utilizar menor quantidade de amostra biológica. Sendo assim, solicito incluir como adendo ao projeto, o teste deste novo método para a detecção dos transcritos quiméricos por PCR quantitativo. Todos estes testes serão feitos no Centro Infantil Boldrini.

Adendo 2: NGS de 1385 genes, FISH e MLPA das LLA ph-like e b-other

O segundo adendo é referente à etapa 3. A melhor forma de confirmarmos a identidade das ph-like será por análise de microarranjos de expressão gênica, que é o método primeiramente usado para a detecção desse subgrupo, embora por ser caro não serve como método rotineiro. Tendo conseguido financiamento do Instituto Ronald McDonald (processo 2016109; R\$ 270.059,75) e do CNPq (processo 429811/2016-0; R\$ 110.000,00) para o projeto, iremos fazer análise de microarranjos de expressão gênica em 120 casos de LLA, incluindo todos os suspeitos ph-like e alguns representantes dos demais 6 subgrupos moleculares: ETV6/RUNX1, TCF3/PBX1, BCR/ABL1, MLL, hiperdiploide, “b-other” (LLA sem subgrupo molecular definido).

A etapa 3 tem também por finalidade caracterizar outras alterações genéticas associadas à LLA ph-like, que também é uma maneira de confirmar por método ortogonal a identidade desses casos de LLA. Para efeito de comparação, as mesmas análises serão feitas nas LLA sem subgrupo molecular definido, isto é nas LLA “b-other”. Entre as várias alterações genéticas das LLA ph-like as mais importantes atualmente são as translocações cromossômicas envolvendo os genes ABL1, ABL2, CSF1R e PDGFRB, pois foi mostrado que as LLA portadoras dessas alterações são sensíveis ao imatinibe, fármaco já aprovado e em uso clínico contra a LLA portadora da t(9;22)(BCR/ABL1).

Uma forma de detectar alterações em ABL1, ABL2, CSF1R e PDGFRB é pelo método de sequenciamento de última geração (NGS, next generation sequencing), que é o que pretendemos realizar. Para tanto utilizaremos o kit TruSight PanCancer (Illumina). O TruSight PanCancer permite a análise NGS de 1.385 genes. Assim, além dos dados de ABL1, ABL2, CSF1R e PDGFRB, propomos incorporar também os dados obtidos para os demais 1.385 genes, analisando a incidência de mutações no grupo das LLA ph-like, bem como das LLA b-other e ETV6/RUNX1 para efeito comparativo.

Ao mesmo tempo, para agilizar o diagnóstico das alterações ABL1, ABL2, CSF1R e PDGFRB, propomos também testar método FISH dessas alterações, que já está disponível comercialmente.

Além disso, além do método PCR convencional inicialmente proposto, iremos padronizar método MLPA (*Multiplex Ligation-dependent Probe Amplification*) de análise de deleções de IKZF1. Utilizaremos para tanto um kit da MRC Holland, que analisa juntamente ao IKZF1 vários outras alterações, as quais incluem CNVs (copy number variation) dos genes IKZF1, PAX5, ETV6, RB1, BTG1, EBF1, CDKN2A-CDKN2B, Xp22.33. A empresa MRC Holland desenvolveu um método denominado “digital MLPA” e combina MLPA seguido de sequenciamento na plataforma NGS. Eventualmente iremos testar esse novo método, que além dos genes acima, inclui alguns outros. Todas estas alterações serão associadas quanto a incidência e valor prognóstico no grupo das LLA ph-like e LLA b-other.

Adendo 3: *Granulos de estresse e resistência ao imatinib na LLA*

Finalmente, apresento em anexo o projeto da aluna de doutorado Natacha Azussa Mígita, cujo nome estou incluindo na equipe de trabalho. O seu projeto de doutorado é “Identificação de fusões gênicas em Leucemia Linfóide Aguda (LLA) pediátrica do subgrupo BCR-ABL1-like e estudo de mecanismos de resistência aos inibidores tirosina quinases”. O projeto quanto sua avaliação positiva pela FAPESP foram adicionados na Plataforma Brasil. Como poderá ser visto no projeto, parte do mesmo é o assunto do adendo 2, isto é, a detecção das alterações em ABL1, ABL2, CSF1R e PDGFRB. A outra parte é um estudo funcional não prevista inicialmente.

Ela irá trabalhar com as análises NGS das LLA ph-like e b-other, acima mencionadas, e além disso irá também estudar o papel de grânulos de estresse na resistência ao imatinib. Até o momento o imatinib é pouco usado na LLA pediátrica e a resistência à droga é devida a mutações no domínio quinase da proteína BCR/ABL1. Acreditamos, porém, que outros mecanismos de resistência possam ser operativos, incluindo pela via de formação de grânulos de estresse, objeto deste adendo. Inicial e majoritariamente propomos estudos com uso de linhagens celulares de LLA. Mas depois, caso encontremos resultados interessantes, gostaríamos de fazer testes confirmatórios com células de LLA ph-like de pelo menos 3 pacientes, que serão cultivadas in vitro ou transplantadas em camundongos. Solicitamos, portanto, autorização do CEP para uso de amostras de pacientes nestes testes.

