UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCIÊNCIAS
PROGRAMA DE PÓS-GRADUAÇÃO EM GENÉTICA

ALEIDE SANTOS DE MELO LIMA

AVALIAÇÃO DO IMPACTO ISOLADO E ADITIVO DE MARCADORES MOLECULARES NO DESFECHO CLÍNICO DE PACIENTES ADULTOS COM LEUCEMIA MIELOIDE AGUDA

Recife
2022

ALEIDE SANTOS DE MELO LIMA

AVALIAÇÃO DO IMPACTO ISOLADO E ADITIVO DE MARCADORES MOLECULARES NO DESFECHO CLÍNICO DE PACIENTES ADULTOS COM LEUCEMIA MIELOIDE AGUDA

Tese apresentada ao Programa de Pós-Graduação em Genética da Universidade Federal de Pernambuco como parte dos requisitos exigidos para obtenção do título de Doutor em Genética. Área de concentração: Genética.

Orientador: Antônio Roberto Lucena de Araújo

Coorientador: Marcos André Cavalcanti Bezerra

Recife

2022

Catalogação na Fonte:
Bibliotecário Bruno Márcio Gouveia, CRB-4/1788

Lima, Aleide Santos de
Avaliação do impacto isolado e aditivo de marcadores moleculares no desfecho clínico de pacientes adultos com leucemia mieloide aguda / Aleide Santos de Lima. - 2022.

229 f. : il.

Orientador: Antônio Roberto Lucena de Araújo.
Coorientador: Marcos André Cavalcanti Bezerra.
Tese (doutorado) – Universidade Federal de Pernambuco. Centro de Biociências. Programa de Pós-graduação em Genética, Recife, 2022.
Inclui referências, apêndices e anexos.

1. Câncer. 2. Leucemia. 3. Sangue – Doenças. I. Araújo, Antônio Roberto Lucena de (orientador). II. Bezerra, Marcos André Cavalcanti (coorientador). III. Título.

616.99419 CDD (22.ed.) UFPE/CB-2022-078

ALEIDE SANTOS DE MELO LIMA

AVALIAÇÃO DO IMPACTO ISOLADO E ADITIVO DE MARCADORES MOLECULARES NO DESFECHO CLÍNICO DE PACIENTES ADULTOS COM LEUCEMIA MIELOIDE AGUDA

Tese apresentada ao Programa de Pós Graduação em Genética da Universidade Federal de Pernambuco, Centro Acadêmico de Biociências, como requisito para a obtenção do título de Doutora em Genética. Área de concentração: Genética

Aprovado em: 29/04/2022.

Banca Examinadora:

_____

Dr. Antônio Roberto Lucena de Araujo
Universidade Federal de Pernambuco

_____

Dra. Lorena Lobo de Figueiredo Pontes
Faculdade de Medicina da Universidade de São Paulo de Ribeirão Preto

_____

Dr. Douglas Rafaele Almeida Silveira
King's College London

_____

Dr. Felipe Saldanha de Araujo
Universidade de Brasília

_____

Dra. Neide Santos
Universidade Federal de Pernambuco

Dedico este trabalho à minha família cujo apoio e suporte foi imprescindível para sua conclusão e a todos os pacientes que gentilmente aceitaram participar da pesquisa.

AGRADECIMENTOS

Agradeço primeiramente aos pacientes que gentilmente aceitaram participar da pesquisa e que sem eles nada disso teria sido possível. Agradeço a Deus pela vida e saúde que permitiram que pudesse desenvolver este projeto.

Agradeço ao meu orientador, Antônio Roberto Lucena de Araujo, e ao meu coorientador, Marcos André Cavalcante Bezerra, pela oportunidade e confiança depositadas em mim e por todo o conhecimento transmitido. Muito obrigada por todos os ensinamentos que foram importantes para meu crescimento acadêmico.

Agradecimento especial faço a Diego Martins e Juan Coelho por toda orientação, conhecimento transmitido e discussões que engradeceram meu conhecimento e, especialmente, a Juan por toda paciência que teve, tantas vezes gentilmente me respondendo até nos fins de semana! Obrigada por todo o apoio!

Agradeço a Matheus Filgueira por todo apoio científico e pela amizade! Obrigada por sua colaboração tão importante para o desenvolvimento desse trabalho e por todas as leves conversas e amizade!

Agradeço à banca examinadora por gentilmente terem aceitado o meu convite e por toda contribuição dada ao trabalho.

Agradeço ao professor Rafael França do Centro de Pesquisas Aggeu Magalhães por ceder o espaço para desenvolver parte desse estudo e a Cassia Docena e Viviane Carvalho do Núcleo de Plataformas Tecnológicas pelo apoio.

Agradeço aos meus amigos de laboratório e, muitos deles, amigos de vida! Nesses cinco anos, muitas alegrias e dificuldades foram divididas com tantos grandes amigos e pessoas maravilhosas! Flavinha, Marianny, Jéssica Maria, Gabi Arcanjo, Betânia, Amanda, Malu, agradeço a todos vocês por estarem juntos comigo

nessa caminhada, essa jornada foi mais leve graças a vocês! Agradecimento especial faço a essa grande amiga que a ciência me deu, Jéssica Vitória! Obrigada Jessiquinha, sem você, sem o seu apoio, amizade e companheirismo eu não teria conseguido!

Agradecimento especial faço às minhas companheiras de rotina, as “meninas da rotina”, Gabriela, Neyla e Caroline. Obrigada por todo apoio nesses cinco anos, sem a força de vocês eu não teria conseguido dar conta de todas as demandas!

Agradeço aos meus amigos que tantas vezes me deram força nos momentos de dificuldades com essa tese. Agradeço à minha família, meus irmãos Daniel e Gabriel, minha mãe Ivaneide e minha avó Aleide, por todo apoio nos momentos de dificuldades e por todo amor, carinho e amizade que deixaram essa caminhada mais leve. Agradeço ao meu companheiro, Luydson, que com todo seu amor, carinho e paciência me apoiou na realização desse trabalho e, também, com as conversas científicas que engradeceram meu conhecimento! Muito obrigada a todos vocês! Nada disso teria sido possível sem a presença de vocês em minha vida!

Agradeço, por fim, às agências de auxílio financeiro, CNPq e FACEPE, que propiciaram financeiramente a execução do projeto.

“*Que nada nos defina, que nada nos sujeite. Que a liberdade seja nossa própria substância, já que viver é ser livre*.” (de Beauvoir, 1949)

RESUMO

Atualmente, a estratificação de risco do grupo *European LeukemiaNet* (ELN) para leucemia mieloide aguda (LMA) é a mais aplicada na prática clínica, mas alguns autores defendem seu refinamento. Este estudo propõe-se a avaliar o impacto isolado e aditivo de marcadores moleculares no desfecho clínico de pacientes adultos com LMA. Para isso, três coortes foram utilizadas: *The Cancer Genome Atlas* (TCGA) como coorte de desenho e duas coortes de validação, uma externa disponível publicamente (GEO: GSE6891) e uma interna de pacientes tratados em hospitais brasileiros. Nós demonstramos que as mutações no gene *DNMT3A* de maneira isolada ou em cooperação com mutações *NPM1* e *FLT3*-ITD impactam de maneira adversa o desfecho de pacientes LMA. A hiperexpressão de *ID1* foi um preditor negativo para a sobrevida global (SG) na coorte TCGA e na coorte de validação interna. Nós propomos, também, um índice prognóstico baseado em transcriptoma (IPT) a partir do impacto aditivo da expressão dos genes *PDE7B*, *IGF2BP3* e *ST6GALNAC4*. O IPT foi um fator prognóstico independente para a SG, sobrevida livre de doença e de eventos de pacientes com LMA, foi capaz de refinar a classificação ELN2010 e foi validado na coorte externa GSE6891. Entretanto, o IPT não foi validado em uma coorte interna de pacientes da “vida real”. O número de pacientes, o curto tempo de seguimento e as diferentes metodologias de análise de expressão gênica podem ter limitado as análises.

Palavras-chave: leucemia mieloide aguda; prognóstico; mutações; transcriptoma.

ABSTRACT

Currently, the risk stratification for acute myeloid leukemia (AML) of the European LeukemiaNet (ELN) group is the most applied in clinical practice, but your refinement is necessary. This study aims to assess the isolated and additive impact of molecular markers on the outcome of adult patients with acute myeloid leukemia. For this, three cohorts were used: The Cancer Genome Atlas (TCGA) as a training cohort and two validation cohorts, one external publicly available (GEO: GSE6891) and one internal of patients treated in Brazilian hospitals. We demonstrate that mutations in the *DNMT3A* gene alone or in cooperation with *NPM1* and *FLT3*-ITD mutations adversely impact the outcome of AML patients. *ID1* overexpression was a negative predictor of overall survival (OS) in the TCGA cohort and in the internal validation cohort. A transcriptome-based prognostic index (TPI) based on the additive impact of the expression of the *PDE7B*, *IGF2BP3* and *ST6GALNAC4* genes was also proposed. TPI was an independent prognostic factor for OS, disease-free and event-free survival of patients with AML, was able to refine the ELN2010 classification, and was validated in the external cohort GSE6891. However, the TPI has not been validated in an internal cohort of “real-life” patients. The number of patients, the short follow-up time and the different methodologies for analyzing gene expression may have limited the analyses.

Key words: acute myeloid leukemia; prognosis; mutations; transcriptome.

LISTA DE ILUSTRAÇÕES

ARTIGO 2 - THREE-GENE PROGNOSTIC INDEX FOR PREDICTING OUTCOME IN ADULT AML PATIENTS

LISTA DE TABELAS

ARTIGO 1 - PROGNOSTIC IMPLICATIONS OF *ID1* EXPRESSION IN ACUTE MYELOID LEUKEMIA PATIENTS TREATED IN A RESOURCE-CONSTRAINED SETTING

ARTIGO 2 - THREE-GENE PROGNOSTIC INDEX FOR PREDICTING OUTCOME IN ADULT AML PATIENTS

LISTA DE ABREVIATURAS E SIGLAS

Item	Definição
AIC	Do inglês, *Akaike Information Criteria*
AML	Do inglês, *acute myeloid leucemia*
APL	Do inglês, *acute promyelocytic leucemia*
ARCH	Do inglês, *age-related clonal hematopoiesis*
AUC	Area under the curve
bHLH	Do inglês, basic helix-to-helix
BM	Do inglês, *bone marrow*
CBF-LMA	Leucemia mieloide aguda com translocação envolvendo CBF, do inglês, *core binding factor*
CCUS	Do inglês, *clonal cytopenia of unknown significance*
cDNA	DNA complementar
$CEBPA^{bi}$	Do inglês, *biallelic CEBPA-mutated*
CEP	Comitê de Ética em Pesquisa
CHIP	Do inglês, *clonal hematopoiesis of indeterminate potential*
CI	Do inglês, *confidence interval*
CIR	Do inglês, *cumulative incidence of relapse*
CNS	Conselho Nacional de Saúde
CR	Do inglês, *complete remission*
CTE	Células tronco embrionárias
CTH	Célula tronco hematopoiética
CTL	Célula tronco leucêmica
DEPC	Dietilpirocarbonato
DFS	Do inglês, *disease-free suvival*

DN	Distribuição normal
DNA	Do inglês, *deoxyribonucleotide acid*
DNMT3A	Do inglês, *DNA methyltransferase 3 alpha*
EFS	Do inglês, *event-free suvival*
ELN	Do inglês, *European LeukemiaNet*
ESC	Do inglês, *embryonic stem cell*
EUA	Estados Unidos da América
FAB	Franco-Americano-Britânico
Fem	Feminino
FLT3	Do inglês, *FMS-like tyrosine kinase 3*
GEO	Do inglês, *Gene Expression Omnibus*
GSEA	Do inglês, *Gene Set Enrichment Analysis*
HCP	Hospital do Câncer de Pernambuco
HDI	Do inglês, *Human Development Index*
HEMOPE	Hemocentro de Pernambuco
HIC	Do inglês, *high-income countries*
HLH	Do inglês, *helix-to-helix*
HR	Do inglês, *hazard ratio*
HSC	Do inglês, *hematopoietic stem cell*
HSCT	Do inglês, *hematopoietic stem cell transplantation*
ICAL	Do inglês, *International Consortium on Acute Leukemia*
IGF2BP	Do inglês *insulin-like growth factor 2 mRNA-binding proteins*
IHDI	Do inglês, I*nequality-adjusted Human Development Index*
IMIP	Instituto de Medicina Integral Professor Fernando Figueira
INCA	Instituto Nacional do Câncer

IPT	Índice de prognóstico baseado em transcriptoma
ITD	Do inglês, *internal tandem duplication*
KS	Kolmogorov-smirnov
LLA	Leucemia linfoide aguda
LLA-B	Leucemia linfoide aguda B
LLC	Leucemia linfoide crônica
LMA	leucemia mieloide aguda
LMA-CN	Leucemia mieloide aguda com cariótipo normal
LMC	Leucemia mieloide crônica
LMIC	Do inglês, *low- and middle-income countries*
LMIC	Do inglês, *low- and middle- income countries*
LPA	Leucemia promielocítica aguda
LPA	Leucemia promielocitica aguda
LSC	Do inglês, *leukemic stem cell*
Masc	Masculino
miRNA	Do inglês, *micro ribonucleotide acid sequencing*
MO	Medula óssea
MRC	Do inglês, *Medical Reserch Council*
NMPC	Neoplasia mieloproliferativa crônica
NPM1	Do inglês, *nucleophosmin*
OMS	Organização Mundial de Saúde
OS	Do inglês, *overall survival*
PH	Do inglês, *proportional hazards*
PLT	Do inglês, *platelets*
qPCR	Do inglês, *quantitative polymerase chain reaction*

RC	Remissão completa
RNA-seq	Do inglês, Ribonucleotide acid sequencing
RPKM	do inglês, *reads per kilo base per million mapped*
RT-qPCR	Do inglês *reverse transcription quantitative polymerase chain reaction*
SG	Sobrevida global
SISNEP	Sistema Nacional de Informação sobre Ética envolvendo Seres Humanos
SLD	Sobrevida livre de doença
SMD	Síndrome mielodisplásica
SP	Sangue periférico
TAE	Tampão tris, acetate, EDTA
TCGA	Do inglês, *The Cancer Genome Atlas*
TCLE	Termo de consentimento livre e esclarecido
TCTH	Transplante de células tronco hematopoiéticas
TF	Antígeno Thomsen-Friedenreich
TF	Do inglês, *Thomsen-Friedenreich*
TPI	Do inglês, *transcriptome-based prognostic index*
UNICAMP	Universidade de Campinas
VAF	Do inglês, *variant allele frequency*
WBC	Do inglês, *white blood cells*
WHO	Do inglês, *World Health Organization*

LISTA DE SÍMBOLOS

α	Alfa
β	Beta

SUMÁRIO

1 INTRODUÇÃO

A leucemia mieloide aguda (LMA) é a leucemia mais frequente em adultos e apresenta alto risco de recaída e elevada taxa de mortalidade. Alterações genéticas sequencialmente adquiridas pelas células tronco e progenitoras hematopoiéticas perturbam a hematopoese por bloqueio da diferenciação, proliferação descontrolada e inibição da apoptose. Com isso, células blásticas imaturas ou parcialmente diferenciadas com capacidade de auto-renovação se acumulam, primeiramente, na medula óssea e então infiltram o sangue periférico e órgãos, impedindo suas funções. Apesar de apresentar sintomas e implicações clínicas similares, a LMA é uma doença heterogênea apresentando um amplo espectro de subtipos com diferenças citogenéticas e moleculares e no desfecho clínico.

As três últimas décadas foram marcadas por um grande esforço para se identificar marcadores biológicos capazes de fornecer informação não redundante para refinamento prognóstico de pacientes com LMA. Consequentemente, uma quantidade notável de alterações citogenéticas e moleculares foram identificadas e usadas para definir grupos de pacientes com prognósticos distintos. Nesse contexto, o grupo colaborativo *European LeukemiaNet* (ELN) propôs uma estratificação de risco para os pacientes com LMA, baseada no impacto de alterações citogenéticas e mutações somáticas bem definidas (*NPM1*, *FLT3*-ITD, *CEBPA*, *RUNX1*, *ASXL1* e *TP53*) no desfecho clínico desses pacientes. Essa proposta corresponde a mais amplamente aceita e aplicada na prática clínica. No entanto, ainda existe uma grande heterogeneidade na resposta ao tratamento dos pacientes em cada um

dos grupos de risco, principalmente no grupo intermediário, o que chama a atenção para o refinamento dessa proposta.

Com o avanço no uso das tecnologias de sequenciamento de nova geração uma ampla gama de novos marcadores e de perfis de cooperação mutacional tem sido descritos e podem contribuir para o aperfeiçoamento dessa classificação. Nesse contexto, foram propostos diversos painéis de estratificação de risco alternativos baseados em novas mutações somáticas e em cooperações mutacionais ou até na expressão alterada de genes. Apesar do uso dos dados de expressão gênica como ferramenta prognóstica ser complexo, pois trata-se de variável contínua, alguns marcadores moleculares como a hiperexpressão dos genes *EVI1* e *BAALC*, são utilizados em painéis de biomarcadores prognósticos ao diagnóstico de forma padronizada e foram validados por diversos centros. Nesse contexto, é possível que biomarcadores prognósticos baseados na expressão gênica possam ser utilizados na rotina diagnóstica com a devida padronização.

Ademais, a maioria dos trabalhos sobre marcadores moleculares prognósticos são realizados em coortes de pacientes de ensaios clínicos controlados e são concentrados em populações dos EUA e Europa, o que significa dizer que a aplicabilidade prática desses estudos é completamente desconhecida para a maioria dos pacientes do mundo. Diante disso, este trabalho tem como objetivo avaliar o impacto isolado e aditivo de marcadores moleculares no desfecho clínico de pacientes adultos com LMA, na tentativa de oferecer informação adicional para o refinamento da estratificação de risco dos pacientes, e validar os achados em uma coorte de pacientes da “vida real”.

2 REVISÃO DA LITERATURA

2.1 LEUCEMIA MIELOIDE AGUDA

A leucemia mielóide aguda (LMA) constitui um grupo heterogêneo de doenças hematológicas caracterizadas pelo acúmulo de células progenitoras hematopoiéticas com alterações genéticas somáticas adquiridas. Tais alterações afetam os mecanismos normais de auto renovação, proliferação e diferenciação, tornando as células progenitoras anormais incapazes de se diferenciarem e com vantagem de crescimento e sobrevivência em relação às normais. Consequentemente, estas células progenitoras anormais se acumulam na medula óssea (MO), levando a falha da MO e morte. O envolvimento do sangue periférico é frequente com um quadro de anemia, neutropenia, plaquetopenia e presença de mieloblastos. A infiltração de órgãos, como fígado, baço e pele, por sua vez, podem ocorrer, mas são infrequentes (JACOMO *et al*., 2017; ESTEY, 2018).

A LMA é a leucemia mais frequente em adultos. Dados epidemiológicos dos Estados Unidos (EUA) mostram que a incidência de LMA entre os adultos tem sido maior do que qualquer um dos outros 3 subtipos de leucemia (leucemia linfóide aguda, leucemia mieloide crônica e leucemia linfoide crônica) até 2017 (Song *et al*., 2018; Shallis *et al*., 2019) (Figura 1). Em 2018, o banco de dados GLOBOCAN produzido pela Agência Internacional de Pesquisa sobre o Câncer estimou 437.033 novos casos de leucemias e 309.006 mortes em todo o mundo (Bray *et al*., 2018). Tem sido estimado que aproximadamente 21.450 adultos serão diagnosticados com LMA em 2019, nos EUA, e que de todos os subtipos de leucemia, a LMA é a maior causa de óbito (Figura 1), apresentando a quinta pior sobrevida global em 5 anos por tipo de câncer nos

EUA. Países europeus têm registrado taxas de incidência semelhante aos EUA (Song *et al*., 2018; Shallis *et al*., 2019).

No Brasil, a incidência de LMA é desconhecida, uma vez que o Instituto Nacional do Câncer (INCA) relata apenas dados referentes às leucemias em geral. Para o triênio 2020-2022, o INCA estima 5.920 casos novos de leucemia em homens e 4.890 em mulheres para cada ano. Esses valores correspondem a um risco estimado de 5,67 casos novos a cada 100 mil homens e 4,56 casos novos para cada 100 mil mulheres, ocupando a nona e a décima posições entre os tipos de câncer mais incidentes entre homens e mulheres, respectivamente. O estado de Pernambuco tem uma taxa estimada de 5,33 casos para cada 100 mil homens e 3,88 casos para cada 100 mil mulheres (INCA, 2019).

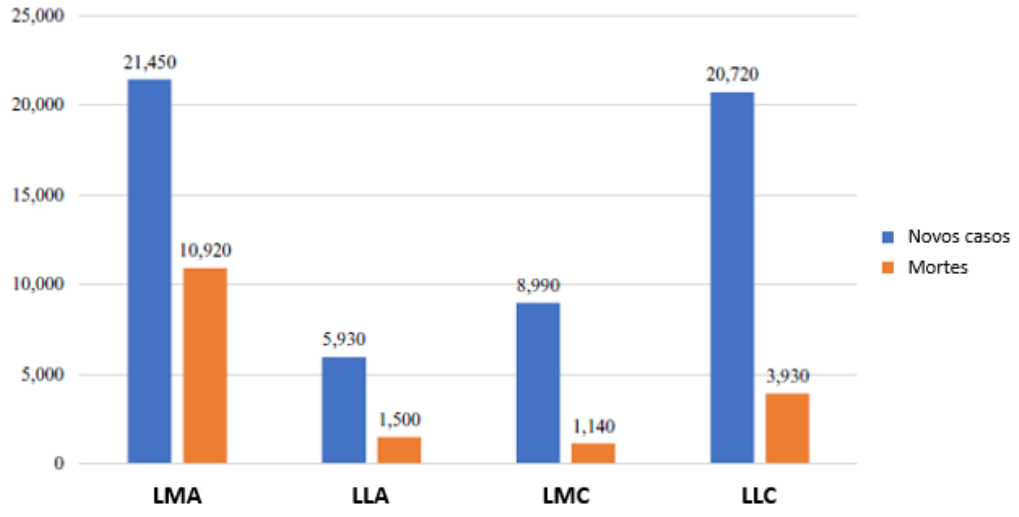


Figura 1. Número estimado de novos casos e de morte dos subtipos de leucemia nos Estados Unidos para 2019. Legenda: LMA: leucemia mieloide aguda; LLA: leucemia linfoide aguda; LMC: leucemia mieloide crônica; LLC: leucemia linfoide crônica (Fonte: Shallis *et al*., 2019 modificado).

2.1.1 Fisiopatologia da LMA

Carcinogênese é um processo gradual em que a aquisição de alterações genéticas, tais como variação no número de cópias, rearranjos gênicos, mutações em genes e alterações epigenéticas, dirige a evolução das células somáticas a um estado pré-neoplásico e, eventualmente, até clones neoplásicos que apresentam distúrbios da proliferação celular, bloqueio da diferenciação e apoptose e estratégias de escape imune (Hartmann e Metzeler, 2019).

Na hematopoese, divisões celulares assimétricas das células tronco hematopoéticas (CTH) resultam na manutenção de um pool de CTH, bem como na formação de células mais diferenciadas. A expansão e diferenciação das CTH para progenitores de linhagens específicas e, por fim, para células maduras ocorre dentro do microambiente da medula óssea sob a influência de múltiplos fatores reguladores: citocinas pleiotrópicas e linhagem específica, fatores de transcrição, o próprio microambiente da medula óssea e numerosas proteínas sinalizadoras. Clones neoplásicos podem se originar de CTHs pluripotentes ou de células progenitoras mieloides mais comprometidas a partir de alterações genéticas e epigenéticas, e o fenótipo resultante de malignidade pode ser determinado por essas alterações, bem como pela célula de origem (Hartmann e Metzeler, 2019; Zeisig *et al*., 2021).

Além dos fatores genéticos, estudos epidemiológicos sugerem que fatores ambientais e ocupacionais também estão envolvidos na etiologia da LMA, embora esse processo não seja inteiramente esclarecido e muitos casos

de LMA permaneçam sem uma etiologia clara (Belson *et al*., 2007; Linet *et al*., 2018; Shallis *et al*., 2019).

A substância química mais consistentemente identificada como leucemogênica é o benzeno, mas outras substâncias como formaldeído e pesticidas também foram associadas a risco para LMA em trabalhadores expostos (Linet *et al*., 2018). O tratamento de outras neoplasias com agentes quimioterápicos e radioterapia está associado a um risco aumentado para a LMA e, nesse caso, levam ao desenvolvimento de um subtipo secundário, de pior prognóstico, classificado, segundo a Organização Mundial de Saúde (OMS) em 2016, no subgrupo de “neoplasias mieloides relacionadas à terapia” (Arber *et al*., 2018). Obesidade e tabagismo também são considerados como fatores de risco para o desenvolvimento da LMA associados ao estilo de vida. Além disso, a predisposição hereditária pela presença de mutações germinativas ou alterações cromossômicas congênitas é descrita. Mais recentemente, a presença de hematopoese clonal tem sido descrita como um fator de risco para a LMA (Boddu *et al*., 2017; Linet *et al*., 2018; Shallis *et al*., 2019).

Mais de 300 translocações e/ou inversões cromossômicas foram descritas e estão envolvidas na fisiopatologia dessa doença (Velloso *et al*., 2011; Dohner *et al*., 2017). Translocações cromossômicas na LMA dão origem a genes de fusão no local das quebras cromossômicas, com isso, os éxons dos genes envolvidos se tornam justapostos e formam um único gene de fusão, que dá origem a uma nova proteína híbrida com características únicas. Os alvos mais frequentes desses eventos na LMA são genes que codificam fatores de transcrição importantes para o desenvolvimento hematopoiético normal,

que, depois da recombinação, dão origem a genes híbridos que codificam proteínas de fusão com funções alteradas, resultando em perda da função e bloqueio da diferenciação (Alcalay *et al*., 2001; Faber *et al*., 2016; Al-Harbi *et al*., 2020).

Os parceiros desses genes envolvidos nas translocações da LMA codificam para proteínas que são muito mais heterogêneas em termos de função. A contribuição desses parceiros para o potencial leucemogênico das proteínas de fusão ainda não é tão claro. Entretanto, sua presença parece ser fundamental para ao menos alguns dos efeitos biológicos dessas proteínas. Muitos deles têm funções fisiológicas associadas com a regulação do ciclo celular e apoptose (Alcalay *et al*., 2001; Yang *et al*., 2016). Essas proteínas de fusão têm se mostrado necessárias, mas não suficientes para a leucemogênese, sugerindo que atuem juntamente com outras classes de mutações para culminar na transformação leucêmica (Hyde *et al*., 2015; Faber *et al*., 2016; Madan *et al*., 2018).

Há aproximadamente duas décadas, Gilliland e Griffin apresentaram um modelo para o processo leucemogênico da LMA, chamado de *two hit model* (Figura 2). Segundo este modelo, o desenvolvimento da LMA é considerado um processo multipasso que requer a colaboração de pelo menos duas classes de mutações. Este modelo compreende as mutações de classe I, que ativam vias de transdução de sinal e conferem vantagem proliferativa às células hematopoiéticas, e as mutações de classe II, que afetam fatores de transcrição e primariamente impedem a diferenciação hematopoiética (Gilliland e Griffin, 2002).

Mutações que afetam os genes *FLT3*, *KIT* e os da família *RAS* são consideradas mutações de classe I, enquanto que alterações cromossômicas como t(8;21), inv(16) e t(15;17), que geram, respectivamente, os rearranjos gênicos *RUNX1*/*RUNX1T1*, *CBFB*/*MYH11* e *PML*/*RARA*, juntamente com as mutações nos genes *RUNX1*, *CEBPA* e *MLL* são consideradas de classe II. De acordo com esse modelo, o acúmulo de mutações de classe I e II culminaria no desenvolvimento de progenitores hematopoéticos transformados, capazes de propagar o fenótipo leucêmico (Gilliland e Griffin, 2002).

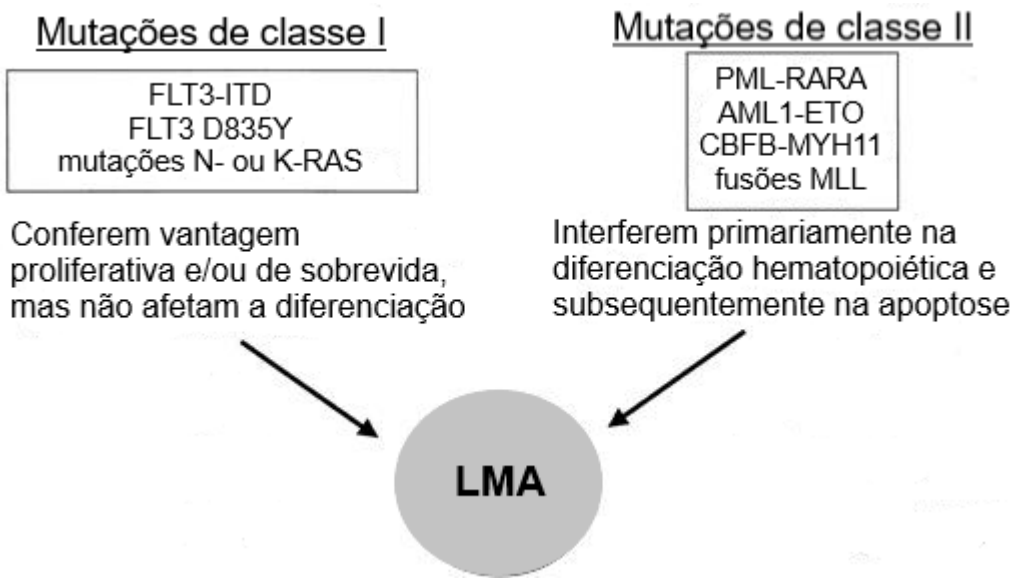


Figura 2. Diagrama esquemático do modelo *Two Hit* de cooperação mutacional (Fonte: Gilliland e Griffin, 2002 modificado)

Nos últimos anos, entretanto, tem sido identificado mutações em genes que não obedecem a nenhuma das duas classes, sugerindo que as mutações de classe I e II são apenas uma parte de um quadro mais complexo e, portanto, o modelo “two-hit” não é mais adequado (Conway O’brien *et al*., 2014). Em 2013, com o uso de tecnologias de sequenciamento de nova geração, o

consórcio *The Cancer Genome Atlas* (TCGA) revelou que a LMA é uma doença heterogênea com quase 2.000 genes apresentando mutações somáticas analisadas em 200 pacientes com LMA. Diante desse panorama, foi proposto uma classificação das mutações condutoras da LMA em nove categorias de genes funcionalmente relacionados e que são certamente relevantes para a patogênese da LMA: fusões gênicas de fatores de transcrição, gene codificante da nucleofosmina (*NPM1*), genes supressores tumorais, genes relacionados à metilação do DNA, genes de vias de sinalização, genes modificadores da cromatina, genes de fatores de transcrição mieloide, genes do complexo da coesina e genes do complexo do spliceossomo (Cancer Genome Atlas, 2013).

Padrões de cooperação e exclusividade mútua entre essas mutações condutoras têm sido relatados e sugerem forte relação entre esses genes e categorias, sendo essa interrelação complexa de eventos genéticos importante para a patogênese da LMA (Metzeler *et al*., 2016; Tyner *et al*., 2018).

Um padrão importante de exclusividade mútua relatado é o que ocorre entre algumas mutações somáticas comuns na LMA (por exemplo, *DNMT3A*, *CEPBA*, *IDH1/2*, e *RUNX1*) e os genes de fusão envolvendo fatores de transcrição (Cancer Genome Atlas, 2013). Como relatado anteriormente, esses genes de fusão envolvendo fatores de transcrição tem se mostrado relevantes para a iniciação da doença em ratos (Hyde *et al*., 2015; Faber *et al*., 2016). Dessa forma, essa exclusividade mútua sugere que essas mutações podem ter funções na iniciação da LMA de maneira semelhante às funções dos genes de fusão (Cancer Genome Atlas, 2013).

Entre os padrões de cooperação mutacional relatados nas amostras de LMA, a co-ocorrência das mutações *FLT3*-ITD, *DNMT3A* e *NPM1* tem sido a

mais frequente e importante cooperação entre três genes. Amostras apresentando mutações nesses três genes apresentam características clínicas típicas, sugerindo que amostras triplo-mutadas representam um novo subtipo da LMA (Guryanova *et al*., 2016; Papaemmanuil *et al*., 2016; Tyner *et al*., 2018; Garg *et al*., 2019).

Outro padrão de cooperação mutacional é o que ocorre entre mutações em genes modificadores epigenéticos, incluindo *DNMT3A*, *IDH1/2*, *ASXL1* e/ou *TET2* (Romer-Seibert e Meyer, 2021). Um estudo utilizando um modelo animal apresentando duplo *knockout* em *DNMT3A* e *TET2* observou que a perda desses dois modificadores epigenéticos levou à expansão dos compartimentos das células tronco e progenitoras hematopoiéticas e redução significativa da sobrevivência em comparação com animais com apenas uma deleção, mostrando assim a natureza cooperativa dessas mutações (Zhang *et al*., 2016).

Embora exista uma hipótese estocástica para a leucemogênese, grande número de evidências sustenta a hipótese de que a LMA apresenta uma organização hierárquica clonal, de maneira semelhante à hematopoese normal. Essa organização hierárquica se reflete na sua heterogeneidade celular, no que diz respeito à morfologia, marcadores de superfície celular ou características funcionais. A LMA é iniciada e sustentada por uma pequena população de células tronco leucêmicas (CTL) com capacidade de auto-renovação, a qual produz uma progênie de progenitores mais maduros. Esses progenitores, apresentam alta taxa proliferativa e limitada capacidade de diferenciação e assim dão origem a uma população de blastos leucêmicos que constituem a maioria das células leucêmicas na medula óssea e sangue periférico dos pacientes LMA (Martelli *et al*., 2009; Nepstad *et al*., 2020).

A CTL pode surgir a partir de um progenitor hematopoiético comprometido, como é o caso da LMA com presença do gene de fusão *MLL-AF9*, ou de uma CTH, como é o caso da LMA com mutação no gene *DNMT3A* (Yang *et al*., 2016; Garg *et al*., 2019). A evolução dessa doença é dinâmica e ordenada seguindo a aquisição hierárquica de mutações condutoras (Loberg *et al*., 2019). Essa afirmativa é reforçada pelo achado de que, frequentemente, nesta doença, tem-se uma mistura heterogênea de clones leucêmicos, com mais da metade dos tumores apresentando um clone fundador, que é o clone que apresenta mutação com maior frequência de alelo variante (VAF – do inglês variant allele frequency), e ao menos um subclone com mutações adicionais (Cancer Genome Atlas, 2013; Bullinger *et al*., 2017).

Mutações com VAF próximo de 50% são mutações provavelmente adquiridas precocemente e, potencialmente, iniciam os eventos durante a leucemogênese. Entre essas mutações, estão aquelas que ocorrem em modificadores epigenéticos, como os genes *DNMT3A*, *ASXL1*, *IDH1/2* e *TET2*. Essas mutações são encontradas no clone fundador e quase sempre associadas a outras mutações. Mutações ativadoras em genes de vias de sinalização, tais como *FLT3*, *KIT*, *NRAS* e *KRAS*, por sua vez, comumente têm VAF menores e, aparentemente, ocorrem mais tarde durante a evolução do clone maligno. Esses conceitos são suportados pela análise de amostras pareadas no diagnóstico e na remissão. Mutações em *DNMT3A*, *ASXL1*, *IDH1/2* e *TET2* permanecem sendo detectadas nas amostras de remissão, enquanto que mutações em *FLT3*, *KIT, KRAS* e outras são perdidas com a remissão da doença (Metzeler *et al*., 2016; Papaemmanuil *et al*., 2016; Yang *et al*., 2016; Bullinger *et al*., 2017).

Mais recentemente, os estudos de sequenciamento do genoma de célula única (sc-DNAseq, do inglês *single-cell DNA-sequencing*), corroboram a afirmação de que cada caso de LMA representa uma combinação de populações clonais geneticamente distintas, resultando em profunda heterogeneidade intratumoral. Demonstraram, ainda, que a heterogeneidade intratumor é ainda mais alta do que anteriormente se pensava, de maneira que a maioria dos pacientes com LMA apresentam entre três e 13 clones únicos com três a sete mutações genéticas diferentes, podendo, em casos extremos, chegar até a 30 clones diferentes. Apesar desse grande número de clones por paciente, geralmente predominam um a dois clones por tumor (Potter *et al*., 2019; Miles *et al*., 2020; Morita *et al*., 2020).

Esses clones apresentam perfiz mutacionais distintos e podem ser classificados em: clone fundador, que dá origem a leucemogênese e apresenta a mutação de iniciação que está presente na grande maioria das células; subclone, que são clones que apresentam outras mutações além da mutação iniciadora; e clone dominante, que é aquele que predomina na população tumoral. Na evolução clonal, clones com mutações em um gene específico ou co-ocorrência de mutações específicas podem conferir vantagem de um clone sobre outros, mas a contribuição genética para clones dominantes, provavelmente, também é influenciada pelo contexto leucêmico, como diversidade clonal e seleção terapêutica (Romer-Seibert e Meyer, 2021).

Algumas dessas mutações presentes na LMA são provavelmente eventos *background* que ocorrem nas células tronco ou progenitores hematopoiéticas antes da iniciação da leucemogênese, como um evento pré-leucêmico (Cancer Genome Atlas, 2013). A análise de dezenas de milhares de

amostras de sangue periférico de indivíduos sem desordens hematológicas, por tecnologias de sequenciamento de nova geração, demonstrou que mutações em genes previamente associados somente com malignidades mieloides podemser detectadas em indivíduos saudáveis sem um diagnóstico de neoplasia hematológica (Jaiswal *et al*., 2014; Genovese *et al*., 2014; Xie *et al*., 2014). Essas mutações somáticas ocorrem em um clone de célula tronco ou progenitora hematopoiética e conferem a ele vantagem proliferativa. Além disso, o clone mutado mantém sua capacidade de diferenciação em múltiplas linhagens, contribuindo substancialmente para a hematopoese fenotipicamente normal, o que é chamado de hematopoese clonal (Bowman *et al*., 2018; Hartmann e Metzeler, 2019).

A hematopoese clonal e sua expansão (geralmente com VAF superior a 2%), sem evidência de displasia ou citopenia, é definida como hematopoese clonal de potencial indeterminado (CHIP – do inglês, *clonal hematopoiesis of indeterminate potential*) e sua frequência é dependente da idade (Shallis *et al*., 2019). Mutações somáticas foram raramente (aproxidamente 1%) identificadas em indivíduos com menos de 50 anos de idade, enquanto isso, foram identificadas em aproximadamente 10% das pessoas com mais de 65 anos e em mais de 18% daqueles com idade superior a 90 anos (Jaiswal *et al*., 2014; Genovese *et al*., 2014; Xie *et al*., 2014).

Indivíduos sem alteração hematológica diagnosticada e com hematopoese clonal têm um risco, aproximadamente, 10 vezes maior de desenvolver uma malignidade hematológica, em relação àquele sem hematopoese clonal. Por sua vez, indivíduos apresentando grandes clones, com VAF superior a 10%, podem ter seu risco aumentado em 50 vezes para

desenvolvimento de malignidade hematológica (Jaiswal *et al*., 2014; Genovese *et al*., 2014; Xie *et al*., 2014).

Existe um espectro contínuo de alterações e fenótipos clínicos associados com a hematopoese clonal, em que o paciente pode progredir ao longo do tempo devido ao processo evolucionário de mutações e à seleção dos clones com maior vantagem adaptativa (Figura 3). De um lado desse espectro estão os indivíduos saudáveis com CHIP, também conhecido como hematopoese clonal relacionado à idade (ARCH – do inglês, *age-related clonal hematopoiesis*), que apresentam contagens e morfologia das células da medula óssea e do sangue periférico normais. Seguindo o espectro, estão os pacientes que apresentam mutações semelhantes a CHIP mas que já desenvolveram citopenias, são o fenótipo chamado citopenia clonal de significado desconhecido (CCUS – do inglês, *clonal cytopenia of unknown significance*) e que podem então progredir para a síndrome mielodisplásica (SMD) e/ou LMA francas (Hartmann e Metzeler, 2019).

O risco de progressão de CCUS para SMD ou LMA é provavelmente maior do que a progressão a partir de CHIP, entretanto alguns desses pacientes podem nunca progredir para essas malignidades hematológicas (Hartmann e Metzeler, 2019). Como essas alterações formam um espectro de condições relacionadas, é possível que muitos dos pacientes LMA classificados como *de novo* sejam, na verdade, uma evolução de uma hematopoese clonal pré-existente ou de uma SMD não diagnosticada (Metzeler *et al*., 2016; Estey. 2018).

Mutações detectadas em pacientes com hematopoese clonal afetam mais comumente modificadores epigenéticos, tais como *DNMT3A*, *ASXL1* e

TET2, mas também outras classes de genes como *JAK2*, *SRF3B1*, *SRSF2* e *TP53* (Hartmann e Metzeler, 2019; Loberg *et al*., 2019). O gene mais frequentemente mutado é o *DNMT3A*, detectado em 60-70% dos pacientes com hematopoese clonal e é mais comumente encontrado associado a outras mutações. Mutações em *TET2* e *ASXL1* seguem o *DNMT3A* em ordem de frequência, enquanto o *TP53* é o quarto gene mais mutado na hematopoese clonal. Além disso, nenhum paciente foi encontrado para ter hematopoese clonal com mutações em *FLT3* ou *NPM1*, o que é consistente com a afirmação que tais mutações são adquiridas mais tardiamente na evolução do clone leucêmico (Genovese *et al*., 2014; Jaiswal *et al*., 2014; Desai *et al*., 2018).

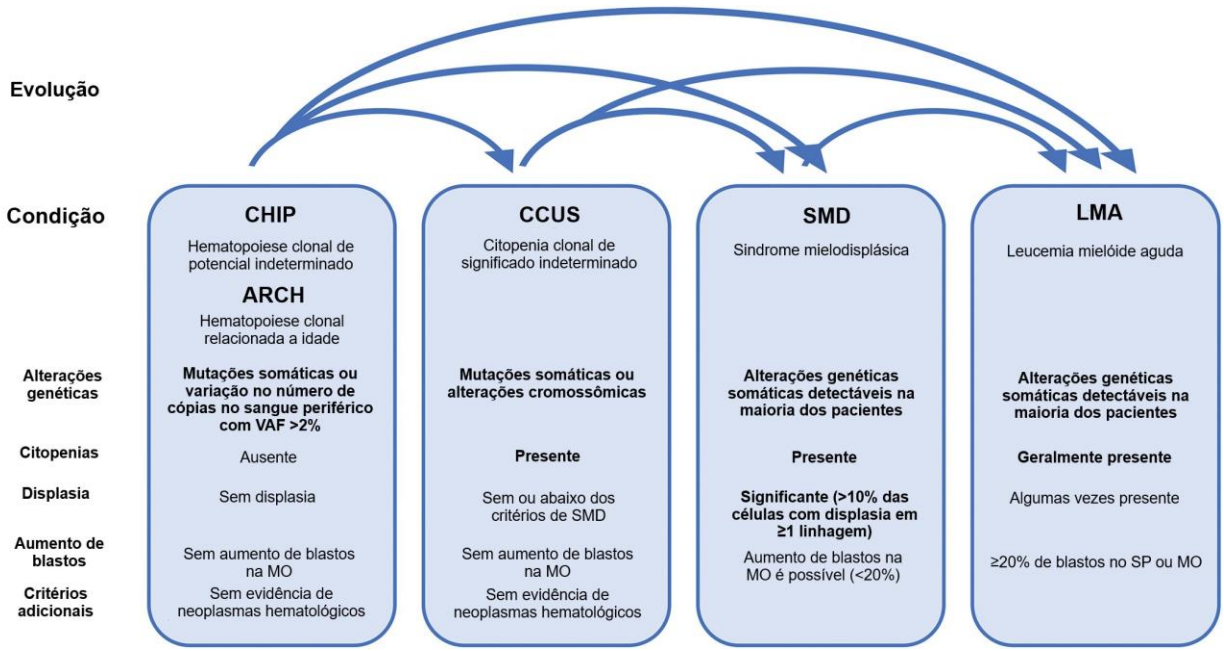


Figura 3: Diagrama esquemático das etapas da evolução da hematopoese clonal. (Fonte: Hartmann e Metzeler, 2019 modificado)

2.1.2 Diagnóstico e classificação da LMA

O diagnóstico da LMA inicia a partir da suspeita clínica com a observação do quadro clínico agudo, que se caracteriza por anemia, sangramento, geralmente do tipo purpúrico devido à plaquetopenia, e febre às vezes sem foco infeccioso, devido à neutropenia. Pode haver infiltração leucêmica em outros órgãos que geralmente causam hepatomegalia, esplenomegalia, hipertrofia de gengivas e dor óssea e, menos frequentemente, infiltração de pele e comprometimento do sistema nervoso central (Estey, 2018; Narayanan e Weinberg, 2020).

Para concluir o diagnóstico e classificação da LMA é necessário a avaliação morfológica da medula óssea e do sangue periférico, a análise de expressão dos marcadores citoplasmáticos e de superfície celular por meio da citometria de fluxo, a avaliação do cariótipo por meio de técnicas de citogenética convencional e, mais recentemente, o *screening* para alterações genéticas moleculares (Dohner *et al*., 2015). Somente a integração de todos esses métodos permite a compreensão e caracterização complementar de cada caso, o que é pré-requisito para o diagnóstico e manejo adequado da LMA (Haferlach e Schmidts, 2020).

No hemograma, ao diagnóstico, é frequente encontrar anemia, plaquetopenia, neutropenia e presença de mieloblastos (Narayanan e Weinberg, 2020). O critério da OMS de 2016 para o diagnóstico da LMA é a contagem de pelo menos 20% de mieloblastos na medula óssea ou sangue periférico. A linhagem mieloide dos blastos é definida por citometria de fluxo, a partir dos marcadores celulares de superfície e citoplasmáticos. Exceções ao critério de ponto de corte de ≥20% de mieloblastos são os casos da LMA com

t(8;21)(q22;q22.1), que corresponde a fusão dos genes *RUNX1-RUNX1T1*, da LMA com inv(16)(p13.1q22) ou t(16;16)(p13.1;q22), que corresponde a fusão dos genes *CBFB-MYH11* e da leucemia promielocítica aguda (LPA). Nesses casos o diagnóstico da LMA é independente da porcentagem de blastos no SP ou MO, no qual apenas a detecção da alteração genética subjacente é suficiente (Arber *et al*., 2016).

A LMA representa um grupo heterogêneo de leucemias que diferem quanto à sua biologia, curso clínico e prognóstico, o que motiva o estabelecimento de uma classificação (Dohner *et al*., 2017). A primeira classificação foi proposta em 1976 pelo grupo cooperativo Franco-Americano-Britânico (FAB), que categorizou seis subtipos de LMA (M1 a M6), baseada estritamente nos aspectos morfológicos e citoquímicos do clone leucêmico (Bennett *et al*., 1976). Essa classificação foi revisada em 1985 pela FAB, dando origem a uma classificação que passou a incluir dois novos subtipos, M0 e M7, aos seis subtipos existentes, cujo diagnóstico passou a incluir o uso de marcadores imunofenotípicos. Essa edição da FAB leva em consideração apenas as características morfológicas, citoquímicas e imunofenotípicas do clone leucêmico, objetivando a determinação da linhagem e do grau de maturação das células blásticas (Bennett *et al*., 1985, 1991).

Apesar de sua importância, a classificação FAB tornou-se obsoleta por não incorporar os achados citogenéticos e moleculares relevantes à fisiopatologia da doença, ficando limitada no significado biológico, prognóstico e terapêutico. Devido a isso, em 2001, a OMS divulgou uma classificação para as LMA que passava a incorporar os achados citogenéticos e, dessa forma, introduziu uma correlação prognóstica importante (Heerema-Mckenney e Arber,

2009). Nesta classificação, as LMA foram divididas em quatro categorias: LMA com anormalidades genéticas repetitivas, LMA com displasia de múltiplas linhagens, LMA associada a tratamento prévio e LMA não categorizada nos ítens anteriores, esta última que se baseou na classificação FAB (Vardiman *et al*., 2002).

Esse sistema logo se tornou insatisfatório, pois em cerca de 50% dos casos não é possível identificar nenhuma anormalidade citogenética, sendo esse grupo definido como citogeneticamente normal (LMA-CN) (Heerema-Mckenney e Arber, 2009). Neste grupo, são identificadas uma série de anormalidades genéticas submicroscópicas, como mutações ou alterações de expressão em alguns genes, que estão associados à leucemogênese e à resposta clínica ao tratamento. Esses marcadores moleculares, juntamente com as alterações cromossômicas, ajudam a caracterizar melhor os casos de LMA e refinam a estratificação de risco dos pacientes. (Dohner *et al*., 2017).

Diante disso, a OMS revisou a classificação de 2001 e, em 2008, lançou um novo sistema. Essa classificação destacou a importância de se associar os achados citogenéticos aos moleculares no diagnóstico e categorização dos subtipos, incluindo duas entidades provisórias caracterizadas pela presença de mutações genéticas (LMA com mutação no gene *NPM1* e LMA com mutação no gene *CEBPA*). Além disso, ampliou o número de subtipos com translocações cromossômicas recorrentes, refinou o diagnóstico de LMA associado à mielodisplasia e passou a classificar separadamente as proliferações mielóides relacionadas à Síndrome de Down (Vardiman *et al*., 2009; Dohner *et al*., 2010).

Desde então, com a evolução dos estudos envolvendo análise de expressão gênica e sequenciamento de nova geração, houve numerosos avanços na identificação de biomarcadores únicos associados com alguns neoplasmas e leucemias agudas (Marcucci *et al*., 2008; Patel *et al*., 2012; Cancer Genome Atlas, 2013; Papaemmanuil *et al*., 2016). Esses novos achados permitiram significativamente melhorar os critérios diagnósticos, bem como, a relevância prognóstica das entidades incluídas na classificação de 2008, sugerindo, ainda, a inclusão de novas entidades. Dessa forma, essa classificação foi revisada em 2016, correspondendo a atualmente utilizada na prática clínica (Quadro 1), e tenta incorporar novos dados clínicos, prognósticos, morfológicos, imunofenotípicos e genéticos à classificação anterior (Arber *et al*., 2016).

LMA com anormalidades genéticas recorrentes
LMA com t(8;21)(q22;q22.1); *RUNX1-RUNX1T1*
LMA com inv(6)(p13.1q22) ou t(16;16)(p13.1;q22); *CBFB-MYH11*
LPA com t(15;17)(q22;q12); *PML-RARA*
LMA com t(9;11)(p21.3;q23.3); *MLLT3-KMT2A*
LMA com t(6;9)(p23;q34.1); *DEK-NUP214*
LMA com inv(3)(q21.3q26.2) ou t(3;3)(q21.3;q26.2); GATA2,MECOM(EVI*1)*
LMA (megacarioblástica) com t(1;22)(p13.3;q13.3); *RBM15-MLK1*
LMA com mutação no *NPM1*
LMA com mutação bialélica do *CEBPA*
Entidade provisória: LMA com *BCR-ABL1*
Entidade provisória: LMA com mutação no *RUNX1*
LMA com alterações relacionadas a mielodisplasia
Neoplasias mieloides relacionadas a terapia
LMA não categorizada nos itens anteriores
LMA com diferenciação minima
LMA sem maturação
LMA com maturação
Leucemia mielomonocítica aguda
Leucemia monoblástica/monocítica aguda

Leucemia eritroide pura
Leucemia megacarioblástica aguda
Leucemia basofílica aguda
Pan-mielose aguda com mielofibrose
Sarcoma mieloide
Proliferações mieloides relacionadas a síndrome de Down
Mielopoese anormal transiente
Leucemia mieloide associada com síndrome de Down
Neoplasma blástico de células dendríticas plasmocitóides

Quadro 1: Classificação da OMS de 2016 para leucemia mieloide aguda e neoplasmas relacionados (Arber et al. 2016 adaptado)

As principais mudanças na classificação da OMS de 2016 para a LMA consistem na introdução de novas entidades. A nova categoria provisória “LMA com *BCR-ABL1*” foi adicionada a fim de guiar a terapia desses pacientes com um inibidor de tirosina quinase. Embora a distinção diagnóstica entre a LMA de novo com *BCR-ABL1* e a crise blástica da LMC possa ser difícil sem a adequada informação clínica, sentiu-se a necessidade de colocá-la como entidade provisória devido ao impacto no tratamento do paciente. Dados preliminares sugerem que deleções dos receptores de antígeno (*IGH* e *TCR*), do *IKZF1* e/ou do *CDKN2A* podem auxiliar no diagnóstico diferencial entre LMA *de novo* ou crise blástica da LMC. Outra entidade provisória incluída foi a “LMA com mutação *RUNX1*”, que é associada a características clínico-patológicas distintas e possivelmente a um pior prognóstico em relação a outros subtipos de LMA. Além disso, com a descoberta de que a melhora no desfecho do paciente com mutação no *CEBPA* é devido à mutação bialélica e não à monoalélica, essa entidade agora é chamada de “LMA com mutações bialélicas do *CEBPA*” e passou a ser uma entidade definitiva, assim como a “LMA com *NPM1* mutado” (Arber *et al*., 2016; Dohner *et al*., 2017).

Outra mudança importante na revisão de 2016 da OMS foi a inclusão de uma nova seção intitulada de “neoplasias mieloides com predisposição da linhagem germinativa”, que inclui casos de SMD, neoplasia mieloproliferativa crônica (NMPc) e LMA que apresentam uma mutação de predisposição na linhagem germinativa. Essa inclusão foi motivada pelo fato de uma minoria desses casos serem associados a uma mutação germinativa e serem familiares, dessa forma a pesquisa dessas alterações genéticas poderiam fazer parte do diagnóstico, sendo necessário o rastreio familiar dessas alterações (Arber *et al*., 2016; Dohner *et al*., 2017; Wartiovaara-Kautto *et al*., 2018).

2.1.3 Fatores prognósticos e estratificação de risco

Os fatores prognósticos podem ser subdivididos entre aqueles relacionados às características do paciente e à sua condição de saúde geral e aqueles relacionados às características do clone leucêmico. Os fatores do primeiro grupo são particularmente relevantes na predição da tolerância e mortalidade relacionada ao tratamento e torna-se mais importante com o aumento da idade. Enquanto os do segundo grupo são preditivos da resistência à terapia convencional (Dohner *et al*., 2010; Estey, 2018).

A idade ao diagnóstico encontra-se entre os fatores de risco pré-tratamento mais importantes. O aumento da idade é independentemente associado a um pior desfecho. Entretanto, outros fatores de risco associados ao paciente como *performance status*, saúde geral e comorbidades específicas modulam o efeito da idade na tolerância à quimioterapia. Além disso, fatores associados ao clone leucêmico como anormalidades genéticas, SMD ou NMPc prévias e exposição anterior a terapias citotóxicas (por exemplo, exposição a

agentes alquilantes ou radiação) para outras doenças aumentam a possibilidade de resistência à quimioterapia. Dessa forma, a idade de forma isolada não deve ser determinante para a tomada de decisão do tratamento (Dohner *et al*., 2017; Song *et al*., 2018; Shallis *et al*., 2019).

Juntamente com a idade, as alterações citogenéticas e moleculares presentes ao diagnóstico são as principais variáveis relacionadas ao prognóstico na LMA (Estey, 2018). Numerosas alterações citogenéticas recorrentes estruturais ou numéricas são identificadas nesta doença. Muitas dessas alterações não somente são marcadores diagnósticos para especificar subtipos de LMA, mas também constituem fatores prognósticos independentes para prever a remissão completa, o risco de recaída e a sobrevida global. Por esse motivo, a análise citogenética tem sido veementemente recomendada por grupos colaborativos multicêntricos e agências regulatórias (Arber *et al*., 2016; Dohner *et al*., 2017).

A estratificação de risco para LMA recomendada pelo grupo colaborativo *Medical Research Council* (MRC) foi a primeira proposta de avaliação prognóstica envolvendo o uso de alterações citogenéticas. Segundo essa proposta os pacientes são estratificados da seguinte forma: 1. prognóstico favorável, que inclui a t(8;21)(q22;q22), inv (16)(p13;q22)/t(16;16)(p13;q22) ou t(15;17)(q22;q21); 2. prognóstico adverso, que inclui inv(3)(q21;q26)/t(3;3)(q21;q26), translocações balanceadas envolvendo 11q23, anormalidades envolvendo o cromossomo 5 ou 7 [add(5q), del(5q), -5, -7, add(7q), del(7q),], t(6;11)(q27;q23), t(10;11)(p11~13;q23), deleção do cromossomo 17 ou anormalidades envolvendo 17p, e cariótipo complexo; 3. prognóstico intermediário, que inclui aqueles sem alterações citogenéticas

(LMA-CN) e os pacientes com LMA portando alguma anormalidade citogenética que não seja considerada no grupo de risco favorável ou desfavorável, totalizando cerca de 60% de todos os casos de LMA (Grinwade *et al*., 2010).

O grupo de LMA-CN corresponde a 40-50% dos casos de LMA e apresenta um prognóstico obscuro, devido à heterogeneidade nas alterações moleculares subjacentes e na responsividade ao tratamento (Grimwade *et al*., 2010; McCurdy e Levis, 2017). Devido a isso, diversos marcadores moleculares foram sendo identificados na LMA, os quais possibilitaram a estratificação dos pacientes com cariótipo normal e o refinamento na estratificação dos demais grupos (Nakao *et al*., 1996; Frohling *et al*., 2004; Falini *et al*., 2005; Small, 2006).

Em consequência, o grupo cooperativo ELN, em 2010, propôs uma estratificação de risco que associava o prognóstico conferido pelo cariótipo ao impacto no desfecho das mutações *FLT3*/ITD e nos genes *CEBPA* e *NPM1* nos casos de LMA-CN, as mutações mais bem estudadas e caracterizadas, até o momento, com maior impacto no prognóstico dos pacientes. Nessa classificação, os casos de LMA-CN que apresentassem mutação em *NPM1* sem a presença concomitante de *FLT3*-ITD ou mutação em *CEBPA* passavam a ser classificadas dentro do grupo de risco favorável. Além disso, o grupo intermediário era dividido em 2 grupos: intermediário I, que compreendia os pacientes com cariótipo normal que não eram classificados no grupo favorável; e o intermediário II, que incluía as anormalidades citogenéticas que não eram classificadas nem no grupo adverso e nem no favorável (Dohner *et al*., 2010).

As alterações moleculares se mostraram importantes também para refinar o prognóstico de subgrupos já bem definidos com alterações citogenéticas, como é o caso do grupo das leucemias com alterações envolvendo os fatores de transcrição chamados *core binding fator* (grupo CBF-LMA), que são as LMA com t(8;21)(q22;q22.1); *RUNX1-RUNX1T1* e LMA com inv(16)(p13.1q22) ou t(16;16)(p13.1;q22); *CBFB-MYH11* (Opatz et al. 2020). Em particular, na LMA com t(8;21), a presença de mutações no gene *KIT* provavelmente está associada a um pior prognóstico, especialmente se o paciente apresentar altos níveis do alelo mutado (Allen *et al*., 2013; Ayatollahi *et al*., 2017; Ishikawa *et al*., 2020). O mesmo acontece com as mutações *FLT3*-ITD, em que pacientes com t(8;21) ou inv(16);t(16;16) apresentando altos níveis da mutação parecem ter um pior desfecho clínico (Allen *et al*., 2013).

Dados provenientes de estudos com o uso de tecnologias de sequenciamento de nova geração forneceram uma visão mais abrangente do espectro e frequência de mutações, seus padrões distintos de cooperatividade e exclusividade mútua, sua arquitetura subclonal, sua evolução durante o curso da doença e seu impacto no prognóstico (Patel *et al*., 2012; Cancer Genome Atlas, 2013; Papaemmanuil *et al*., 2016). De fato, mutações genéticas são identificadas em mais de 97% dos casos, frequentemente na ausência de grandes anormalidades cromossômicas (Patel *et al*., 2012; Cancer Genome Atlas, 2013; Papaemmanuil *et al*., 2016). Com isso, veio a necessidade de revisar a estratificação de risco proposta pelo grupo ELN e, em 2017, foi proposta uma nova edição (Quadro 2), que corresponde a atualmente utilizada na prática clínica (Dohner *et al*., 2017).

As principais mudanças dessa edição compreendem a unificação dos grupos intermediários I e II em apenas um grupo, totalizando três grupos de risco (favorável, intermediário e adverso) e a inclusão de mudanças no status mutacional dos genes *NPM1*, *CEBPA* e *FLT3*-ITD e da análise de mutações nos genes *RUNX1* e *ASXL1*. Quanto a mutação *FLT3*-ITD, deve-se levar em consideração a sua razão alélica. Pacientes com baixa razão alélica do *FLT3*-ITD apresentam o mesmo impacto no prognóstico dos pacientes com *FLT3*-ITD negativo, enquanto pacientes com *NPM1* selvagem e alta razão alélica do *FLT3*-ITD passam, agora, a fazer parte do grupo de risco adverso. Para as mutações *CEBPA*, apenas as mutações bialélicas tem impacto favorável no prognóstico, enquanto as mutações nos genes *RUNX1*, *ASXL1* e *TP53* são incluídas no grupo adverso (Dohner *et al*., 2017).

Categoria de risco	**Anormalidade Genética**
Favorável	t(8;21)(q22;q22.1); *RUNX1-RUNX1T1*
	inv(16)(p13.1q22) ou t(16;16)(p13.1;q22); *CBFB-MYH11*
	NPM1 mutado sem *FLT3-ITD* ou com *FLT3-ITD*^{baixo}
	Mutação bialélica no *CEBPA*
Intermediário	*NPM1* mutado com *FLT3-ITD*^{alto}
	NPM1 selvagem sem *FLT3-ITD* ou com *FLT3-ITD*^{baixo} (sem alterações genéticas de risco adverso)
	t(9;11)(p21.3;q23.3); *MLLT3-KMT2A*
	Anormalidades citogenéticas não classificadas em risco favorável ou adverso
Adverso	t(6;9)(p23;q34.1); *DEK-NUP214*
	t(v;11q23.3); rearranjo envolvendo o gene *KMT2A*
	t(9;22)(q34.1;q11.2); *BCR-ABL1*
	inv(3)(q21.3q26.2) ou t(3;3)(q21.3;q26.2); *GATA2,MECOM(EVI1)*
	-5 ou del(5q); -7; -17/abn(17p)
	Cariótipo complexo, cariótipo monossômico

	NPM1 selvagem com *FLT3-ITD*^{alto}
	Mutação no *RUNX1*
	Mutação no *ASXL1*
	Mutação no *TP53*

Quadro 2: Estratificação de risco do grupo ELN 2017 segundo alterações genéticas (Fonte: Dohner *et al*., 2017 modificado).

Para pacientes adultos com idade inferior a 60 anos essa categorização é primariamente utilizada para guiar o protocolo terapêutico a ser escolhido após a primeira remissão. Os pacientes de grupo de risco favorável que são submetidos a quimioterapia convencional, como terapia pós-remissão, tendem a apresentar resultados semelhantes àqueles submetidos a transplante de células tronco hematopoiéticas (TCTH). Por sua vez, para pacientes do grupo de risco intermediário e adverso, o TCTH é indicado como terapia de consolidação na primeira remissão completa, uma vez que a quimioterapia convencional como terapia pós-remissão apresenta resultados inferiores para esses pacientes (Lowenberg *et al*., 2011; Cornelissen e Blaise, 2016; Dohner *et al*., 2017).

A estratificação de risco do grupo ELN é a mais amplamente utilizada e aceita na prática clínica, entretanto, nos últimos anos, diversos autores têm defendido a importância de se examinar uma gama mais ampla de mutações do que a proposta no ELN2017. Ademais, diversas cooperações mutacionais têm sido descritas e a avaliação isolada de marcadores genéticos tem se demonstrado limitado no seu valor prognóstico (Patel *et al*., 2012; Metzeler *et al*., 2016; Papaemmanuil *et al*., 2016; Estey, 2018; Tyner *et al*., 2018).

A co-ocorrência das mutações *FLT3*-ITD, *DNMT3A* e *NPM1* tem se revelado a mais frequente cooperação entre 3 genes. O triplo mutado além de

apresentar características biológicas distintas, o que faz dele um provável novo subtipo de LMA, tem demonstrado um prognóstico particular. Pacientes triplo mutados parecem ter um prognóstico bastante adverso em relação aos pacientes com duas ou uma dessas mutações, daí a importância de se avaliar a co-ocorrência dessas mutações (Guryanova *et al*., 2016; Papaemmanuil *et al*., 2016; Tyner *et al*., 2018).

Além dessa interação, tem sido demonstrado que pacientes com mutação no gene *NPM1* e em algum gene *RAS* tem sobrevida superior quando essas alterações estão associadas à presença de mutação no gene *DNMT3A*. A presença simultânea de mutações nos genes *ASXL1* e *SRSF2* é um outro exemplo de associação aditiva, neste caso, o efeito deletério das mutações se somam e a co-ocorrência indica um prognóstico ainda mais adverso. No entanto, não se pode superestimar que o efeito prognóstico de uma dada mutação sempre vai depender da presença de outra mutação. Maiores estudos são necessários para elucidar o efeito biológico e prognóstico dessas interações (Papaemmanuil *et al*., 2016; Eisfeld *et al*., 2018; Estey, 2018).

2.1.4 O transcriptoma como ferramenta de avaliação prognóstica para estratificação de risco da LMA

Apesar do grande esforço realizado nos últimos anos para melhorar a estratificação de risco da LMA e sua resposta ao tratamento, essa doença (com exceção da leucemia promielocítica aguda - LPA) ainda carece de tratamento eficaz (Estey, 2018). Vemos um grande progresso na descoberta de mutações na LMA, sua associação com diagnóstico e prognóstico e sua aplicação rotineira na clínica (Papaemmanuil *et al*., 2016; Dohner *et al*., 2017; Tyner *et*

al., 2018). Entretanto, estudos de transcriptoma permanecem pouco traduzidos para a prática clínica, mesmo que diversos estudos de expressão gênica tenham contribuído significativamente para a elucidação da patogênese da LMA e indiquem possíveis marcadores prognósticos e, consequentemente, direções terapêuticas (Li *et al*., 2013; Metzeler *et al*., 2013; Shi *et al*., 2014; Ng *et al*., 2016; Huang *et al*., 2017).

Na última década tem sido demonstrado que a expressão de diversos genes, de forma isolada, tem impacto no prognóstico da LMA e poderiam ser considerados para estratificar o risco dos pacientes com essa doença. De fato, genes como *BAALC*, *WT1*, *ERG*, *EVI1* e *ID1* foram associados com o desfecho clínico da LMA, ainda que os resultados sejam controversos (Langabeer *et al*., 2001; Damm *et al*., 2011; Damm *et al*., 2012; Weber *et al*., 2014, Vazquez *et al*., 2016; Hinai e Valk, 2016). Entretanto, tem sido demonstrado que a expressão de grupos de genes, ao invés da expressão isolada, carrega maior robustez para caracterizar a LMA, permitindo maior refinamento na sua classificação biológica e prognóstica. O progresso tecnológico iniciado no final do milênio passado com as plataformas de microarray e, mais recentemente, com o sequenciamento de nova geração, permitiu a transição da análise de um único gene para toda a escala do transcriptoma (Handschuh, 2019).

Inicialmente, as assinaturas de expressão gênica foram aplicadas para elucidar a fisiopatologia da LMA, bem como, na tentativa de diferenciar subgrupos com anormalidades citogenéticas e moleculares. A partir disso, diversos autores demonstraram que alguns dos subgrupos de LMA possuíam uma assinatura de expressão gênica característica e poderiam ser distinguidos

uns dos outros (Bullinger *et al*., 2004; Verhaak *et al*., 2009; De la Bletiere *et al*., 2012).

Um estudo desenvolvido por Valk e colaboradores, em 2004, foi capaz de identificar 16 grupos de pacientes com base nas suas assinaturas de expressão gênica em uma análise de clusterização não supervisionada. Todos os pacientes com as alterações citogenéticas recorrentes inv(16)/t(16;16), t(8;21) e t(15;17) foram agrupados nos clusters 9, 13 e 12, respectivamente, reforçando o fato de que definem entidades biológicas distintas. Pacientes com outras alterações citogenéticas e moleculares também definiram *clusters*, como mutações no gene *CEBPA* e t(11q23) e, além disso, diversos outros novos *clusters* foram definidos consistindo principalmente de pacientes com cariótipo normal (Figura 4A). Ademais, esses *clusters* definiam grupos de pacientes com desfechos clínicos distintos e foi capaz de definir um novo *cluster* que apresentou um desfecho bastante adverso (Figura 4B) (Valk *et al*., 2004).

Desde então, diversas assinaturas de expressão gênica capazes de diferenciar grupos prognósticos distintos têm sido propostas (Li *et al*., 2013; Yang *et al*., 2015; Ng *et al*., 2016; Huang *et al*., 2017; Horibata *et al*., 2019; Wagner *et al*., 2019). Entre elas, diversos autores foram capazes de definir assinaturas de expressão gênica que caracterizavam amostras de LMA enriquecidas com células tronco leucêmicas. Pacientes com esse perfil de expressão gênica apresentavam maior resistência à quimioterapia e, por tanto, um pior desfecho com altas taxas de recaída (Metzeler *et al*., 2013; Rapin *et al*., 2014; Yang *et al*., 2015; Ng *et al*., 2016; Horibata *et al*., 2019).

Utilizando ferramentas de bioinformática e bioestatística, alguns índices de prognóstico baseados no efeito aditivo da expressão de genes no desfecho

dos pacientes com LMA foram apresentados e tem trazido informação adicional para a estratificação de risco desses pacientes (Li *et al*., 2013; Huang *et al*., 2014; Shi *et al*., 2014; Xie *et al*., 2017). Mais recentemente, Wagner e colaboradores, em 2019, desenvolveram um índice de prognóstico baseado na expressão de três genes (*CALCRL*, *CD109*, e *LSP1*), dos quais apenas a expressão do *CD109* já tinha sido associado à LMA, utilizando uma ferramenta de redes neurais artificiais baseada em aprendizagem de máquina. Esse índice foi capaz de refinar a estratificação de risco ELN2017 e significativamente melhorou a previsão de resultados (Wagner *et al*., 2019).

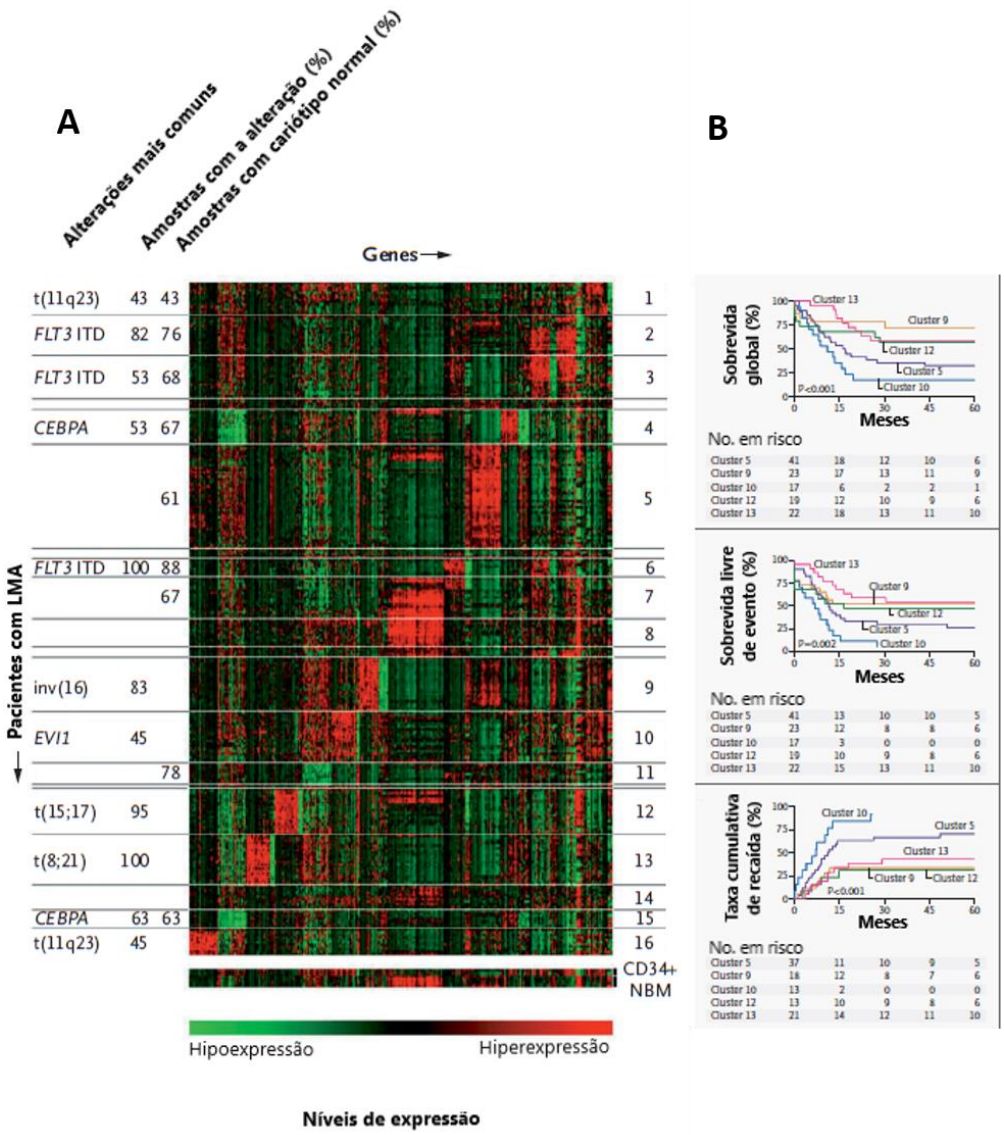


Figura 4. (A) *Heatmap* apresentando os níveis de expressão dos 40 principais genes identificados pela análise de significância de *microarrays* de cada um dos 16 *clusters*, bem como na medula óssea normal (NBM, do inglês *normal bone marrow*) e células CD34+. A barra de escala, na parte inferior da figura, indica um aumento (vermelho) ou diminuição (verde) no nível de expressão por um fator de pelo menos 4 em relação à média geométrica de todas as amostras. As porcentagens das alterações mais comuns (aquelas presentes em mais de 40 por cento das amostras) e as porcentagens de amostras com um cariótipo normal, em cada um dos clusters, são indicadas. (B) Estimativa de Kaplan-Meier para as curvas de sobrevida sobrevida global (painel superior), de sobrevida livre de evento (painel central) e de taxa cumulativa de recaída após remissão completa (painel inferior) entre os pacientes com LMA dos clusters 5, 9, 10, 12 e 13. (Fonte: Valk *et al*., 2004 modificado)

Essas propostas de estratificação de risco baseadas em transcriptoma tem se mostrado promissoras para prever o desfecho dos pacientes com LMA. Entretanto, os resultados ainda são bastante conflitantes no que diz respeito aos marcadores prognósticos, pontos de corte ideal para o valor de expressão gênica e ferramentas e metodologias de bioinformática e bioestatística utilizadas. Além disso, utilizar dados de expressão gênica para avaliação prognóstica tem como limitação a padronização do ponto de corte ideal para dicotomização da variável, além das variabilidades entre as diversas técnicas que podem ser utilizadas. Entretanto, plataformas de microarray para a detecção simultânea de mutações, rearranjos gênicos e expressão alteradas de genes, como o *BAALC* e o *EVI1,* têm sido validados em estudos retrospectivos e sugeridos para estratificação de ensaios clínicos prospectivos (Nomdedeu *et al*. 2017; Kappala *et al*. 2017).

2.1.5 Bancos de dados de coortes públicas

Atualmente está disponível publicamente diversas coortes de pacientes LMA com seus dados clínicos, genômicos e transcriptômicos (Gene Expression Omnibus, National Center for Biotechnology Information, 2021; cBioPortal for Cancer Genomics, 2021). No presente trabalho utilizaremos duas coortes

disponíveis publicamente que chamaremos aqui de GSE6891 e TCGA (do inglês, *The Cancer Genome Atlas*), por isso faremos nessa seção uma breve revisão sobre esses dois bancos de dados.

A coorte TCGA é composta por dados clínicos e dados genômicos e transcriptômicos, obtidos pela tecnologia de sequenciamento de nova geração, de 200 pacientes com LMA. As 200 amostras foram selecionadas de um conjunto de mais de 400 amostras, de um único centro, para refletir uma distribuição de subtipos do “mundo real” (Cancer Genome Altas, 2013). Esse estudo fez parte do programa *The Cancer Genome Atlas* (TCGA), criado em 2006, como um esforço conjunto entre o *National Cancer Institute* e o *National Human Genome Research Institute*. O TCGA é um programa de referência em genômica do câncer que caracterizou molecularmente mais de 20.000 cânceres primários e amostras normais correspondentes abrangendo 33 tipos de câncer. Os dados gerados têm contribuído para melhorias no diagnóstico, tratamento e prevenção do câncer e estão publicamente disponíveis (National Cancer Institute, 2021).

A coorte GSE6891 é composta por dados clínicos e transcriptômicos, obtidos pela tecnologia de *microarray*, de 537 pacientes com LMA e síndrome mielodisplásica, proveniente do estudo desenvolvido por Verhaak e colaboradores, em 2009 (Verhaak *et al*., 2009). Essa coorte está disponível no banco de dados *Gene Expression Omnibus* (GEO) que é um repositório público de dados genômicos funcionais e que suporta dados de *microarray* ou sequenciamento. As ferramentas do GEO são fornecidas para ajudar os usuários a consultar e baixar dados de expressão gênica curados (Gene Expression Omnibus, National Center for Biotechnology Information, 2021). O

banco de dados GEO é uma seção do *National Center for Biotechnology Information* (NCBI), criado em 1988 como uma divisão da *National Library of Medicine* (NLM) no *National Institutes of Health* (NIH) (National Center for Biotechnology Information, 2021).

2.1.6 Panorama da LMA no Brasil

Diversos trabalhos têm demonstrado que o desfecho dos pacientes com LMA em países em desenvolvimento, especialmente no Brasil, é bastante inferior ao de países desenvolvidos. Enquanto no Brasil a taxa de sobrevida global (SG) em 5 anos é de cerca de 25% e a taxa de morte precoce (no primeiro mês do diagnóstico) em torno de 29% (Bittencourt *et al*., 2003; Fagundes *et al*., 2006; Lima *et al*., 2015; Benício *et al*., 2017; Silveira *et al*., 2021), nos países desenvolvidos a SG é em torno dos 45% e a taxa de morte precoce, em torno de 10%, mesmo fora de ensaios clínicos controlados (Appelbaum *et al*., 2006; Lowenberg *et al*., 2011; Silveira *et al*., 2021).

O cenário é ainda mais difícil quando avaliada a região Nordeste do Brasil. Benício *et al*. (2017) demonstraram que pacientes tratados na cidade de Recife (PE), apresentaram desfecho inferior ao de pacientes tratados na cidade de São Paulo (SP), e na cidade de Belo Horizonte (MG), ambas cidades do Sudeste brasileiro. Recife possuía, no período do estudo, Índice de Desenvolvimento Humano (IDH) inferior ao das outras duas cidades (Benício *et al*. 2017).

Em um estudo brasileiro, desenvolvido por Fagundes *et al*. (2006), foi demonstrado que os pacientes com LMA que viviam em cidades com IDH baixo eram menos submetidos a quimioterapia com altas doses de citarabina e

apresentaram menores sobrevida global e livre de doença). O Índice de Desenvolvimento Humano (IDH) é desenvolvido pela Organização das Nações Unidas e reflete as condições socioeconômicas de um país sob três aspectos básicos: longevidade, conhecimento e padrão de vida (Human Development Report, 2020). As condições socioeconômicas são fatores que podem influenciar diretamente o desfecho de pacientes com LMA. De fato, diversos trabalhos (Rego *et al*., 2003; Lucena-Araujo *et al*., 2010; Lima *et al*., 2015; Benício *et al*., 2017; Silveira *et al*., 2021) têm demonstrado a associação das condições socioeconômicas da população brasileira com o desfecho da LMA.

A alta taxa de mortalidade precoce descrita no Brasil tem ocorrido, principalmente, devido a infecções bacterianas e fúngicas fatais (Rego *et al*., 2003; Benício *et al*., 2017; Silveira *et al*., 2021). Ainda que não tenha sido analisado diretamente, os autores relacionam essa alta taxa de infecções fatais a diversos fatores socioeconômicos, como dificuldades no acesso a atendimento médico urgente, taxas mais altas de comorbidade, a baixa escolaridade dos doentes de serviços públicos que dificulta o entendimento e gravidade dos sintomas e da doença, além da falta de infraestrutura adequada (Fagundes *et al*., 2006; Nucci *et al*., 2013; Silveira *et al*., 2021). Esse excesso de mortalidade devido à infecção, entretanto, é potencialmente evitável por meio de melhor educação do paciente, acesso a cuidados de saúde de emergência de alta qualidade e administração imediata de antibióticos com atividade contra os organismos causadores (da Silva *et al*., 2019)

Um estudo realizado por Nucci *et al.* (2013) relatou uma taxa de infecções fúngicas invasivas muito superiores entre os pacientes que receberam apenas tratamento quimioterápico convencional comparado aos

pacientes que receberam TCTH alogênico. Eles associaram isso ao fato de que o sistema de saúde público brasileiro reembolsa muito menos dinheiro para o tratamento quimioterápico da LMA em comparação com o TCTH, que tem um programa e financiamento separados. Como consequência, por exemplo, os pacientes LMA que fazem apenas tratamento quimioterápico são normalmente atendidos em quartos compartilhados e sem filtros HEPA, enquanto os receptores de TCTH alogênicos são atendidos em quartos individuais, com filtros HEPA e pressão positiva. Eles ponderam, entretanto, que essas diferenças também podem ser devido ao fato de que o TCTH é geralmente um procedimento eletivo em que os pacientes são cuidadosamente selecionados, enquanto os pacientes com LMA representam uma população não selecionada (Nucci *et al*., 2013).

A alta taxa de mortalidade precoce é considerada uma das principais causas da baixa taxa de sobrevida global, entretanto, outros fatores também têm sido relatados como importantes para levar às baixas taxas de SG e sobrevida livre de doença (SLD) e às altas taxas de incidência cumulativa de recaída (ICR). Entre eles, a superlotação dos leitos e a falta de quimioterápicos que levam a atrasos no tratamento, disposição inadequada dos cuidados de suporte, baixo nível socioeconômico da população, incluindo más condições de moradia, baixa renda per capita e consumo de energia, além da baixa quantidade de pacientes encaminhados para TCTH na primeira remissão completa e longo tempo de espera para realização do TCTH (Rego *et al*., 2003; Fagundes *et al*., 2006; Lima *et al*., 2015; Benício *et al*., 2017; Silveira *et al*., 2021).

Em um estudo brasileiro, entretanto, foi demonstrado que os pacientes que atingiram a remissão completa e receberam tratamento pós-remissão com altas doses de citarabina tiveram probabilidades de SG e SLD semelhantes aos de pacientes tratados em países desenvolvidos. No entanto, apenas 37% dos pacientes receberam uma terapia potencialmente curativa, o que contrasta com 47 a 61% relatados em ensaios clínicos prospectivos. Além disso, pacientes de cidades com baixo IDH apresentaram menor probabilidade de serem incluídos em tratamento com altas doses de citarabina (Fagundes *et al*., 2006)

As condições socioeconômicas de uma população, que pode ser medida pelo IDH, impactam até no desfecho de pacientes LMA tratados em países que apresentam alto IDH, como os países europeus. Em um estudo em que foram comparados os desfechos de pacientes com leucemias agudas submetidos a TCTH de diferentes países europeus, em análise multivariada, os pacientes de países com IDH superiores foram associados a maior SLD e menores taxas de ICR em relação a pacientes de países com IDH inferiores. Além disso, em países com IDH superiores a taxa de pacientes submetidos a TCTH é maior do que os de países com IDH inferiores (Giebel *et al*., 2010). Essa realidade se torna bastante alarmante quando avaliamos a realidade brasileira, em que um estudo multicêntrico relatou que apenas 5% dos pacientes de uma coorte “da vida real” foram submetidos a TCTH na primeira remissão completa (Benício *et al*., 2017).

Embora o TCTH seja a única opção curativa para pacientes com LMA de alto risco e seja indicado como consolidação na primeira remissão completa para pacientes da categoria de risco intermediário e adverso do grupo ELN (Dohner *et al*., 2017), ele é bastante caro. O TCTH requer unidades de

transplante bem organizadas e equipadas, cuidados de suporte avançados e equipe médica e de enfermagem altamente qualificada. Além disso, o monitoramento e tratamento adequados são necessários por muitos meses a anos após o transplante, pois complicações com risco de vida podem aparecer tardiamente. Dessa maneira, pode-se esperar que o acesso ao TCTH, bem como o resultado do procedimento, possa depender de fatores socioeconômicos que variam entre e dentro dos países (Giebel *et al*., 2010).

Vários estudos realizados tanto em países em desenvolvimento, como o Brasil, quanto em países desenvolvidos, como Estados Unidos ou Reino Unido, demonstraram a associação das condições socioeconômica com o acesso ao TCTH e mortalidade após o transplante (Serna *et al*., 2003; Silla *et al*., 2009; Giebel *et al*., 2010). Em um estudo recente, que comparou o desfecho de pacientes LMA tratados no Brasil com uma coorte de pacientes do Reino Unido, relatou que os pacientes brasileiros tiveram menor taxa de SG e maior incidência cumulativa de recaída e foram menos propensos a se submeter ao TCTH e esperaram mais para o TCTH do que os pacientes do Reino Unido. Além disso, a maioria dos TCTH, no Brasil, são autólogos, enquanto aproximadamente metade dos transplantes no Reino Unido são alogênicos (Silveira *et al*., 2021).

Diante desse cenário vê-se que os esforços para melhorar os resultados da LMA no Brasil devem se concentrar na prevenção e controle de infecção e aumento do acesso ao TCTH. Isso pode ser alcançado através da implementação de melhor educação do paciente e da equipe, profilaxia antimicrobiana apropriada, implementação de protocolos de neutropenia febril e medidas de controle de infecção para isolar e prevenir a propagação de

organismos multirresistentes. Além disso, é necessário investir em infraestrutura, treinamento e fontes alternativas de doadores para aumentar a frequência e melhorar a oportunidade do TCTH (Silveira *et al*., 2021).

Usando a LPA como um modelo de doença, Rego e colaboradores, através do Consórcio Internacional para Leucemia Promielocítica Aguda (IC-APL, do inglês *International Consortium on Acute Promyelocytic Leukemia*), demonstraram que uma rede cooperativa internacional pode resultar em uma melhora significativa no resultado do paciente (Rego *et al*., 2013). O IC-APL foi estabelecido para criar uma rede de instituições em países em desenvolvimento que trocariam experiências e dados e receberia apoio de grupos cooperativos bem estabelecidos dos Estados Unidos e da Europa. O IC-APL formulou um diagnóstico rápido, tratamento e diretrizes de suporte que foram adaptadas às condições locais (Rego *et al*., 2013).

Antes do estabelecimento do IC-APL, a SG em 2 anos de pacientes adultos com LPA no Brasil era de cerca de 50% e a taxa de mortalidade precoce era de 30% (Jácomo *et al*., 2007). Como resultado do IC-APL, a SG em 2 anos aumentou para 80% e a taxa de mortalidade precoce foi reduzida pela metade. Além disso, a SG e SLD em 2 anos e a taxa cumulativa de recaída foram comparáveis às relatadas em estudos realizados em países desenvolvidos (Rego *et al*., 2013).

Os resultados obtidos pelo IC-APL demonstraram que é possível melhorar o prognóstico de doenças malignas graves e rapidamente fatais, como a LPA, em países em desenvolvimento, por meio de colaboração internacional, com troca de expertise, a um custo relativamente baixo. Atualmente está em vigência um consórcio para o tratamento da LMA, com o

objetivo de ampliar para os outros subtipos de LMA a experiência adquirida com o IC-APL (Rego *et al*., 2013).

3 OBJETIVOS

3.1 OBJETIVO GERAL

Avaliar o impacto isolado e aditivo de marcadores moleculares no desfecho clínicos de pacientes adultos com leucemia mieloide aguda.

3.2 OBJETIVOS ESPECÍFICOS

1. Determinar o significado prognóstico isolado e aditivo de mutações nos genes *NPM1*, *FLT3*, *CEBPA*, *WT1*, *IDH1*, *IDH2*, *DNMT3A*, *ASXL1* e *TP53* e da expressão dos genes *BAALC*, *ERG*, *EVI1*, *KMT2E*, *ID1* e *MN1*, utilizando a sobrevida global como desfecho clínico a partir de uma coorte pública;
2. Selecionar, a partir da análise de transcritptoma e da análise de sobrevida, utilizando como variável de desfecho a sobrevida global, os marcadores prognósticos moleculares que serão utilizados para o desenho de um índice de prognóstico;
3. Desenvolver um índice de prognóstico com base no valor aditivo dos marcadores moleculares selecionados para a predição dos principais desfechos clínicos da doença (morte e recaída).

4 ARTIGO 1 – PROGNOSTIC IMPLICATIONS OF ID1 EXPRESSION IN ACUTE MYELOID LEUKEMIA PATIENTS TREATED IN A RESOURCE-CONSTRAINED SETTING

Artigo foi submetido à revista *Blood* – Fator de impacto: 22,113; Qualis/CAPES Ciências Biológicas I: A1

Aleide S Lima,[1] Matheus F Bezerra,[2] Amanda Moreira-Aguiar,[1] Isabel Weinhäuser,[3] Bianca L Santos,[1] Raul M Falcão,[4] Maria L Salustiano-Bandeira,[1] Pedro L Franca-Neto,[1] Marinus M Lima,[1] Felipe Saldanha-Araujo,[5] Juan L Coelho-Silva,[6] Diego A Pereira-Martins,[3] Marcos A Bezerra,[1] Antonio R Lucena-Araujo[1]*

Affiliations: [1]Department of Genetics, Federal University of Pernambuco, Recife, Brazil; [2]Department of Microbiology, Aggeu Magalhães Institute, Fiocruz, Recife, Pernambuco, Brazil; [3]Department of Hematology, Cancer Research Centre Groningen, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands [4]Bioinformatics Multidisciplinary Environment (BioME), Metrópole Digital Institute, Federal University of Rio Grande do Norte; [5]Laboratório de Hematologia e Células-Tronco, Universidade de Brasília, Av. L2 Norte, Brasília, DF, 70.910-900, Brazil. [6]Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA.