Solicitamos também a inclusão de Natacha Azussa Migita na equipe do projeto.

Pelo exposto, cabe ressaltar que nenhuma das mudanças irá requerer nova coleta de amostra dos pacientes. Será usada amostra obtida ao diagnóstico, já armazenada em Biobanco. Considerando que os objetivos primários do estudo não foram alterados e que os pacientes e/ou seus responsáveis assinaram TCLE para armazenamento das amostras no Biobanco, solicitamos dispensa de novo TCLE. Esclarecemos que o TCLE do Biobanco contempla a possibilidade de uso da amostra em novo estudo aprovado pelo CEP, com dispensa ou não de novo consentimento. Caso a opção "Quero ser consultado antes de cada pesquisa" tenha sido assinalada, a amostra não será usada sem previa autorização do paciente/responsável.

Atenciosamente,



José Andrés Yunes

**CENTRO INFANTIL DE
INVESTIGAÇÕES
HEMATOLÓGICAS DR.**



PARECER CONSUBSTANCIADO DO CEP

DADOS DA EMENDA

Título da Pesquisa: Método para identificar a Leucemia Linfóide Aguda (LLA) pediátrica do subgrupo BCR-ABL1-like

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: 2

CAAE: 51991015.0.0000.5376

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

Patrocinador Principal: MINISTERIO DA CIENCIA, TECNOLOGIA E INOVACAO

DADOS DO PARECER

Número do Parecer: 2.802.372

Apresentação do Projeto:

Método para identificar a Leucemia Linfóide Aguda (LLA) pediátrica do subgrupo BCR-ABL1-like

Objetivo da Pesquisa:

Objetivos:

Parte I – Estabelecer método diagnóstico das mutações e fusões gênicas comuns a pacientes BCR-ABL1-like.

1.1 Desenvolver método de enriquecimento e sequenciamento de RNAs dos genes ABL1, ABL2, BRAF, CRLF2, CSF1R, DGKH, DYRK1A, EBF1, EPOR, FLT3, IKZF1, IL2RB, IL7R, JAK1, JAK2, JAK3, KRAS, NF1, NRAS, NTRK3, PAX5, PDGFRB, PTK2B, PTPN11, SH2B3, TSLP e TYK2; em 31 amostras de pacientes classificados como BCR-ABL1-like.

1.2 Padronizar análise de bioinformática para detectar mutações e/ou ocorrência de transcritos quiméricos nos genes ABL1, ABL2, BRAF, CRLF2, CSF1R, DGKH, DYRK1A, EBF1, EPOR, FLT3, IKZF1, IL2RB, IL7R, JAK1, JAK2, JAK3, KRAS, NF1, NRAS, NTRK3, PAX5, PDGFRB, PTK2B, PTPN11, SH2B3, TSLP e TYK2.

1.3 Havendo fusões gênicas ainda não descritas, verificar se são sensíveis ao imatinib em modelo xenográfico (camundongos NOD/SCID transplantados com a amostra de LLA em questão).

Parte II – Estudo da resistência ao imatinib pela formação de grânulos de estresse.

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.



Continuação do Parecer: 2.802.372

1.1 Verificar formação de grânulos de estresse em diferentes linhagens celulares de LLA (REH, RS4:11, 697, JURKAT, DND41 e SIL) tratadas com imatinib, tecendo associações com as características de cada linhagem (se LLA B-derivada ou T-ALL; presença de fusões do tipo ABL-fusion; presença de outras fusões gênicas ou mutações).

1.2 Verificar se a formação de grânulos de estresse leva a aumento de resistência ao imatinib.

1.3 Determinar a composição dos grânulos de estresse em linhagens celulares de LLA tratadas com imatinib, tendo como controle outro agente indutor de grânulos de estresse (arsenito).

Avaliação dos Riscos e Benefícios:

Riscos:

Os pesquisadores julgam que não é previsto riscos.

Benefícios:

Os pesquisadores julgam não haver benefícios diretos. Os benefícios são de cunho científico na identificação de casos de LLA BCR-ABL1-like.

Comentários e Considerações sobre a Pesquisa:

O adendo está bem escrito, detalhado e claro. A metodologia é adequada e factível.