The aberrant expression of the inhibitor of DNA binding (*ID1*) gene has been frequently associated with the initiation and progression of human tumors.[1–3] Particularly in acute myeloid leukemia (AML), a growing body of evidence suggest that *ID1* overexpression can impact both on AML leukemogenesis[4–7] and prognostication,[8–10] although these findings remain controversial. Most important, the clinical importance of the *ID1* expression in patients treated outside well-controlled clinical trials has never been evaluated, which means that its potential prognostic value remains unknown. Hence, we investigated the role of *ID1* expression on clinical outcomes of non-selected AML patients treated in a single center from northeast Brazil.

To test the feasibility of the study, we took advantage of The Cancer Genome Atlas database (154 non-acute promyelocytic leukemia patients, hereafter called “training cohort”)[11] and explored the association of the *ID1* expression with clinical and laboratory features. Details of the statistical analysis, clinical endpoints and strategies for dichotomization can be found in the Supplemental data. After classifying patients according to cytogenetic abnormalities at diagnosis, we noticed that *ID1* expression was significantly lower in patients with CBF-leukemia (*P*=0.014; Figure 1A). Moreover, *ID1* was overexpressed in *IDH1*- (*P*=0.022; Figure 1B) and *TP53*-mutated patients (*P*=0.0094; Figure 1C), while biallelic *CEBPA*-mutated patients presented lower *ID1* expression (*P*=0.0332; Figure 1D). Next, we determined the prognostic impact of *ID1* expression on treatment outcomes of patients. Using the median value of *ID1* expression, patients were dichotomized into low and high expression groups. In univariate analysis, high *ID1* expression was associated

with poor 5-years OS rate (*P*=0.013; Figure 1E). However, *ID1* expression did not retain its prognostic impact after adjustment for age (continuous variable), leukocyte counts (continuous variable), and cytogenetic risk stratification (Supplemental table 1). Also, there was no impact of *ID1* expression on post-remission outcomes (disease-free survival, *P*=0.615; cumulative incidence of relapse, *P*=0.52). To translate these findings into resource-constrained scenario, we enrolled 128 adult *de novo* AML patients diagnosed and treated in northeast Brazil (here referred as "resource-constrained cohort"). Details for treatment protocols, and inclusion criteria were published elsewhere.[12] All samples used were obtained at diagnosis from bone marrow aspirates. Details can be found in the supplemental data. Ethical approval was obtained from the local Research Ethics Board (CAAE: 77527717.4.0000.5208). In accordance with the Declaration of Helsinki, informed written consent for sample collection as well as permission for its use in research was obtained from patients or their relatives.

Figure 1F shows that patients with CBF-leukemia from the resource-constrained cohort also exhibited low *ID1* expression compared to other cytogenetic risk groups (*P*=0.042). Because of a limited number of patients, the performance of the receiver operating characteristic (ROC) curve analysis was unsatisfactory to define the optimal cut-off point for *ID1* expression (area under the curve: 0.525, 95% confidence interval, CI: 0.411-0.639). Therefore, we opted to adopt the same strategy of dichotomization used in the training cohort. Based on the median value of *ID1* expression (median value: 0.485, range: 0.1-8.9), high *ID1* patients had low platelets counts (*P*=0.04) and higher levels of lactate dehydrogenase (*P*=0.021) (Table 1). Of the 128 enrolled patients, 6

(5%) patients did not have available information on the treatment procedures or their respective follow-up and were considered ineligible for the induction therapy analyses. The median follow-up was 35 months (95% confidence interval, CI: 30–40 months) and the estimated 3-years OS rate was 15% (95% CI: 8-23%). Overall, 58/122 (47%) patients achieved complete remission. *ID1* expression had no impact on CR achievement (*P*=0.276). Patients with a high *ID1* expression had a lower 3-year OS (9%, 95% CI: 3-20%) compared to patients with a low *ID1* expression (22%, 95% CI: 11-34%) (*P*=0.037) (Figure 1G). The best-fitted multivariate Cox proportional hazards model for OS was age (continuous variable), leukocyte counts (continuous variable), and *ID1* expression (Supplemental table 2). This model showed that *ID1* expression was not independently associated with poor OS (hazard ratio, HR: 1.5, 95% CI: 0.98-2.28; *P*=0.057). *ID1* expression had no impact on disease-free survival (*P*=0.648) or cumulative incidence of relapse (P=0.584).

Three issues require particular attention. Contrary to Tang et al.,[8] we and other[9,10] have failed to demonstrate an independent association between high *ID1* expression and poor induction outcomes in AML. Reasons that may contribute for this disparity are multifactorial and may range from patient-related features (such as age and ethnic differences) to differences in treatment protocols, in particular post-remission therapy. Furthermore, we cannot rule out that methodological differences between studies (different standards for ID1 quantification or different definitions for “high” and “low” ID1 expression) could lead to different results and conclusions. These arguments are frequently used to justify differences between studies. But because our study has included consecutive and non-selected patients treated outside well-controlled clinical

trials, an additional topic needs to addressed. Our data are confronted with many variables that cannot be fully controlled, including drug unavailability, risk-adapted treatment, comorbidities, and time from diagnosis to treatment initiation.[12,13] Despite this obvious limitation involving clinical studies conducted in a real-world setting, we are firm believers that such initiative can provide more realistic data on understudied populations, in particular those from low- and middle-income countries.

Regardless whether *ID1* is an independent prognostic factor in AML or not, an outstanding question is whether its implementation could improve the current scheme for AML risk stratification.[14] Based on its biological importance, it is conceivable that *ID1* gene or its encoded protein may be useful in future revised versions of the scheme for AML risk stratification. Initially, the encoded Id1 protein seems to be a key transcriptional regulator of hematopoietic stem cell (HSC) lineage commitment, and the absence of Id1 may compromise the self-renewal capacity of HSCs.[15] In human tumors, Id1 protein can control cell proliferation, self-renewal capacity of cancer stem cells and disease aggressiveness via Myc.[16] Furthermore, *ID1* overexpression induces cell proliferation and invasion, and also protects cells against drug-induced apoptosis.[17] In a hematological context, the *ID1* overexpression is able to immortalize myeloid progenitors *in vitro* and promote myeloproliferatine-like phenotype *in vivo.*[7] Conversely, its deletion in hematopoietic stem cell decreased cell proliferation, mitochondrial biogenesis, metabolic activity, and ribosomal biogenesis.[18] It is important to note that the encoded Id1 protein is a common downstream target of constitutively activated oncogenic tyrosine kinase in AML[5] and can be upregulated during CEBPA-induced myeloid

differentiation.[19] Additionally, the forced expression of *MLL-AF9* fusion gene in fetal liver cells leads to an increase of *ID1* expression and the rapid development of AML with maturation in a fetal liver transplant model.[4] Together, these pieces of evidence support two main conclusions. First, the encoded Id1 protein plays an important role in pathophysiology of human tumors, including hematological malignancies, and the deregulated *ID1* expression does not seem to be a primary genetic event in leukemic blasts but a surrogate marker depending on other genetic aberrations. If so, it is conceivable that *ID1* overexpression may be associated with specific AML subtypes, which could explain the downregulation of *ID1* expression in CBF-leukemia reported in both training and resource-constrained cohorts.

In summary, *ID1* expression was associated with induction outcomes in patients treated outside well-controlled clinical trials, although not in an independent manner. With respect to its potential use in clinical practice (either as a potential therapeutic target or prognostic factor), further studies are necessary to validate this hypothesis.

Table 1. Baseline characteristics according to the *ID1* expression.

Characteristics	All patients		ID1 expression[1]				*P* value[2]
			Low		High		
	No.	%	No.	%	No.	%	
Age, years							0.574
< 60 years	84	65.6	42	65.6	42	65.6	
60 years and older	44	34.4	22	34.4	22	34.4	
Median (range)	51.2 (18 to 93.7)		50.3 (18.4 to 87.7)		51.5 (19 to 93.7)		0.651
Sex							0.111
Female	60	46.9	35	54.7	25	39.1	
Male	68	53.1	29	45.3	39	60.9	
FAB subtype							0.671
M0	6	5.1	3	5.3	3	4.9	
M1	18	15.3	7	12.3	11	18	
M2	45	38.1	23	40.4	22	36.1	
M4	35	29.7	15	26.3	20	32.8	
M5	13	11	8	14	5	8.2	
M6	1	0.8	1	1.8	-	-	
M7	-	-	-	-	-	-	
Missing data	10	-	7	-	64	-	
Cytogenetic risk stratification[3]							0.378
Favorable	17	30.9	10	38.5	7	24.1	
Intermediate	31	56.4	14	53.8	17	58.6	
Adverse	7	12.7	2	7.7	5	17.2	
Missing data[4]	73	-	38	-	35	-	
FLT3-ITD							0.532
Mutated	30	23.4	13	28.3	17	26.6	
Non-mutated	98	76.6	51	79.7	47	73.4	
NPM1							0.842
Mutated	34	26.6	18	28.1	16	25	
Non-mutated	94	73.4	46	71.9	48	75	
Bone marrow blasts, %, median (range)	54 (20 to 98)		57 (20 to 98)		47 (20 to 97)		0.53
Leukocyte counts, $\times 10^9$/L, median (range)	49.2 (760 to 435)		50.2 (1.3 to 435)		45.3 (0.76 to 420)		0.746
Platelet counts, $\times 10^9$/L, median (range)	51 (6 to 404)		54 (10 to 404)		42.5 (6 to 212)		0.04*
Hemoglobin, g/dL, median (range)	8.1 (3 to 11.7)		8 (3.7 to 11.4)		8.1 (3 to 11.7)		0.848
LDH level, U/l, median (range)	697 (116 to 4,192)		543 (116 to 3,891)		910 (129 to 4,192)		0.021*

NOTE:
* Indicates statistically significant differences.
1: Patients with low or high ID1 expression were defined according to the median value of *ID1* transcript levels.
2: Missing values were excluded for the calculation of *P* values.
3: The cytogenetic risk groups were defined according to Medical Research Council criteria [20].
4: Material not available or no metaphases detected.

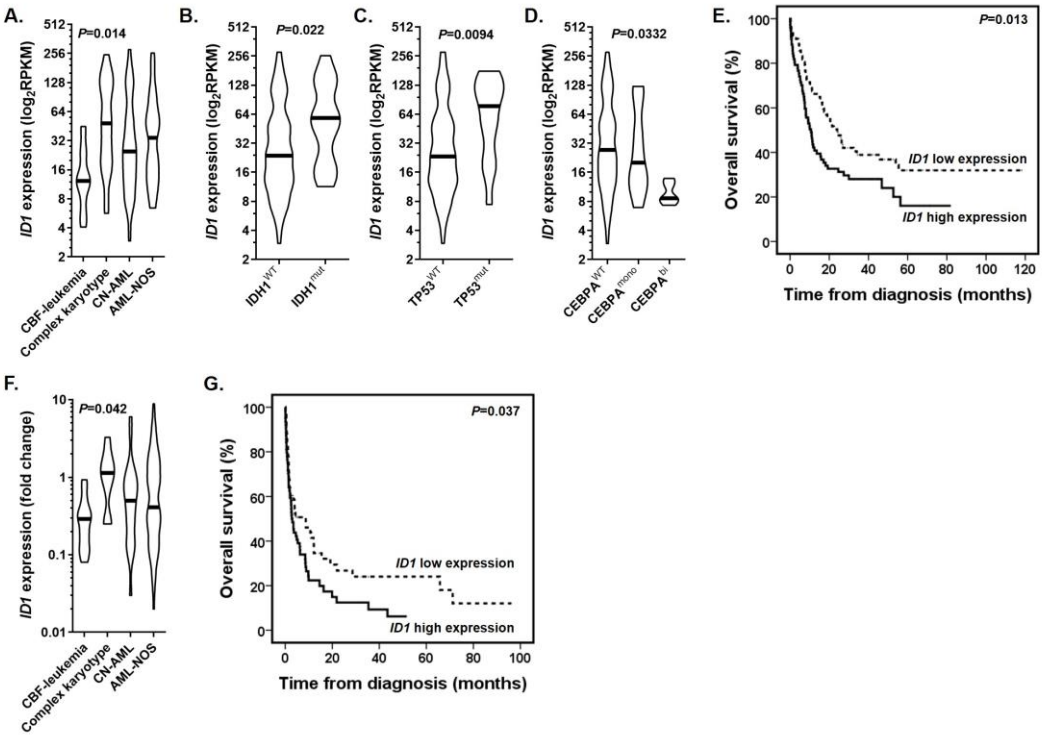


Figure 1. (A) *ID1* expression according to the cytogenetics abnormalities at diagnosis of AML patients from the training cohort. (B-D) *ID1* expression according to mutational status of (B) *IDH1*, (C) *TP53*, and (D) *CEBPA* genes. (E) Comparison of OS curves for patients from the training cohort with low and high *ID1* expression. (F) *ID1* expression according to the cytogenetics abnormalities at diagnosis of AML patients from the resource-constrained cohort (G) Comparison of OS curves for patients from the resource-constrained cohort with low and high *ID1* expression.

References

1. Zhao G-S, Zhang Q, Cao Y, et al. High expression of ID1 facilitates metastasis in human osteosarcoma by regulating the sensitivity of anoikis via PI3K/AKT depended suppression of the intrinsic apoptotic signaling pathway. *Am. J. Transl. Res.* 2019;11(4):2117–2139.
2. Wu Z, Li J, Zhang Y, Hu L, Peng X. Synchronous co-expression of Id-1 and nuclear NF-κB p65 promotes cervical cancer progression and malignancy, and is associated with a poor prognosis and chemosensitivity. *Oncol. Rep.* 2019;
3. Hu X, Chen M, Li Y, et al. Overexpression of ID1 promotes tumor progression in penile squamous cell carcinoma. *Oncol. Rep.* 2018;
4. Man N, Sun X-J, Tan Y, et al. Differential role of Id1 in MLL-AF9–driven leukemia based on cell of origin. *Blood*. 2016;127(19):2322–2326.
5. Tam WF, Gu T-L, Chen J, et al. Id1 is a common downstream target of oncogenic tyrosine kinases in leukemic cells. *Blood*. 2008;112(5):1981–1992.
6. Wang L, Man N, Sun X-J, et al. Regulation of AKT signaling by Id1 controls t(8;21) leukemia initiation and progression. *Blood*. 2015;126(5):640–650.
7. Suh HC, Leeanansaksiri W, Ji M, et al. Id1 immortalizes hematopoietic progenitors in vitro and promotes a myeloproliferative disease in vivo. *Oncogene*. 2008;27(42):5612–5623.
8. Tang R, Hirsch P, Fava F, et al. High Id1 expression is associated with poor prognosis in 237 patients with acute myeloid leukemia. *Blood*.

2009;114(14):2993–3000.

9. Damm F, Wagner K, Görlich K, et al. ID1 expression associates with other molecular markers and is not an independent prognostic factor in cytogenetically normal acute myeloid leukaemia. *Br. J. Haematol.* 2012;158(2):208–215.
10. Zhou J-D, Yang L, Zhu X-W, et al. Clinical significance of up-regulated ID1 expression in Chinese de novo acute myeloid leukemia. *Int. J. Clin. Exp. Pathol.* 2015;8(5):5336–44.
11. Cancer Genome Atlas Research Network, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 2013;368(22):2059–74.
12. Lima AS, De Mello MR, Fernandes E, et al. Clinical outcomes of patients with acute myeloid leukemia: Evaluation of genetic and molecular findings in a real-life setting. *Blood*. 2015;126(15):.
13. Benicio MTDLMT de LMTDL, Ribeiro AFTAFT, Américo AD, et al. Evaluation of the European LeukemiaNet recommendations for predicting outcomes of patients with acute myeloid leukemia treated in low- and middle-income countries (LMIC): A Brazilian experience. *Leuk. Res.* 2017;60:109–114.
14. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424–447.
15. Lasorella A, Benezra R, Iavarone A. The ID proteins: master regulators of cancer stem cells and tumour aggressiveness. *Nat. Rev. Cancer*. 2014;14(2):77–91.

16. Sun Y, Lai X, Yu Y, et al. Inhibitor of DNA binding 1 (Id1) mediates stemness of colorectal cancer cells through the Id1-c-Myc-PLAC8 axis via the Wnt/β-catenin and Shh signaling pathways. *Cancer Manag. Res.* 2019;Volume 11:6855–6869.

17. Ling M-T. Id-1 expression induces androgen-independent prostate cancer cell growth through activation of epidermal growth factor receptor (EGF-R). *Carcinogenesis*. 2003;25(4):517–525.

18. Singh SK, Singh S, Gadomski S, et al. Id1 Ablation Protects Hematopoietic Stem Cells from Stress-Induced Exhaustion and Aging. *Cell Stem Cell*. 2018;23(2):252-265.e8.

19. Wagner K, Zhang PP, Rosenbauer F, et al. Absence of the transcription factor CCAAT enhancer binding protein alpha results in loss of myeloid identity in bcr/abl-induced malignancy. *Proc. Natl. Acad. Sci. U. S. A.* 2006;103(16):6338–6343.

20. Grimwade D, Hills RK, Moorman A V., et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood*. 2010;116(3):354–65.

5 ARTIGO 2 – THREE-GENE PROGNOSTIC INDEX FOR PREDICTING OUTCOME IN ADULT AML PATIENTS

Artigo será submetido à revista *Blood* – Fator de impacto: 22,113; Qualis/CAPES Ciências Biológicas I: A1

Aleide S Lima,[1] Juan L Coelho-Silva,[2] Matheus F Bezerra,[3] Diego A Pereira-Martins,[4] Amanda M G de Aguiar,[1] Raul M Falcão,[5] Maria L S Bandeira,[1] Pedro L Franca-Neto,[1] Marinus M Lima,[1] Reijane A de Assis,[6] Marcos A Bezerra,[1] Antonio R Lucena-Araujo[1]*

Affiliations: [1]Department of Genetics, Federal University of Pernambuco, Recife, Brazil; [2]Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA; [3]Department of Microbiology, Aggeu Magalhães Institute, Fiocruz, Recife, Pernambuco, Brazil; [4]Department of Hematology, Cancer Research Centre Groningen, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands; [5]Bioinformatics Multidisciplinary Environment (BioME), Metrópole Digital Institute, Federal University of Rio Grande do Norte; [6]Cancer Hospital of Pernambuco, Recife, Pernambuco, Brazil.

Abstract

Refinement of risk stratification and therapy-decision algorithms for acute myeloid leukemia (AML) is urgent. Here, we developed a transcriptome-based prognostic index (TPI) based on the expression of 3 genes (*IGF2BP3*, *PDE7B* and *ST6GALNAC4*) able to further refine the European LeukemiaNet (ELN) risk stratification for newly diagnosed non-acute promyelocytic leukemia (APL) AML. Using the Cancer Genome Atlas (TCGA) database, we compared the transcriptome of two groups of patients (disease-free and relapsed) classified according to their outcome status after hematopoietic stem cell transplantation. The leading edge genes identified by the Gene-Set Enrichment Analysis were screened by Cox proportional hazard model according to its overall survival predictive capacity. The TPI was defined by computing the weight of each variable according to the univariate hazard ratio. TPI was an was an independent predictor for both OS and DFS curves (both $p<0.001$) in the training cohort (n=121, TCGA database) and was able to refine the ELN classification. The TPI OS and DFS prognostic capacity was validated in an external cohort publicly available (n=408, GSE6891; both $p<0.001$), but not in a cohort of 108 AML accompanied in a reference center in Brazil ($p=0.279$ and $p=0.103$, respectively). The divergent results may be explained by the number of patients enrolled, the quantification gene expression method and treatment heterogeneity. The TPI proved to be an independent prognostic factor for predicting the outcome of newly diagnosed non-APL AML patients and was able to refine the ELN risk stratification, although it has not been validated in a real-life cohort.

Introduction

Acute myeloid leukemia (AML) is a heterogeneous group of clonal diseases characterized by the accumulation of immature blast cells initially in the bone marrow (BM) and eventually in peripheral blood and other organs[1]. Its malignant phenotype derives from the sequential acquisition of somatic mutations that disrupts cell differentiation, proliferation and apoptosis mechanisms in haematopoietic precursors. [1] Given its wide morphological, molecular and clinical heterogeneity[2-4], there has been massive efforts to identify genetic prognostic markers that can improve the algorithms for treatment decision.[2, 5-9]

Currently, the most well accepted prognostic risk classification in clinical practice is the one proposed by the collaborative group European LeukemiaNet (ELN).[3] While valuable for outcome prediction in the good and adverse prognosis groups, further refinement of this algorithm is necessary to reduce the intragroup heterogeneity that still exists in the intermediate group[10], that pools together patients with rather distinct clinical and molecular backgrounds[2, 3, 5]. Refining the prediction for this particular group is also important to identify which patients might benefit from allogeneic hematopoietic stem cell transplantation (HSCT) or consolidation chemotherapy as a post-remission treatment.[3, 11, 12]

Despite the current advances in molecular methods[2, 5, 7, 13], transcriptome studies and gene expression data are still difficult to translate into prognosis evaluation for clinical practice, as there is significant variation across distinct quantitative methodologies, with differences in sensitivity and difficulties in defining the ideal cut-off points. Nevertheless, a panel of prognostic biomarkers

that includes the expression of the *EVI1* and *BAALC* genes has been validated by several centers and has been shown to refine the risk stratification of AML.[14-17] Thus, it is possible that prognostic biomarkers based on gene expression can be used in the diagnostic routine with due standardization.

In this context we propose a transcriptome-based prognostic index (TPI) from the additive impact of the expression of three genes (*PDE7B*, *IGF2BP3*, *ST6GALNAC4*), in an attempt to propose a refinement of the ELN risk stratification. Of these three genes, only the expression of the *PDE7B* gene has been reported to be associated with the clinical outcome of AML patients.[18] To date, for the first time the *IGF2BP3* and *ST6GALNAC4* genes are associated with the outcome of AML patients.

Methods

Public datasets

The training cohort for the TPI design were obtained from publicly available dataset The Cancer Genome Atlas database[5] (http://cbioportal.org) and were composed of 121 *de novo* AML patients aged 18 or over, excluding APL. Gene expression data were obtained by the RNA sequencing methodology. As a validation cohort, 408 adult patients with *de novo* AML from the public microarray database GSE6891[19] (https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=gse6891) were included, after excluding patients without complete data or diagnosed with APL (will be named here GSE6891 validation cohort).

Patients

The TPI was further validated in a cohort from a real-life setting (will be named here Brazilian validation cohort), that consisted of 108 AML patients diagnosed

from October 2010 to October 2020 at three reference centers for the treatment of AML patients, located in Recife, Brazil. The patients were diagnosed according to the cytomorphological, immunophenotypic and genetic World Health Organization (WHO) criteria. Only patients with *de novo* AML and aged 18 or over were considered, excluding APL. All patients were treated with cytotoxic chemotherapy following the "7+3" treatment protocol and twelve patients underwent to allogeneic HSCT after remission. Further details on the treatment have been published earlier.[20] The study was approved by the Research Ethics Board of each participating institute (CAAE: 77527717.4.0000.5208) and adhered to the tenets of the Declaration of Helsinki.

Candidate genes selection

The TPI design was divided into two stages: the first stage consisted of pre-selection of candidate genes according to Gene Set Enrichment Analysis (GSEA) after differential expression analysis between the two groups described below, and the second stage consisted of selection of genes with the greatest impact on patient outcome (Figure 1).

In the first stage, only patients from the TCGA cohort assigned to the intermediate I and II risk groups of the 2010 European LeukemiaNet (ELN) risk classification[21] were included. From those, only patients who received a standard “7+3” chemotherapy regimen[3] as induction therapy and who underwent allogeneic HSCT in the first complete remission as consolidation therapy were selected. Thus, 16 patients were divided into two groups: relapse group, which included patients who relapsed after HSCT (n=7), and disease-free group, which included patients who were disease-free for more than 24

months (n=7). Two patients with DFS less than 24 months were excluded from the disease-free group (Supplemental table 1). GSEA was performed using the preranked GSEA tool (GSEA v.4.1.0 software). The ranked list of genes obtained from the differential expression analysis between the relapsed and disease-free groups, calculated using the metric of the T test, was used as input data. The hallmark and curated gene sets included in the analyzes (including only those that used Homo sapiens as the study organism) were obtained from the MSigDB v7.2 database.

After preranked GSEA, only the gene sets enriched in the relapsed group with FDR<0.05 were considered. The leading-edge genes from these sets were selected for the second stage of candidate genes selection. This step included 121 *de novo* AML patients (non-APL) from the TCGA cohort who received standard “7+3” chemotherapy as induction therapy[3]. Survival analysis by uni and multivariate Cox regression models was applied to test the association with OS of the 2999 leading-edge genes. Cox regression models were calculated by the *SurvivalAnalysis* package of the R program. The continuous variable of gene expression was dichotomized into high and low expression groups based on choosing the best cut-off point to predict overall survival (OS) using the *cutpointr* package of R 3.3.2. In the multivariate analysis, the confounding variables were age, white blood cell count (WBC) and ELN2010 risk classification. After survival analyses, the following filters were applied to select the genes with the greatest impact on the OS of AML patients: 1. univariate Cox analysis p-value <0.001; 2. p-value of multivariate Cox analysis <0.005; 3. Hazard ratio (HR) from univariate Cox analysis ≥ 1.5; 4. AUC > 0.60; 5. mode

percentage <1.0. Three genes with the greatest impact on OS in AML patients were selected: *IGF2BP3*, *PDE7B* e *ST6GALNAC4* (Supplemental figure 1).

TPI design strategy

The same cohort used in the second step of selection of candidate genes (n=121, TCGA) was used for TPI design. The gene expression values were normalized by the median, so that it was possible to maintain the same cut-off point between the different cohorts. The HR values from the univariate Cox analysis were used to assign the weight of each variable (*IGF2BP3*, *PDE7B* e *ST6GALNAC4*) in the elaboration of the index. As the HR values from the three variables were very similar to each other (Supplemental table 2), we assigned a weight of 1 to the 3 variables when the gene was highly expressed. The TPI value for each patient was given from the sum of the weights of each variable (low expression: 0; high expression: 1). Thus, the TPI corresponded to a continuous variable, ranging from 0 to 3. After inspection of the Kaplan-Meyer curves (Supplemental figure 2), using overall survival as the outcome variable, the continuous variable of the TPI was categorized into 3 risk groups: favorable (no genes with high expression), intermediate (1 or 2 genes with high expression) and adverse (3 genes with high expression).

RT-qPCR assay

To validation of TPI in the Brazilian validation cohort, we evaluated the expression of the 3 selected genes through the RT-qPCR method. The material used for RT-qPCR analysis were obtained at diagnosis from bone marrow aspirates. Gene expression levels were determined by SYBR®Green chemistry (Applied Biosystems, Foster City, CA) and the relative expression of the genes

was determined by the 2-ΔΔCq comparative method. For more details on the qPCR methodology see the Supplementary methods.

Validation of TPI

To validate the TPI in the GSE6891 validation cohort and in the Brazilian validation cohort, the determination of the TPI was performed following the criteria defined in the training cohort. The gene expression values were normalized to the median so that it was possible to use the same cutoff point as the training cohort (IGF2BP3: 0.8303; PDE7B: 0.7450; ST6GALNAC4: 1.1458).

Statistics analysis and clinical endpoints

Fisher's exact test or Chi-square test, as appropriate, was used to compare categorical variables. Mann-Whitney, t-test, Kruskal-Wallis or ANOVA test, as appropriate, were used to compare continuous variables. Overall survival (OS) and disease-free survival (DFS) were estimated using the Kaplan–Meier method. OS was defined as the time from diagnosis to death from any cause; those alive or lost to follow-up were censored at the date last known alive. Early mortality was defined as death occurring within 30 days from diagnosis. For patients who achieved CR, DFS was defined as the time from CR achievement to the first adverse event: relapse, or death from any cause, whichever occurred first. The log-rank test was used for comparisons of Kaplan–Meier curves. The Cox proportional hazards regression model for multivariate analysis was used to assess the independence of TPI variable for the clinical outcome variables. As evidence of the validity of the prognostic index, we used the area under the curve (AUC) and the kappa agreement coefficient. For the calculation of AUC value, Survival and pROC R libraries were used to fit Cox proportional hazards regression model and to build ROC curves respectively. Two linear models

were created, and ROC curve was associated for each model under 95% confidence interval (CI) to calculate the area under the curve (AUC) for each variable (ELN2010 and TPI). To evaluate which potential indicator (ELN2010 or TPI) had a better performance, we used 10,000 bootstrap resampling and calculated in each iteration the difference between ROC AUC defined as ΔAUC = AUC_{TPI} – $AUC_{ELN2010}$. All *P*-values were two sided with a significance level of 0.05. All calculations were performed using Stata Statistic/Data Analysis version 9 (Stata Corporation, USA), IBM SPSS Statistics version 19 and R 3.3.2 (The CRAN project, www.r-project.org) software.

Results

Transcriptomic prognostic index modeling

There was no statistically significant difference in any of the demographic, clinical and laboratory variables analyzed between the relapse and disease-free groups of the first stage of gene selection (Table 1). In the *preranked* GSEA, 521 gene sets were positively enriched in the relapse group and were composed of 2999 leading edge genes. The group of patients who relapsed after allogeneic HSCT were positively enriched with a transcriptional program of hematopoietic stem cells (HSC) (eg. JAATINEN_HEMATOPOIETIC_STEM_CELL_UP) and embryonic stem cells (ESC) (eg. WONG_EMBRYONIC_STEM_CELL_CORE). In addition, they were positively enriched in gene sets of the myc oncogene targets (eg. HALLMARK_MYC_TARGETS_V1), gene sets associated with control and progression of cell cycle (eg. HALLMARK_E2F_TARGETS; REACTOME_S_PHASE; REACTOME_G2_M_MARKPOINTS) and mitochondrial metabolism (eg.

HALLMARK_OXIDATIVE_PHOSPHORYLATION), Krebs cycle (eg. KEGG_CITRATE_CYCLE_TCA_CYCLE), fatty acid oxidation (eg. REACTOME_MITOCHONDRIAL_FATTY_ACID_BETA_OXIDATION) and mitochondrial translation (eg. REACTOME_MITOCHONDRIAL_TRANSLATION) (Figure 2; Supplemental table 4)

The overexpression of the 3 genes (*IGF2BP3*, *PDE7B*, *ST6GALNAC4*) selected to compose the TPI was individually associated with a worse OS and DFS (Supplemental figure 1). The TPI was composed of three risk groups: the favorable group (14.9% of patients; 18/121) included patients with low expression of the three genes, while the intermediate group (66.1% of patients; 80/121) had one or two of the three genes with high expression, and the adverse group (19% of patients; 23/121) had high expression of the three genes.

Clinical and laboratory features and outcome of training cohort patients

There was no statistical difference between the three TPI risk groups regarding any of the demographic and clinical-laboratory variables analyzed in the TCGA training cohort (Table 2). The three TPI risk groups were different regarding OS%3y ($p<0.001$) and 3 years DFS rate (DFS%3y) ($p<0.001$) in the training cohort. The TPI adverse risk group had lower OS%3y (4.3%) and DFS%3y (0.0%) than the intermediate (OS%3y= 41.6%; DFS%3y= 29.4%, respectively) and favorable groups (OS%3y= 82.5%; DFS%3y= 68%, respectively) (Figure 3A and 3B, respectively). Furthermore, in multivariate analysis, TPI was an independent variable for predicting both OS and DFS when adjusted to age, WBC and ELN2010[21] (Figure 3G and 3H, respectively).

ROC curves showed that the TPI had a higher AUC (0.73; 95%CI, 0.65-0.80) than ELN2010 (0.60; 95%CI, 0.50-0.70). Also, the ROC AUC difference between ELN2010 and TPI was significant (ΔAUC, 0.13; 95%CI, 0.02-0.24; p, 0.018), thus we observe that the later model had a better performance of a prediction model (Supplemental figure 3). Furthermore, the kappa agreement coefficient between the TPI and ELN indices was 0.132 (standard deviation = 0.064), demonstrating low redundance among them. Next, we assessed whether the TPI was able to refine the prognosis of each ELN2010 risk groups (Figure 4A). Due to the low number of patients in the intermediate I and II risk groups, we merged those into a single group. Remarkably, the TPI was able to identify a group of patients who had adverse outcome in favorable and intermediate ELN2010 risk groups. For patients in the favorable ELN2010 group (n= 25 patients), who had OS%3y and DFS%3y of 50.3% and 41.7% respectively (Figure 4B and 4F, respectively), the TPI was able to identify three different risk groups for OS (OS%3y= favorable TPI: 83.3%, intermediate TPI: 44.3%, adverse TPI: 0.0%; p<0.001) and DFS (DFS%3y= favorable TPI: 65.5%, intermediate TPI: 34.1%, adverse TPI: 0.0%; p=0.001) (Figure 4C and 4G, respectively). When we evaluated patients in the intermediate ELN2010 risk group (n= 47 patients), who had OS%3y and DFS%3y of 35.4% and 22.8%, respectively (Figure 4B and 4F, respectively), the TPI was also able to identify three different risk groups for OS (OS%3y= favorable TPI: 75.0%, intermediate TPI: 38.0%, adverse TPI: 10.0%; p= 0.002), but it was not observed difference between the TPI groups in relation to DFS (DFS%3y= favorable TPI: 50.0%, intermediate TPI: 24.5%, adverse TPI: 0.0%; p= 0.073) (Figure 4D and 4H, respectively). Similarly, when evaluating the adverse risk group ELN2010 (n=

25 patients; OS%3y= 32%; DFS%3y= 36.4%) (Figure 4B and 4F, respectively), the TPI was able to identify three different risk groups regarding OS (OS%3y= favorable TPI: 100%, intermediate TPI: 40.0%, adverse TPI: 0.0%; p<0.001), but no difference was observed between the TPI risk groups in relation to DFS (DFS%3y= favorable TPI: 50.0%, intermediate TPI: 31.7%, adverse TPI: censored at 17 months with 25% probability; p=0.280) (Figure 4E and 4I, respectively).

Application of the TPI in the GSE6891 validation cohort

The favorable risk group comprised 13.5% (55/408) of patients, while 74% (302/408) of patients were classified in the intermediate group and 12.5% (51/408) of patients in the adverse group (Table 2).

Patients in the adverse risk group had worse OS%3y and 3-year event-free survival (EFS%3y) (23.7% and 22.5%) compared to patients in the favorable risk groups (OS%3y: 64.4%; EFS%3y: 63.0%) and intermediate (OS%3y: 40.6%; EFS%3y: 43.8%) (p<0.001, both) (Figures 3C e 3 D, respectively). In addition, the TPI was an independent variable for predicting a worse outcome for both OS and EFS (Figures 3I and 3J, respectively) and was able to refine the ELN2010 risk classification in the GSE6891 validation cohort. (Supplemental figure 4).

Application of the TPI in the Brazilian validation cohort

In the favorable, intermediate and adverse TPI group, 18 (16.7%), 64 (59.3%) and 26 (24.1%) patients were included, respectively. The median follow-up of the Brazilian validation cohort was 33 months. The 3-years OS rate was 20.9% and a median OS of only 8.8 months (95%CI). Overall, 70/108 (65.4%) patients achieved CR. . Evaluating patients from each of the risk groups, 14/18 (77.8%)

of patients in the favorable risk group, 42/63 (66.6%) of patients in the intermediate risk group and 14/26 (53.8%) of patients in the adverse risk group achieved complete remission (P=0.247). There was no difference between the three TPI risk groups regarding OS (favorable: OS%3y= 34%; intermediate: OS%3y= 14.5%; adverse: 28.1%; p=0.279)(Figure 3E), neither DFS (favorable: DFS%3y= 49.5%; intermediate: DFS%3y= 11.6%; adverse: 52.9%; p= 0.103) (Figure 3F).

Gene set enrichment analysis

To understand the transcriptional mechanisms involved in these three prognostic risk groups, we performed the *preranked* GSEA comparing the transcriptomes of these three groups. We observed that increased TPI risk was associated with upregulation of gene sets enriched in leukemic stem cell (LSC)/HSC and with transcriptional programs of important pathways for LSCs survival and self-renewal as well as cell proliferation and differentiation. Furthermore, the increased risk was positively enriched with gene sets associated with translation and metabolism (Figure 5).

More specifically, the adverse group was positively enriched with gene sets expressed in LSC/HSC (eg. EPPERT_CE_HSC_LSC) when compared to the intermediate or favorable group. The adverse group was also positively enriched with gene sets involved in the JAK/STAT signaling pathway and target genes of the myc transcription factor (eg. HALLMARK_IL6_JAK_STAT3_SIGNALING; BILD_MYC_ONCOGENIC_SIGNATURE). Furthermore, the adverse group showed high metabolic activity represented by the enrichment of gene sets associated with metabolism, translation and mitochondrial function. The

intermediate group, when compared to the favorable group, similarly to the adverse group, was positively enriched with gene sets involved in the JAK/STAT, PI3K/AKT, NOTCH and target genes of the myc transcription factor signaling pathways, and with metabolic process (Figure 5).

Discussion

Here, we proposed a prognostic score based on the expression profile of three genes (*PDE7B*, *IGF2BP3*, *ST6GALNAC4*), which was able to define three risk groups (favorable, intermediate and adverse) according to OS and DFS outcomes and it was validated in a cohort of 408 adult non-APL AML patients. This scoring system was an independent prognosis factor when adjusted for confounders in the TCGA training and GSE6891 validation cohort.

The TPI showed a predictive ability that was able to refine the ELN2010 classification in the training and GSE6891 validation cohorts. Notably, the TPI was able to identify a subset of patients with an adverse outcome within both the favorable and intermediate ELN risk groups. Therefore, it is possible that patients with an adverse outcome for TPI, but classified as favorable by ELN, could benefit more from allogeneic HSCT than the regular maintenance chemotherapy after the first CR.[3, 11, 22]

The genes selection that composed the TPI consisted of selecting targets that could be related to the most aggressive phenotype of relapse after allogeneic HSCT, using the GSEA. The choice for the GSEA tool, instead of selecting the most differentially expressed genes, for the subsequent analysis of survival, lies in the fact that the analysis by gene sets also allows the detection of genes with subtle changes in expression, but which may play a relevant role on the biology of AML.[23] The group of patients who relapsed after allogeneic HSCT were

positively enriched with both HSC and ESC transcriptional programs, MYC oncogene target gene sets and gene sets associated with cell cycle control and progression. There is a great body of evidence showing that AML with HSC transcriptional signature are associated with worse prognosis and higher risk of relapse.[24-27] The MYC oncogene had be shown to play an important role in the proliferation, self-renewal and survival of stem cell, like ESC and LSC, and chemotherapy resistance in AML.[28-33] Furthermore, *MYC* is commonly overexpressed in several AML subtypes with different mutational profiles[34, 35] and it's overexpression is capable of inducing an AML-like disease, with similar transcriptional programs, in murine model[36-38]. The relapse group was also enriched with gene sets associated with mitochondrial metabolism. Recent studies have shown that leukemic cells depend mainly on oxidative phosphorylation for survival and the ability increase mitochondrial metabolism is possibly a hallmark of chemotherapy resistance in vivo in AML.[39-41]

Among TPI genes, only *PDE7B* had already been associated with outcome in AML patients. In a study involving 632 AML patients with normal karyotype, overexpression of *PDE7B* was an independent predictor for a poor outcome.[18] The *PDE7B* gene encodes an phosphohydrolase that acts on the hydrolysis of cAMP to AMP, being one of the responsible for the reduction of intracellular cAMP levels.[42, 43] In turn, cAMP is one of the oldest known cell signaling molecules[44] and is involved in several cellular processes such as proliferation control, apoptosis, anti-inflammatory response and homing.[45, 46] In an *in vitro* study using PDE7B inhibitors in cells from chronic lymphocytic leukemia patients, a pro-apoptotic effect of inhibition of this protein was observed, which was associated with an intracellular cAMP increased.[47] Furthermore, in an study

using both trans-retinoic acid-sensitive and resistant APL cell lines and murine models, the use of a cAMP analog caused cell growth arrest and increased cell differentiation.[45]

The *IGF2BP3* gene encodes an oncofetal protein that regulates the stability, degradation and localization of target mRNAs, as well as miRNA biogenesis.[48] Recent data shows that *IGF2BP3* is play a role in the disruption of cell cycle, inhibition of apoptosis, CSC self-renewal and cell migration.[49-55] *IGF2BP3* overexpression has been associated with adverse prognosis in several solid tumors,[56-59] but only one study demonstrated the impact of this gene expression on the outcome of pediatric acute lymphoblastic leukemia .[60] To date, no studies has been published associating the expression of this gene with the outcome of AML patients.

The human *ST6GALNAC4* gene was first identified and characterized in 2000 by Harduin-Lepers et al.[61] This gene encodes a glycosyltransferase that transfers sialic acid to residues containing N-acetylgalactosamine.[61] These sialylated sugar chains are present in the cell membrane and can contribute to evasion of tumor cells from cytotoxic mechanisms or serve as ligands for a wide variety of molecules, such as the Thomsen-Friedenreich antigen (TF).[62-64] Several studies have shown that TF overexpression on the cancer cell surface plays an active role in tumor progression and metastasis through interaction with galectins.[63, 65] *ST6GALNAC4* overexpression has been associated with tumor progression and invasiveness in lung cancer and follicular thyroid carcinoma. [65, 66] To date, this is the first work that describes the association of the *ST6GALNAC4* overexpression with a poor clinical outcome in AML patients. However, an *in vitro* study using doxorubicin-resistant AML cell lines indicated

that sialylation alterations were involved in the development of multidrug resistance.[67]

In the GSEA, by comparing the transcriptome of the three TPI risk groups, we observed that increased risk was positively enriched with HSC/LSC gene sets. As previously mentioned, several reports in the literature have shown that AML with a transcriptomic signature of HSC/LSC tend to have a worse outcome, with greater chances of relapse and death, in both adult and pediatric patients.[6, 24, 68-72] Furthermore, the increased TPI risk was positively enriched with important transduction pathways (JAK/STAT, MYC oncogene target genes, PI3KAKT and NOTCH) for cell proliferation, LSC self-renewal, control of apoptosis and metabolism[26, 38, 73-77]. Constitutive activation of pathways such as JAK/STAT, MYC oncogene target genes, PI3KAKT and NOTCH tends to be associated with a worse outcome in patients with AML.[26, 75, 77]

Although the TPI was able to define distinct risk groups and refine the ELN2010 risk classification in the training and GSE6891 validation cohort, we were not able to validate this score using the qPCR methodology in the Brazilian validation cohort. Several factors may have limited the analyzes. First, different methodologies were used to determine the gene expression levels. In the training, GSE6891 validation and Brazilian validation cohorts was used RNA sequencing, microarray and qPCR methodology, respectively. The use of gene expression data as a prognostic tool is complex because it is a continuous variable, which makes it difficult to compare different methodologies. However, some prognostic markers based on gene expression, such as the overexpression of the *EVI1* and *BAALC* genes, are used in panels of prognostic biomarkers at diagnosis in a standardized way and have already been validated

by several centers.[14-17] In this context, it is possible that prognostic biomarkers based on gene expression can be used in the diagnostic routine with due standardization.

Second, this study was retrospective with small number of patients involved, a short follow-up time and was used a “real life” cohort of AML patients treated in Brazil, a low- and middle-income countries (LMIC). We advocate that clinical data obtained from real-world studies, if properly registered and with guaranteed accessibility, can provide representative evidence from routine practice about the clinical outcomes of patients[20]. However, real-world results should be analyzed with caution because many variables (e.g. drug unavailability, inadequate infrastructure, risk-adapted treatment, comorbidities, time from diagnosis to treatment initiation and several others socio-economic problems) [78-83] cannot be fully controlled may bias the results. In such cases, the studies must involve a large number of patients, which was not the case in our study. In addition, prospective studies are necessary to validate these results.

References

1. Estey EH. Acute myeloid leukemia: 2019 update on risk-stratification and management. *Am J Hematol.* 2018;93(10)**:**1267-1291.
2. Papaemmanuil E, Gerstung M, Bullinger L et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N Engl J Med.* 2016;374(23)**:**2209-2221.
3. Dohner H, Estey E, Grimwade D et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood.* 2017;129(4)**:**424-447.
4. Kishtagari A, Levine RL, Viny AD. Driver mutations in acute myeloid leukemia. *Curr Opin Hematol.* 2020;27(2)**:**49-57.
5. Cancer Genome Atlas Research N, Ley TJ, Miller C et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N Engl J Med.* 2013;368(22)**:**2059-2074.