Considerações sobre os Termos de apresentação obrigatória:

Foram encaminhados ao CEP o adendo ao projeto datado de 19.06.2018, o Despacho da FAPESP sobre o projeto supracitado e o projeto de pesquisa "Identificação de fusões gênicas em Leucemia Linfóide Aguda (LLA) pediátrica do subgrupo BCR-ABL1-like e estudo de mecanismos de resistência aos inibidores tirosina quinases"

Recomendações:

- O pesquisador deve desenvolver a pesquisa conforme delineada no protocolo aprovado. - Se o pesquisador considerar a descontinuação do estudo, esta deve ser justificada e somente ser realizada após análise das razões da descontinuidade pelo CEP que o aprovou. O pesquisador deve aguardar o parecer do CEP quanto à descontinuação, exceto quando perceber risco ou dano não previsto ao participante ou quando constatar a superioridade de uma estratégia diagnóstica ou terapêutica oferecida a um dos grupos da pesquisa, isto é, somente em caso de necessidade de ação imediata com intuito de proteger os participantes.
- Eventuais modificações ou emendas ao protocolo devem ser apresentadas ao CEP de forma clara

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

Página 02 de 04

CENTRO INFANTIL DE INVESTIGAÇÕES HEMATOLÓGICAS DR.



Continuação do Parecer: 2.802.372

e sucinta, identificando a parte do protocolo a ser modificada e suas justificativas e aguardando a aprovação do CEP para continuidade da pesquisa. Em caso de projetos do Grupo I ou II apresentados anteriormente à ANVISA, o pesquisador ou patrocinador deve enviá-las também à mesma, junto com o parecer aprovatório do CEP, para serem juntadas ao protocolo inicial.

- Relatórios parciais e final devem ser apresentados ao CEP e ao término do estudo.

- Lembramos que segundo a Resolução 466/2012, item XI.2 letra e, "cabe ao pesquisador apresentar dados solicitados pelo CEP ou pela CONEP a qualquer momento".

- O pesquisador deve manter os dados da pesquisa em arquivo, físico ou digital, sob sua guarda e responsabilidade, por um período pré definido ou 5 anos após o término da pesquisa.

Conclusões ou Pendências e Lista de Inadequações:

Aprovado

Considerações Finais a critério do CEP:

Aprovado sem pendências ou inadequações

Este parecer foi elaborado baseado nos documentos abaixo relacionados:

Tipo Documento	Arquivo	Postagem	Autor	Situação
Informações Básicas do Projeto	PB_INFORMAÇÕES_BÁSICAS_1152215_E1.pdf	20/06/2018 17:40:27		Aceito
Outros	DespachoFAPESPprojetoNatacha031116.pdf	20/06/2018 17:22:03	Jose Andres Yunes	Aceito
Outros	ProjetoDRNatachaMigita.pdf	20/06/2018 17:21:38	Jose Andres Yunes	Aceito
Outros	AdendosProjetoBCRABlike200618.pdf	20/06/2018 17:20:08	Jose Andres Yunes	Aceito
Outros	termosDeConcessao_bolsa_CNPq.pdf	11/12/2015 16:47:31	Jose Andres Yunes	Aceito
Outros	Parecer_FAPESP.pdf	11/12/2015 16:45:37	Jose Andres Yunes	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Carta_dispenza_TCLE.doc	11/12/2015 16:41:24	Jose Andres Yunes	Aceito
Projeto Detalhado / Brochura Investigador	Projeto_ABL1_cep.pdf	11/12/2015 16:37:22	Jose Andres Yunes	Aceito
Declaração de Pesquisadores	Termo_responsabilidade_e_comprometimento.pdf	11/12/2015 16:34:29	Jose Andres Yunes	Aceito

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

Página 03 de 04

**CENTRO INFANTIL DE
INVESTIGAÇÕES
HEMATOLÓGICAS DR.**



Continuação do Parecer: 2.802.372

TCLE / Termos de Assentimento / Justificativa de Ausência	Termo_de_assentimento_12_17_anos.doc	11/12/2015 16:33:45	Jose Andres Yunes	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	Termo_de_assentimento_6_11_anos.doc	11/12/2015 16:32:59	Jose Andres Yunes	Aceito
TCLE / Termos de Assentimento / Justificativa de Ausência	TCLE_ABL1_CEP.doc	11/12/2015 16:31:57	Jose Andres Yunes	Aceito
Folha de Rosto	Folha_de_rosto0001.pdf	11/12/2015 16:29:15	Jose Andres Yunes	Aceito

Situação do Parecer:

Aprovado

Necessita Apreciação da CONEP:

Não

CAMPINAS, 06 de Agosto de 2018

Assinado por:
Maristela Amaral Palazzi
(Coordenador)

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

Página 04 de 04

13.2 Anexo II: Declaração sobre 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 **"Mutações genéticas na Leucemia Linfóide Aguda pediátrica do subgrupo 'B-other'"**, não infringem os dispositivos da Lei n.º 9.610/98, nem o direito autoral de qualquer editora.

Campinas, 21 de março de 2022



Assinatura:

Nome do(a) autor(a): **Natacha Azussa Migita**

RG n.º 44.513.306-5



Assinatura:

Nome do(a) orientador(a): **José Andrés Yunes**

RG n.º 54.134.637-4