6. Ng SW, Mitchell A, Kennedy JA et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. *Nature.* 2016;540(7633)**:**433-437.
7. Patel JP, Gonen M, Figueroa ME et al. Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. *N Engl J Med.* 2012;366(12)**:**1079-1089.
8. Wagner S, Vadakekolathu J, Tasian SK et al. A parsimonious 3-gene signature predicts clinical outcomes in an acute myeloid leukemia multicohort study. *Blood Adv.* 2019;3(8)**:**1330-1346.
9. DiNardo CD, Tiong IS, Quaglieri A et al. Molecular patterns of response and treatment failure after frontline venetoclax combinations in older patients with AML. *Blood.* 2020;135(11)**:**791-803.
10. Herold T, Rothenberg-Thurley M, Grunwald VV et al. Validation and refinement of the revised 2017 European LeukemiaNet genetic risk stratification of acute myeloid leukemia. *Leukemia.* 2020;34(12)**:**3161-3172.
11. Cornelissen JJ, Blaise D. Hematopoietic stem cell transplantation for patients with AML in first complete remission. *Blood.* 2016;127(1)**:**62-70.
12. Tallman MS, Wang ES, Altman JK et al. Acute Myeloid Leukemia, Version 3.2019, NCCN Clinical Practice Guidelines in Oncology. *J Natl Compr Canc Netw.* 2019;17(6)**:**721-749.
13. Tyner JW, Tognon CE, Bottomly D et al. Functional genomic landscape of acute myeloid leukaemia. *Nature.* 2018;562(7728)**:**526-531.
14. Alessandrini M, Kappala SS, Pepper MS. AMLprofiler: A Diagnostic and Prognostic Microarray for Acute Myeloid Leukemia. *Methods Mol Biol.* 2017;1633**:**101-123.
15. Brand J, van Vliet MH, de Best L et al. A standardized microarray assay for the independent gene expression markers in AML: EVI1 and BAALC. *Exp Hematol Oncol.* 2013;2(1)**:**7.
16. Kappala SS, Alessandrini M, Matlhako T, Beltchev E, Pool R, Pepper MS. Application of the AMLprofiler Diagnostic Microarray in the South African Setting. *Stem Cells Int.* 2017;2017**:**2560191.
17. Nomdedeu JF, Puigdecanet E, Bussaglia E et al. Feasibility of the AML profiler (Skyline Array) for patient risk stratification in a multicentre trial: a preliminary comparison with the conventional approach. *Hematol Oncol.* 2017;35(4)**:**778-788.
18. Cao L, Zhang W, Liu X et al. The Prognostic Significance of PDE7B in Cytogenetically Normal Acute Myeloid Leukemia. *Sci Rep.* 2019;9(1)**:**16991.
19. Verhaak RG, Wouters BJ, Erpelinck CA et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica.* 2009;94(1)**:**131-134.
20. Bezerra MF, Lima AS, Pique-Borras MR et al. Co-occurrence of DNMT3A, NPM1, FLT3 mutations identifies a subset of acute myeloid leukemia with adverse prognosis. *Blood.* 2020;135(11)**:**870-875.
21. Dohner H, Estey EH, Amadori S et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood.* 2010;115(3)**:**453-474.

22. Pratcorona M, Brunet S, Nomdedeu J et al. Favorable outcome of patients with acute myeloid leukemia harboring a low-allelic burden FLT3-ITD mutation and concomitant NPM1 mutation: relevance to post-remission therapy. *Blood.* 2013;121(14)**:**2734-2738.
23. Subramanian A, Tamayo P, Mootha VK et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci U S A.* 2005;102(43)**:**15545-15550.
24. Eppert K, Takenaka K, Lechman ER et al. Stem cell gene expression programs influence clinical outcome in human leukemia. *Nat Med.* 2011;17(9)**:**1086-1093.
25. Duy C, Li M, Teater M et al. Chemotherapy Induces Senescence-Like Resilient Cells Capable of Initiating AML Recurrence. *Cancer Discov.* 2021;11(6)**:**1542-1561.
26. Rodrigues A, Costa RGA, Silva SLR, Dias I, Dias RB, Bezerra DP. Cell signaling pathways as molecular targets to eliminate AML stem cells. *Crit Rev Oncol Hematol.* 2021;160**:**103277.
27. Zeisig BB, Fung TK, Zarowiecki M et al. Functional reconstruction of human AML reveals stem cell origin and vulnerability of treatment-resistant MLL-rearranged leukemia. *Sci Transl Med.* 2021;13(582).
28. Abraham SA, Hopcroft LE, Carrick E et al. Dual targeting of p53 and c-MYC selectively eliminates leukaemic stem cells. *Nature.* 2016;534(7607)**:**341-346.
29. Ge Y, Schuster MB, Pundhir S et al. The splicing factor RBM25 controls MYC activity in acute myeloid leukemia. *Nat Commun.* 2019;10(1)**:**172.
30. Call SG, Duren RP, Panigrahi AK et al. Targeting Oncogenic Super Enhancers in MYC-Dependent AML Using a Small Molecule Activator of NR4A Nuclear Receptors. *Sci Rep.* 2020;10(1)**:**2851.
31. Cartwright P, McLean C, Sheppard A, Rivett D, Jones K, Dalton S. LIF/STAT3 controls ES cell self-renewal and pluripotency by a Myc-dependent mechanism. *Development.* 2005;132(5)**:**885-896.
32. Wong DJ, Liu H, Ridky TW, Cassarino D, Segal E, Chang HY. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell.* 2008;2(4)**:**333-344.
33. Varlakhanova NV, Cotterman RF, deVries WN et al. myc maintains embryonic stem cell pluripotency and self-renewal. *Differentiation.* 2010;80(1)**:**9-19.
34. Mughal MK, Akhter A, Street L, Pournazari P, Shabani-Rad MT, Mansoor A. Acute myeloid leukaemia: expression of MYC protein and its association with cytogenetic risk profile and overall survival. *Hematol Oncol.* 2017;35(3)**:**350-356.
35. Ohanian M, Rozovski U, Kanagal-Shamanna R et al. MYC protein expression is an important prognostic factor in acute myeloid leukemia. *Leuk Lymphoma.* 2019;60(1)**:**37-48.
36. Luo H, Li Q, O'Neal J, Kreisel F, Le Beau MM, Tomasson MH. c-Myc rapidly induces acute myeloid leukemia in mice without evidence of lymphoma-associated antiapoptotic mutations. *Blood.* 2005;106(7)**:**2452-2461.

37. Beverly LJ, Varmus HE. MYC-induced myeloid leukemogenesis is accelerated by all six members of the antiapoptotic BCL family. *Oncogene.* 2009;28(9)**:**1274-1279.
38. Bulaeva E, Pellacani D, Nakamichi N et al. MYC-induced human acute myeloid leukemia requires a continuing IL-3/GM-CSF costimulus. *Blood.* 2020;136(24)**:**2764-2773.
39. Farge T, Saland E, de Toni F et al. Chemotherapy-Resistant Human Acute Myeloid Leukemia Cells Are Not Enriched for Leukemic Stem Cells but Require Oxidative Metabolism. *Cancer Discov.* 2017;7(7)**:**716-735.
40. Baccelli I, Gareau Y, Lehnertz B et al. Mubritinib Targets the Electron Transport Chain Complex I and Reveals the Landscape of OXPHOS Dependency in Acute Myeloid Leukemia. *Cancer Cell.* 2019;36(1)**:**84-99 e88.
41. Bosc C, Broin N, Fanjul M et al. Autophagy regulates fatty acid availability for oxidative phosphorylation through mitochondria-endoplasmic reticulum contact sites. *Nat Commun.* 2020;11(1)**:**4056.
42. Hetman JM, Soderling SH, Glavas NA, Beavo JA. Cloning and characterization of PDE7B, a cAMP-specific phosphodiesterase. *Proc Natl Acad Sci U S A.* 2000;97(1)**:**472-476.
43. Safavi M, Baeeri M, Abdollahi M. New methods for the discovery and synthesis of PDE7 inhibitors as new drugs for neurological and inflammatory disorders. *Expert Opin Drug Discov.* 2013;8(6)**:**733-751.
44. Dumaz N, Marais R. Integrating signals between cAMP and the RAS/RAF/MEK/ERK signalling pathways. Based on the anniversary prize of the Gesellschaft fur Biochemie und Molekularbiologie Lecture delivered on 5 July 2003 at the Special FEBS Meeting in Brussels. *FEBS J.* 2005;272(14)**:**3491-3504.
45. Guillemin MC, Raffoux E, Vitoux D et al. In vivo activation of cAMP signaling induces growth arrest and differentiation in acute promyelocytic leukemia. *J Exp Med.* 2002;196(10)**:**1373-1380.
46. Perez DR, Smagley Y, Garcia M et al. Cyclic AMP efflux inhibitors as potential therapeutic agents for leukemia. *Oncotarget.* 2016;7(23)**:**33960-33982.
47. Zhang L, Murray F, Zahno A et al. Cyclic nucleotide phosphodiesterase profiling reveals increased expression of phosphodiesterase 7B in chronic lymphocytic leukemia. *Proc Natl Acad Sci U S A.* 2008;105(49)**:**19532-19537.
48. Mancarella C, Scotlandi K. IGF2BP3 From Physiology to Cancer: Novel Discoveries, Unsolved Issues, and Future Perspectives. *Front Cell Dev Biol.* 2019;7**:**363.
49. Liao B, Hu Y, Herrick DJ, Brewer G. The RNA-binding protein IMP-3 is a translational activator of insulin-like growth factor II leader-3 mRNA during proliferation of human K562 leukemia cells. *J Biol Chem.* 2005;280(18)**:**18517-18524.
50. Jonson L, Christiansen J, Hansen TVO, Vikesa J, Yamamoto Y, Nielsen FC. IMP3 RNP safe houses prevent miRNA-directed HMGA2 mRNA decay in cancer and development. *Cell Rep.* 2014;7(2)**:**539-551.
51. Rivera Vargas T, Boudoukha S, Simon A et al. Post-transcriptional regulation of cyclins D1, D3 and G1 and proliferation of human cancer

cells depend on IMP-3 nuclear localization. *Oncogene.* 2014;33(22)**:**2866-2875.
52. Palanichamy JK, Tran TM, Howard JM et al. RNA-binding protein IGF2BP3 targeting of oncogenic transcripts promotes hematopoietic progenitor proliferation. *J Clin Invest.* 2016;126(4)**:**1495-1511.
53. Chen CL, Tsukamoto H, Liu JC et al. Reciprocal regulation by TLR4 and TGF-beta in tumor-initiating stem-like cells. *J Clin Invest.* 2013;123(7)**:**2832-2849.
54. Sheen YS, Liao YH, Lin MH et al. IMP-3 promotes migration and invasion of melanoma cells by modulating the expression of HMGA2 and predicts poor prognosis in melanoma. *J Invest Dermatol.* 2015;135(4)**:**1065-1073.
55. Samanta S, Sun H, Goel HL et al. IMP3 promotes stem-like properties in triple-negative breast cancer by regulating SLUG. *Oncogene.* 2016;35(9)**:**1111-1121.
56. Yan J, Wei Q, Jian W et al. IMP3 Predicts Invasion and Prognosis in Human Lung Adenocarcinoma. *Lung.* 2016;194(1)**:**137-146.
57. Ohashi R, Sangen M, Namimatsu S et al. Prognostic value of IMP3 expression as a determinant of chemosensitivity in triple-negative breast cancer. *Pathol Res Pract.* 2017;213(9)**:**1160-1165.
58. Liu H, Zeng Z, Afsharpad M et al. Overexpression of IGF2BP3 as a Potential Oncogene in Ovarian Clear Cell Carcinoma. *Front Oncol.* 2019;9**:**1570.
59. Wakita A, Motoyama S, Sato Y et al. IGF2BP3 Expression Correlates With Poor Prognosis in Esophageal Squamous Cell Carcinoma. *J Surg Res.* 2021;259**:**137-144.
60. Makinen A, Nikkila A, Haapaniemi T et al. IGF2BP3 Associates with Proliferative Phenotype and Prognostic Features in B-Cell Acute Lymphoblastic Leukemia. *Cancers (Basel).* 2021;13(7).
61. Harduin-Lepers A, Stokes DC, Steelant WF et al. Cloning, expression and gene organization of a human Neu5Ac alpha 2-3Gal beta 1-3GalNAc alpha 2,6-sialyltransferase: hST6GalNAcIV. *Biochem J.* 2000;352 Pt 1**:**37-48.
62. Schauer R. Sialic acids as regulators of molecular and cellular interactions. *Curr Opin Struct Biol.* 2009;19(5)**:**507-514.
63. Sindrewicz P, Lian LY, Yu LG. Interaction of the Oncofetal Thomsen-Friedenreich Antigen with Galectins in Cancer Progression and Metastasis. *Front Oncol.* 2016;6**:**79.
64. Wang L, Li S, Yu X et al. alpha2,6-Sialylation promotes immune escape in hepatocarcinoma cells by regulating T cell functions and CD147/MMP signaling. *J Physiol Biochem.* 2019;75(2)**:**199-207.
65. Reticker-Flynn NE, Bhatia SN. Aberrant glycosylation promotes lung cancer metastasis through adhesion to galectins in the metastatic niche. *Cancer Discov.* 2015;5(2)**:**168-181.
66. Miao X, Jia L, Zhou H et al. miR-4299 mediates the invasive properties and tumorigenicity of human follicular thyroid carcinoma by targeting ST6GALNAC4. *IUBMB Life.* 2016;68(2)**:**136-144.
67. Ma H, Zhou H, Song X, Shi S, Zhang J, Jia L. Modification of sialylation is associated with multidrug resistance in human acute myeloid leukemia. *Oncogene.* 2015;34(6)**:**726-740.

68. Gentles AJ, Plevritis SK, Majeti R, Alizadeh AA. Association of a leukemic stem cell gene expression signature with clinical outcomes in acute myeloid leukemia. *JAMA.* 2010;304(24)**:**2706-2715.
69. Metzeler KH, Maharry K, Kohlschmidt J et al. A stem cell-like gene expression signature associates with inferior outcomes and a distinct microRNA expression profile in adults with primary cytogenetically normal acute myeloid leukemia. *Leukemia.* 2013;27(10)**:**2023-2031.
70. Jung N, Dai B, Gentles AJ, Majeti R, Feinberg AP. An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. *Nat Commun.* 2015;6**:**8489.
71. Duployez N, Marceau-Renaut A, Villenet C et al. The stem cell-associated gene expression signature allows risk stratification in pediatric acute myeloid leukemia. *Leukemia.* 2019;33(2)**:**348-357.
72. Elsayed AH, Rafiee R, Cao X et al. A six-gene leukemic stem cell score identifies high risk pediatric acute myeloid leukemia. *Leukemia.* 2020;34(3)**:**735-745.
73. Heuser M, Sly LM, Argiropoulos B et al. Modeling the functional heterogeneity of leukemia stem cells: role of STAT5 in leukemia stem cell self-renewal. *Blood.* 2009;114(19)**:**3983-3993.
74. Oancea C, Ruster B, Brill B et al. STAT activation status differentiates leukemogenic from non-leukemogenic stem cells in AML and is suppressed by arsenic in t(6;9)-positive AML. *Genes Cancer.* 2014;5(11-12)**:**378-392.
75. Tabe Y, Tafuri A, Sekihara K, Yang H, Konopleva M. Inhibition of mTOR kinase as a therapeutic target for acute myeloid leukemia. *Expert Opin Ther Targets.* 2017;21(7)**:**705-714.
76. Mudgapalli N, Nallasamy P, Chava H et al. The role of exosomes and MYC in therapy resistance of acute myeloid leukemia: Challenges and opportunities. *Mol Aspects Med.* 2019;70**:**21-32.
77. Nepstad I, Hatfield KJ, Gronningsaeter IS, Reikvam H. The PI3K-Akt-mTOR Signaling Pathway in Human Acute Myeloid Leukemia (AML) Cells. *Int J Mol Sci.* 2020;21(8).
78. Rego MF, Pinheiro GS, Metze K, Lorand-Metze I. Acute leukemias in Piaui: comparison with features observed in other regions of Brazil. *Braz J Med Biol Res.* 2003;36(3)**:**331-337.
79. Lucena-Araujo AR, Souza DL, Morato de Oliveira F et al. Results of FLT3 mutation screening and correlations with immunophenotyping in 169 Brazilian patients with acute myeloid leukemia. *Ann Hematol.* 2010;89(2)**:**225-228.
80. Nucci M, Garnica M, Gloria AB et al. Invasive fungal diseases in haematopoietic cell transplant recipients and in patients with acute myeloid leukaemia or myelodysplasia in Brazil. *Clin Microbiol Infect.* 2013;19(8)**:**745-751.
81. Lima AS, de Mello MR, Fernandes E et al. Clinical outcomes of patients with acute myeloid leukemia: evaluation of genetic and molecular findings in a real-life setting. *Blood.* 2015;126(15)**:**1863-1865.
82. Benicio MTL, Ribeiro AFT, Americo AD et al. Evaluation of the European LeukemiaNet recommendations for predicting outcomes of patients with acute myeloid leukemia treated in low- and middle-income countries (LMIC): A Brazilian experience. *Leuk Res.* 2017;60**:**109-114.

83. Silveira DRA, Coelho-Silva JL, Silva WF et al. A multicenter comparative acute myeloid leukemia study: can we explain the differences in the outcomes in resource-constrained settings? *Leuk Lymphoma.* 2021;62(1)**:**147-157.
84. Grimwade D, Hills RK, Moorman AV et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood.* 2010;116(3)**:**354-365.

Table 1: Demographic, laboratory and clinical characteristics of the relapsed and disease-free patient groups.

Characteristics	**All cohort**	**Disease-free**	**Relapsed**	
	N= 14 (%)	**N= 7 (%)**	**N= 7 (%)**	***p-value***
Age (years), median	58.5	60.0	57	0.371
Range	23-69	35-69	23-63	
Gender				1.000
Female	8 (57.1)	4 (57.1)	4 (57.1)	
Male	6 (42.9)	3 (42.9)	3 (42.9)	
WBC ($x10^9$/L), median	11.95	11.5	27.6	0.655
Range	0.6-80.5	1.2-72.1	0.6-80.5	
BM Blasts (%), median	64.5	86.0	59.0	0.230
Range	35-99	42-99	35-89	
FAB				0.087
M0	1 (7.1)	1 (14.3)	0 (0.0)	
M1	5 (35.7)	4 (57.1)	1 (14.3)	
M2	4 (28.6)	0 (0.0)	4 (57.1)	
M4	3 (21.4)	1 (14.3)	2 (28.6)	
M5	1 (7.1)	1 (14.3)	0 (0.0)	
M6	-	-	-	
M7	-	-	-	
Cytogenetic[1]				0.280
Intermediate risk abnormality	6 (42.9)	4 (57.1)	2 (28.6)	
Normal karyotype	8 (57.1)	3 (42.9)	5 (71.4)	
Matched allogeneic HSCT type				0.577
Sibling donor	9 (64.3)	5 (71.4)	4 (57.1)	
Unrelated donor	5 (35.7)	2 (28.6)	3 (42.9)	
FLT3-ITD				0.577
Non-mutated	9 (64.3)	5 (71.4)	4 57.1)	
Mutated	5 (35.7)	2 (28.6)	3 (42.9)	
NPM1				0.237
Wild type	10 (71.4)	4 (57.1)	6 (85.7)	
Mutated	4 (28.6)	3 (42.9)	1 (14.3)	
Molecular risk[2]				0.299
Low risk	1 (7.1)	1 (14.3)	0 (0.0)	
High risk	13 (92.9)	6 (85.7)	7 (100.0)	

CEBPA				-
Non-biallelic	14 (100.0)	7 (100.0)	7 (100.0)	
Biallelic mutation	0 (0.0)	0 (0.0)	0 (0.0)	
ASXL1				0.299
Wild type	13 (92.9)	6 (85.7)	7 (100.0)	
Mutated	1 (7.1)	1 (14.3)	0 (0.0)	
TP53				-
Wild type	14 (100.0)	7 (100.0)	7 (100.0)	
Mutated	0 (0.0)	0 (0.0)	0 (0.0)	
RUNX1				1.0
Wild type	12 (85.7)	6 (85.7)	6 (85.7)	
Mutated	2 (14.3)	1 (14.3)	1 (14.3)	

Abbreviation: BM, bone marrow; WBC, White blood cells; FAB, French-American-British Classification.
1: Cytogenetic categories were defined following the MRC criteria.
2: Low risk comprises patients with mutated *NPM1* gene and non-mutated *FLT3*-ITD, while the high risk group comprises the genotypes *NPM1* non-mutated/*FLT3*-ITD non-mutated, *NPM1* non-mutated/*FLT3*-ITD mutated, *NPM1* mutated/*FLT3*-ITD mutated.

Table 2: Demographic, laboratory, and clinical characteristics of training cohort (TCGA), external validation cohort (GSE6891) and internal validation cohort (Brazilian) patients, as well as favorable, intermediate, and adverse risk groups for TPI.

	Training cohort					Brazilian cohort					GSE68914 cohort				
Characteristics		**TPI**			***p***		**TPI**			***p***		**TPI**			***p***
	All cohort	**Favorable**	**Intermediate**	**Adverse**		**All cohort**	**Favorable**	**Intermediate**	**Adverse**		**All cohort**	**Favorable**	**Intermediate**	**Adverse**	
	N= 121 (%)	**N= 18 (%)**	**N= 80 (%)**	**N= 23 (%)**		**N= 108 (%)**	**N= 11 (%)**	**N= 85 (%)**	**N= 12 (%)**		**N=408 (%)**	**N=39 (%)**	**N= 289 (%)**	**N=80 (%)**	
Age (years), median	55	53	55	58	0.389	47	50	46	48	0.515	45	38	46	44.5	0.056
Range	18-81	21-76	18-81	21-77		19-71	24-69	19-71	25-63		18-77	19-57	18-77	18-70	
Gender					0.731					0.906					0.697
Female	56 (46.3)	9 (50.0)	35 (43.8)	12 (52.2)		51 (47.2)	5 (45.5)	41 (48.2)	5 (41.7)		204 (50.0)	17 (43.6)	146 (50.5)	41 (51.3)	
Male	65 (53.7)	9 (50.0)	45 (56.3)	11 (47.8)		57 (52.8)	6 (54.5)	44 (51.8)	7 (58.3)		204 (50.0)	22 (56.4)	143 (49.5)	39 (48.8)	
WBC (x10^9/L), median	29.4	12.3	32.1	14.5	0.254	48.8	54.7	47.7	18.7	0.542	-	-	-	-	-
Range	0.6-297,4	2.5-297.4	0.6-223.8	0.8-171.9		1.2-300.0	5.7-24.3	1.2-300.0	1.3-175.1		-	-	-	-	-
BM Blasts (%), median	74.0	72.0	71.5	79.0	0.244	69.0	85.0	69.0	61.5	0.965	-	-	-	-	-
Range	30.0-100.0	40.0-99.0	30.0-100.0	43.0-97.0		20.0-97.0	26.0-97.0	20.0-97.0	36.0-97.0		-	-	-	-	-
FAB					0.382					0.925					0.050
M0	12 (10.0)	2 (11.1)	5 (6.3)	5 (21.7)		5 (5.3)	-	5 (6.8)	-		16 (4.1)	-	12	4 (5.1)	
M1	37 (30.8)	7 (38.9)	24 (30.4)	6 (26.1)		10 (10.6)	1 (10.0)	7 (9.5)	2 (20.0)		93 (23.5)	8 (20.5)	68	17 (21.8)	
M2	27 (22.5)	5 (27.8)	19 (24.1)	3 (13.0)		38 (40.4)	5 (50.0)	29 (39.2)	4 (40.0)		99 (25.1)	18 (46.2)	64	17 (21.8)	
M4	29 (24.2)	3 (16.7)	21 (26.6)	5 (21.7)		30 (31.9)	2 (20.0)	24 (32.4)	4 (40.0)		81 (20.5)	8 (20.5)	60	13 (16.7)	
M5	12 (10.0)	1 (5.6)	9 (11.4)	2 (8.7)		9 (9.6)	2 (20.0)	7 (9.5)	-		100 (25.3)	5 (12.8)	68	27 (34.6)	

M6	2 (1.7)	-	1 (1.3)	1 (4.3)		1 (1.1)	-	1 (1.4)	-		6 (1.5)	-	6	-	
M7	1 (0.8)	-	-	1 (4.3)		1 (1.1)	-	1 (1.4)	-		-	-	-	-	
Missing data[1]	1	-	1	1		14	6	11	6		13	-	11	2	
Cytogenetics[2]					0.132					0.369					<0.001
t(8;21)(q22;q22)	7 (5.9)	3 (16.7)	4 (5.1)	-		9 (16.7)	2 (40.0)	7 (16,3)	-		32 (8.2)	17 (43.6)	15 (5.5)	-	
inv(16)/t(16/16) (p16;q22)	10 (8.4)	1 (5.6)	8 (10.1)	1 (4.5)		4 (7.4)	1 (20.0)	3 (7.0)	-		31 (7.9)	6 (15.4)	21 (7.7)	4 (5.1)	
Complex karyotype	15 (12.6)	2 (11.1)	8 (10.1)	5 (22.7)		8 (14.8)	-	8 (18.6)	-		19 (4.8)	-	17 (6.2)	2 (2.5)	
Normal karyotype	64 (53.8)	11 (61.1)	45 (57.0)	8 (36.4)		27 (50.0)	2 (40.0)	21 (48.8)	4 (66.7)		195 (49.7)	12 (30.8)	134 (48.9)	49 (62.0)	
Intermediate risk abnormality	14 (11.8)	1 (5.6)	8 (10.1)	5 (22.7)		4 (7.4)	-	3 (7.0)	1 (16.7)		74 (18.9)	3 (7.7)	60 (21.9)	11 (13.9)	
Other poor risk abnormality	9 (7.6)	-	6 (7.6)	3 (13.6)		2 (3.7)	-	1 (2.3)	1 (16.7)		41 (10.5)	1 (2.6)	27 (9.9)	13 (16.5)	
Missing data[1]	2	-	1	1		54	6	42	6		16	-	15	1	
MRC risk[2]					0.178					0.185					<0.001
Favorable	17 (14.3)	4 (22.2)	12 (15.2)	1 (4.5)		13 (24.1)	3 (60.0)	10 (23.3)	-		63 (16.1)	23 (59.0)	36 (13.1)		
Intermediate	78 (65.5)	12 (66.7)	53 (67.1)	13 (59.1)		31 (57.4)	2 (40.0)	24 (55.8)	5 (83.3)		269 (68.6)	15 (38.5)	194 (70.8)		
Adverse	24 (20.2)	2 (11.1)	14 (17.7)	8 (36.4)		10 (18.5)	-	9 (20.9)	1 (16.7)		60 (15.3)	1 (2.5)	44 (16.1)		
Missing data[1]	2	-	1	1		54	6	42	6		16	-	15		
ELN 2010					0.057					0.492					<0.001
Favorable	47 (39.5)	12 (66.7)	31 (39.2)	4 (18.2)		18 (45.0)	4 (80.0)	12 (38.7)	2 (50.0)		121 (30.9)	29 (74.4)	80 (29.2)		
Intermediate I	30 (25.2)	3 (16.7)	22 (27.8)	5 (22.7)		8 (20.0)	1 (20.0)	7 (22.6)	-		137 (34.9)	6 (15.4)	90 (32.8)		
Intermediate II	17 (14.3)	1 (5.6)	11 (13.9)	5 (22.7)		4 (10.0)	-	3 (9.7)	1 (25.0)		69 (17.6)	3 (7.7)	55 (20.1)		
Adverse	25 (21.0)	2 (11.1)	15 (19.0)	8 (36.4)		10 (25.0)	-	9 (29.0)	1 (25.0)		65 (16.6)	1 (2.6)	49 (17.9)		

Missing data[1]	2	-	1	1	68	6	54	8	16	-	15	

Abbreviation: BM, bone marrow; FAB, French-American-British Classification; MRC: Medical Research Council; ELN: European Leukemia Net classification risk.
1: Missing data was excluded from p-value calculation
2: Cytogenetic categories were defined following the MRC criteria.[84]

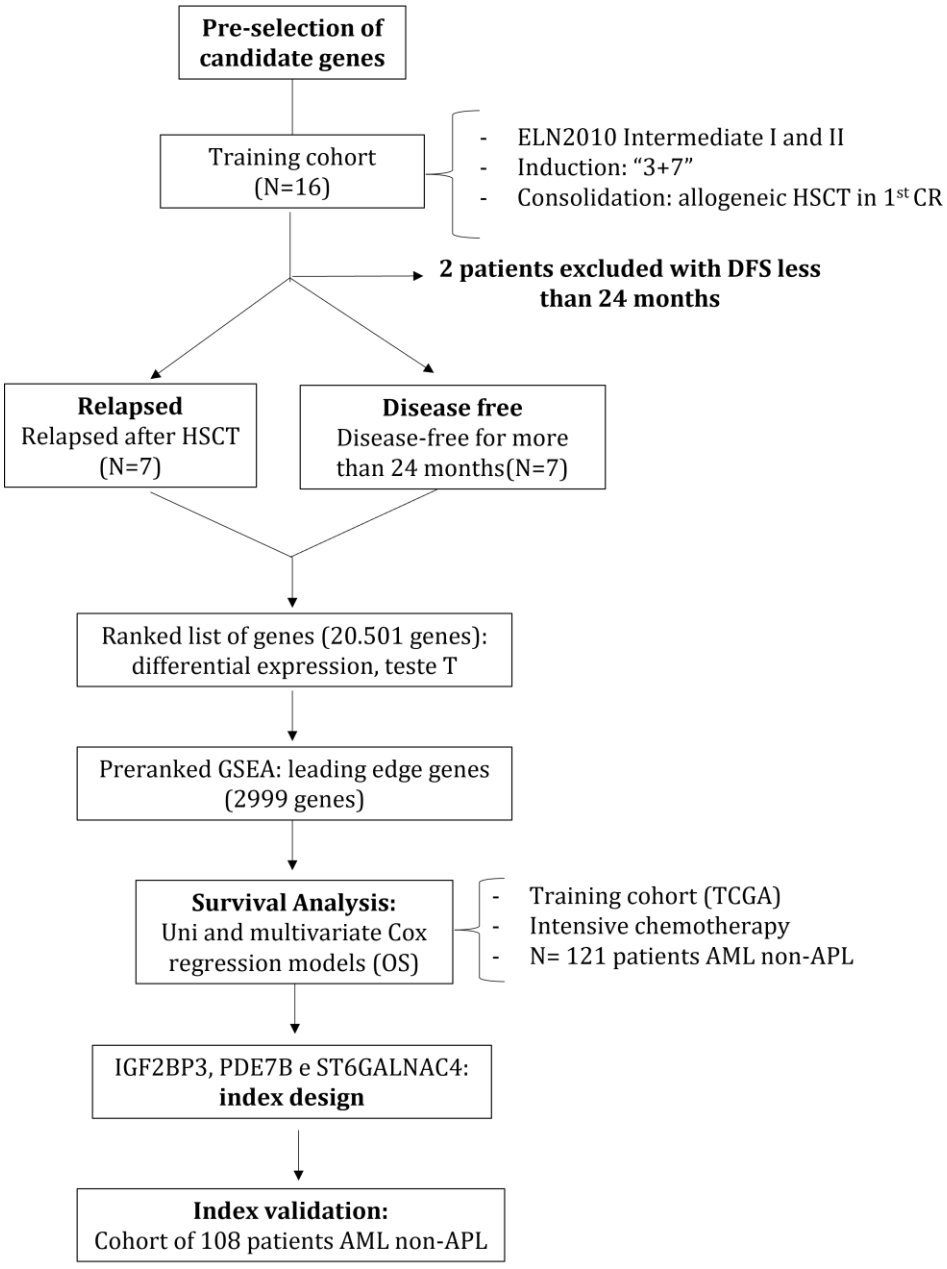


Figure 1: Study design flowchart.

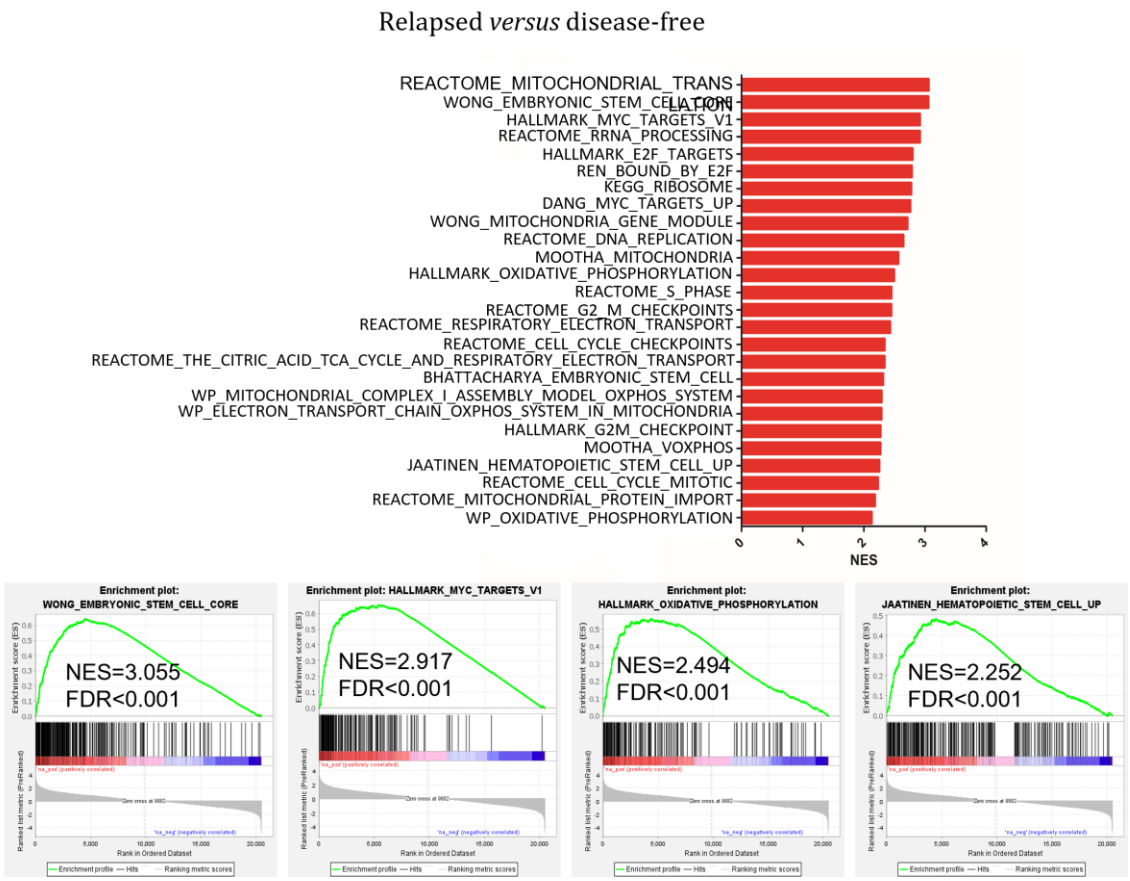


Figure 2: Gene Set Enrichment Analysis (GSEA) Preranked comparing the transcriptome of relapsed and disease-free groups.

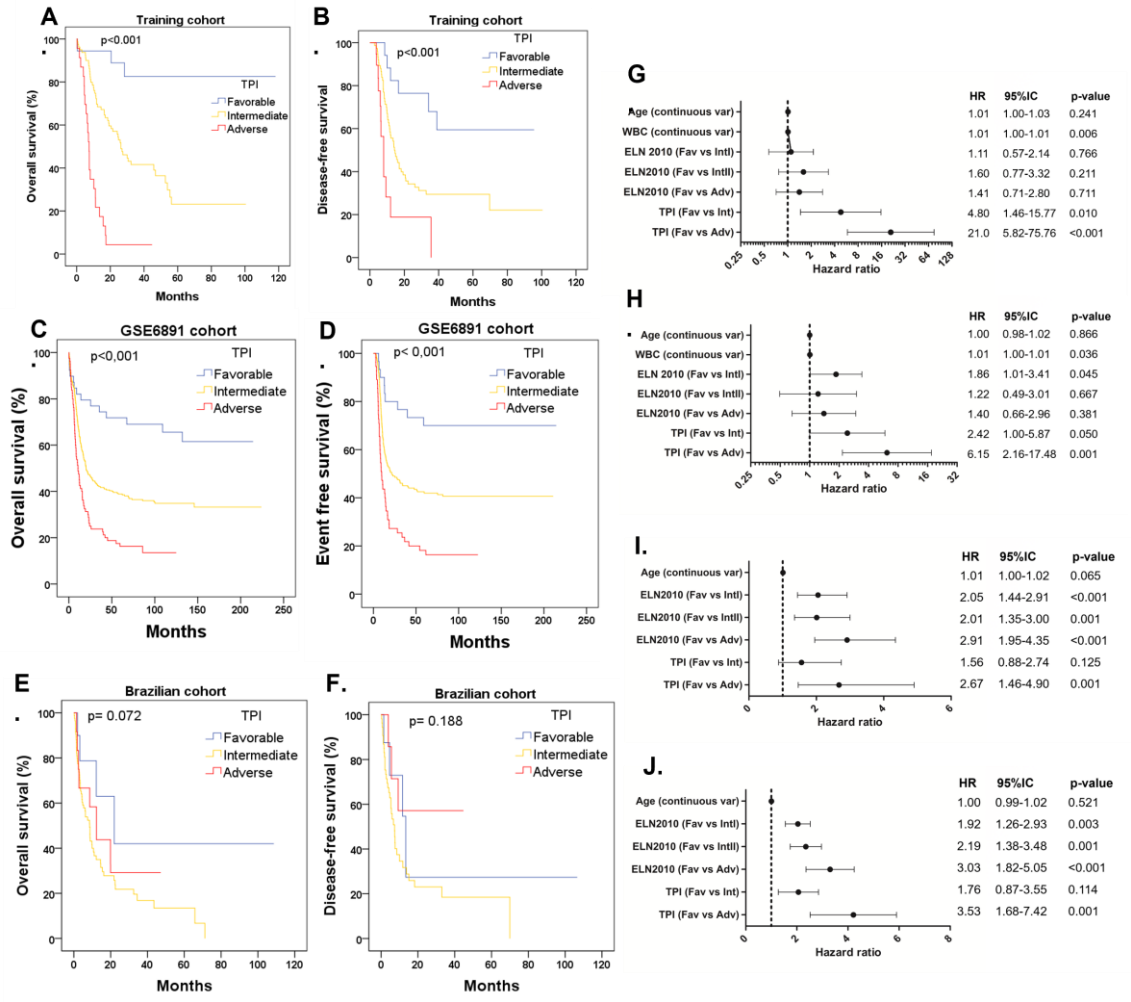


Figure 3: Comparison of OS (A) and DFS (B) curves for patients in the favorable, intermediate and adverse TPI risk group of the training cohort. Comparison of OS (C) and EFS (D) curves for patients in the favorable, intermediate and adverse TPI risk group of the GSE6891 external validation cohort. Comparison of OS (E) and DFS (F) curves for patients in the favorable, intermediate and adverse TPI risk group of the Brazilian internal validation cohort. Multivariate Cox model for OS (G) and DFS (H) of the training cohort and for OS (I) and EFS (J) of the GSE6891 external validation cohort.

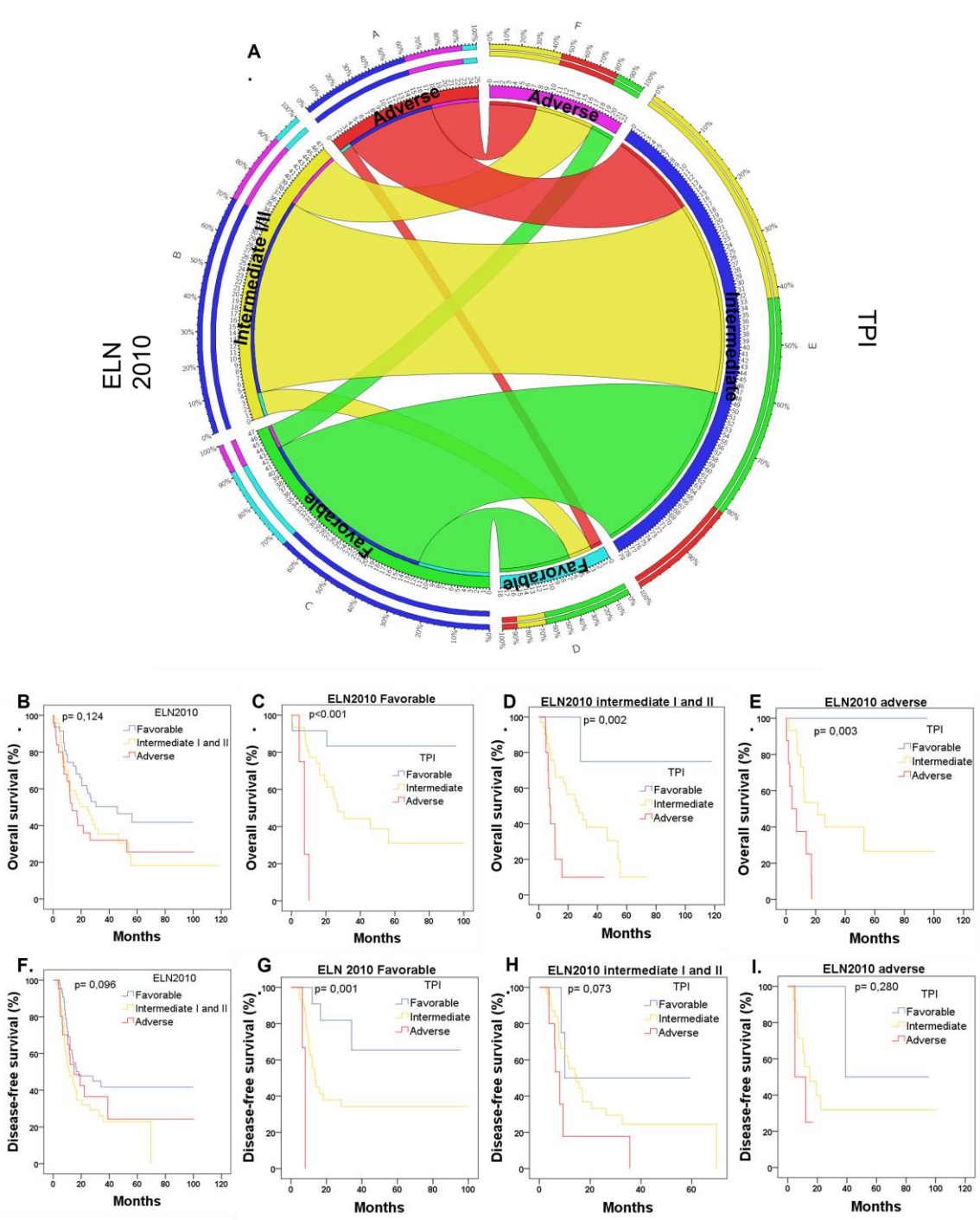


Figure 4: Circus plot (A) showing the relationship between risk groups ELN2010 and TPI. Comparison between the OS (B) and DFS (F) curves of the favorable, intermediate I/II and adverse ELN2010 risk groups of the training cohort. Comparison of OS and DFS curves between favorable, intermediate and adverse TPI risk groups of patients in the favorable (C and G, respectively), intermediate I/II (D and H, respectively) and adverse (E and I, respectively) ELN2010 risk group. Due to the small number of patients, we merged intermediate ELN2010 risk groups I and II into just one intermediate group.

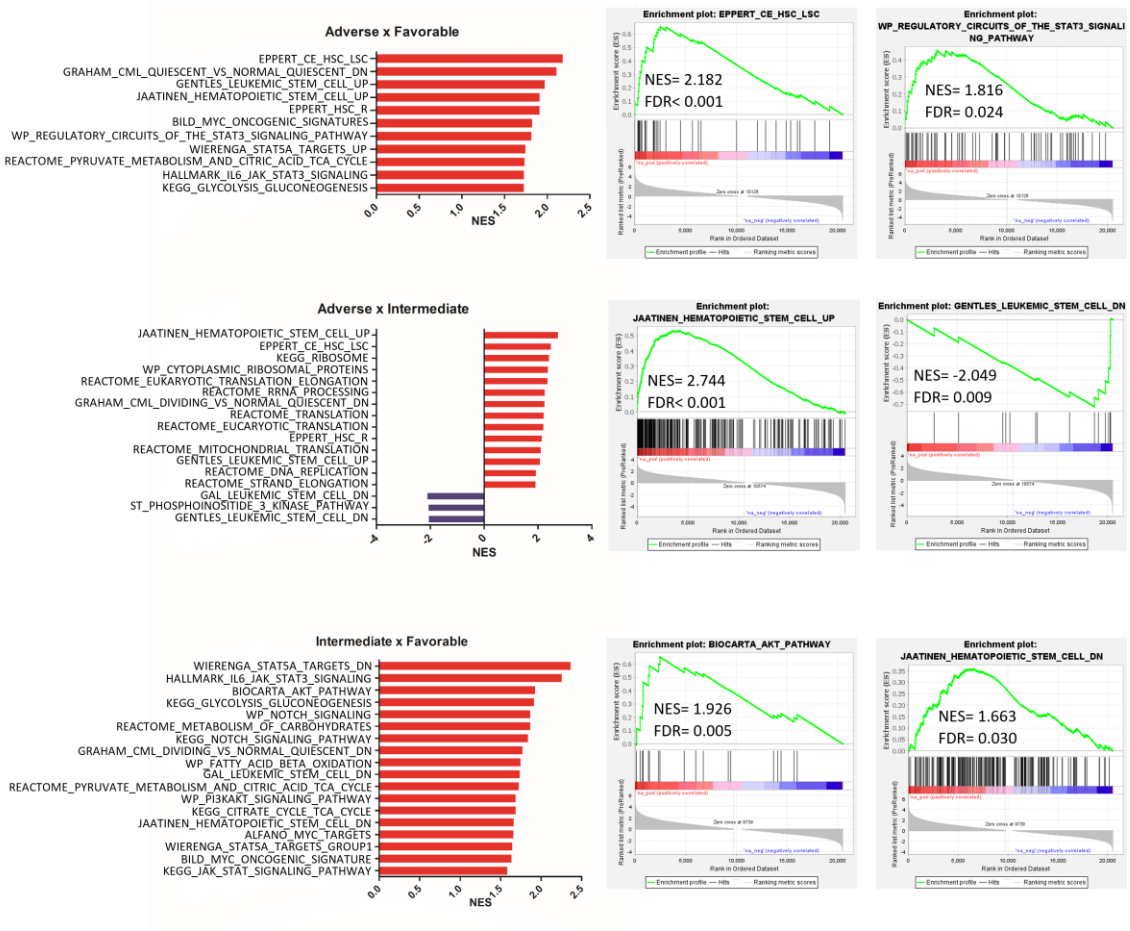


Figure 5: Gene Set Enrichment Analysis (GSEA) Preranked comparing the transcriptome of favorable, intermediate and adverse TPI risk groups.

6 DISCUSSÃO GERAL

No presente trabalho nos propomos a avaliar o impacto isolado e aditivo de marcadores moleculares no desfecho clínico de pacientes adultos com LMA. Aqui demonstramos o impacto isolado de mutações no gene *DNMT3A* e da hiperexpressão do gene *ID1*, bem como da co-ocorrência de mutações nos genes *DNMT3A*, *FLT3* e *NPM1* no desfecho de pacientes LMA. Além disso, propomos um índice de prognóstico baseado no impacto aditivo da expressão aberrante de três genes (*IGF2BP3*, *PDE7B* e *ST6GALNAC4*) para estratificação de risco dos pacientes LMA, na tentativa de refinar a estratificação de risco do grupo colaborativo ELN2010 (Dohner et al. 2010).

Para a análise das mutações no gene *DNMT3A* nos detivemos às mutações no códon R882, localizada no exon 23. Esta decisão foi baseada no fato desse ser um *hotspot* mutacional na LMA e por apresentar impacto prognóstico particular nessa doença e papel fisiopatológico distinto, em comparação com as mutações não-R882 (Thol et al. 2011; Rennevile et al. 2012; Gaidzik et al. 2013; Ley et al.2013; Guryanova et al. 2016; Emperle et al. 2019; Venugopal et al. 2021; Huang et al. 2022). O fato de nos restringirmos a mutações no códon R882 deve ter sido o motivo de termos relatado uma menor frequência de mutações no gene *DNMT3A* em comparação com outros estudos (Ley et al. 2010; Thol et al. 2011; Rennevile et al. 2012; Gaidzik et al. 2013; Roller et al. 2013).

As mutações no gene *DNMT3A* estiveram associadas com pior desfecho em uma coorte de 507 pacientes adultos não selecionados com LMA tratados em cinco centros especializados no Brasil. Entretanto, os resultados apresentados na literatura ainda são conflitantes. Diversos grupos especulam

que as características relacionadas aos pacientes, critérios de inclusão e exclusão dos pacientes no estudo e as diferenças nos protocolos de tratamento poderiam explicar os resultados contraditórios (Ley et al. 2010; Thol et al. 2011; Rennevile et al. 2012; Gaidzik et al. 2013).

Além disso, avaliamos o impacto prognóstico da co-ocorrência das mutações *DNMT3A* com mutações *NPM1* e *FLT3-ITD*, uma vez que essa co-ocorrência é a mais frequente (Cancer Genome Atlas, 2013; Papaemmanuil et al. 2016) e, como também, a LMA triplo-mutada tende a apresentar impacto prognóstico adverso particular (Papaemmanuil et al. 2016) e parece definir um novo subgrupo de LMA com assinatura de RNAm, microRNA e de metilação do DNA característicos (Cancer Genome Atlas, 2013; Garg et al. 2019).

A frequência da co-ocorrência dessas três mutações foi bem similar ao observado em outros trabalhos (Cancer Genome Atlas, 2013; Papaemmanuil et al. 2016). Os pacientes triplo-mutados tiveram piores sobrevida global e sobrevida livre de doença e maior taxas de incidência cumulativa de recaída do que os pacientes não triplo-mutados. Os resultados foram validados em duas coortes de dados públicos, a coorte TCGA (Cancer Genome Atlas, 2013) e a GSE6891 (Verhaak et al. 2009), que também demonstraram o impacto negativo da co-ocorrência dessas mutações no desfecho dos pacientes LMA. Vemos, por tanto, que esses relatos estão de acordo com a literatura (Papaemmanuil et al. 2016; Wang et al. 2020). A explicação fisiopatológica para o impacto adverso dessas mutações na LMA ainda é incerta, mas um estudo recente demonstrou que a expressão do alelo *DNMT3A* R882 induziu expansão de células tronco hematopoiéticas e que cooperando com mutações

FLT3-ITD e *NPM1* mutado induziu LMA *in vivo* e promoveu resistência a quimioterapia com antraciclina (Guryanova et al. 2016).

Na tentativa de propor um índice de prognóstico baseado no impacto aditivo de alterações genéticas no desfecho de pacientes LMA, avaliamos primeiramente o impacto isolado de diversos marcadores genéticos, já descritos como associados com LMA, na sobrevida global de pacientes de uma coorte de dados públicos, TCGA (Cancer Genome Atlas, 2013). Os resultados estão apresentados no Apêndice 1. Apenas as mutações no gene *TP53* e a hiperexpressão do gene *ID1* estiveram associadas com pior SG. O índice baseado no impacto aditivo dessas duas variáveis não se mostrou uma ferramenta robusta para estratificação de risco da LMA. Entretanto, nos estimulou a aprofundar as análises do impacto da hiperexpressão do gene *ID1* no desfecho dos pacientes LMA.

Nós demonstramos que a hiperexpressão do gene *ID1* foi um fator preditor para um pior desfecho em uma coorte pública e que manteve sua capacidade preditiva em uma coorte de pacientes da “vida-real”. O que está de acordo com o relatado na literatura (Tang et al. 2009; Damm et al. 2012; Zhou et al. 2015). Entretanto, em ambas as coortes, a hiperexpressão de *ID1* não foi um fator preditor independente de outras variáveis de confusão para o desfecho dos pacientes LMA (Tang et al. 2009; Damm et al. 2012; Zhou et al. 2015). Na literatura essa questão é controversa, enquanto dois estudos distintos (Tang et al. 2009; Zhou et al. 2015) demonstraram que a hiperexpressão de *ID1* foi um fator prognóstico independente para um pior desfecho em pacientes adultos jovens com LMA, um outro estudo (Damm et al. 2012) demonstrou que a expressão do gene *ID1* não foi variável independente,

mesmo quando avaliado apenas pacientes adultos jovens. Diferenças nas características e nos critérios de inclusão dos pacientes e nos protocolos de tratamento poderiam explicar os resultados contraditórios.

Ademais, vimos que a hiperexpressão de *ID1* está associada a anormalidades inegavelmente associadas a pior evolução da LMA, como as mutações em *TP53*, enquanto sua hipoexpressão está associada a anormalidades sabidamente associadas a um desfecho favorável na LMA, como as mutações CEBPA e anormalidades no *core binding fator*. Além disso tem sido demonstrado, por ensaios funcionais, que *ID1* tem sua expressão controlada por algumas dessas mutações condutoras (Wagner et al, 2006; Tam et al. 2008). Dessa maneira, é possível que o efeito da expressão do gene *ID1* no desfecho dos pacientes LMA seja secundário a ação de outras alterações condutoras da LMA. Entretanto, a sua expressão tem-se demonstrado importante para fisiopatologia da LMA e de outros tumores (Suh et al, 2008; Wang et al, 2015; Man et al, 2016; Singh et al, 2018; Zhao et al, 2019; Wu et al, 2019). Sendo assim, se a expressão do gene *ID1* é útil como um marcador prognóstico para o refinamento de estratificações de risco da LMA ou como alvo terapêutico é uma questão que merece maiores estudos.

Na tentativa de propor refinamento para estratificação de risco ELN2010, desenvolvemos um índice de prognóstico (TPI, do inglês *transcriptome-based prognostic index*) baseado no perfil de expressão de três genes (*PDE7B*, *IGF2BP3*, *ST6GALNAC4*), que foi capaz de caracterizar três grupos de risco (favorável, intermediário e adverso). Dos três genes selecionados, apenas o *PDE7B* já havia sido associado diretamente com o desfecho de pacientes LMA. Em um estudo que envolveu 632 pacientes LMA com cariótipo normal (LMA-

CN) a hiperexpressão do *PDE7B* foi um fator preditor independente para um pior desfecho (Cao et al, 2019). Até onde temos conhecimento, este é o primeiro trabalho que associa a hiperexpressão dos genes *IGF2BP3* e *ST6GALNAC4* com o desfecho de pacientes adultos com LMA.

O TPI foi um fator prognóstico independente para predição de SG e SLD numa coorte com 121 pacientes adultos com LMA não-LPA e foi validado em uma coorte pública com 408 pacientes adultos LMA não-LPA. Ademais, o TPI apresentou capacidade preditiva para a SG superior à classificação ELN2010 e foi capaz de refinar esta classificação na coorte de desenho e de validação externa. O TPI foi capaz de definir um grupo com desfecho adverso para a SG dentro dos grupos de risco favorável e intermediário da classificação ELN2010 e para a SLD dentro do grupo de risco favorável ELN2010. Dessa maneira, é possível que pacientes com desfecho adverso para o TPI, que estavam classificados no grupo favorável ELN2010, possam se beneficiar do TCTH alogênico após a primeira remissão completa. Uma vez que os pacientes no grupo favorável da classificação ELN2010 não são inicialmente indicados para TCTH alogênico na primeira remissão completa (Cornelissen et al, 2016; Dohner et al, 2017). Entretanto, esse estudo é retrospectivo e envolve pacientes não uniformemente tratados, dessa maneira, ensaios clínicos prospectivos são necessários para validar esses achados.

Na tentativa de reproduzir o TPI em uma coorte da vida real, utilizando a metodologia de RT-qPCR, aplicamos o TPI em uma coorte composta por 108 pacientes adultos com LMA diagnosticados e tratados em 3 centros especializados no Nordeste do Brasil. Entretanto, nessa coorte, o TPI não foi capaz de distinguir os três grupos de risco com relação a taxa de RC, SG e

SLD. Vários fatores podem ter limitado as análises. Primeiramente, diferentes metodologias foram usadas para determinar os níveis de expressão gênica. Na coorte de desenho, foi utilizado o sequenciamento de RNA, enquanto nas coortes de validação externa e interna, foram utilizadas as metodologias de *microarray* e RT-qPCR, respectivamente. Assim, devido a diferentes grandezas e dispersão de valores e diferentes sensibilidades do teste, não foi possível determinar exatamente o mesmo ponto de corte para ambas as coortes, sendo necessário definir um ponto de corte específico para cada uma das situações.

Por ser uma variável contínua, dados de expressão gênica são complexos para serem utilizados como marcador prognóstico, sendo difícil a padronização e reprodução em metodologias diferentes. No entanto, alguns marcadores prognósticos baseados na expressão gênica, como a superexpressão dos genes *EVI1* e *BAALC*, são utilizados em painéis de biomarcadores prognósticos ao diagnóstico de forma padronizada e já foram validados por diversos centros (Brand et al. 2013; Alessandrini et al. 2017; Kappala et al. 2017; Nomdedeu et al. 2017). Dessa maneira, é possível que, com a devida padronização, biomarcadores prognósticos baseados na expressão gênica possam ser utilizados na rotina diagnóstica.

Em segundo lugar este estudo foi retrospectivo usando uma coorte da "vida real" de pacientes com LMA tratados no Brasil, um país em desenvolvimento. Dessa maneira, os pacientes não foram tratados de maneira uniforme e diversos problemas socioeconômicos podem ter interferido no desfecho: falta de infraestrutura adequada e tratamento de suporte adequado que levam a altas taxas de infecções bacterianas e fúngicas fatais, leitos superlotados e falta de medicamentos quimioterápicos que levam a atrasos no

tratamento, baixo nível socioeconômico da população, incluindo más condições de moradia, baixa renda per capita e baixa escolaridade (Rego et al. 2003; Lucena-Araujo et al. 2010; Nucci et al. 2013; Lima et al. 2015; Benicio et al. 2017). Esses problemas socioeconômicos em geral acarretam resultados adversos, bastante inferiores ao de países desenvolvidos (Applebaum et al. 2006; Lowenberg et al. 2011; Silveira et al. 2021). Portanto, é possível que a dificuldade no tratamento de pacientes com LMA tenha levado a resultados muito insatisfatórios que superaram as diferenças biológicas dos pacientes em resposta à terapia.

Essa realidade alarmante chama a atenção para a necessidade urgente de melhoria no suporte clínico e tratamento de pacientes com LMA em nosso país, a fim de melhorar o desfecho dos pacientes e obter resultados comparáveis aos de países desenvolvidos. Tais melhorias podem ser alcançadas por meio de esforços de colaboração internacional. Um exemplo disso é o Consórcio Internacional para Leucemia Promielocítica Aguda (IC-APL, do inglês *International Consortium on Acute Promyelocytic Leukemia*), que estabeleceu um diagnóstico rápido, tratamento e diretrizes de suporte para a LPA que foram adaptadas às condições locais, com o apoio e troca de expertise com equipes dos Estados Unidos e da Europa. As melhorias alcançadas com esse consórcio demonstram que é possível melhorar o prognóstico de malignidades graves em países em desenvolvimento por meio de colaboração internacional a um custo relativamente baixo (Rego et al. 2013).

7 CONCLUSÕES

- Mutações no gene *DNMT3A* foram associadas com piores taxas de incidência cumulativa de recaída e SLD e foram independentemente associadas com pior taxa de SG.
- Mutações *DNMT3A* foram frequentemente associadas com mutações *FLT3*-ITD e no gene *NPM1*.
- Pacientes triplo-mutados *DNMT3A*/*FLT3*-ITD/*NPM1* apresentaram menores taxas de SG e SLD e maior taxa de incidência cumulativa de recaída em comparação com os pacientes não triplo-mutados, enquanto não apresentaram impacto na taxa de RC.
- A hiperexpressão do gene *ID1* foi um fator preditor negativo para a sobrevida global dos pacientes com LMA em uma coorte pública e uma coorte da "vida real" de pacientes LMA, entretanto não foi uma variável independente de outras variáveis de confusão.
- A hiperexpressão do gene *ID1* não esteve associada com RC ou com SLD.
- A hiperexpressão dos genes *PDE7B*, *ST6GALNAC4* e *IGF2BP3* foi selecionada como marcador prognóstico molecular para desenho do TPI.
- O aumento do risco do TPI foi um preditor independente para uma pior SG e SLD na coorte de desenho e de validação externa.
- O TPI não esteve associado com RC, SG e SLD na coorte de validação interna.

REFERÊNCIAS

ALCALAY M et al. **Common themes in the pathogenesis of acute myeloid leukemia**. Oncogene, 20, 5680-5694, 2001

ALESSANDRINI M; Kappala SS; Pepper MS. **AMLprofiler:** A Diagnostic and Prognostic Microarray for Acute Myeloid Leukemia. Methods Mol Biol, 1633, 101-123, 2017.

ALLEN C et al. **The importance of relative mutant level for evaluating impact on outcome of KIT, FLT3 and CBL mutations in core-binding factor acute myeloid leukemia**. Leukemia, 27, 1891-1901, 2013.

AL-HARBI S et al. **An update on the molecular pathogenesis and potential therapeutic targeting of AML with t(8;21)(q22;q22.1);RUNX1-RUNX1T1.** Blood Adv, 4, 1, 229-238, 2020.

APPELBAUM FR et al. **Age and acute myeloid leukemia**. Blood, 107, 9, 3481-5, 2006

ARBER DA et al. **The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia**. Blood, 127, 20, 2391-2405, 2016.

ARNOLD N et al. **Establishing a control population to screen for the occurrence of nineteen unclassified variants in the BRCA1 gene by denaturing high-performance liquid chromatography**. J Chromatogr B Analyt Technol Biomed Life Sci, 782, 99-104, 2002.

AYATOLLAHI H et al. **Prognostic Importance of C-KIT Mutations in Core Binding Factor Acute Myeloid Leukemia:** A Systematic Review. Hematol Oncol Stem Cell Ther, 10, 1-7, 2017

BELSON M; KINGSLEY B; HOLMES. **A Risk Factors for Acute Leukemia in Children:** A Review. Environ Health Perspect, 115, 138-145, 2007.

BENICIO MTL et al. **Evaluation of the European LeukemiaNet recommendations for predicting outcomes of patients with acute myeloid leukemia treated in low- and middle-income countries (LMIC)**: A Brazilian experience. Leuk Res, 60, 109-114, 2017

BENNETT JM et al. **Criteria for the Diagnosis of Acute Leukemia of Megakaryocyte Lineage (M7):** A Report of the French-American-British Cooperative Group. Ann Intern Med, 103, 460-462, 1985

BENNETT JM et al. **Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group**. Br J Haematol, 33, 451-458, 1976.

BENNETT JM et al. **Proposal for the recognition of minimally differentiated acute myeloid leukaemia (AML-MO)**. Br J Haematol, 78, 325-329, 1991

BITTENCOURT R et al. **Leucemia Mielóide Aguda:** perfil de duas décadas do Serviço de Hematologia do Hospital das Clínicas de Porto Alegre – RS. Rev. Bras. Hematol. Hemoter, 25, 1, 17-24, 2003

BODDU P et al. **Treated secondary acute myeloid leukemia:** a distinct high-risk subset of AML with adverse prognosis. Blood Adv, 1, 17, 1312-1323, 2017

BOWMAN RL et al. **Clonal Hematopoiesis and Evolution to Hematopoietic Malignancies**. Cell Stem Cell. 22, 157-170, 2018.

BRAND J et al. **A standardized microarray assay for the independent gene expression markers in AML:** EVI1 and BAALC. Exp Hematol Oncol, 2, 1, 7, 2013.

BRAY F et al. **Global cancer statistics 2018:** GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin, 68, 6, 394-424, 2018.

BULLINGER L et al. **Use of gene-expression profiling to identify prognostic subclasses in adult acute myeloid leukemia**. N Engl J Med, 350, 16, 1605-16, 2004

BUSCARLET M et al. **Lineage restriction analyses in CHIP indicate myeloid bias for TET2 and multipotent stem cell origin for DNMT3A**. Blood, 132, 3, 277-280, 2018

CANCER GENOME ATLAS RESEARCH NETWORK. **Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia**. N Engl J Med, 368, 22, 2059-74, 2013.

CAO L et al. **The Prognostic Significance of PDE7B in Cytogenetically Normal Acute Myeloid Leukemia**. Sci Rep, 9, 1, 16991, 2019

CONWAY O'BRIEN E; PRIDEAUX S; CHEVASSUT, T. **The Epigenetic Landscape of Acute Myeloid Leukemia**. Adv Hematol. 2014, 1–15, 2014

CORCES-ZIMMERMAN MR, MAJETI R. **Pre-leukemic evolution of hematopoietic stem cells:** the importance of early mutations in leukemogenesis. Leukemia, 28, 2276–2282, 2014.

CORNELISSEN JJ; BLAISE D. **Hematopoietic stem cell transplantation for patients with AML in first complete remission**. Blood, 127, 1, 62-70, 2016.

DAMM F et al. **Integrative prognostic risk score in acute myeloid leukemia with normal karyotype**. Blood, 117, 17, 4561-4568, 2011.

DAMM F et al. **ID1 expression associates with other molecular markers and is not an independent prognostic factor in cytogenetically normal acute myeloid leukaemia**. Br J Haematol, 158, 208-215, 2012.

SILVA WF et al. **Real-life Outcomes on Acute Promyelocytic Leukemia in Brazil - Early Deaths Are Still a Problem**. Clin Lymphoma Myeloma Leuk, 19, 2, e116-e122, 2019

DE LA BLÉTIÈRE DR et al. **Routine use of microarray-based gene expression profiling to identify patients with low cytogenetic risk acute myeloid leukemia: accurate results can be obtained even with suboptimal samples.** BMC Med Genomics, 30, 5-6, 2012.

DEBERNARDI S et al. **Genome-wide analysis of acute myeloid leukemia with normal karyotype reveals a unique pattern of homeobox gene expression distinct from those with translocation-mediated fusion events**. Genes Chromosomes Cancer, 37, 2, 149-58, 2003.

DESAI P et al. **Somatic mutations precede acute myeloid leukemia years before diagnosis**. Nat Med, 24, 7, 1015–1023, 2018.

DOHNER H; ESTEY E; GRIMWADE D. **Diagnosis and management of AML in adults:** 2017 ELN recommendations from an international expert panel. Blood, 129, 4, 424-447, 2017.

DÖHNER H et al. **Diagnosis and management of acute myeloid leukemia in adults:** recommendations from an international expert panel, on behalf of the European LekemiaNet. Blood, 115, 453-474, 2010

DÖHNER H et al. **Acute Myeloid Leukemia**. N Engl J Med, 373, 12, 1136-1152, 2015.

EISFELD AK et al. **Mutation patterns identify adult patients with de novo acute myeloid leukemia aged 60 years or older who respond favorably to standard chemotherapy:** an analysis of alliance studies. Leukemia, 32, 1338-1348, 2018.

EMPERLE M et al. **Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns.** Nucleic Acids Res, 47, 21, 11355-11367, 2019

ESTEY EH. **Acute myeloid leukemia:** 2019 update on risk-stratification and management. Am J Hematol, 93, 10, 1267-1291, 2018.

FABER ZJ et al. **The Genomic Landscape of Core-Binding Factor Acute Myeloid Leukemias**. Nat Genet, 48, 12, 1551–1556, 2016.

FAGUNDES EM et al. **De novo acute myeloid leukemia in adults younger than 60 years of age:** socioeconomic aspects and treatment results in a Brazilian university center. Leuk Lymphoma, 47, 8, 1557-64, 2006.

FALINI B et al. **Cytoplasmic Nucleophosmin in Acute Myelogenous Leukemia with a Normal Karyotype**. N Engl J of Med, 352, 254-266, 2005.

FRÖHLING S et al. **Prognostic significance of activating FLT3 mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML Study Group Ulm**. Blood, 100, 4372-4380, 2002.

GAIDZIK VI et al. **Clinical impact of DNMT3A mutations in younger adult patients with acute myeloid leukemia:** results of the AML Study Group (AMLSG). Blood, 121, 23, 4769-77, 2013.

GARG S et al. **Hepatic leukemia factor is a novel leukemic stem cell regulator in DNMT3A, NPM1, and FLT3-ITD triple-mutated AML**. Blood, 134, 3, 263-276, 2019.

GENOVESE G et al. **Clonal Hematopoiesis and Blood-Cancer Risk Inferred from Blood DNA Sequence**. N Engl J Med, 371, 26, 2477–2487, 2014.

GIEBEL S et al. **Association of Human Development Index with rates and outcomes of hematopoietic stem cell transplantation for patients with acute leukemia**. Blood, 116, 1, 122-8, 2010

GILLILAND DG, GRIFFIN JD. **The roles of FLT3 in hematopoiesis and leukemia**. Blood, 100, 1532-1542, 2002

GRAPHODATSKY AS et al (2011) **The genome diversity and karyotype evolution of mammals**. Mol Cytogen 4:1-16.

GRIMWADE D et al. **Refinement of cytogenetic classification in acute myeloid leukemia:** determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood, 116, 354-365, 2010.

GURYANOVA OA et al. **DNMT3A R882 mutations promote anthracycline resistance in acute myeloid leukemia through impaired nucleosome remodeling**. Nat Med, 22, 12, 1488–1495, 2016.

HAFERLACH T, SCHMIDTS I **The power and potential of integrated diagnostics in acute myeloid leukaemia**. Br J Haematol, 188, 1, 36-48, 2020.

HANDSCHUH L. **Not Only Mutations Matter:** Molecular Picture of Acute Myeloid Leukemia Emerging from Transcriptome Studies. J Oncol, 2019, 1-36, 2019.

HARTMANN L, METZELER KH. **Clonal hematopoiesis and preleukemia-Genetics, biology, and clinical implications**. Genes Chromosomes Cancer. 58, 12, 828-838, 2019.

HARTMANN L, METZELER KH. **Clonal hematopoiesis and preleukemia**—Genetics, biology and clinical implications. Genes Chromosomes Cancer, 58, 12, 828-838, 2019.

HEEREMA-MCKENNEY A, ARBER DA. **Acute Myeloid Leukemia**. Hematol Oncol Clin North Am, 23, 633-654, 2009

HINAI AA E VALK PJM. **Review: Aberrant EVI1 expression in acute myeloid leukaemia**. Br J Haematol, 172, 6, 870-8, 2016.

HORIBATA S et al. **Heterogeneity in refractory acute myeloid leukemia**. Proc Natl Acad Sci U S A, 116, 21, 10494-10503, 2019.

HUANG R, LIAO X, LI Q. **Identification and validation of potential prognostic gene biomarkers for predicting survival in patients with acute myeloid leukemia**. Onco Targets Ther, 10, 5243-5254, 2017.

HUANG YH et al. **Systematic Profiling of DNMT3A Variants Reveals Protein Instability Mediated by the DCAF8 E3 Ubiquitin Ligase Adaptor**. Cancer Discov, 12, 1, 220-235, 2022.

HUMAN DEVELOPMENT REPORTS – **United Nations Development Programme**. DIsponível em: http://hdr.undp.org/en/humandev. Acesso em 5 de Novembro de 2021

HYDE RK, ZHAO L, ALEMU L, LIU PP. **Runx1 is required for hematopoietic defects and leukemogenesis in Cbfb-MYH11 knockin mice**. Leukemia, 29, 8, 1771–1778, 2015.

INSTITUTO NACIONAL DO CÂNCER (INCA). **Estimativa 2020: Incidência de Câncer no Brasil. Instituto Nacional de Câncer José Alencar Gomes da Silva, Coordenação Geral de Ações Estratégicas, Coordenação de Prevenção e Vigilância**, 40-73, 2019

ISHIKAWA Y et al. **Prospective evaluation of prognostic impact of KIT mutations on acute myeloid leukemia with RUNX1-RUNX1T1 and CBFB-MYH11.** Blood Adv, 4, 1, 66-75, 2020.

JACAMO R et al. **Tumor Trp53 status and genotype affect the bone marrow microenvironment in acute myeloid leukemia**. Oncotarget, 8, 48, 83354-83369, 2017.

JÁCOMO RH et al. **Clinical features and outcomes of 134 Brazilians with acute promyelocytic leukemia who received ATRA and anthracyclines**. Haematologica, 92, 10, 1431-2, 2007

JAISWAL S et al. **Age-Related Clonal Hematopoiesis Associated with Adverse Outcomes**. N Engl J Med, 371, 26, 2488–2498, 2014.

KAPPALA SS et al. **Application of the AMLprofiler Diagnostic Microarray in the South African Setting**. Stem Cells Int, 2017, 2560191, 2017

LANGABEER SE et al. **EVI1 expression in acute myeloid leukaemia**. Br J Haematol, 112, 1, 208-11, 2001

LEY TJ et al. **DNMT3A mutations in acute myeloid leukemia**. N Engl J Med, 363, 25, 2424-33.

LIMA AS et al. **Clinical outcomes of patients with acute myeloid leukemia: evaluation of genetic and molecular findings in a real-life setting**. Blood, 126, 15, 1863-5, 2015

LI Z et al. **Identification of a 24-Gene Prognostic Signature That Improves the European LeukemiaNet Risk Classification of Acute Myeloid Leukemia:** An International Collaborative Study. J Clin Oncol, 31, 9, 1172-1181, 2013

LINET MS et al. **Benzene Exposure Response and Risk of Myeloid Neoplasms in Chinese Workers**: A Multicenter Case-Cohort Study. J Natl Cancer Inst, 111, 5, 465-474, 2019.

LOBERG MA et al. **Sequentially inducible mouse models reveal that Npm1 mutation causes malignant transformation of Dnmt3a-mutant clonal hematopoiesis.** Leukemia, 33, 1635–1649, 2019.

LÖWENBERG B et al. **Cytarabine Dose for Acute Myeloid Leukemia**. N Engl J of Med, 364, 11, 1027-1036, 2011

LUCENA-ARAUJO AR et al. **Results of FLT3 mutation screening and correlations with immunophenotyping in 169 Brazilian patients with acute myeloid leukemia**. Ann Hematol, 89, 2, 225-8, 2010

MADAN V et al. **ASXL2 regulates hematopoiesis in mice and its deficiency promotes myeloid expansion**. Haematologica, 103,12, 1980-1990, 2018.

MARCUCCI G et al. **Prognostic Significance of, and Gene and MicroRNA Expression Signatures Associated With, CEBPA Mutations in Cytogenetically Normal Acute Myeloid Leukemia With High-Risk Molecular Features:** A Cancer and Leukemia Group B Study. J Clin Oncol, 26, 31, 5078-5087, 2008.

MCCURDY SR, Levis MJ **Emerging molecular predictive and prognostic factors in acute myeloid leukemia**. Leuk Lymphoma, 59, 9, 2021-2039, 2017.

METZELER KH et al. **Spectrum and prognostic relevance of driver gene mutations in acute myeloid leukemia**. Blood, 128, 5, 686-698, 2016.

METZELER KH et al. (2013) **A stem cell-like gene expression signature associates with inferior outcomes and a distinct microRNA expression**

profile in adults with primary cytogenetically normal acute myeloid leukemia. Leukemia, 27, 2023-2031, 2016.

MILES LA et al. **Single-cell mutation analysis of clonal evolution in myeloid malignancies**. Nature, 587, 7834, 477-482, 2020.

MORITA K et al. **Clonal evolution of acute myeloid leukemia revealed by high-throughput single-cell genomics**. Nat Commun, 11, 1, 5327.

NAKAO M et al. **Internal tandem duplication of the flt3 gene found in acute myeloid leukemia**. Leukemia, 10, 1911-1918, 1996

NARAYANAN D, Weinberg OK. **How I investigate acute myeloid leukemia**. Int J Lab Hematol, 42, 1, 3-15, 2020

NG SWK et al. **A 17-gene stemness score for rapid determination of risk in acute leukaemia**. Nature, 000, 1-17, 2016

NOMDEDÉU JF et al. **Feasibility of the AML profiler (Skyline™ Array) for patient risk stratification in a multicentre trial:** a preliminary comparison with the conventional approach. Hematol Oncol, 5, 4, 778-788, 2017

NUCCI M et al. **Invasive fungal diseases in haematopoietic cell transplant recipients and in patients with acute myeloid leukaemia or myelodysplasia in Brazil**. Clin Microbiol Infect, 19, 8, 745-51, 2013

OPATZ S et al. **The clinical mutatome of core binding factor leukemia**. Leukemia, 34, 6, 1553-1562, 2020

PAPAEMMANUIL E et al. **Genomic Classification and Prognosis in Acute Myeloid Leukemia**. N Engl J Med, 374, 23, 2209-2221, 2016.

PATEL JP et al. **Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia**. N Engl J Med, 366, 12, 1079-1089, 2012.

POTTER N et al. **Single cell analysis of clonal architecture in acute myeloid leukaemia**. Leukemia, 33, 5, 1113-1123, 2019.

RAPIN N et al. **Comparing cancer vs normal gene expression profiles identifies new disease entities and common transcriptional programs in AML patients**. Blood, 123, 6, 894-904, 2014.

REGO EM et al. **Improving acute promyelocytic leukemia (APL) outcome in developing countries through networking, results of the International Consortium on APL**. Blood, 121, 11, 1935-43, 2013

REGO MFN et al. **Acute leukemias in Piauí: comparison with features observed in other regions of Brazil**. Braz J Med Biol Res, 36, 3, 331-7, 2003

RENNEVILLE A et al. **Prognostic significance of DNA methyltransferase 3A mutations in cytogenetically normal acute myeloid leukemia**: a study by the Acute Leukemia French Association. Leukemia, 26, 6, 1247-54, 2012.

ROLLER A et al. **Landmark analysis of DNMT3A mutations in hematological malignancies**. Leukemia, 27, 7, 1573-8, 2013.

ROMER-SEIBERT JS, MEYER SE **Genetic heterogeneity and clonal evolution in acute myeloid leukemia**. Curr Opin Hematol, 28, 1, 64-70, 2021

SERNA DS et al. **Trends in survival rates after allogeneic hematopoietic stem-cell transplantation for acute and chronic leukemia by ethnicity in the United States and Canada**. J Clin Oncol, 21, 20, 3754-60, 2003.

SHALLIS RM et al. **Epidemiology of acute myeloid leukemia: Recent progress and enduring challenges**. Blood Rev, 36, 70-87, 2019.

SHI M et al. **Network-based sub-network signatures unveil the potential for acute myeloid leukemia therapy**. Mol Biosyst, 10, 12, 3290-3297, 2014.

SHLUSH LI et al. **Identification of pre-leukaemic haematopoietic stem cells in acute leukaemia**. Nature, 506, 328–33, 2014.

SILLA L et al. **Patient socioeconomic status as a prognostic factor for allo-SCT**. Bone Marrow Transplant, 43, 7, 571-7, 2009

SILVEIRA DRA et al. **A multicenter comparative acute myeloid leukemia study: can we explain the differences in the outcomes in resource-constrained settings?** Leuk Lymphoma, 62, 1, 147-157, 2021

SMALL D. **FLT3 mutations:** biology and treatment. Hematology Am Soc Hematol Educ Program, 178-184, 2006.

SONG X et al. **Incidence, Survival, and Risk Factors for Adults with Acute Myeloid Leukemia Not Otherwise Specified and Acute Myeloid Leukemia with Recurrent Genetic Abnormalities:** Analysis of the Surveillance, Epidemiology, and End Results (SEER) Database, 2001-2013. Acta Haematol, 139, 2, 115-127, 2018.

TANG R et al. **High Id1 expression is associated with poor prognosis in 237 patients with acute myeloid leukemia**. Blood, 114, 14, 2993-3000, 2009.

THE CANCER GENOME ATLAS RESEARCH NETWORK. **Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia**. N Engl J Med, 368, 22, 2059-2074, 2013.

THOL F et al. **Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia**. J Clin Oncol, 29, 21, 2889-96, 2011.

TYNER JW et al. **Functional genomic landscape of acute myeloid leukaemia**. Nature, 562, 526-546, 2018.

VALK PJ1 et al. **Prognostically useful gene-expression profiles in acute myeloid leukemia**. N Engl J Med, 350, 16, 1617-28, 2004.

VARDIMAN JW, HARIA NL, BRUNNING RD. **TheWorld Health Organization (WHO) classification of the myeloid neoplasms**. Blood, 100, 2292-2302, 2002

VARDIMAN JW et al. **The 2008 revision of the World Health Organization (WHO) classification of myeloid neoplasms and acute leukemia:** rationale and important changes. Blood, 114, 937-951, 2009

VÁZQUEZ I et al. **Down-regulation of EVI1 is associated with epigenetic alterations and good prognosis in patients with acute myeloid leukemia.** Haematologica, 96, 10, 1448-1456, 2011.

VELLOSO EDRP. **Molecular and cytogenetic abnormalities in acute myeloid leukemia:** review and case studies. Einstein, 9, 184-189, 2011.

VERHAAK R et al. **Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling**. Haematologica, 94, 1, 131-134, 2009.

VENUGOPAL K et al. **Alterations to DNMT3A in Hematologic Malignancies**. Cancer Res, 81, 2, 254-263, 2021.

VICENTE C et al. **The role of the GATA2 transcription factor in normal and malignant hematopoiesis**. Crit Rev Oncol Hematol, 82, 1, 1-17, 2012.

WAGNER S et al. **A parsimonious 3-gene signature predicts clinical outcomes in an acute myeloid leukemia multicohort study**. Blood Adv, 3, 8, 1330-1346, 2019.

WANG R-Q et al. **Characteristics and prognostic significance of genetic mutations in acute myeloid leukemia based on a targeted next-generation sequencing technique**. Cancer Med, 9, 22, 8457-8467, 2020.

WARTIOVAARA-KAUTTO U et al **Germline alterations in a consecutive series of acute myeloid leukemia**. Leukemia, 32, 10, 2282-2285, 2018.

WEBER S et al. **BAALC expression:** a suitable marker for prognostic risk stratification and detection of residual disease in cytogenetically normal acute myeloid leukemia. Blood Cancer J, 4, 1-12, 2014.

XIE F et al. **Bipartite network analysis reveals metabolic gene expression profiles that are highly associated with the clinical outcomes of acute myeloid leukemia**. Comput Biol Chem, 67, 150-157, 2017.

XIE M et al. **Age-related mutations associated with clonal hematopoietic expansion and malignancies**. Nat Med, 20, 12, 1472-1480, 2014.

YANG L et al. **DNMT3A loss drives enhancer hypomethylation in FLT3-ITD-associated leukemias**. Cancer Cell, 29, 6, 922–934, 2016.

ZHANG X et al. **DNMT3A and TET2 compete and cooperate to repress lineage-specific transcription factors in hematopoietic stem cells**. Nat Genet, 48, 9, 1014-23, 2016.

ZHOU J-D et al. **Clinical significance of up-regulated ID1 expression in Chinese de novo acute myeloid leukemia**. Int J Clin Exp Pathol, 8, 5, 5336-44, 2015.

APÊNDICE A – CO-OCCURRENCE OF DNMT3A, NPM1, FLT3 MUTATIONS IDENTIFIES A SUBSET OF ACUTE MYELOID LEUKEMIA WITH ADVERSE PROGNOSIS

Artigo publicado em março de 2020 na revista *Blood* – Letter to Blood, Fator de impacto: 22,113; Qualis/CAPES Ciências Biológicas I: A1; DOI: 10.1182/blood.2019003339

Matheus F Bezerra,[1,2] Aleide S Lima,[3] Maria-Riera Piqué-Borràs,[4] Douglas R Silveira,[3,4] Juan L Coelho-Silva,[3,7] Diego A Pereira-Martins,[3,7] Isabel Weinhäuser,[7] Pedro L Franca-Neto,[3] Lynn Quek,[8,9] Anna Corby,[9] Mayara M Oliveira,[3] Marinus M Lima,[10] Reijane A de Assis,[10] Paula de Melo Campos,[11] Bruno K Duarte,[12,13] Israel Bendit,[5] Vanderson Rocha,[5] Eduardo M Rego,[5] Fabiola Traina,[14] Sara T Saad,[12,13] Marcos A Bezerra,[2] Antonio R Lucena-Araujo[2]*

Affiliations: [1]Keizo Asami Immunopathology Laboratory, Federal University of Pernambuco, Recife, Brazil; [2]Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, Brazil; [3]Department of Genetics, Federal University of Pernambuco, Recife, Brazil; [4]Department of Biomedicine, University of Basel, Switzerland; [5]Hematology Division, LIM31, Faculdade de Medicina, University of Sao Paulo, Sao Paulo, Brazil; [6]Hematology Department, AC Camargo Cancer Center, São Paulo, Brazil; [7]Department of Internal Medicine, Medical School of Ribeirao Preto, Ribeirao Preto, Brazil; [8]Department of Hematology, Oxford University Hospitals NHS Foundation Trust, Oxford, UK; [9]MRC Molecular Hematology Unit, WIMM, University of Oxford, Oxford, UK; [10]Department of Internal Medicine, Hematology and Hemotherapy Foundation of Pernambuco, Recife, Brazil; [11]Hematology and Transfusion Medicine Center, University of Campinas, Campinas, Sao Paulo, Brazil; [12]Department of Internal Medicine, University of Campinas, Campinas, Brazil; [13]Hematology and Hemotherapy Center, University of Campinas, Campinas, Brazil; [14]Department of Medical Images, Hematology and Clinical Oncology, Medical School of Ribeirao Preto, University of Sao Paulo.

The impact of mutations in *DNA methyltransferase 3 alpha* (*DNMT3A*) at diagnosis as a prognostic marker in acute myeloid leukemia (AML) has been contradictory so far.[1–3] These discrepancies most likely arise from differences of therapeutic protocols used. Most important, few if any studies have evaluated the clinical importance of the 3-way co-occurrence of mutations affecting *DNMT3A*, *nucleophosmin* (*NPM1*) and *fms-like tyrosine kinase 3* (*FLT3*) genes (in particular *FLT3* length mutations or *FLT3*-ITD for "internal tandem duplication") in patients treated outside well-controlled clinical trials, a real-life setting that represents most low- and middle-income countries. Hence, we assessed the frequency and clinical impact of *DNMT3A* mutations and the co-occurrence of *DNMT3A*/*NPM1*/*FLT3*-ITD mutations on treatment outcomes of non-selected AML patients, followed from June 2003 to January 2019 at five Brazilian reference centers specialized on AML treatment.

Bone marrow samples from 507 consecutive patients with *de novo* AML (median age: 51 years, range: 18-94 years; 47% male) were obtained at diagnosis. Patients with acute promyelocytic leukemia, therapy-related AML, or with a previous history of myelodysplastic syndrome were excluded. Details for treatment protocols can be found in the Supplemental data. The study adhered to the tenets of the Declaration of Helsinki and informed consents were obtained from all patients or their relatives. The local Research Ethics Board of each participating center approved the study.

The *DNMT3A* and *NPM1* mutations were analyzed by standard sequencing techniques. Details are described in the Supplemental data. Screening for *FLT3*-ITD mutations was performed by PCR according to the method of Kiyoi *et al*.[4] In parallel, we explored the *FLT3* allelic ratio in patients with the *FLT3*-ITD mutated

status. *FLT3*-tyrosine kinase domain mutations were not evaluated in this study. Since most of *DNMT3A* mutations in myeloid neoplasms occur at exon 23, with a significant enrichment for mutations at codon R882,[5,6] we evaluated the mutational and phenotypic profile of patients harboring *DNMT3A*-R882 and non-R882 mutations using The Cancer Genome Atlas database (TCGA) dataset.[5] We observed that *NPM1* and *FLT3*-ITD mutations were significantly enriched in *DNMT3A*-R882 when compared to non-R882 mutations or *DNMT3A* wildtype (Supplemental figure 1). Therefore, based on its biological and clinical significance[3,7–10] and our own experience, screening for *DNMT3A* mutations was restricted to the codon R882.

DNMT3A-R882 mutations were detected in 64/507 patients (13%), most of them identified as R882H (49/64; 76%), followed by R882C (12/64; 19%), and R882P (3/64; 5%). Samples without detectable *DNMT3A*-R882 mutations or carrying single nucleotide polymorphisms are referenced here as '*DNMT3A* non-mutated'. To decide which variables to include in the multivariate Cox proportional hazard model, we performed a backward elimination analysis using the Akaike Information Criteria (AIC) as fitness measure and getting the best-fitted model (Supplemental table 1). The basis (indispensable) variable used was cytogenetic risk stratification. Treatment-related variables were not included in the multivariate model due to the biased nature of a retrospective study.

The clinical and baseline characteristics are summarized in the Table 1. Overall, 302/507 (60%) patients achieved complete remission (CR), of whom 37/64 (53%) and 265/443 (60%) were assigned to the *DNMT3A* mutated and *DNMT3A* non-mutated groups, respectively (P=0.786). The median follow-up among survivals was 39 months (95% confidence interval [CI]: 26–53 months).

Patients with *DNMT3A* mutations had significantly lower 5-year overall survival (OS) (9%, 95%CI: 3–18%) compared to those without *DNMT3A* mutations (22%, 95%CI: 17–27%) (*P*=0.0035) (Figure 1A). The best-fitted multivariate Cox proportional hazards model for OS was age (>60-years-old), leukocyte counts (>50×10^9/L), *DNMT3A* status, and cytogenetic risk stratification. This model showed that *DNMT3A* mutational status was independently associated with poor OS (hazard ratio, HR: 1.4, 95%CI: 1.01-2.1; *P*=0.04) (Figure 1G). Of the 302 patients who achieved CR, 133 patients (44%) relapsed. Considering non-relapse death as a competing cause of failure, the 5-year cumulative incidence of relapse (CIR) rate was 54% (95%CI: 48-60%). CIR rates for patients assigned to the *DNMT3A* mutated and non-mutated groups were 72% (95%CI: 58-86%) and 50% (95%CI: 42-57%), respectively (*P*<0.0001; Figure 1B). Patients with *DNMT3A* mutations had a significantly lower disease-free survival (DFS) rate (19%, 95%CI: 7-34%) in comparison to patients without *DNMT3A* mutations (42%, 95%CI: 34-49%) (*P*<0.0001; Figure 1C).

Considering its clinical[11] and biological[9,10,12] relevance and the strikingly high frequency of 3-way co-occurrence mutations in AML,[5] we evaluated the clinical relevance of the co-occurrence *DNMT3A/NPM1/FLT3*-ITD in a real-life setting. The frequency of triple-mutated patients in our cohort (35/507; 7%) was very similar to other studies.[5,11] Table 1 summarized the main baseline and clinical characteristics. Triple-mutated patients had significantly lower OS (4%, 95%CI: 2-15% *versus* 21%, 95%CI: 17-26%; *P*=0.011), higher CIR rate (85%, 95%CI: 71-98% *versus* 50%, 95%CI: 43-56%; *P*<0.0001), and lower DFS rates (5%, 95%CI: 1-20% *versus* 42%, 95%CI: 34-49%; *P*<0.0001) compared to non-triple-mutated patients (Figure 1D-F). In multivariate analysis, the lowest AIC for DFS was achieved when

FLT3-ITD and *NPM1* status, triple-mutant AML, and cytogenetic risk stratification were included. In this model, triple-mutant AML presented a higher relapse risk (HR: 2.49, 95%CI: 1.3-5.5; *P*=0.02) (Figure 1H). The co-occurrence of *DNMT3A/NPM1/FLT3*-ITD mutations had no impact on CR achievement (*P*=0.725).

To validate our findings, we took advantage of two publicly available AML datasets (TCGA[5] and Gene Expression Omnibus,[13] GEO; www.ncbi.nlm.nih.gov/geo, accession number: GSE6891). For survival analysis, only *de novo* AML patients submitted to intensive therapy were included. *DNMT3A*-R882 mutations were reported in 23/139 (17%) and 64/479 (13%) patients included in the TCGA and GEO cohorts, respectively. *DNMT3A* mutations were significantly associated with poor OS (HR: 1.61; 95%CI: 1.1–2.53; *P*=0.037), but not with DFS (HR: 1.48; 95%CI: 0.95–2.44; *P*=0.177) in TCGA patients. In contrast, *DNMT3A* mutational status had no impact on OS (HR: 1.22, 95%CI: 0.95-1.58; *P*=0.115), but was significantly associated with poor DFS for patients included in the GEO cohort (HR: 1.47, 95%CI: 1.1-2; *P*=0.015). The co-occurrence of *DNMT3A/NPM1/FLT3*-ITD mutations was significantly associated with poor DFS (HR: 1.3, 95%CI: 1.1-1.68; *P*=0.038), but not with OS (HR: 1.16, 95%CI: 0.92-1.48; *P*=0.19) in TCGA patients. Similar results were obtained for GEO patients: DFS (HR: 2.02, 95%CI: 1.3-3.17; *P*=0.002) and OS (HR: 1.5, 95%CI: 1.01-2.24; *P*=0.48).

In summary, we demonstrated that *DNMT3A* mutations might be useful for AML outcome prediction, although results remain conflicting. Several groups speculate that patient-related features and differences in treatment protocols could explain the contradictory results.[1–3,7] In fact, the functional genomic landscape of AML suggest that the response to drugs is specific to combinatorial mutational events.[14] Therefore, the impact of *DNMT3A* mutations in clinical

decision-making remains disputable. Importantly, we restricted the screening for *DNMT3A* mutations to codon R882, which probably explains our lower rate of *DNMT3A* mutations in comparison to other studies.[1,2] Although, one may argue that restricting our analysis to the codon R882 may limit our study, it is important to notice that current literature supports the idea that only *DNMT3A*-R882 mutations contribute to prognostication[3,7,15] and pathophysiology of AML.[9,10,16] For instance, non-R882 mutations found on the DNMT3A enzyme affecting different domains showed few biochemical consequences.[17] Furthermore, *DNMT3A*-R882 mutations (but not non-R882 mutations) may have an impact on clonal hematopoiesis.[18–20] Finally, a remarkable difference in DNA methylation signatures between samples with *DNMT3A*-R882 and non-R882 mutations has been reported,[8] suggesting that these mutations should not be pooled together. It is possible that mutations in different *DNMT3A* domains lead to different neomorphic functions, resulting in pathogenetic variabilities. Nevertheless, whether *DNMT3A* non-R882 mutations harbor biological or clinical importance in AML requires further studies.

Yet, the co-occurrence *DNMT3A/NPM1/FLT3*-ITD mutations is more representative regarding the biology of the disease and may constitute a more robust strategy for outcome prediction in AML. Reasons for such robustness remain to be elucidated, although functional studies have identified a link between the co-occurrence of *DNMT3A*/*NPM1*/*FLT3*-ITD mutations and AML resistance to anthracycline based-chemotherapy.[9] In agreement, triple-mutated AML showed a unique differentiation response to the FLT3 inhibitor AC220[21] and increased sensitivity to the Food and Drug Administration-approved drug Ibrutinib.[14] More recently, transcriptomic and immunophenotypic data describe triple-mutant blasts

to be associated with high leukemia stem cell frequency, and synergistic upregulation of specific leukemia stem cell regulator.[12] Finally, specific DNA methylation signatures were characterized in triple-mutated patients.[22] We speculate that these findings may help us to better understand the poor prognosis of this specific AML subtype.

This is one of the first studies to describe the prognostic importance of *DNMT3A* mutations and the co-occurrence of *DNMT3A*/*NPM1*/*FLT3*-ITD mutations involving consecutive non-selected patients treated outside well-controlled clinical trials. As such, any conclusion drawn from this "real-world" data should be interpreted with caution. In contrast to previous studies that draw their conclusions based on a uniformly treated patient population,[11] our study is confronted with many variables (including drug unavailability, risk-adapted treatment, comorbidities and time from diagnosis to treatment initiation) that cannot be fully controlled. Nevertheless, we are firm believers that clinical data obtained from real-world studies, if properly registered and with guaranteed accessibility, can provide representative evidence from routine practice about the clinical outcomes of patients, without the classical selection criteria of clinical trials.[23] Most importantly, these findings can serve as a more reliable basis for extrapolation of data to understudied populations.

Table 1. Clinical and baseline characteristics according to the *DNMT3A* mutational status and according to the *DNMT3A*/*NPM1*/*FLT3*-ITD mutations.

Characteristics	All patients		*DNMT3A* mutated		*DNMT3A* non-mutated		*P* value[1]	Triple-mutated		Non-triple-mutated		*P* value[1]
	No.	%	No.	%	No.	%		No.	%	No.	%	
Age, years							< 0.001*					0.003*
18–40 years	157	31	6	9.4	151	34.1		3	8.6	154	32.6	
41–60 years	185	36.5	34	53.1	151	34.1		21	60	164	34.7	
60 years and older	165	32.5	24	37.5	141	31.8		11	31.4	154	32.6	
Median (range)	50.6 (18, 93.8)		54.4 (27, 91)		49.2 (18, 93.8)		0.003*	54.8 (27, 77.5)		49.9 (18, 93.8)		0.08
Sex							0.033*					0.292
Female	269	53.1	42	66.6	227	51.2		22	62.9	247	52.3	
Male	238	46.9	22	34.4	216	48.8		13	37.1	225	47.7	
FAB subtype							0.756					0.928
M0	22	4.7	2	3.4	20	4.9		1	3.1	21	4.8	
M1	90	19.3	9	15.3	81	19.9		7	21.9	83	19.1	
M2	150	32.1	17	28.8	133	32.6		12	37.5	138	31.7	
M4	150	32.1	23	39	127	31.1		10	31.3	140	32.2	
M5	41	8.8	7	11.9	34	8.3		2		39	9	
M6	10	2.1	1	1.7	9	2.2		-	-	10	2.3	
M7	4	0.9	-	-	4	1		-	-	4	0.9	
Missing data	40	-	5	-	35	-		3	-	37	-	
Cytogenetic risk stratification[2]							< 0.001*					0.006*
Favorable	69	19.5	-	-	69	22.6		-	-	69	21.1	
Intermediate	227	64.3	43	89.6	184	60.3		24	92.3	203	62.1	
Adverse	57	16.1	5	10.4	52	17		2	7.7	55	16.8	
Missing data[3]	154	-	16	-	138	-		-	-	145	-	
FLT3-ITD							< 0.001*					
Mutated	134	26.4	42	65.6	94	20.8		-	-	-	-	
Non-mutated	373	73.1	22	34.4	351	79.2		-	-	-	-	
NPM1							< 0.001*					
Mutated	145	28.6	44	68.8	103	22.8		-	-	-	-	
Non-mutated	362	71.4	20	31.3	342	77.2		-	-	-	-	
Bone marrow blasts, %, median (range)	67 (20, 100)		74 (20, 96)		66 (10, 100)		0.143	77 (21, 96)		66 (20, 100)		0.117
Leukocyte counts, $\times 10^9$/L, median (range)	27.3 (0.28, 790)		61 (0.28, 435)		25.6 (0.6, 790)		0.009*	67 (0.86, 435)		26 (0.28, 790)		0.001*
Platelet counts, $\times 10^9$/L, median (range)	45 (3, 600)		59 (5, 404)		44 (3, 600)		0.818	60 (10, 196)		45 (3, 600)		0.763
Hemoglobin, g/dL, median (range)	8.1 (3, 16.2)		8 (3.1, 11.6)		8.1 (3, 16.2)		0.727	7.8 (3.9, 13.2)		8.1 (3, 16.2)		0.534

LDH level, U/I, median (range)	682.5 (116, 11, 722)	775 (136, 5,289)	662 (116, 11,722)	0.188	767 (136, 2,789)	672 (116, 11,722)	0.607
Complete remission, %	60	53	60	0.786	62	59	0.725
Overall survival, % (95% CI)	20 (16, 24)	9 (3, 18)	22 (17, 27)	0.003*	3 (0-14)	22 (17, 26)	0.011*
Disease-free survival, % (95% CI)	38 (31, 45)	19 (7, 34)	42 (34, 49)	< 0.001*	4 (0-19)	42 (35, 50)	< 0.001*
Cumulative incidence of relapse, % (95% CI)	54 (48, 60)	72 (58, 86)	50 (42, 57)	< 0.001*	85 (71, 98)	50 (43, 56)	< 0.001*

NOTE: * Indicates statistically significant differences.

1: Missing values were excluded for the calculation of *P* values.

2: The cytogenetic risk groups were defined according to Medical Research Council criteria.[24]

3: Material not available or no metaphases detected.

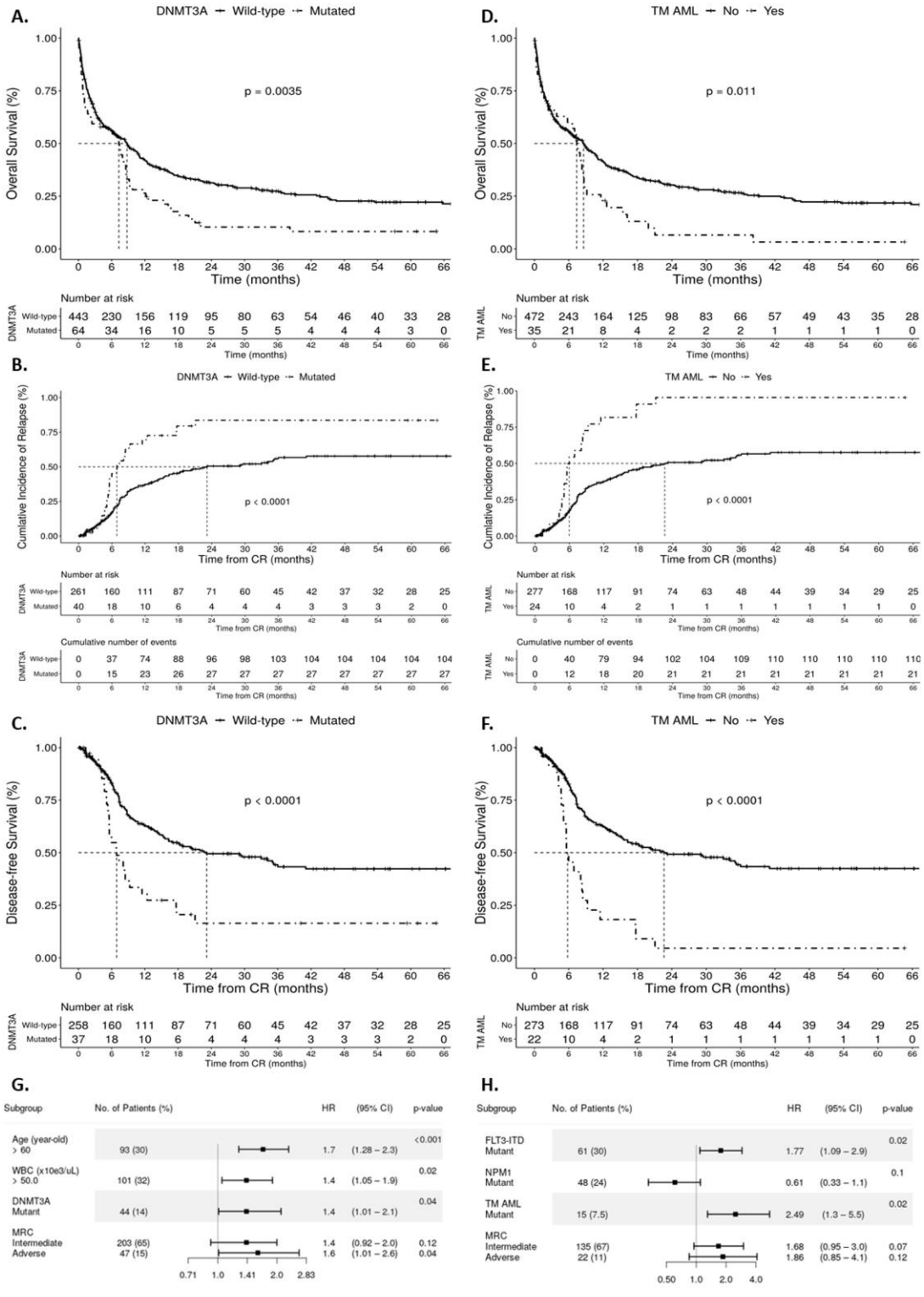


Figure 1. The probability of overall survival (A), cumulative incidence of relapse (B) and disease-free survival (C) in patients with AML according to *DNMT3A* mutations. Overall survival (D), cumulative incidence of relapse (E) and disease-

free survival (F) in triple-mutated (TM-AML) patients Survival curves were estimated using the Kaplan–Meier method, and the log-rank test was used for comparison. Cumulative incidence curves for non-relapse death and relapse with or without death were constructed to reflect time to relapse and time to non-relapse death as competing risks. Time to relapse and time to non-relapse death were measured from the date of complete remission. (G) Multivariate Cox model for overall survival and (H) disease-free survival.

References

1. Thol F, Damm F, Lüdeking A, et al. Incidence and prognostic influence of DNMT3A mutations in acute myeloid leukemia. *J. Clin. Oncol.* 2011;29(21):2889–96.
2. Renneville A, Boissel N, Nibourel O, et al. Prognostic significance of DNA methyltransferase 3A mutations in cytogenetically normal acute myeloid leukemia: a study by the Acute Leukemia French Association. *Leukemia*. 2012;26(6):1247–54.
3. Gaidzik VI, Schlenk RF, Paschka P, et al. Clinical impact of DNMT3A mutations in younger adult patients with acute myeloid leukemia: results of the AML Study Group (AMLSG). *Blood*. 2013;121(23):4769–77.
4. Kiyoi H, Naoe T, Nakano Y, et al. Prognostic implication of FLT3 and N-RAS gene mutations in acute myeloid leukemia. *Blood*. 1999;93(9):3074–80.
5. Cancer Genome Atlas Research Network, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 2013;368(22):2059–74.
6. Roller A, Grossmann V, Bacher U, et al. Landmark analysis of DNMT3A mutations in hematological malignancies. *Leukemia*. 2013;27(7):1573–8.
7. Ley TJ, Ding L, Walter MJ, et al. DNMT3A mutations in acute myeloid leukemia. *N. Engl. J. Med.* 2010;363(25):2424–33.
8. Russler-Germain DA, Spencer DH, Young MA, et al. The R882H DNMT3A mutation associated with AML dominantly inhibits wild-type DNMT3A by blocking its ability to form active tetramers. *Cancer Cell*. 2014;25(4):442–54.
9. Guryanova OA, Shank K, Spitzer B, et al. DNMT3A mutations promote anthracycline resistance in acute myeloid leukemia via impaired nucleosome remodeling. *Nat. Med.* 2016;22(12):1488–1495.
10. Loberg MA, Bell RK, Goodwin LO, et al. Sequentially inducible mouse models reveal that Npm1 mutation causes malignant transformation of Dnmt3a-mutant clonal hematopoiesis. *Leukemia*. 2019;33(7):1635–1649.
11. Papaemmanuil E, Gerstung M, Bullinger L, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. *N. Engl. J. Med.* 2016;374(23):2209–2221.
12. Garg S, Reyes-Palomares A, He L, et al. Hepatic leukemia factor is a novel

leukemic stem cell regulator in DNMT3A, NPM1, and FLT3-ITD triple-mutated AML. *Blood*. 2019;134(3):263–276.
13. Verhaak RGW, Wouters BJ, Erpelinck CAJ, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica*. 2009;94(1):131–4.
14. Tyner JW, Tognon CE, Bottomly D, et al. Functional genomic landscape of acute myeloid leukaemia. *Nature*. 2018;562(7728):526–531.
15. Marcucci G, Metzeler KH, Schwind S, et al. Age-related prognostic impact of different types of DNMT3A mutations in adults with primary cytogenetically normal acute myeloid leukemia. *J. Clin. Oncol.* 2012;30(7):742–50.
16. Emperle M, Adam S, Kunert S, et al. Mutations of R882 change flanking sequence preferences of the DNA methyltransferase DNMT3A and cellular methylation patterns. *Nucleic Acids Res.* 2019;
17. Holz-Schietinger C, Matje DM, Reich NO. Mutations in DNA methyltransferase (DNMT3A) observed in acute myeloid leukemia patients disrupt processive methylation. *J. Biol. Chem.* 2012;287(37):30941–51.
18. Genovese G, Kähler AK, Handsaker RE, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. *N. Engl. J. Med.* 2014;371(26):2477–87.
19. Zink F, Stacey SN, Norddahl GL, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. *Blood*. 2017;130(6):742–752.
20. Buscarlet M, Provost S, Zada YF, et al. DNMT3A and TET2 dominate clonal hematopoiesis and demonstrate benign phenotypes and different genetic predispositions. *Blood*. 2017;130(6):753–762.
21. Nybakken GE, Canaani J, Roy D, et al. Quizartinib elicits differential responses that correlate with karyotype and genotype of the leukemic clone. *Leukemia*. 2016;30(6):1422–5.
22. Yang L, Rodriguez B, Mayle A, et al. DNMT3A Loss Drives Enhancer Hypomethylation in FLT3-ITD-Associated Leukemias. *Cancer Cell*. 2016;29(6):922–934.
23. Juliusson G, Lazarevic V, Hörstedt A-S, et al. Acute myeloid leukemia in the real world: why population-based registries are needed. *Blood*. 2012;119(17):3890–9.
24. Grimwade D, Hills RK, Moorman A V, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood*. 2010;116(3):354–65.

APÊNDICE B

O apêndice I compreende a metodologia, os resultados, a discussão e a bibliografia correspondentes aos objetivos 1, 3 e 6 que não apresentaram resultados satisfatórios para publicação.

1. Materiais e métodos

1.1. Casuística e tratamento

O presente trabalho foi realizado após a submissão e aprovação pelo Sistema Nacional de Informação sobre Ética envolvendo Seres Humanos (SISNEP-Plataforma Brasil) e pelos Comitês de Ética em Pesquisa (CEP) da Universidade Federal de Pernambuco (CAAE: 77527717.4.0000.5208), Fundação HEMOPE/PE (CAAE: 77527717.4.3001.5195), do Hospital do Câncer de Pernambuco (HCP) (CAAE: 77527717.4.3003.5205) e do Instituto de Medicina Integral Professor Fernando Figueira (IMIP) (CAAE: 77527717.4.3002.5201) (Anexos II, III, IV e V respectivamente) sendo desenvolvido obedecendo integralmente os princípios éticos estabelecidos na resolução 466/12 do Conselho Nacional de Saúde (CNS).

Para o desenho do IP foi utilizada uma coorte de dados públicos disponível no website https://www.cbioportal.org/. Essa coorte, que será chamada de The Cancer Genome Atlas (TCGA) apresenta dados clínicos e de sequenciamento do genoma e transcriptoma completos de 179 pacientes (The Cancer Genome Atlas, 2013). Para esse estudo excluímos os pacientes classificados como FAB M3, com detecção do transcrito *BCR-ABL1* ou aqueles que não tinham os dados de sequenciamento de RNA. Dessa maneira, essa coorte foi composta por 154 pacientes.

Para a validação do IP será utilizada uma coorte interna, que será chamada de Recife, de pacientes com LMA *de novo* não-LPA, diagnosticados de março de 2010 a junho de 2020 em hospitais da cidade de Recife, Pernambuco (HEMOPE, HCP e IMIP). Até o momento essa coorte é composta de 128 pacientes diagnosticados até novembro de 2018. O material de escolha para análises moleculares são amostras de medula óssea (MO) de blastos leucêmicos circulantes. É utilizado somente o material excedente àquele utilizado ao diagnóstico, com autorização prévia dos participantes após leitura e consentimento do termo de consentimento livre e esclarecido (TCLE) (Apêndice V), seguindo as recomendações da declaração de Helsinki. Os pacientes são maiores de 18 anos e submetidos ao esquema de tratamento 3+7, como descrito anteriormente, ou tratamento paliativo com baixas doses de Ara-C (Lima *et al*., 2015).

1.2. Validação do IP na coorte de validação e análises moleculares

As análises moleculares realizadas nas amostras dos pacientes da coorte de validação foram desenvolvidas no setor de Hematologia do LabCen/CB/UFPE e no Centro de Pesquisas Aggeu Magalhães. Como o IP não se mostrou uma ferramenta robusta para estratificação de risco dos pacientes LMA, decicidimos não validar o IP na coorte de validação.

1.2.1. Extração do material genético

A extração do material genético (DNA genômico) dos leucócitos totais das amostras de MO foram realizadas a partir da metodologia de fenol-clorofórmio modificado (Davis *et al*., 1986).

1.2.2. Análise das mutações *FLT3*-ITD e de mutações no gene *NPM1*

A análise das mutações *FLT3*-ITD e no gene *NPM1* foi realizada por PCR convencional, seguido de eletroforese em gel de agarose a 4%, como previamente descrito (Lima *et al*., 2015). A sequência dos *primers* está descrita na Tabela 1.

Tabela 1: Sequência dos *primers* utilizados nas reações de PCR convencional para pesquisa das mutações *FLT3*-ITD e no gene *NPM1.*

Primers	**Sentido**	**Sequência**
FLT3-ITD	Forward	5’ GCAATTTAGGTATGAAAGCCAGC 3’
	Reverse	5’ CTTTCAGCATTTTGACGGCAACC 3’
NPM1	Forward	5’ GTGGTAGAATGAAAAATAGAT 3’
	Reverse	5’ CTTGGCAATAGAACCTGGAC 3’

1.3. Desfechos

O seguimento dos pacientes foi censurado/atualizado em março de 2020. Os pacientes em que o seguimento foi perdido foram censurados na última data em que sabidamente estavam vivos. Remissão hematológica completa (RC) foi definida como aspirados de medula óssea normocelulares, contendo menos de 5% de blastos leucêmicos e com evidências de maturação normal dos outros elementos, bem como a não visualização de blastos leucêmicos circulantes ou evidência de leucemia extramedular. Recaída foi definida como a presença de 5% ou mais de blastos na medula óssea, presença de blastos leucêmicos circulantes no sangue periférico, ou o desenvolvimento de leucemia extramedular. Sobrevida global e sobrevida livre de doença serão calculadas usando o método Kaplan-Meier. Sobrevida global (SG) refere-se ao tempo compreendido entre o diagnóstico e o óbito/censura. Sobrevida livre de doença (SLD) é o tempo entre a remissão completa e o primeiro evento: recaída ou óbito/censura.

1.4. Análise estatística

A análise estatística foi realizada com o auxílio dos softwares SPSS Statistics 19.0, STATA versão 14.9 e GraphPad Prism 5. As variáveis contínuas, inicialmente, foram submetidas ao teste de Kolmogorov-smirnov (KS), a fim de analisar se a variável apresentava distribuição normal (DN). Para as variáveis que apresentaram DN, a comparação entre os grupos propostos foi realizada utilizando o teste T não pareado ou ANOVA. Para as variáveis que não apresentavam DN, a comparação entre os grupos propostos foi realizada utilizando o teste Mann-Whitney ou Kruskal-Wallis. As variáveis categóricas foram comparadas entre os grupos utilizando o teste Chi-quadrado ou o teste exato de Fisher. As probabilidades estimadas, a média e a mediana para curvas de sobrevida foram calculadas pelo método de Kaplan-Meier e o teste

de log-rank foi utilizado para comparar as diferenças entre as curvas. Valores de *p* menores que 0,05 foram considerados significativos. Para a análise de expressão gênica os valores contínuos da expressão relativa foram dicotomizados na mediana definindo dois grupos: baixa expressão e alta expressão.

Para elaboração do índice prognóstico (IP), os pesos de cada variável foram derivados do modelo de risco proporcional de Cox aplicado às variáveis dos pacientes da coorte de desenho, utilizando a sobrevida global como variável do desfecho clínico. A proporção de risco (do inglês, *hazard ratio*, HR) foi calculada para cada variável molecular separadamente e apenas as variáveis com valor de *P* menores que 0,05 na análise univariada para a sobrevida global foram incluídas no modelo. O valor do HR foi convertido em número inteiro. Como apenas duas variáveis, mutações no gene *TP53* (*TP53*^{mut}) e alta expressão do gene *ID1* (*ID1*^{alta}), apresentaram *p-value* menores que 0,05 e o HR do *TP53*^{mut} foi aproximadamente o dobro do valor do HR de *ID1*^{alta}, atribuímos peso 1 para a variável *ID1*^{alta} e peso 2 para a variável *TP53*^{mut}.

Cada paciente representou a soma dos pesos de cada variável que ele apresentou, produzindo uma variável contínua com intervalo de 0 a 3. Posteriormente, o ponto de corte ideal para dicotomizar os pacientes em dois grupos (baixo risco e alto risco) foi determinado pela curva ROC. Dessa forma, o IP foi desenvolvido e sua capacidade de predição de sobrevida foi calculada a partir da curva ROC através da medida da área sobre a curva (AUC, do inglês *area under the curve*), utilizando o *status* de sobrevida global (vivo na censura ou morto) como padrão ouro. AUC é a medida de desempenho de um teste e é interpretada como a probabilidade de um indivíduo com o evento ter um risco preditivo maior do que uma pessoa sem o evento. Um valor de AUC maior que 0,5 significa que o teste apresenta capacidade de discriminação e quanto maior esse valor, maior essa capacidade, sendo o valor máximo igual a 1,0 (Martinez *et al*., 2003). Uma vez desenvolvido o IP, ele será aplicado na coorte de validação e apenas as variáveis incluídas no IP serão analisadas na validação.

2. Resultados

2.1. Caracterização das coortes de desenho e de validação do IP

A tabela 5 descreve as características demográficas, clínico-laboratoriais, citogenéticas e moleculares das coortes de desenho (TCGA) e de validação (Recife) do IP. A coorte TCGA foi composta por 154 pacientes LMA não-M3 e sem detecção de transcrito *BCR-ABL1*. A mediana de idade dessa coorte foi de 59 anos (intervalo: 18-88 anos) e 53,2% dos pacientes eram do sexo masculino. Houve uma maior frequência do grupo FAB M1 (28,1%), seguido pelo grupo FAB M2 (24,2%).

O resultado da análise citogenética da coorte TCGA estava disponível em 151 pacientes, sendo a maioria dos casos (52,3%) correspondente ao cariótipo normal. Nesse contexto, o risco citogenético proposto pelo grupo colaborativo *Medical Research Council* (MRC) com maior frequência foi o intermediário (64,9%), seguido pelo grupo de risco adverso (22,1%) e, por fim, o grupo favorável (11,0%). Segundo a estratificação de risco proposta pelo

grupo colaborativo *European LeukemiaNet* em 2010 (ELN2010), 34,4% dos pacientes estão no grupo de risco favorável, 25,8% estão no grupo intermediário I, 15,9% estão no grupo intermediário II e 23,8% no grupo adverso.

Tabela 5: Características demográficas, clínico-laboratoriais, citogenéticas e moleculares das coortes TCGA e Recife.

	TCGA		**Recife**	
Características	**Freq. (N=154)**	**%**	**Freq. (N=128)**	**%**
Mediana de idade, anos	59	-	51	-
Intervalo	18-88	-	18-94	-
Sexo				
Masculino	82	53,2	68	53,1
Feminino	72	46,8	30	46,9
Hb (g/dL), mediana	-	-	8,1	-
Intervalo	-	-	3,0-11,7	-
Blastos MO (%), mediana	72	-	69	-
Intervalo	30-100	-	20-97	-
Leucócitos (x10^9/L), mediana	19,75	-	49,2	-
Intervalo	0,6-297,4	-	0,76-435,94	-
Plaquetas (x10^9/L), mediana	-	-	51,0	-
Intervalo	-	-	6,0-404,0	-
Subtipo FAB				
M0	16	10,5	6	5,1
M1	43	28,1	18	15,3
M2	37	24,2	45	38,1
M4	34	22,2	35	29,7
M5	18	11,8	13	11,0
M6	2	1,3	-	
M7	3	2,0	1	0,8
Dados perdidos	1		10	
Citogenética				
t(8;21)	7	4,6	12	21,8
t(16;16)/inv(16)	10	6,6	5	9,1
Translocação MLL	3	2,0	-	-
Cariótipo complexo	22	14,6	7	12,7
Cariótipo normal	79	52,3	25	45,5
Anormalidade de risco intermediário	21	13,9	6	10,9
Outras anormalidades de risco adverso	9	6,0	-	-
Dados perdidos	3		73	

Risco MRC				
Favorável	17	11,0	17	13,3
Intermediário	100	64,9	31	56,4
Adverso	34	22,1	7	12,7
Dados perdidos	3		73	
Risco molecular				
Baixo	17	11,3	16	12,6
Alto	92	60,9	111	87,4
Dados perdidos	3		1	
ELN 2010				
Favorável	52	34,4	22	40,7
Intermediário I	39	25,8	19	35,2
Intermediário II	24	15,9	6	11,1
Adverso	36	23,8	7	13,0
Dados perdidos	3		74	

Freq.: frequência absoluta. Hb: hemoglobina.

Na coorte Recife, 288 pacientes com LMA *de novo* não-LPA foram diagnosticados no período do estudo, entretanto apenas 143 pacientes apresentavam dados clínicos completos e com ambas as amostras de RNA e DNA, coletadas ao diagnóstico, disponíveis. Desses 143 pacientes, o RNA não apresentava boa qualidade para as análises em 15 amostras, ficando a coorte composta por 128 pacientes. A mediana de idade dessa coorte foi de 51 anos e 53,1% dos pacientes eram do sexo masculino. Diferentemente da coorte TCGA, o grupo FAB M2 foi o mais frequente (38,1%), seguido pelo grupo FAB M4 (29,7%).

A análise citogenética foi realizada com sucesso em 55 (43%) pacientes. Desses, 45,5% dos pacientes apresentaram cariótipo normal. Nesse contexto, 56,4% dos pacientes estavam no grupo de risco intermediário, 13,3% no grupo favorável e 12,7% no grupo adverso. No que diz respeito a estratificação de risco ELN2010, 40,7% dos pacientes fazem parte do grupo de risco favorável, 35,2% dos pacientes são do grupo de risco intermediário I, 11,1% do grupo intermediário II e 13% estão no grupo adverso, de maneira semelhante a coorte TCGA.

2.2. Desenho do índice de prognóstico

A coorte TCGA foi utilizada para o desenho do IP por apresentar dados clínicos e de sequenciamento do genoma e exoma completos disponíveis publicamente. A tabela 6 mostra o resultado do modelo do risco proporcional de Cox univariado, utilizando como variável de desfecho a sobrevida global, aplicada às variáveis moleculares. Apenas as mutações no gene *TP53* (*TP53*^{mut}) e a alta expressão do gene *ID1* (*ID1*^{alta}) estiveram associadas com pior sobrevida global (*TP53*^{mut}: p<0,001; HR=3,78 / *ID1*^{alta}: p= 0,023; HR= 1,56).

A figura 1 apresenta as curvas de sobrevida global dos grupos segundo *status* mutacional do gene *TP53* e níveis de expressão do gene *ID1*.

Tabela 6: Modelo de risco proporcional de Cox univariado para as variáveis moleculares utilizando a SG como variável de desfecho.

SG (n=154)	***p-value***	**HR**	**IC95%**
***CEBPA* mutação bialélica, ausente *versus* presente**	0,359	0,584	0,185-1,844
***NPM1,* selvagem *versus* mutado**	0,995	0,999	0,660-1,512
***FLT3*-ITD, ausente *versus* presente**	0,878	0,960	0,570-1,618
***WT1,* selvagem *versus* mutado**	0,764	0,889	0,413-1,916
***IDH1,* selvagem *versus* mutado**	0,251	0,670	0,338-1,328
***IDH2,* selvagem *versus* mutado**	0,868	0,950	0,520-1,737
***DNMT3A* R882, ausente *versus* presente**	0,144	1,476	0,876-2,486
***ASXL1,* selvagem *versus* mutado**	0,163	2,270	0,718-7,179
***TP53,* selvagem *versus* mutado**	<0,001	3,781	2,068-6,914
***BAALC,* baixa *versus* alta expressão**	0,636	0,912	0,623-1,335
***ERG,* baixa *versus* alta expressão**	0,062	0,694	0,472-1,018
***MECOM,* baixa *versus* alta expressão**	0,546	0,889	0,607-1,303
***KMT2E,* baixa *versus* alta expressão**	0,180	0,770	0,525-1,128
***ID1,* baixa *versus* alta expressão**	0,023	1,559	1,062-2,289
***MN1,* baixa *versus* alta expressão**	0,835	0,960	0,656-1,406

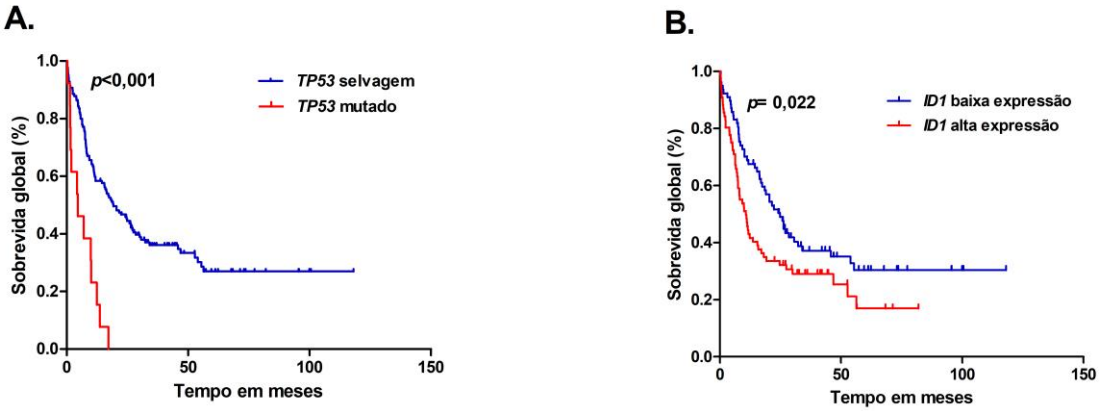


Figura 1: Impacto prognóstico da alta expressão do gene *ID1* e de mutações no gene *TP53* em pacientes da coorte de desenho do IP (TCGA). (A) Sobrevida global de acordo com o status mutacional do gene *TP53*. (B) Sobrevida global de acordo com os níveis de expressão do gene *ID1*.

Diante desses resultados, o IP foi composto por apenas duas variáveis moleculares, *TP53*^{mut} e *ID1*^{alta}. Como o HR do *TP53*^{mut} foi aproximadamente o dobro do valor do HR do *ID1*^{alta}, foi atribuído peso 1 para a variável *ID1*^{alta} e peso 2 para a variável *TP53*^{mut}. Para cada paciente foi feita a soma dos pesos atribuídos para cada variável apresentada: *TP53* não mutado foi atribuído peso 0 *versus TP53*^{mut}, peso 2 e baixa expressão do gene *ID1*, peso 0 *versus ID1*^{alta}, peso 1. Como resultado, foi criada uma variável contínua (IP) com intervalo de 0 a 3. O ponto de corte ideal para definir os grupos de alto e baixo risco foi determinado pela curva ROC: pacientes com IP menor que 1 foram considerados de baixo risco e pacientes com IP maior ou igual a 1 foram considerados de alto risco.

Os pacientes considerados de alto risco para o IP apresentaram taxa de SG em 5 anos inferior aos de baixo risco (16,3% *versus* 31,6%, respectivamente; *p*= 0,005) (Figura 2A). Na análise multivariada, utilizando como covariáveis a idade, contagem de leucócitos e a ELN2010, o IP se mostrou uma variável independente para a SG (Tabela 7). Em relação a SLD, o IP não foi capaz de distinguir os grupos de alto e baixo risco (25,1% *versus* 34,8%, respectivamente; *p*= 0,989) (Figura 2D). O modelo de estratificação de risco proposto pelo grupo colaborativo European LeukemiaNet em 2010 (Figura 2B) também foi capaz de diferenciar os grupos de risco quanto a sobrevida global (*p*= 0,027). Entretanto, em relação a sobrevida livre de doença, não foi observado diferença estatística entre os grupos de risco ELN2010 (*p*=0,188).

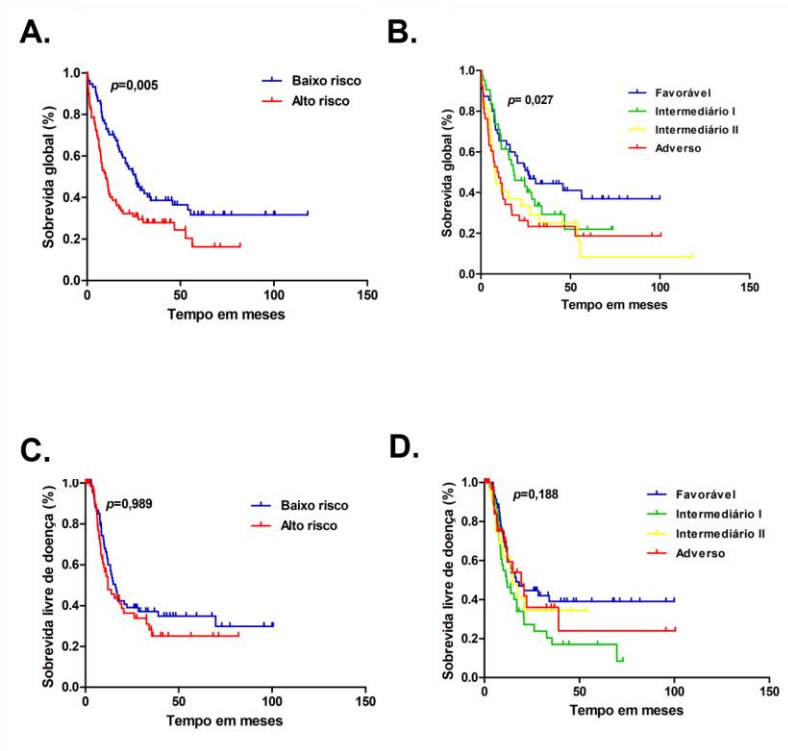


Figura 2: Estratificação de risco dos pacientes LMA segundo a proposta IP e do grupo colaborativo ELN2010. (A) Sobrevida global e (C) sobrevida livre de doença de acordo com a estratificação de risco proposta pelo IP. (B) Sobrevida global e (D) sobrevida livre de doença de acordo com a estratificação de risco proposta pelo ELN2010.

Tabela 7: Modelo de risco proporcional de Cox multivariado utilizando a SG e SLD como variável de desfecho clínico.

Variáveis	***p-value***	**HR**	**IC95%**
SG (n= 154)			
IP, baixo *versus* alto risco	0,046	1,509	1,008-2,260
Idade, <60 *versus* ≥60 anos	<0,001	1,032	1,016-1,049
Leucócitos, <50 *versus* ≥ 50x10^9/L	0,004	1,007	1,002-1,012
ELN 2010, fav. *vs* int.I	0,900	1,037	0,591-1,817
ELN 2010, fav. *vs* int II	0,045	1,865	1,013-3,432
ELN 2010, fav. *vs* adv	0,015	2,017	1,145-3,552

Fav.: favorável. Int.: intermediário. Adv.: adverso

A partir da medida de AUC, calculada pela curva ROC, podemos comparar o desempenho do IP em predizer sobrevida com o da proposta de estratificação de risco ELN2010. Nesse contexto, o IP apresentou um desempenho inferior (AUC=0,568) à proposta ELN2010 (AUC=0,630) (Figura 3).

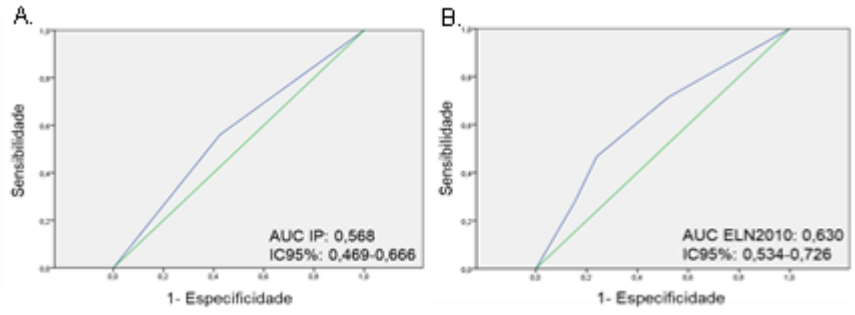


Figura 3: Curva ROC e medida de AUC para a proposta de estratificação de risco IP (A), ELN2010 (B).

3. Discussão

Embora a heterogeneidade citogenética e molecular da LMA e sua importância prognóstica sejam reconhecidas e bem aceitas, a tradução dessas informações para a prática clínica ainda é obscura. Diversos índices de prognóstico, baseados em marcadores moleculares, têm sido propostos na tentativa de complementar a informação da proposta, atualmente aceita e aplicada na prática clínica, do grupo colaborativo ELN (Damm *et al*., 2011; Horibata *et al.*, 2019; Ng *et al*., 2018; Wagner *et al*., 2019). Entretanto, os resultados ainda são bastante conflitantes no que diz respeito aos marcadores prognósticos, pontos de corte ideal para o valor de expressão gênica e ferramentas e metodologias de bioinformática e bioestatística utilizadas.

Nesse contexto, na tentativa de oferecer informação adicional para o refinamento da estratificação de risco dos pacientes com LMA, nos propusemos a desenvolver de um índice de prognóstico baseado no impacto aditivo de marcadores moleculares no desfecho clínico de pacientes com LMA. Escolhemos marcadores prognósticos já descritos na literatura como associados com o prognóstico da LMA (Damm *et al*., 2011; Dohner *et al*., 2017; Papaemmanuil *et al*., 2016; Patel *et al*., 2012). Entretanto, ao aplicarmos na coorte de desenho TCGA, apenas as mutações no gene *TP53* e a alta expressão de *ID1* estiveram associadas com uma pior sobrevida global, na análise do modelo do risco proporcional de Cox univariado. Dessa forma, o IP foi desenhado baseado na soma dos pesos atribuídos a apenas essas duas variáveis a partir dos seus HR.

Pacientes com IP de alto risco apresentaram uma pior sobrevida global em relação aos de baixo risco, entretanto, não houve diferença entre esses grupos de risco quanto a sobrevida livre doença. Em análise multivariada, o IP se mostrou uma variável independente para predição de sobrevida global. Entretanto, o IP se mostrou uma proposta de estratificação de risco com capacidade discriminatória inferior à proposta ELN2010, observada a partir do valor de AUC, calculado a partir da curva ROC. Em contrapartida, outros autores foram capazes de desenvolver índices com elevado poder discriminatório (Huang *et al.*, 2017; Ng *et al*., 2016; Wagner *et al*., 2019). Dessa maneira, vemos que o IP não se mostrou uma ferramenta robusta para a estratificação de risco da LMA.

4. Referências bibliográficas

Ahn J, Kim J, Kim H, Kim Y, Lee S, Jung S, Yang D, Lee J, Kim NY, Choi SH et al (2016) Normal karyotype acute myeloid leukemia patients with CEBPA double mutation have a favorable prognosis but no survival benefit from allogeneic stem cell transplant. Ann Hematol. 95(2):301-10.

Ashraf S, Noguera NI, Di Giandomenico J, Zaza S, Hasan SK, Lo-Coco F (2013) Rapid detection of IDH2 (R140Q and R172K) mutations in acute myeloid leukemia. Ann Hematol. 92(10):1319-23.

cBioPortal for Cancer Genomics, https://www.cbioportal.org/ (16 de novembro de 2021)

Chou W, Huang Y, Huang C, Tseng M, Tien H (2010) A single-tube, sensitive multiplex method for screening of isocitrate dehydrogenase 1 (IDH1) mutations. Blood. 116(3):495-6.

Damm F, Heuser M, Morgan M, et al (2011) Integrative prognostic risk score in acute myeloid leukemia with normal karyotype. Blood. 117(17): 4561-4568.

Dohner H, Estey E, Grimwade D (2017) Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 129(4): 424-447.

Gelsi-Boyer V, Trouplin V, Adélaïde J, Bonansea J, Cervera N, Carbuccia N, Lagarde A, Prebet T, Nezri M, Sainty D et al (2009) Mutations of polycomb-associated gene ASXL1 in myelodysplastic syndromes and chronic myelomonocytic leukaemia. Br J Haematol. 145(6):788-800.

Gene Expression Omnibus, National Center for Biotechnology Information, https://www.ncbi.nlm.nih.gov/geo/ (16 de novembro de 2021)

Gröschel S, Lugthart S, Schlenk RF, Valk PJM, Eiwen K, Goudswaard C, van Putten WJL, Kayser S, Verdonck LF, Lübbert M et al (2010) High EVI1 expression predicts outcome in younger adult patients with acute myeloid leukemia and is associated with distinct cytogenetic abnormalities. J Clin Oncol. 28(12):2101-7.

Horibata S, Gui G, Lack J, DeStefano CB, Gottesman MM, Hourigan CS (2019) Heterogeneity in refractory acute myeloid leukemia. Proc Natl Acad Sci U S A. 116(21):10494-10503.

Huang R, Liao X, Li Q (2017) Identification and validation of potential prognostic gene biomarkers for predicting survival in patients with acute myeloid leukemia. Onco Targets Ther. 10: 5243-5254.

Lima AS, de Mello MR, Fernandes E, Bezerra MF, Oliveira MM, Duarte BK, de Assis RA, Souto FR, Ramos CF, Machado CG et al. Clinical outcomes of patients with acute myeloid leukemia: evaluation of genetic and molecular findings in a real-life setting. Blood. 126(15):1863-5.

Martinez EZ, Louzada Neto F, Pereira B de B (2003) A curva ROC para testes diagnósticos. Cadernos de Saúde Coletiva. 11(1): 7-31.

National Cancer Institute, The Cancer Genome Atlas Program, https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga (16 de novembro de 2021)

National Cancer for Biotechnology Information, https://www.ncbi.nlm.nih.gov/ (16 de novembro de 2021)

Ng SWK, Mitchell A, Kennedy JA, et al (2016) A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature. 000: 1-17

Ng CWS, Kosmo B, Lee P, Lee CK, Guo J, Chen Z, Chiu L, Lee HK, Ho S, Zhou J et al (2018) CEBPA mutational analysis in acute myeloid leukaemia by a laboratory-developed next-generation sequencing assay. J Clin Pathol 71(6):522-531.

Papaemmanuil E, Gerstung M, Bullinger L, et al (2016) Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 374(23): 2209-2221.

Patel JP, Gönen M, Figueroa ME, et al (2012) Prognostic Relevance of Integrated Genetic Profiling in Acute Myeloid Leukemia. N Engl J Med. 366(12): 1079-1089.

Renneville A, Boissel N, Zurawski V, Llopis L, Biggio V, Nibourel O, Philippe N, Thomas X, Dombret H, Preudhomme C (2009) Wilms tumor 1 gene mutations are associated with a higher risk of recurrence in young adults with acute myeloid leukemia: a study from the Acute Leukemia French Association. Cancer. 115(16):3719-27.

The Cancer Genome Atlas Research Network (2013) Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia. N Engl J Med. 368(22): 2059-2074.

Wagner S, Vadakekolathu J, Tasian SK, Altmann H, Bornhäuser M, Pockley AG, Ball GR, Rutella S (2019) A parsimonious 3-gene signature predicts clinical outcomes in an acute myeloid leukemia multicohort study. Blood Adv. 3(8):1330-1346.

APÊNDICE C

Material suplementar do artigo referente ao capítulo I

METHODS

Patients and study design

We performed a retrospective cohort study investigating the prevalence of *DNMT3A* mutations and the 3-way co-occurrence of *DNMT3A*/*NPM1*/*FLT3*-ITD mutations in patients with acute myeloid leukemia (AML) at diagnosis. Between June 2003 to January 2019, 590 consecutive patients were diagnosed with AML at five Brazilian reference centers specialized on AML treatment. AML diagnosis was determined by cytomorphology, immunophenotyping and genetic studies according to WHO criteria. In this study only patients diagnosed with *de novo* AML were included (507 patients), while those with acute promyelocytic leukemia (36 patients), therapy-related AML (25 patients), or a previous history of myelodysplastic syndrome (22 patients) were excluded. An overview of patient characteristics according to the *DNMT3A* mutational status can be found in Table 1. All patients or their relatives gave their written informed consent for scientific evaluations. The study was approved by the Internal Review Board (#7147/2005) and adhered to the tenets of the Declaration of Helsinki.

Treatment protocol

For five patients out of 507, there was no available information on the treatment procedures. The majority of the patients (405 patients, 80%) were intensively treated with cytotoxic chemotherapy. Induction therapy consisted of daunorubicin (60 or 90 mg/m^2 daily for 3 days) and cytarabine (100 or 200 mg/m^2 daily for 7 days) or thioguanine, cytosine arabinoside, and daunorubicin[1], followed by two or three cycles of consolidation therapy with high doses cytarabine (1.5g/m^2 or 3g/m^2 for 3 days). According to clinical judgment and donor availability, a post-remission therapy based on autologous or allogeneic transplantation was performed in only 19 patients. Complete remission (CR) was assessed by bone marrow examination 28 days after each course of chemotherapy. Patients that did not achieve complete remission after one course of chemotherapy, received a second course. Of the 97 patients, which did not receive intensive treatment, 85 were 60 years old or above. These patients were treated with low-dose of ARA-C in combination with etoposide, thioguanine, and idarubicin, or exposed to palliative care.

Cytogenetic and molecular analyses

All materials used for genetic analyses were obtained at diagnosis and were processed in the reference laboratories of each participating center. Cytogenetic analysis was performed from bone marrow aspirates according to standard techniques for chromosomal banding.

For molecular analyses, genomic DNA was extracted using the Puregene kit (Gentra System) according to the manufacturer's protocol. Standard polymerase chain reaction and sequencing techniques were performed for the detection of *NPM1,* and *FLT3*-ITD mutations as described previously.[2–4] *FLT3* allelic ratio was determined in patients with the *FLT3*-ITD mutated status, following the 2017

European LeukemiaNet recommendations for AML.[5,6] *DNMT3A* mutations on exon 23 were obtained by Sanger sequencing(flanking the hotspot R882 region). PCR products were purified with ExoSAP-IT™ PCR Product Cleanup Reagent (Applied Biosystems). Sequencing reactions were performed with BigDye Terminator v3.1 (Applied Biosystems) and run in capillary electrophoresis (ABI 3500, Applied Biosystems). Primer set for exon 23 amplification and sequencing were: *DNMT3A*-EXON23 FW 5'-GTGTGGTTAGACGGCTTCC-3' and *DNMT3A*-EXON23 RV 5'-CTCTCCCACC TTTCCTCTG–3'.

Statistics analysis and clinical endpoints

Descriptive analyses were performed for patient baseline features. Fisher's exact test or Chi-square test, as appropriate, was used to compare categorical variables. Kruskal-Wallis test was used to compare continuous variables. Details of the statistical analysis and clinical endpoints were described elsewhere.[7] All *P*-values were two sided with a significance level of 0.05. All calculations were performed using Stata Statistic/Data Analysis version 12 (Stata Corporation, USA), and R 3.3.2 (The CRAN project, www.r-project.org) software.

Public datasets

All clinical, cytogenetic and molecular data from the Gene Expression Omnibus were obtained from http://www.ncbi.nlm.nih.gov/geo, GSE6891.[8]

Patient's clinical features, mutational and transcriptome data from The Cancer Genome Atlas (TCGA)[9] project were obtained from http://cancergenome.nih.gov. For mutational landscape visualization, the Oncoprinter tool, available at www.cbioportal.org was used. For the RNAseq analysis, normalization of the counts by log2 of Counts Per Million (CPM) was used for PCA and expression plots. EdgeR was used to detect differential expression signatures between two conditions. To test the differentially expressed genes of R882-only compared to the triple-mutated patients, we used the design ~ 0 + group. Gene set enrichment analysis (GSEA) of differentially expressed genes was preformed using fGSEA (nperm=100000, minSize=8, maxSize=1000).[10] Gene expression signatures from MsigDB[11] and Gene Ontology (The Gene Ontology Consortium, 2017) were used for the analysis.

RESULTS

Clinical outcomes of intensively treated patients

This study aimed to present an additional analysis of clinical outcomes related to the *DNMT3A* mutations and the 3-way co-occurrence *DNMT3A/NPM1/FLT3*-ITD mutations. For this purpose, only patients submitted to intensive treatment were included. The clinical and baseline characteristics are summarized in the Supplemental table 1.

Of the 405 intensively treated patients, 271 (67%) patients achieved complete remission (CR), of whom 35/51 (68%) and 236/354 (66%) were assigned to the *DNMT3A* mutated and *DNMT3A* non-mutated groups, respectively. *DNMT3A* mutational status was not associated with CR achievement (*P*=0.874). With a median follow-up of 38 months (95% confidence interval [CI]: 26–51 months), the estimated 5-year overall survival (OS) and disease-free survival (DFS) values were 23% (95%CI: 18%–28%) and 37% (95%CI: 20%–45%), respectively. Patients with *DNMT3A* mutations had significantly lower OS rate (*P*=0.0012) and

DFS (*P*=0.0002) than those without *DNMT3A* mutations. The corresponding picture for the 3-way co-occurrence *DNMT3A/NPM1/FLT3*-ITD mutations presents the following information: CR achievement (*P*=0.682), OS (*P*=0.009) and DFS (*P*<0.001). Details for patient outcomes are summarized in Supplemental table 2.

Supplemental table 1. Clinical and baseline characteristics according to the *DNMT3A* mutational status and according to the *DNMT3A*/*NPM1*/*FLT3*-ITD mutations in intensively treated patients.

Characteristics	All patients		*DNMT3A* mutated		*DNMT3A* non-mutated		*P* value[1]	Triple-mutated		Non-triple-mutated		*P* value[1]
	No.	%	No.	%	No.	%		No.	%	No.	%	
Age, years							< 0.001*					0.007*
Median (range)	46.7 (18, 86.6)		53.6 (27, 77)		44.7 (18, 86.6)			54 (27, 69)		45.8 (18, 86.6)		
Sex							0.072					0.565
Female	212	52.3	33	67.7	179	50.6		17	58.6	195	51.9	
Male	193	47.7	18	35.3	175	49.4		12	41.4	181	48.1	
FAB subtype							0.861					0.999
M0	17	4.6	2	4.3	15	4.6		1	3.8	16	4.6	
M1	71	19	6	13	65	19.9		5	19.2	66	19	
M2	118	31.6	14	30.4	104	31.8		9	34.6	109	31.4	
M4	124	33.2	17	37	107	32.7		9	34.6	115	33.1	
M5	33	8.8	6	13	27	8.3		2	7.7	31	8.9	
M6	8	2.1	1	2.2	7	2.1		-	-	8	2.3	
M7	2	0.5	-	-	2	0.6		-	-	2	0.6	
Missing data	32	-	5	-	27	-		3	-	29	-	
Cytogenetic risk stratification[2]							0.002*					0.024*
Favorable	58	19.2	-	-	58	22.2		-		58	20.7	
Intermediate	198	65.6	36	87.8	162	62.1		20	90.9	178	63.6	
Adverse	46	15.2	5	12.2	41	15.7		2	9.1	44	15.7	
Missing data[3]	103	-	10	-	93	-		7	-	96	-	
FLT3-ITD							< 0.001*					
Mutated	111	27.4	35	68.6	76	21.5		-	-	-	-	
Non-mutated	294	72.6	16	31.4	278	78.5		-	-	-	-	
NPM1							< 0.001*					
Mutated	113	27.9	34	66.7	79	22.3		-	-	-	-	
Non-mutated	292	72.1	17	33.3	275	77.7		-	-	-	-	
Bone marrow blasts, %, median (range)	67 (20, 100)		69 (21, 96)		67 (20, 100)		0.308	75 (21, 96)		66 (20, 100)		0.211
Leukocyte counts, $\times 10^9$/L, median (range)	26.4 (0.3, 51.9)		59.8 (0.3, 204)		24.2 (0.6, 51.9)		0.035*	61.7 (1.8, 20.4)		24.4 (0.3, 51.9)		0.002*
Platelet counts, $\times 10^9$/L, median (range)	46 (4, 600)		60 (5, 404)		45 (4, 600)		0.682	65 (1, 196)		45 (4, 600)		0.453
Hemoglobin, g/dL, median (range)	8.1 (3, 16)		8.1 (3.9, 13.2)		8 (3, 16)		0.452	8 (3.9, 13.2)		8.1 (3, 16)		0.947
LDH level, U/I, median (range)	643.5 (116, 11,722)		771 (136, 5,289)		624 (116, 11,722)		0.153	767 (136, 2550)		628 (116, 11,722)		0.544

NOTE: * Indicates statistically significant differences.

1: Missing values were excluded for the calculation of *P* values.
2: The cytogenetic risk groups were defined according to Medical Research Council criteria.[12]
3: Material not available or no metaphases detected.

Supplemental table 2. Summary of outcomes of intensively treated AML patients according to *DNMT3A* and the 3-way co-occurrence *DNMT3A/NPM1/FLT3*-ITD mutations.

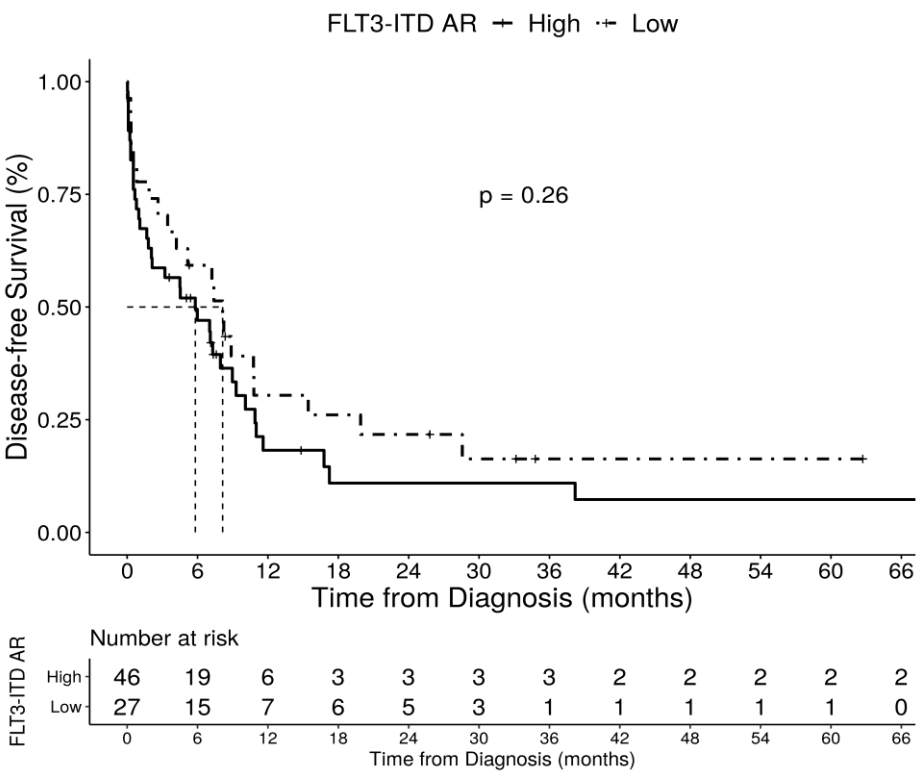
***DNMT3A* status**	CR %	OS		DFS	
		No.	5-yr % (95% CI)	No.	5-yr % (95% CI)
Mutated, No. (%): 51 (13)	68	51	10 (3, 21)	35	15 (5, 30)
Non-mutated, No. (%): 354 (87)	66	354	25 (20, 30)	241	42 (34, 49)
***P*-value**	0.874		0.0012*		0.0002*
***DNMT3A/NPM1/FLT3*-ITD mutations**	CR %	OS		DFS	
		No.	5-yr % (95% CI)	No.	5-yr % (95% CI)
Triple mutated, No. (%): 29 (7)	72	29	4 (1, 16)	21	5 (1, 20)
Non-triple mutated, No. (%): 376 (93)	66	376	24 (19, 30)	255	42 (34, 49)
***P*-value**	0.682		0.009*		< 0.0001*

Abbreviations: CR: complete remission; OS: overall survival; DFS: disease-free survival; CI: confidence interval. * Indicates differences that are statistically significant (*P* < 0.05).

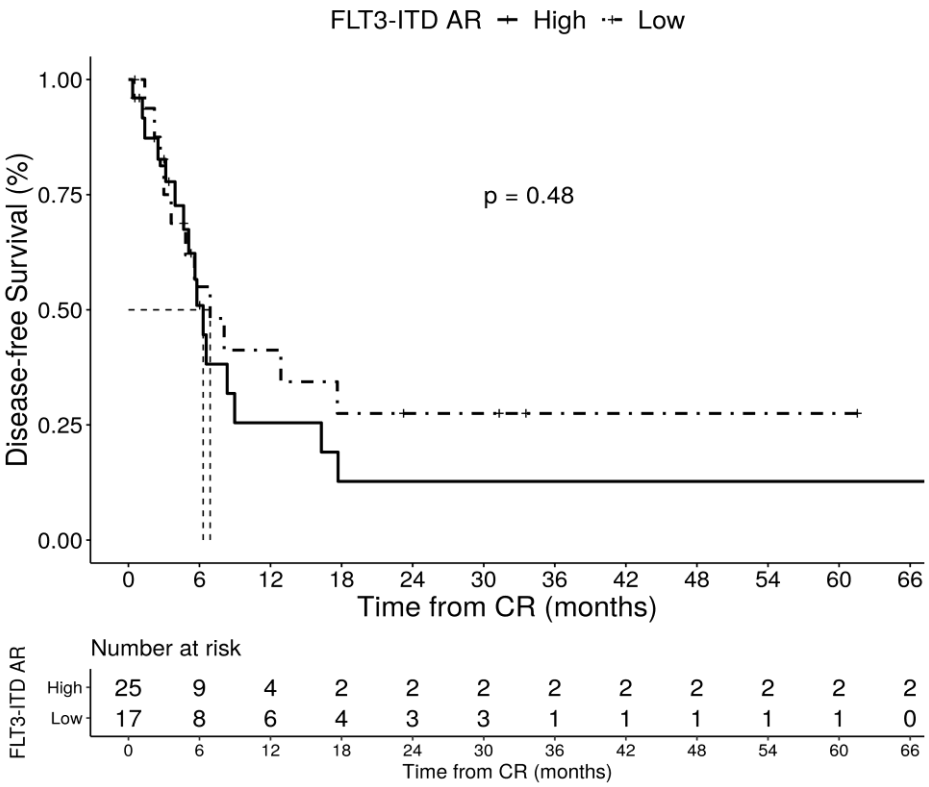
Clinical outcomes of patients according to the FLT3 allelic ratio

The median follow-up of patients with high and low *FLT3* allelic ratio was 6 months (2-9 months) and 8 months (4-20 months), respectively. Although patients with high *FLT3* allelic ratio had a lower 5-year OS (7%, 95% CI: 2-26%) compared to patients with a low *FLT3* allelic ratio (16%, 95% CI: 6–42%), this difference did not reach significance (*P*=0.26) (Figure 1A). Similar results were obtained for DFS. While patients with a high *FLT3* allelic ratio had a 5-year DFS of 13% (95% CI: 3–46%), patients with low *FLT3* allelic ratio had a 5-year DFS of 27% (95% CI: 12–63%) (*P*=0.48) (Figure 1B).

A.



B.



Supplemental figure 1. The probability of overall survival (A) and disease-free survival (B) in patients with AML according to *FLT3 allelic ratio*. Survival curves were estimated using the Kaplan–Meier method, and the log-rank test was used for comparison.

Cox proportional hazard modeling

Supplemental table 3 shows the backward elimination output. The action field shows would happen to Akaike Information Criteria (AIC) if we took the given action with the variable.

Supplemental table 3. Backward elimination for multivariate Cox proportional hazard model.

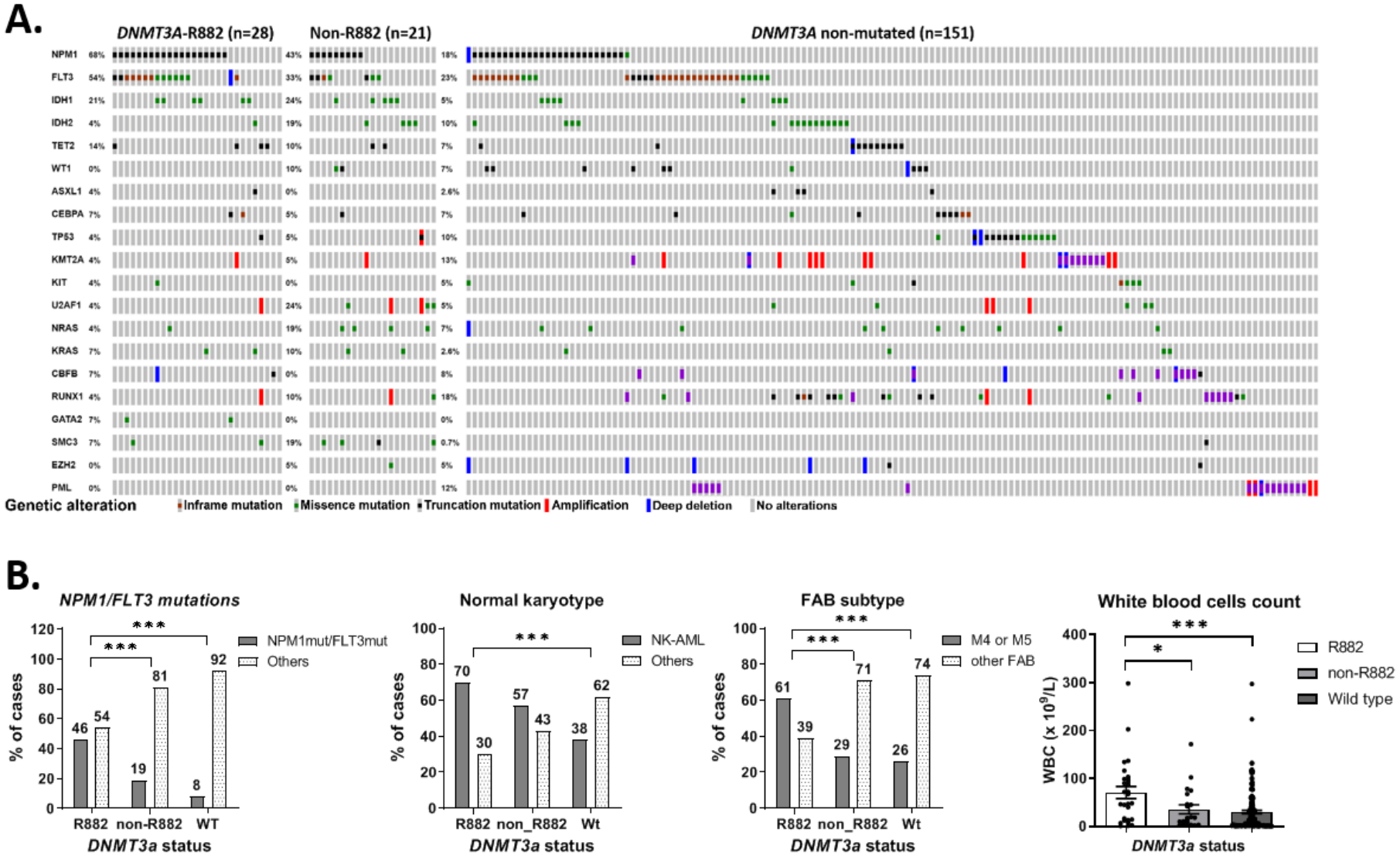
Overall survival			Disease-free survival		
Action	Variable	AIC	Action	Variable	AIC
None	Age + WBC + *DNMT3A* + MRC	2264,7	None	*FlT3*-ITD + *NPM1* + TM.AML + MRC	890,2
Add	*FLT3*-ITD	2265,3	Remove	*NPM1*	891,09
Add	Gender	2265,7	Add	WBC	891,14
Remove	*DNMT3A*	2266,2	Add	Age	891,93
Add	*NPM1*	2266,7	Add	*DNMT3A*	891,96
Add	TM.AML	2266,7	Add	Gender	892,07
Remove	WBC	2267,8	Remove	TM.AML	893,21
Remove	Age	2275,4	Remove	*FLT3*-ITD	893,25

Abbreviation: AIC, Akaike Information Criteria; TM.AML, triple-mutated acute myeloid leukemia; WBC, white blood cells; MRC, Medical Research Council.[12]

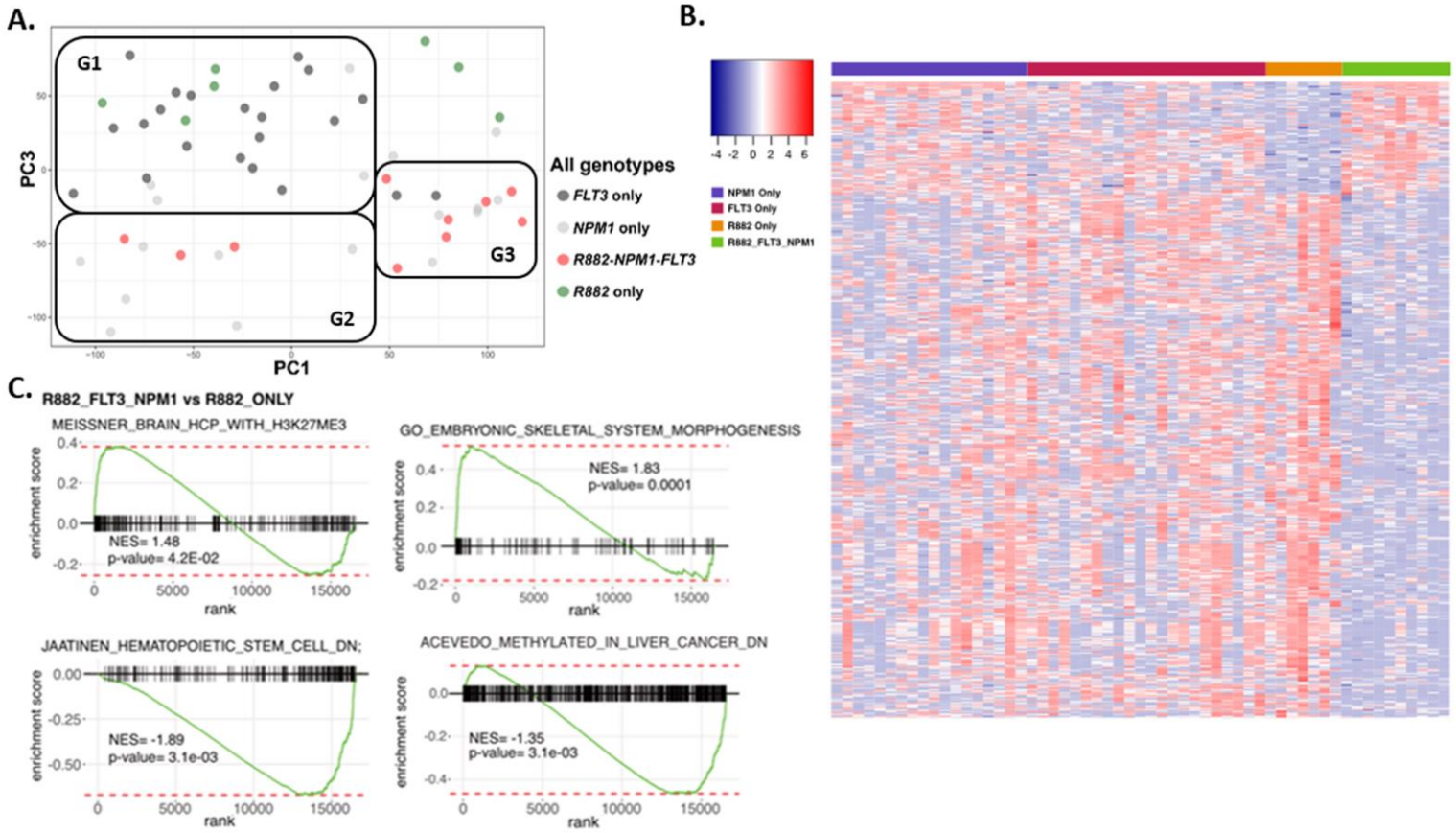
Gene expression analyses of DNMT3A-R882 versus non-R882 mutations in AML

Supplemental figure 2 illustrates mutational and phenotypic differences between patients with either *DNMT3A*-R882 or non-R882 mutations. The *NPM1* and *FLT3* mutations were significantly enriched in *DNMT3A*-R882 mutations.

Analysis of TCGA patient's transcriptome data revealed that triple-mutated patients had distinct gene expression signatures in comparison to their counterparts with isolated mutations (Figure 3A-B). In addition, gene sets containing dense histone H3K27me3 methylation in cancer, tumor-related DNA-hypomethylation and overexpression of the HOX genes cluster were enriched in triple-mutated cases. On the other hand, downregulation of hematopoietic "stemness" genes were enriched in patients harboring isolated R882 mutations (Figure 3C).


A.
DNMT3A-R882 (n=28)
Non-R882 (n=21)
DNMT3A non-mutated (n=151)
NPM1 68% 43% 18%
FLT3 54% 33% 23%
IDH1 21% 24% 5%
IDH2 4% 19% 10%
TET2 14% 10% 7%
WT1 0% 10% 7%
ASXL1 4% 0% 2.6%
CEBPA 7% 5% 7%
TP53 4% 5% 10%
KMT2A 4% 5% 13%
KIT 4% 0% 5%
U2AF1 4% 24% 5%
NRAS 4% 19% 7%
KRAS 7% 10% 2.6%
CBFB 7% 0% 8%
RUNX1 4% 10% 18%
GATA2 7% 0% 0%
SMC3 7% 19% 0.7%
EZH2 0% 5% 5%
PML 0% 0% 12%
Genetic alteration
Inframe mutation
Missence mutation
Truncation mutation
Amplification
Deep deletion
No alterations
B.
NPM1/FLT3 mutations
% of cases
NPM1mut/FLT3mut
Others
46
54
19
81
8
92
R882
non-R882
WT
DNMT3a status
Normal karyotype
NK-AML
Others
70
30
57
43
38
62
R882
non_R882
Wt
DNMT3a status
FAB subtype
M4 or M5
other FAB
61
39
29
71
26
74
R882
non_R882
Wt
DNMT3a status
White blood cells count
WBC (x 10^9/L)
R882
non-R882
Wild type
DNMT3a status

Supplemental figure 2. Clinical and biological aspects of *DNMT3A* mutations in TCGA patients. (**A**) Mutational landscape of TCGA AML patients according to *DNMT3A* status reveals that *FLT3* and *NPM1* mutations are enriched in the *DNMT3A*-R882 group. (**B**) Normal karyotype (NK), as well as monocytic FAB subtypes and higher white blood cells counts are more frequent in patients with *DNMT3A*-R882 compared to patients with non-R882 or wildtype *DNMT3A*.



Supplemental figure 3. (A) Principal Component Analysis (PCA) from RNAseq data shows that DNMT3A forms distinct clusters (G1, G2 and G3) depending on the presence of the partner mutations in *FLT3* or *NPM1*. (B) Heatmap of differentially expressed genes (supervised clustering) in triple-mutated patients when compared to single-occurrence of R882. (C) Gene set enrichment analysis (GSEA) of differential expression between triple-mutated patients and single occurrence of R882 mutations.

REFERENCES

1. Pagnano KB, Traina F, Takahashi T, et al. Conventional chemotherapy for acute myeloid leukemia: a Brazilian experience. *Sao Paulo Med. J.* 2000;118(6):173–8.
2. Kiyoi H, Naoe T, Yokota S, et al. Internal tandem duplication of FLT3 associated with leukocytosis in acute promyelocytic leukemia. Leukemia Study Group of the Ministry of Health and Welfare (Kohseisho). *Leukemia*. 1997;11(9):1447–52.
3. Noguera NI, Ammatuna E, Zangrilli D, et al. Simultaneous detection of NPM1 and FLT3-ITD mutations by capillary electrophoresis in acute myeloid leukemia. *Leukemia*. 2005;19(8):1479–82.
4. Falini B, Mecucci C, Tiacci E, et al. Cytoplasmic nucleophosmin in acute myelogenous leukemia with a normal karyotype. *N. Engl. J. Med.* 2005;352(3):254–66.
5. Stone RM, Mandrekar SJ, Sanford BL, et al. Midostaurin plus Chemotherapy for Acute Myeloid Leukemia with a FLT3 Mutation. *N. Engl. J. Med.* 2017;377(5):454–464.
6. Döhner H, Estey E, Grimwade D, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. *Blood*. 2017;129(4):424–447.
7. Benicio MTDLMT de LMTDL, Ribeiro AFTAFT, Américo AD, et al. Evaluation of the European LeukemiaNet recommendations for predicting outcomes of patients with acute myeloid leukemia treated in low- and middle-income countries (LMIC): A Brazilian experience. *Leuk. Res.* 2017;60:109–114.
8. Verhaak RGW, Wouters BJ, Erpelinck CAJ, et al. Prediction of molecular subtypes in acute myeloid leukemia based on gene expression profiling. *Haematologica*. 2009;94(1):131–4.
9. Cancer Genome Atlas Research Network, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 2013;368(22):2059–74.
10. Sergushichev AA, Loboda AA, Jha AK, et al. GAM: a web-service for integrated transcriptional and metabolic network analysis. *Nucleic Acids Res.* 2016;44(W1):W194-200.
11. Liberzon A, Birger C, Thorvaldsdóttir H, et al. The Molecular Signatures Database (MSigDB) hallmark gene set collection. *Cell Syst.* 2015;1(6):417–425.
12. Grimwade D, Hills RK, Moorman A V, et al. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. *Blood*. 2010;116(3):354–65.

APÊNDICE D

Material suplementar do artigo referente ao capítulo II

Prognostic implications of *ID1* expression in acute myeloid leukemia patients treated in a real-life setting

Aleide S Lima,[1] Matheus F Bezerra,[2] Amanda Moreira-Aguiar,[1] Isabel Weinhäuser,[3] Bianca L Santos,[1] Raul M Falcão,[4] Maria L Salustiano-Bandeira,[1] Pedro L Franca-Neto,[1] Marinus M Lima,[1] Felipe Saldanha-Araujo,[5] Juan L Coelho-Silva,[6] Diego A Pereira-Martins,[3] Marcos A Bezerra,[1] Antonio R Lucena-Araujo[1]*

Affiliations: [1]Department of Genetics, Federal University of Pernambuco, Recife, Brazil; [2]Department of Microbiology, Aggeu Magalhães Institute, Fiocruz, Recife, Pernambuco, Brazil; [3]Department of Hematology, Cancer Research Centre Groningen, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands [4]Bioinformatics Multidisciplinary Environment (BioME), Metrópole Digital Institute, Federal University of Rio Grande do Norte; [5]Laboratório de Hematologia e Células-Tronco, Universidade de Brasília, Av. L2 Norte, Brasília, DF, 70.910-900, Brazil. [6]Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA.

Supplemental methods

Patients and study design

We performed a retrospective cohort study investigating the clinical impact of ID1 expression in patients with acute myeloid leukemia (AML) at diagnosis. Between January 2011 to November 2019, 128 consecutive patients were diagnosed with AML at three Brazilian reference centers specialized on AML treatment located in northeast Brazil. AML diagnosis was determined by cytomorphology, immunophenotyping and genetic studies according to WHO criteria. In this study only patients diagnosed with *de novo* AML were included (128 patients), while those with acute promyelocytic leukemia (31 patients), therapy-related AML (18 patients), or a previous history of myelodysplastic syndrome (14 patients) were excluded. An overview of patient characteristics according to the *ID1* expression can be found in Table 1. All patients or their relatives gave their written informed consent for scientific evaluations. The study was approved by the Internal Review Board (#7147/2005) and adhered to the tenets of the Declaration of Helsinki.

Treatment protocol

For six patients out of 128, there was no available information on the treatment procedures or their respective follow up. Consequently, they were excluded from induction and post-inductions analyses. The majority of the patients (91 patients, 74%) were intensively treated with cytotoxic chemotherapy. Induction therapy consisted of daunorubicin (60 or 90 mg/m^2 daily for 3 days) and cytarabine (100 or 200 mg/m^2 daily for 7 days) or thioguanine, cytosine arabinoside, and daunorubicin,[1] followed by two or three cycles of consolidation therapy with high-dose cytarabine (1.5g/m^2 or 3g/m^2 for 3 days). Post-remission therapy based on autologous or allogeneic transplantation was not performed in only ten patients. Complete remission (CR) was assessed by bone marrow examination 28 days after each course of chemotherapy. Patients that did not achieve complete remission after one course of chemotherapy, received a second course. Of the 31 patients, which did not receive intensive treatment, 24 were 60 years old or above. These patients were treated with low-dose of ARA-C in combination with etoposide, thioguanine, and idarubicin, or exposed to palliative care.

Cytogenetic and molecular analyses

All materials used for genetic analyses were obtained at diagnosis and were processed in the reference laboratories of each participating center. Cytogenetic analysis was performed from bone marrow aspirates according to standard techniques for chromosomal banding.

The real-time quantitative polymerase chain reaction (RT-qPCR) methodology was used to determine the *ID1* transcript levels. Following total RNA extraction, RT-qPCR assays were performed in duplicate using sample-derived complementary DNA (cDNA) on MicroAmp optical 96-well plates using a QuantStudio™ 5 Dx Real-Time PCR System (Applied Biosystems, Foster City, CA) with the human *GAPDH* pre-developed Taqman assay as the endogenous control. *ID1* gene expression was determined using the TaqMan Gene Expression Assay (Hs03676575_s1; Applied Biosystems) following the manufacturer's instructions. The gene expression of *ID1* was calculated relative to a reference cDNA obtained from OCI-AML3 cell line. Importantly, the same reference cDNA was used as an

internal control in all experiments to ensure that the results of different experiments could be comparable. *ID1* relative gene expression were quantified using the Δ quantification cycle (Cq) method, and the results were expressed using $2^{-\Delta\Delta Ct}$.

Statistics analysis and clinical endpoints

Descriptive analyses were performed for patient baseline features. Fisher's exact test or Chi-square test, as appropriate, was used to compare categorical variables. Kruskal-Wallis test was used to compare continuous variables. Details of the statistical analysis and clinical endpoints were described elsewhere.[2] All *P*-values were two sided with a significance level of 0.05. All calculations were performed using Stata Statistic/Data Analysis version 12 (Stata Corporation, USA), and R 3.3.2 (The CRAN project, www.r-project.org) software.

Public datasets

Patient's clinical features, mutational and transcriptome data from TCGA[3] were obtained from http://cancergenome.nih.gov. For the RNAseq analysis, normalization of the counts by RPKM and, for microarray data, the normalized fluorescence of probe 208937_s_at, were used for expression analysis.

Gene set enrichment analysis (GSEA) were conducted using *preranked* GSEA[4] from TCGA cohort data. The gene set database (all gene sets from hallmark collection and gene sets from curated collection with homo sapiens organism) used for the analysis was obtained from MsigDB v6.2. The input data for the *preranked* GSEA was the ranked list of genes based on Pearson's correlation coefficient for all genes after filtering (19,882 genes).

Supplemental results

Supplemental table 1 summarizes the multivariate analysis for overall survival of patients enrolled in the training cohort.

Table 1. Multivariate analysis

End point	Model variables	HR	95% CI	P-value
OS, 154 patients	ID1 expression (low vs. high)	1.36	0.91-2.03	0.133
	Age (continuous variable)	1.03	1.02-1.05	<0.001
	WBC (continuous variable)	1.00	1.00-1.01	0.021
	Cytogenetic risk (fav vs. int)	2.73	1.10-6.81	0.031
	Cytogenetic risk (fav vs. adv)	4.49	1.72-11.75	0.002

Abbreviations: OS, overall survival; HR, hazard ratio; WBC, white blood cells.

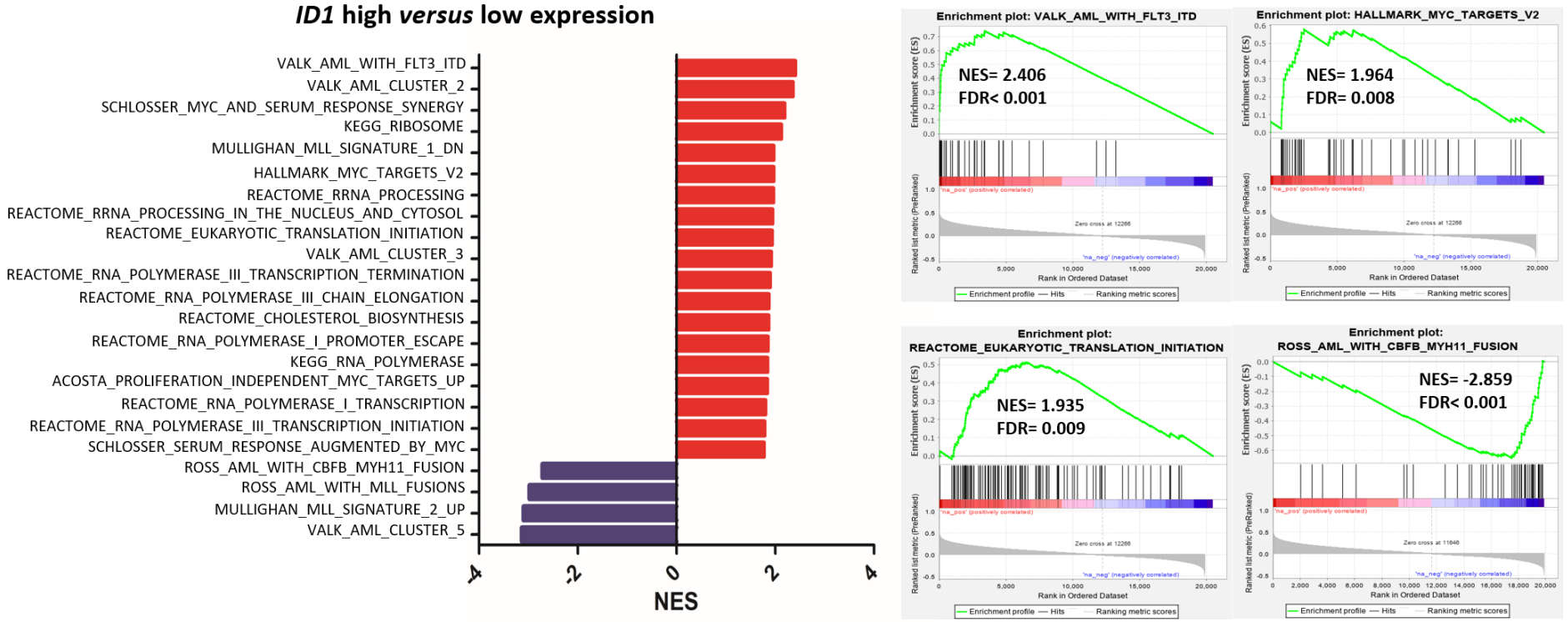
Supplemental table 2 summarizes the multivariate analysis for overall survival of patients enrolled in the real-life cohort.

Table 2. Multivariate analysis

End point	Model variables	HR	95% CI	P-value
OS, 122 patients	ID1 expression (low vs. high)	1.5	0.98-2.28	0.057
	Age (continuous variable)	1.3	1.01-1.6	0.002
	WBC (continuous variable)	1.2	1.07-1.3	0.014

Abbreviations: OS, overall survival; HR, hazard ratio; WBC, white blood cells.

Supplemental figure 1. Shows the *Gene Set Enrichment Analysis* preranked comparing the transcriptome of patients with high and low expression of the *ID1* gene. GSEA revealed that samples with high *ID1* expression were significantly enriched for gene sets associated with *FLT3*-ITD mutated AMLs, cell proliferation, ribosomal biogenesis and translation. Patients with low *ID1* expression were enriched for gene sets characteristic for AMLs with *MLL* translocations and *CBFB-MYH11* fusion gene.



Supplemental figure 1. *Gene Set Enrichment Analysis* preranked comparing the transcriptome of patients with high and low expression of the *ID1* gene. The input data was the ranked list of genes according to the Pearson's correlation coefficient between each gene and the *ID1* gene. All hallmark and curated gene sets from the MsigDB 1.0.0 database were used in the enrichment analysis.

References

1. Pagnano KB, Traina F, Takahashi T, et al. Conventional chemotherapy for acute myeloid leukemia: a Brazilian experience. *Sao Paulo Med. J.* 2000;118(6):173–8.
2. Benicio MTDLMT de LMTDL, Ribeiro AFTAFT, Américo AD, et al. Evaluation of the European LeukemiaNet recommendations for predicting outcomes of patients with acute myeloid leukemia treated in low- and middle-income countries (LMIC): A Brazilian experience. *Leuk. Res.* 2017;60:109–114.
3. Cancer Genome Atlas Research Network, Ley TJ, Miller C, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. *N. Engl. J. Med.* 2013;368(22):2059–74.
4. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci.* 2005;102(43):15545–15550.

APÊNDICE E

Material suplementar do artigo referente ao capítulo III

MYELOID NEOPLASIA

Three-gene prognostic index for predicting outcome in adult AML patients

Aleide S Lima,[1] Juan L Coelho-Silva,[2] Diego A Pereira-Martins,[3] Amanda M G de Aguiar,[1] Raul M Falcão,[4] Maria L S Bandeira,[1] Matheus F Bezerra,[5] Pedro L Franca-Neto,[1] Marinus M Lima,[1] Reijane A de Assis,[6] Marcos A Bezerra,[1] Antonio R Lucena-Araujo[1]*
Amanda, Elton

Affiliations: [1]Department of Genetics, Federal University of Pernambuco, Recife, Brazil; [2]Hematology Branch, National Heart, Lung, and Blood Institute, National Institutes of Health, Bethesda, MD, USA; [3]Department of Hematology, Cancer Research Centre Groningen, University Medical Centre Groningen, University of Groningen, Groningen, The Netherlands; [4]Bioinformatics Multidisciplinary Environment (BioME), Metrópole Digital Institute, Federal University of Rio Grande do Norte; [5]Department of Microbiology, Aggeu Magalhães Institute, Fiocruz, Recife, Pernambuco, Brazil; [6]Cancer Hospital of Pernambuco, Recife, Pernambuco, Brazil.

Supplemental methods

RT-qPCR assay

To internal validation of TPI, we evaluated the expression of the 3 selected genes through the RT-qPCR method in samples from the Brazilian cohort. The material used for RT-qPCR analysis were obtained at diagnosis from bone marrow aspirates. To perform RT-qPCR, complementary DNA (cDNA) was synthesized from 1µg of RNA, using the High-Capacity cDNA Archive Kit (Applied Biosystems, Foster City, CA), following the manufacturer's recommendation. RT-qPCR assays were performed in duplicate using specific primers (Supplemental table 3), SYBR®Green PowerUp MasterMix (Applied Biosystems, Foster City, CA) and reactions were run in a QuantStudio™ 5 Dx Real-Time PCR System (Applied Biosystems, Foster City, CA). The ACTB gene expression was used as an endogenous control. The gene expressions of *IGF2BP3*, *PDE7B*, *ST6GALNAC4 and ACTB* were calculated relative to a reference cDNA obtained from OCI-AML3 cell line. Importantly, the same reference cDNA was used as an internal control in all experiments to ensure that the results of different experiments could be comparable. The gene expression assays were only performed after determining the amplification efficiency for each primer pair, according to the equation $E=10(-1/slope)$ -1. The amplification efficiency values of each primer pair and the melting curves are shown in supplementary table 3. The relative expression of the *IGF2BP3*, *PDE7B* and *ST6GALNAC4* genes was determined by the $2^{-\Delta\Delta Cq}$ comparative method.

Supplemental table 1: Demographic, clinical and genetic characteristics of patients included in risk groups according to relapse status after transplantation.

Patients	Risk group	DFS (months)	Age	Gender	FAB	Mutational profile	Type of HSCT	Amount of HSCT
2821	Disease-free	27.4	64	Male	M1	NPM1^{wt}/non-FLT3-ITD /ASXL1^{mut}/TP53^{wt}/RUNX1^{mut}	Sibling	1
2822	Disease-free	32.3	65	Male	M0	NPM1^{wt}/non-FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Sibling	2
2877	Disease-free	41.4	60	Fem	M1	NPM1^{mut}/ FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Sibling	1
2898	Disease-free	36.1	69	Fem	M1	NPM1^{wt}/non-FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Sibling	1
2916	Disease-free	45.3	49	Fem	M4	NPM1^{wt}/non-FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Sibling	1
2963	Disease-free	53.9	56	Male	M1	NPM1^{mut}/FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Unrelated	2
2981	Disease-free	44.4	35	Fem	M5	NPM1^{mut}/ FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Unrelated	1
2808	Relapsed	10	23	Male	M2	NPM1^{wt}/non-FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Sibling	2
2918	Relapsed	6	47	Fem	M1	NPM1^{wt}/FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Unrelated	1
2921	Relapsed	6.3	55	Fem	M4	NPM1^{wt}/FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Sibling	1
2965	Relapsed	9.9	60	Male	M4	NPM1^{mut}/FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Sibling	1
2966	Relapsed	8.6	57	Fem	M2	NPM1^{wt}/non-FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Unrelated	1
2978	Relapsed	35.6	61	Fem	M2	NPM1^{wt}/non-FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{mut}	Unrelated	1
2995	Relapsed	69.7	63	Male	M2	NPM1^{wt}/non-FLT3-ITD/ASXL1^{wt}/TP53^{wt}/RUNX1^{wt}	Sibling	1

Fem: female; FAB: French-American-British Classification; wt: wild type; mut: mutated

Supplemental table 2: Results of univariate and multivariate Cox regression analyzes for the 3 genes with the greatest impact on OS of patients in the TCGA training cohort.

Gene	Univariate analysis			Multivariate analysis			AUC	%mode
	HR	95%CI	*p*	HR	95%CI	*p*		
IGF2BP3	2,448	1,495-4,010	<0,001	2,857	1,701-4,800	<0,001	0.653	0,826
PDE7B	2,890	1,720-4,855	<0,001	2,537	1,458-4,415	0,001	0.657	0,826
ST6GALNAC4	2,331	1,479-3,673	<0,001	2,349	1,370-4,029	0,002	0.630	0,826

HR: hazard ratio; 95%CI: 95% confidence interval; AUC: area under the curve.

Supplemental table 3: Sequence and amplification efficiency of the primers used in the qPCR assay and expected sizes of the amplicons for each of the genes.

Gene	**Primers sequence**	**Expected amplicon size**	**Amplification efficiency**
IGF2BP3			
Forward	TCACGATATCTCCATTGCAGG	82	102%
Reverse	GCACATGTCTCAACATTGCCT		
PDE7B			
Forward	CAGGCAAGGTGAACTTGAAC	113	99%
Reverse	GCTCCACGATGTAGCTCATG		
ST6GALNAC4			
Forward	CTACTGCAGGGAGAAGAGCC	89	98%
Reverse	CGTGTGCCAGGTACATCTGA		
ACTB			
Forward	AGGCCAACCGCGAGAAG	79	98%
Reverse	ACAGCCTGGATAGCAACGTACA		

Supplemental table 4: Enriched gene sets (FDR<0.05) in the relapsed group

NAME	NES	FDR
REACTOME_TRANSLATION	3.260088	0
REACTOME_MITOCHONDRIAL_TRANSLATION	3.058203	0
WONG_EMBRYONIC_STEM_CELL_CORE	3.054996	0
GRAHAM_CML_DIVIDING_VS_NORMAL_QUIESCENT_UP	2.948079	0
HALLMARK_MYC_TARGETS_V1	2.916697	0
REACTOME_RRNA_PROCESSING	2.915847	0
RHEIN_ALL_GLUCOCORTICOID_THERAPY_DN	2.915507	0
SOTIRIOU_BREAST_CANCER_GRADE_1_VS_3_UP	2.910718	0
CAIRO_HEPATOBLASTOMA_CLASSES_UP	2.905527	0
REACTOME_REGULATION_OF_EXPRESSION_OF_SLITS_AND_ROBOS	2.864644	0
REACTOME_SRP_DEPENDENT_COTRANSLATIONAL_PROTEIN_TARGETING_TO_MEMBRANE	2.859634	0
REACTOME_INFLUENZA_INFECTION	2.849825	0
WINNEPENNINCKX_MELANOMA_METASTASIS_UP	2.846155	0
ROSTY_CERVICAL_CANCER_PROLIFERATION_CLUSTER	2.839429	0
HALLMARK_E2F_TARGETS	2.796649	0
GRAHAM_NORMAL_QUIESCENT_VS_NORMAL_DIVIDING_DN	2.790318	0
REN_BOUND_BY_E2F	2.783128	0
REACTOME_EUKARYOTIC_TRANSLATION_ELONGATION	2.778406	0
KEGG_RIBOSOME	2.775524	0
REACTOME_EUKARYOTIC_TRANSLATION_INITIATION	2.7648	0
DANG_MYC_TARGETS_UP	2.75858	0
WP_CYTOPLASMIC_RIBOSOMAL_PROTEINS	2.750303	0
MANALO_HYPOXIA_DN	2.721336	0
WONG_MITOCHONDRIA_GENE_MODULE	2.711624	0
WHITEFORD_PEDIATRIC_CANCER_MARKERS	2.699196	0
CHIANG_LIVER_CANCER_SUBCLASS_UNANNOTATED_DN	2.656971	0

REACTOME_DNA_REPLICATION	2.644807	0
REACTOME_NONSENSE_MEDIATED_DECAY_NMD_	2.637991	0
REACTOME_RESPONSE_OF_EIF2AK4_GCN2_TO_AMINO_ACID_DEFICIENCY	2.604656	0
CROONQUIST_IL6_DEPRIVATION_DN	2.604007	0
REACTOME_SELENOAMINO_ACID_METABOLISM	2.60103	0
MOOTHA_HUMAN_MITODB_6_2002	2.597703	0
MISSIAGLIA_REGULATED_BY_METHYLATION_DN	2.577464	0
KOBAYASHI_EGFR_SIGNALING_24HR_DN	2.576315	0
EPPERT_PROGENITOR	2.565586	0
MOOTHA_MITOCHONDRIA	2.563844	0
SHEDDEN_LUNG_CANCER_POOR_SURVIVAL_A6	2.554479	0
RHODES_UNDIFFERENTIATED_CANCER	2.549017	0
CROONQUIST_NRAS_SIGNALING_DN	2.546317	0
MUELLER_PLURINET	2.542347	0
ZHOU_CELL_CYCLE_GENES_IN_IR_RESPONSE_24HR	2.537978	0
ALCALAY_AML_BY_NPM1_LOCALIZATION_DN	2.536948	0
BENPORATH_PROLIFERATION	2.531874	0
DUTERTRE_ESTRADIOL_RESPONSE_24HR_UP	2.52857	0
GARCIA_TARGETS_OF_FLI1_AND_DAX1_DN	2.521196	0
BLANCO_MELO_BRONCHIAL_EPITHELIAL_CELLS_INFLUENZA_A_DEL_NS1_INFECTION_DN	2.519411	0
KAUFFMANN_MELANOMA_RELAPSE_UP	2.511782	0
REACTOME_CHROMOSOME_MAINTENANCE	2.502484	0
HALLMARK_OXIDATIVE_PHOSPHORYLATION	2.494492	0
ZHAN_MULTIPLE_MYELOMA_PR_UP	2.484989	0
REACTOME_ORC1_REMOVAL_FROM_CHROMATIN	2.484267	0
REACTOME_DNA_REPLICATION_PRE_INITIATION	2.475051	0
FEVR_CTNNB1_TARGETS_DN	2.470634	0
REACTOME_SIGNALING_BY_ROBO_RECEPTORS	2.468112	0

SONG_TARGETS_OF_IE86_CMV_PROTEIN	2.466296	0
CHNG_MULTIPLE_MYELOMA_HYPERPLOID_UP	2.464167	0
REACTOME_SWITCHING_OF_ORIGINS_TO_A_POST_REPLICATIVE_STATE	2.459824	0
REACTOME_S_PHASE	2.450628	0
REACTOME_G2_M_CHECKPOINTS	2.447114	0
GRAHAM_CML_QUIESCENT_VS_NORMAL_QUIESCENT_UP	2.440643	0
REACTOME_METABOLISM_OF_AMINO_ACIDS_AND_DERIVATIVES	2.438571	0
REACTOME_MITOTIC_G1_PHASE_AND_G1_S_TRANSITION	2.433808	0
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT	2.430348	0
PUJANA_XPRSS_INT_NETWORK	2.428121	0
PUJANA_BRCA2_PCC_NETWORK	2.426084	0
SCHLOSSER_MYC_TARGETS_REPRESSED_BY_SERUM	2.424706	0
PUJANA_BRCA_CENTERED_NETWORK	2.421937	0
SCHUHMACHER_MYC_TARGETS_UP	2.417617	0
REACTOME_ASSEMBLY_OF_THE_PRE_REPLICATIVE_COMPLEX	2.415454	0
ROSS_AML_OF_FAB_M7_TYPE	2.409592	0
REACTOME_APC_C_MEDIATED_DEGRADATION_OF_CELL_CYCLE_PROTEINS	2.408526	0
TARTE_PLASMA_CELL_VS_PLASMABLAST_DN	2.402261	0
KAUFFMANN_DNA_REPAIR_GENES	2.397446	0
REACTOME_THE_ROLE_OF_GTSE1_IN_G2_M_PROGRESSION_AFTER_G2_CHECKPOINT	2.395647	0
MOLENAAR_TARGETS_OF_CCND1_AND_CDK4_DN	2.391533	0
WHITFIELD_CELL_CYCLE_LITERATURE	2.382051	0
FUJII_YBX1_TARGETS_DN	2.378279	0
REACTOME_G1_S_DNA_DAMAGE_CHECKPOINTS	2.376668	0
LINDGREN_BLADDER_CANCER_CLUSTER_3_UP	2.368319	9.10E-06
WANG_RESPONSE_TO_GSK3_INHIBITOR_SB216763_DN	2.365719	8.98E-06
BLUM_RESPONSE_TO_SALIRASIB_DN	2.362384	8.87E-06
LI_WILMS_TUMOR_VS_FETAL_KIDNEY_1_DN	2.359352	8.77E-06

GARY_CD5_TARGETS_DN	2.356132	1.75E-05
FLORIO_NEOCORTEX_BASAL_RADIAL_GLIA_DN	2.352498	1.72E-05
HORIUCHI_WTAP_TARGETS_DN	2.344924	1.70E-05
KANG_DOXORUBICIN_RESISTANCE_UP	2.343677	1.68E-05
ZHOU_CELL_CYCLE_GENES_IN_IR_RESPONSE_6HR	2.342102	1.66E-05
REACTOME_CELL_CYCLE_CHECKPOINTS	2.339241	1.65E-05
REACTOME_THE_CITRIC_ACID_TCA_CYCLE_AND_RESPIRATORY_ELECTRON_TRANSPORT	2.339203	1.63E-05
REACTOME_ACTIVATION_OF_THE_MRNA_UPON_BINDING_OF_THE_CAP_BINDING_COMPLEX_AND_EIFS_AND_SUBSEQUENT_BINDING_TO_43S	2.331826	1.61E-05
MOREAUX_B_LYMPHOCYTE_MATURATION_BY_TACI_DN	2.327483	1.59E-05
TIEN_INTESTINE_PROBIOTICS_24HR_UP	2.326875	1.57E-05
REACTOME_RESPIRATORY_ELECTRON_TRANSPORT_ATP_SYNTHESIS_BY_CHEMIOSMOTIC_COUPLING_AND_HEAT_PRODUCTION_BY_UNCOUPLING_PROTEINS_	2.322006	1.56E-05
REACTOME_BASE_EXCISION_REPAIR	2.317344	1.54E-05
REACTOME_RRNA_MODIFICATION_IN_THE_NUCLEUS_AND_CYTOSOL	2.317225	1.52E-05
LEE_EARLY_T_LYMPHOCYTE_UP	2.315407	1.51E-05
REACTOME_HOMOLOGY_DIRECTED_REPAIR	2.314595	1.49E-05
SCHLOSSER_MYC_TARGETS_AND_SERUM_RESPONSE_DN	2.314379	1.48E-05
BHATTACHARYA_EMBRYONIC_STEM_CELL	2.313633	2.20E-05
REACTOME_MRNA_SPLICING_MINOR_PATHWAY	2.313104	2.18E-05
REACTOME_TRNA_PROCESSING	2.307982	2.15E-05
WP_DNA_MISMATCH_REPAIR	2.307738	2.13E-05
ZHANG_RESPONSE_TO_CANTHARIDIN_DN	2.304818	2.11E-05
REACTOME_RESOLUTION_OF_AP_SITES_VIA_THE_MULTIPLE_NUCLEOTIDE_PATCH_REPLACEMENT_PATHWAY	2.30369	2.09E-05
REACTOME_SCF_SKP2_MEDIATED_DEGRADATION_OF_P27_P21	2.301359	2.07E-05
WP_RETINOBLASTOMA_GENE_IN_CANCER	2.299394	2.05E-05
REACTOME_AUF1_HNRNP_D0_BINDS_AND_DESTABILIZES_MRNA	2.291397	2.03E-05
PID_MYC_ACTIV_PATHWAY	2.290897	2.01E-05
WP_MITOCHONDRIAL_COMPLEX_I_ASSEMBLY_MODEL_OXPHOS_SYSTEM	2.290405	2.00E-05
REACTOME_EXTENSION_OF_TELOMERES	2.288753	1.98E-05

WP_ELECTRON_TRANSPORT_CHAIN_OXPHOS_SYSTEM_IN_MITOCHONDRIA	2.287757	2.62E-05
REACTOME_DNA_STRAND_ELONGATION	2.285835	2.60E-05
REACTOME_STABILIZATION_OF_P53	2.284682	2.57E-05
REACTOME_PCNA_DEPENDENT_LONG_PATCH_BASE_EXCISION_REPAIR	2.279471	2.55E-05
REACTOME_DEFECTIVE_CFTR_CAUSES_CYSTIC_FIBROSIS	2.275405	3.15E-05
REACTOME_APC_C_CDH1_MEDIATED_DEGRADATION_OF_CDC20_AND_OTHER_APC_C_CDH1_TARGETED_PROTEINS_IN_LATE_MITOSIS_EARLY_G1	2.274141	3.13E-05
HALLMARK_G2M_CHECKPOINT	2.272826	3.10E-05
REACTOME_TELOMERE_MAINTENANCE	2.272498	3.07E-05
REACTOME_PROCESSING_OF_CAPPED_INTRON_CONTAINING_PRE_MRNA	2.270511	3.05E-05
MOOTHA_VOXPHOS	2.270476	3.02E-05
RUIZ_TNC_TARGETS_DN	2.263414	3.63E-05
REACTOME_RESOLUTION_OF_ABASIC_SITES_AP_SITES_	2.262976	3.60E-05
BASAKI_YBX1_TARGETS_UP	2.259546	3.57E-05
REACTOME_TERMINATION_OF_TRANSLESION_DNA_SYNTHESIS	2.256636	3.54E-05
RHODES_CANCER_META_SIGNATURE	2.256138	3.51E-05
JAATINEN_HEMATOPOIETIC_STEM_CELL_UP	2.251836	3.49E-05
REACTOME_ACTIVATION_OF_ATR_IN_RESPONSE_TO_REPLICATION_STRESS	2.25125	3.46E-05
REACTOME_DEGRADATION_OF_DVL	2.243511	3.99E-05
REACTOME_DNA_REPAIR	2.241005	3.96E-05
MALONEY_RESPONSE_TO_17AAG_DN	2.238992	4.48E-05
REACTOME_GLOBAL_GENOME_NUCLEOTIDE_EXCISION_REPAIR_GG_NER_	2.237939	4.45E-05
BIDUS_METASTASIS_UP	2.23258	4.96E-05
REACTOME_CYCLIN_A_CDK2_ASSOCIATED_EVENTS_AT_S_PHASE_ENTRY	2.229463	4.93E-05
REACTOME_CELL_CYCLE_MITOTIC	2.22889	4.89E-05
REACTOME_HOST_INTERACTIONS_OF_HIV_FACTORS	2.228346	4.85E-05
HALLMARK_MYC_TARGETS_V2	2.222627	5.36E-05
SARRIO_EPITHELIAL_MESENCHYMAL_TRANSITION_UP	2.222424	5.32E-05
REACTOME_SNRNP_ASSEMBLY	2.221265	5.28E-05

MEBARKI_HCC_PROGENITOR_FZD8CRD_UP	2.221258	5.24E-05
VECCHI_GASTRIC_CANCER_EARLY_UP	2.217858	5.20E-05
REACTOME_GAP_FILLING_DNA_REPAIR_SYNTHESIS_AND_LIGATION_IN_GG_NER	2.216704	5.17E-05
KEGG_PROTEASOME	2.212416	5.13E-05
REACTOME_DNA_DOUBLE_STRAND_BREAK_REPAIR	2.208567	5.60E-05
REACTOME_PROTEIN_LOCALIZATION	2.205572	5.57E-05
KEGG_MISMATCH_REPAIR	2.202509	5.53E-05
WP_BASE_EXCISION_REPAIR	2.196756	5.49E-05
HALLMARK_DNA_REPAIR	2.196263	5.45E-05
REACTOME_DEPOSITION_OF_NEW_CENPA_CONTAINING_NUCLEOSOMES_AT_THE_CENTROMERE	2.195967	5.42E-05
ODONNELL_TARGETS_OF_MYC_AND_TFRC_DN	2.195462	5.38E-05
REACTOME_COMPLEX_I_BIOGENESIS	2.193187	5.34E-05
WP_G1_TO_S_CELL_CYCLE_CONTROL	2.190055	6.27E-05
REACTOME_COOPERATION_OF_PREFOLDIN_AND_TRIC_CCT_IN_ACTIN_AND_TUBULIN_FOLDING	2.18934	6.23E-05
ZHAN_MULTIPLE_MYELOMA_SUBGROUPS	2.188784	6.18E-05
BOYAULT_LIVER_CANCER_SUBCLASS_G3_UP	2.186902	7.09E-05
CAFFAREL_RESPONSE_TO_THC_DN	2.185645	7.04E-05
REACTOME_TRNA_PROCESSING_IN_THE_NUCLEUS	2.185486	7.00E-05
REACTOME_MITOCHONDRIAL_PROTEIN_IMPORT	2.18086	8.34E-05
POMEROY_MEDULLOBLASTOMA_PROGNOSIS_DN	2.178578	8.29E-05
KEGG_PARKINSONS_DISEASE	2.17341	8.69E-05
REACTOME_HDR_THROUGH_HOMOLOGOUS_RECOMBINATION_HRR_	2.170991	8.64E-05
CHEMNITZ_RESPONSE_TO_PROSTAGLANDIN_E2_UP	2.170795	8.59E-05
WP_PURINE_METABOLISM_AND_RELATED_DISORDERS	2.169488	8.53E-05
REACTOME_SEPARATION_OF_SISTER_CHROMATIDS	2.166123	8.94E-05
REACTOME_MRNA_SPLICING	2.161617	8.88E-05
FOURNIER_ACINAR_DEVELOPMENT_LATE_2	2.159027	9.27E-05
REACTOME_RECOGNITION_OF_DNA_DAMAGE_BY_PCNA_CONTAINING_REPLICATION_COMPLEX	2.155545	1.01E-04

KEGG_DNA_REPLICATION	2.155414	1.00E-04
PUJANA_BREAST_CANCER_WITH_BRCA1_MUTATED_UP	2.153326	9.97E-05
BORCZUK_MALIGNANT_MESOTHELIOMA_UP	2.148377	1.03E-04
REACTOME_TRANSLESION_SYNTHESIS_BY_Y_FAMILY_DNA_POLYMERASES_BYPASSES_LESIONS_ON_DNA_TEMPLATE	2.147373	1.11E-04
FARMER_BREAST_CANCER_CLUSTER_2	2.147133	1.10E-04
REACTOME_NUCLEOTIDE_EXCISION_REPAIR	2.146957	1.14E-04
MOREAUX_MULTIPLE_MYELOMA_BY_TACI_DN	2.146937	1.13E-04
GRADE_COLON_AND_RECTAL_CANCER_UP	2.14457	1.17E-04
REACTOME_REGULATION_OF_RAS_BY_GAPS	2.141448	1.16E-04
REACTOME_CROSS_PRESENTATION_OF_SOLUBLE_EXOGENOUS_ANTIGENS_ENDOSOMES_	2.135016	1.24E-04
RODRIGUES_THYROID_CARCINOMA_POORLY_DIFFERENTIATED_UP	2.131676	1.27E-04
PID_ATR_PATHWAY	2.128369	1.35E-04
WP_OXIDATIVE_PHOSPHORYLATION	2.12779	1.34E-04
REACTOME_DNA_DAMAGE_BYPASS	2.123692	1.41E-04
REACTOME_GLYOXYLATE_METABOLISM_AND_GLYCINE_DEGRADATION	2.114812	1.56E-04
WP_DNA_REPLICATION	2.113802	1.56E-04
YAMASHITA_LIVER_CANCER_WITH_EPCAM_UP	2.112361	1.59E-04
REACTOME_TRANSLESION_SYNTHESIS_BY_POLK	2.111775	1.58E-04
BASSO_B_LYMPHOCYTE_NETWORK	2.110456	1.57E-04
REACTOME_METABOLISM_OF_POLYAMINES	2.108401	1.60E-04
REACTOME_NEGATIVE_REGULATION_OF_NOTCH4_SIGNALING	2.106241	1.63E-04
REACTOME_REGULATION_OF_RUNX3_EXPRESSION_AND_ACTIVITY	2.104636	1.66E-04
REACTOME_MITOTIC_METAPHASE_AND_ANAPHASE	2.101546	1.73E-04
LEE_LIVER_CANCER_SURVIVAL_DN	2.099373	1.80E-04
ODONNELL_TFRC_TARGETS_DN	2.098558	1.79E-04
KEGG_SPLICEOSOME	2.098481	1.78E-04
REACTOME_ABC_TRANSPORTER_DISORDERS	2.095477	1.88E-04
REACTOME_TRANSCRIPTION_COUPLED_NUCLEOTIDE_EXCISION_REPAIR_TC_NER_	2.09435	1.95E-04

LI_AMPLIFIED_IN_LUNG_CANCER	2.089011	2.24E-04
HU_ANGIOGENESIS_DN	2.087674	2.22E-04
KEGG_BASE_EXCISION_REPAIR	2.083956	2.32E-04
DANG_REGULATED_BY_MYC_UP	2.08234	2.38E-04
REACTOME_FANCONI_ANEMIA_PATHWAY	2.08173	2.37E-04
DELPUECH_FOXO3_TARGETS_DN	2.075893	2.61E-04
REACTOME_LAGGING_STRAND_SYNTHESIS	2.075156	2.60E-04
REACTOME_REGULATION_OF_RUNX2_EXPRESSION_AND_ACTIVITY	2.071986	2.66E-04
PID_AURORA_B_PATHWAY	2.071224	2.72E-04
OUELLET_OVARIAN_CANCER_INVASIVE_VS_LMP_UP	2.070775	2.74E-04
ANDERSEN_LIVER_CANCER_KRT19_UP	2.069763	2.76E-04
WP_METABOLIC_REPROGRAMMING_IN_COLON_CANCER	2.068463	2.86E-04
LY_AGING_OLD_DN	2.066657	2.95E-04
KIM_MYC_AMPLIFICATION_TARGETS_UP	2.065765	3.00E-04
BIOCARTA_G1_PATHWAY	2.065611	2.99E-04
WP_PROTEASOME_DEGRADATION	2.064639	3.08E-04
YIH_RESPONSE_TO_ARSENITE_C3	2.064242	3.06E-04
REACTOME_HOMOLOGOUS_DNA_PAIRING_AND_STRAND_EXCHANGE	2.063611	3.05E-04
REACTOME_DEGRADATION_OF_GLI1_BY_THE_PROTEASOME	2.061221	3.03E-04
REACTOME_DUAL_INCISION_IN_GG_NER	2.05417	3.50E-04
REACTOME_TRANSLESION_SYNTHESIS_BY_POLH	2.051963	3.86E-04
FERREIRA_EWINGS_SARCOMA_UNSTABLE_VS_STABLE_UP	2.051308	3.84E-04
FISCHER_G1_S_CELL_CYCLE	2.045143	4.02E-04
WP_PYRIMIDINE_METABOLISM	2.044087	4.00E-04
HSIAO_HOUSEKEEPING_GENES	2.041646	4.05E-04
REACTOME_SUMOYLATION_OF_DNA_REPLICATION_PROTEINS	2.03557	4.66E-04
BENPORATH_ES_1	2.034804	4.70E-04
REACTOME_INTERACTIONS_OF_VPR_WITH_HOST_CELLULAR_PROTEINS	2.034139	4.75E-04

PUJANA_BREAST_CANCER_LIT_INT_NETWORK	2.03285	4.79E-04
REACTOME_REGULATION_OF_MRNA_STABILITY_BY_PROTEINS_THAT_BIND_AU_RICH_ELEMENTS	2.031158	4.90E-04
LI_WILMS_TUMOR_ANAPLASTIC_UP	2.030532	4.91E-04
KEGG_PROTEIN_EXPORT	2.030388	4.92E-04
REACTOME_TELOMERE_C_STRAND_LAGGING_STRAND_SYNTHESIS	2.029562	5.03E-04
PID_FANCONI_PATHWAY	2.029472	5.01E-04
REACTOME_INTERACTIONS_OF_REV_WITH_HOST_CELLULAR_PROTEINS	2.028434	5.11E-04
LUI_THYROID_CANCER_CLUSTER_3	2.025883	5.21E-04
WAKASUGI_HAVE_ZNF143_BINDING_SITES	2.023147	5.57E-04
KAUFFMANN_DNA_REPLICATION_GENES	2.019526	5.92E-04
PYEON_HPV_POSITIVE_TUMORS_UP	2.019507	5.90E-04
REACTOME_DECTIN_1_MEDIATED_NONCANONICAL_NF_KB_SIGNALING	2.015638	6.21E-04
REACTOME_TRNA_AMINOACYLATION	2.015463	6.19E-04
REACTOME_G1_S_SPECIFIC_TRANSCRIPTION	2.015306	6.22E-04
TIEN_INTESTINE_PROBIOTICS_6HR_UP	2.014947	6.23E-04
SASAKI_ADULT_T_CELL_LEUKEMIA	2.013365	6.23E-04
TANG_SENESCENCE_TP53_TARGETS_DN	2.012827	6.27E-04
REACTOME_NS1_MEDIATED_EFFECTS_ON_HOST_PATHWAYS	2.009235	6.45E-04
REACTOME_MEIOTIC_RECOMBINATION	2.008805	6.42E-04
CHICAS_RB1_TARGETS_GROWING	2.006941	6.52E-04
REACTOME_DUAL_INCISION_IN_TC_NER	2.003587	6.58E-04
WP_NUCLEOTIDE_EXCISION_REPAIR	2.002459	6.67E-04
REACTOME_CELLULAR_RESPONSES_TO_EXTERNAL_STIMULI	2.002357	6.68E-04
PID_BARD1_PATHWAY	2.001575	6.89E-04
SCIAN_CELL_CYCLE_TARGETS_OF_TP53_AND_TP73_DN	1.999502	7.27E-04
REACTOME_MITOTIC_G2_G2_M_PHASES	1.999048	7.30E-04
VERNELL_RETINOBLASTOMA_PATHWAY_UP	1.996698	7.53E-04
COLLER_MYC_TARGETS_UP	1.996019	7.59E-04

FRASOR_RESPONSE_TO_SERM_OR_FULVESTRANT_DN	1.99356	7.99E-04
REACTOME_DNA_DAMAGE_RECOGNITION_IN_GG_NER	1.991727	8.31E-04
REACTOME_VIRAL_MESSENGER_RNA_SYNTHESIS	1.988155	8.76E-04
REACTOME_NUCLEAR_IMPORT_OF_REV_PROTEIN	1.988043	8.73E-04
REACTOME_PROCESSING_OF_DNA_DOUBLE_STRAND_BREAK_ENDS	1.987527	8.72E-04
KEGG_HUNTINGTONS_DISEASE	1.986736	8.80E-04
MACAEVA_PBMC_RESPONSE_TO_IR	1.985919	8.85E-04
PECE_MAMMARY_STEM_CELL_UP	1.984634	8.93E-04
REACTOME_M_PHASE	1.981187	9.29E-04
REACTOME_RUNX1_REGULATES_TRANSCRIPTION_OF_GENES_INVOLVED_IN_DIFFERENTIATION_OF_HSCS	1.981157	9.25E-04
HONMA_DOCETAXEL_RESISTANCE	1.980709	9.30E-04
WU_APOPTOSIS_BY_CDKN1A_VIA_TP53	1.980461	9.30E-04
WP_PRADERWILLI_AND_ANGELMAN_SYNDROME	1.978757	9.48E-04
REACTOME_MISMATCH_REPAIR	1.978355	9.53E-04
WEST_ADRENOCORTICAL_TUMOR_UP	1.977966	9.49E-04
TOYOTA_TARGETS_OF_MIR34B_AND_MIR34C	1.975869	9.62E-04
REACTOME_ACTIVATION_OF_THE_PRE_REPLICATIVE_COMPLEX	1.975221	9.69E-04
EGUCHI_CELL_CYCLE_RB1_TARGETS	1.973588	0.001006
REACTOME_DEGRADATION_OF_AXIN	1.972209	0.001016
CAFFAREL_RESPONSE_TO_THC_24HR_5_DN	1.969365	0.001058
WP_PARKINUBIQUITIN_PROTEASOMAL_SYSTEM_PATHWAY	1.968589	0.00107
LUI_TARGETS_OF_PAX8_PPARG_FUSION	1.968429	0.001074
REACTOME_REGULATION_OF_TP53_ACTIVITY_THROUGH_PHOSPHORYLATION	1.967435	0.001086
STEIN_ESRRA_TARGETS_RESPONSIVE_TO_ESTROGEN_DN	1.961886	0.001159
WP_DNA_IRDAMAGE_AND_CELLULAR_RESPONSE_VIA_ATR	1.961465	0.001165
REACTOME_METABOLISM_OF_NUCLEOTIDES	1.96132	0.001164
REACTOME_HIV_INFECTION	1.952748	0.001288
KEGG_RNA_DEGRADATION	1.950817	0.001315

REACTOME_FORMATION_OF_TUBULIN_FOLDING_INTERMEDIATES_BY_CCT_TRIC	1.950739	0.00131
BOYAULT_LIVER_CANCER_SUBCLASS_G23_UP	1.948627	0.00134
FAELT_B_CLL_WITH_VH_REARRANGEMENTS_DN	1.945221	0.001392
LY_AGING_MIDDLE_DN	1.941101	0.001446
REACTOME_TRANSCRIPTIONAL_REGULATION_BY_SMALL_RNAS	1.940872	0.001446
GAZDA_DIAMOND_BLACKFAN_ANEMIA_PROGENITOR_DN	1.94059	0.001446
GOLUB_ALL_VS_AML_UP	1.94013	0.001454
FURUKAWA_DUSP6_TARGETS_PCI35_DN	1.937827	0.001497
KEGG_CITRATE_CYCLE_TCA_CYCLE	1.935803	0.001527
SAKAI_TUMOR_INFILTRATING_MONOCYTES_DN	1.933522	0.001565
SANA_RESPONSE_TO_IFNG_DN	1.93089	0.001605
KEGG_VALINE_LEUCINE_AND_ISOLEUCINE_DEGRADATION	1.929926	0.001622
BENPORATH_ES_2	1.929844	0.001621
MITSIADES_RESPONSE_TO_APLIDIN_DN	1.927219	0.001673
REACTOME_UCH_PROTEINASES	1.924767	0.001722
REACTOME_EXPORT_OF_VIRAL_RIBONUCLEOPROTEINS_FROM_NUCLEUS	1.921591	0.00179
SPIELMAN_LYMPHOBLAST_EUROPEAN_VS_ASIAN_UP	1.920411	0.001821
ABRAMSON_INTERACT_WITH_AIRE	1.920072	0.001827
WP_CELL_CYCLE	1.918704	0.00185
WP_TUMOR_SUPPRESSOR_ACTIVITY_OF_SMARCB1	1.914718	0.001964
REACTOME_REGULATION_OF_PTEN_STABILITY_AND_ACTIVITY	1.913116	0.002008
KEGG_CELL_CYCLE	1.910445	0.002094
KEGG_PURINE_METABOLISM	1.909865	0.002114
REACTOME_TELOMERE_EXTENSION_BY_TELOMERASE	1.90294	0.002278
REACTOME_POLYMERASE_SWITCHING_ON_THE_C_STRAND_OF_THE_TELOMERE	1.902787	0.002275
SCIBETTA_KDM5B_TARGETS_DN	1.901021	0.00231
REACTOME_ASYMMETRIC_LOCALIZATION_OF_PCP_PROTEINS	1.900826	0.002305
KEGG_NUCLEOTIDE_EXCISION_REPAIR	1.900041	0.002317

KEGG_PYRIMIDINE_METABOLISM	1.89997	0.002309
REACTOME_G2_M_DNA_DAMAGE_CHECKPOINT	1.899165	0.00233
LE_NEURONAL_DIFFERENTIATION_DN	1.896629	0.002391
REACTOME_G0_AND_EARLY_G1	1.894549	0.002449
NAKAMURA_TUMOR_ZONE_PERIPHERAL_VS_CENTRAL_UP	1.894177	0.002448
YU_BAP1_TARGETS	1.892944	0.002487
BIOCARTA_CELLCYCLE_PATHWAY	1.888014	0.002627
REACTOME_HEDGEHOG_LIGAND_BIOGENESIS	1.883764	0.002809
REACTOME_TRANSPORT_OF_MATURE_TRANSCRIPT_TO_CYTOPLASM	1.880356	0.002901
REACTOME_TRNA_MODIFICATION_IN_THE_NUCLEUS_AND_CYTOSOL	1.879894	0.002915
REACTOME_NUCLEAR_PORE_COMPLEX_NPC_DISASSEMBLY	1.879234	0.002922
REACTOME_BRANCHED_CHAIN_AMINO_ACID_CATABOLISM	1.87505	0.003065
REACTOME_ABC_FAMILY_PROTEINS_MEDIATED_TRANSPORT	1.874653	0.003062
CAFFAREL_RESPONSE_TO_THC_24HR_5_UP	1.87217	0.003139
NAKAMURA_CANCER_MICROENVIRONMENT_DN	1.871568	0.003157
KEGG_ONE_CARBON_POOL_BY_FOLATE	1.871414	0.003163
NADERI_BREAST_CANCER_PROGNOSIS_UP	1.871388	0.003153
WP_MRNA_PROCESSING	1.866147	0.003377
REACTOME_POSTMITOTIC_NUCLEAR_PORE_COMPLEX_NPC_REFORMATION	1.859229	0.003621
KEGG_OXIDATIVE_PHOSPHORYLATION	1.856226	0.003759
REACTOME_REGULATION_OF_GLUCOKINASE_BY_GLUCOKINASE_REGULATORY_PROTEIN	1.853138	0.003879
REACTOME_HEDGEHOG_OFF_STATE	1.85202	0.003916
REACTOME_TRANSPORT_OF_THE_SLBP_DEPENDANT_MATURE_MRNA	1.850498	0.003959
BIOCARTA_P53_PATHWAY	1.849571	0.003976
REACTOME_CELLULAR_RESPONSE_TO_HYPOXIA	1.849293	0.003975
KEGG_PRIMARY_IMMUNODEFICIENCY	1.848324	0.003998
CREIGHTON_ENDOCRINE_THERAPY_RESISTANCE_1	1.847793	0.004016
LI_LUNG_CANCER	1.843856	0.004192

GARGALOVIC_RESPONSE_TO_OXIDIZED_PHOSPHOLIPIDS_TURQUOISE_DN	1.842947	0.004218
REACTOME_SUMOYLATION_OF_SUMOYLATION_PROTEINS	1.841238	0.00429
REACTOME_MITOTIC_SPINDLE_CHECKPOINT	1.84094	0.004299
WP_NUCLEOTIDE_METABOLISM	1.840697	0.004307
REACTOME_TRANSPORT_OF_MATURE_MRNAS_DERIVED_FROM_INTRONLESS_TRANSCRIPTS	1.838799	0.004378
REACTOME_MITOCHONDRIAL_FATTY_ACID_BETA_OXIDATION	1.838323	0.004389
REACTOME_INTERCONVERSION_OF_NUCLEOTIDE_DI_AND_TRIPHOSPHATES	1.836	0.004508
FOURNIER_ACINAR_DEVELOPMENT_LATE_DN	1.835035	0.004569
SHAFFER_IRF4_TARGETS_IN_ACTIVATED_DENDRITIC_CELL	1.834867	0.004565
CHIANG_LIVER_CANCER_SUBCLASS_PROLIFERATION_UP	1.833644	0.004615
MOOTHA_PGC	1.832675	0.004656
KRIEG_KDM3A_TARGETS_NOT_HYPOXIA	1.83191	0.004691
WP_TCA_CYCLE_AND_DEFICIENCY_OF_PYRUVATE_DEHYDROGENASE_COMPLEX_PDHC	1.827096	0.004989
FINETTI_BREAST_CANCER_KINOME_RED	1.825746	0.005061
REACTOME_MAPK6_MAPK4_SIGNALING	1.823874	0.00519
REACTOME_CLASS_I_PEROXISOMAL_MEMBRANE_PROTEIN_IMPORT	1.823565	0.005185
REACTOME_CITRIC_ACID_CYCLE_TCA_CYCLE_	1.821111	0.005363
HONRADO_BREAST_CANCER_BRCA1_VS_BRCA2	1.819767	0.005437
REACTOME_DNA_DAMAGE_TELOMERE_STRESS_INDUCED_SENESCENCE	1.819441	0.005436
PURBEY_TARGETS_OF_CTBP1_AND_SATB1_DN	1.81898	0.005435
BIOCARTA_ATRBRCA_PATHWAY	1.813579	0.005788
REACTOME_HDR_THROUGH_SINGLE_STRAND_ANNEALING_SSA_	1.810239	0.006005
KEGG_HOMOLOGOUS_RECOMBINATION	1.807956	0.006122
REACTOME_FCERI_MEDIATED_NF_KB_ACTIVATION	1.807384	0.006154
SCHLOSSER_MYC_AND_SERUM_RESPONSE_SYNERGY	1.804976	0.006307
MATTIOLI_MGUS_VS_PCL	1.804757	0.006298
FISCHER_G2_M_CELL_CYCLE	1.80378	0.006359
REACTOME_DISEASES_OF_MITOTIC_CELL_CYCLE	1.803415	0.006369

REACTOME_RESOLUTION_OF_SISTER_CHROMATID_COHESION	1.80267	0.006406
REACTOME_BBSOME_MEDIATED_CARGO_TARGETING_TO_CILIUM	1.800755	0.006532
REACTOME_TRANSCRIPTIONAL_REGULATION_BY_RUNX2	1.800285	0.006556
SLEBOS_HEAD_AND_NECK_CANCER_WITH_HPV_UP	1.799285	0.006615
WP_MIRNA_REGULATION_OF_DNA_DAMAGE_RESPONSE	1.798153	0.006654
REACTOME_PYRUVATE_METABOLISM_AND_CITRIC_ACID_TCA_CYCLE	1.797449	0.0067
HOLLEMAN_VINCRISTINE_RESISTANCE_ALL_DN	1.797194	0.006705
DAIRKEE_CANCER_PRONE_RESPONSE_BPA	1.795912	0.006769
REACTOME_NUCLEOTIDE_SALVAGE	1.794482	0.006873
MOREIRA_RESPONSE_TO_TSA_UP	1.794201	0.006874
GEORGES_CELL_CYCLE_MIR192_TARGETS	1.791727	0.007034
REACTOME_ABERRANT_REGULATION_OF_MITOTIC_G1_S_TRANSITION_IN_CANCER_DUE_TO_RB1_DEFECTS	1.790902	0.007091
REACTOME_NUCLEAR_ENVELOPE_BREAKDOWN	1.790508	0.007099
CHUNG_BLISTER_CYTOTOXICITY_UP	1.790217	0.007111
REACTOME_APC_CDC20_MEDIATED_DEGRADATION_OF_NEK2A	1.788425	0.007229
REACTOME_HIV_LIFE_CYCLE	1.787927	0.007246
WP_DNA_DAMAGE_RESPONSE	1.786613	0.007344
SCHLOSSER_MYC_TARGETS_AND_SERUM_RESPONSE_UP	1.786573	0.00733
SHIPP_DLBCL_VS_FOLLICULAR_LYMPHOMA_UP	1.783721	0.007532
REACTOME_RNA_POLYMERASE_III_TRANSCRIPTION_INITIATION_FROM_TYPE_1_PROMOTER	1.781251	0.007738
CONCANNON_APOPTOSIS_BY_EPOXOMICIN_DN	1.780537	0.007783
DAZARD_UV_RESPONSE_CLUSTER_G1	1.776104	0.008151
KEGG_ALANINE_ASPARTATE_AND_GLUTAMATE_METABOLISM	1.776103	0.00813
REACTOME_DOWNSTREAM_SIGNALING_EVENTS_OF_B_CELL_RECEPTOR_BCR_	1.774612	0.008267
WHITFIELD_CELL_CYCLE_G1_S	1.774136	0.008289
PELLICCIOTTA_HDAC_IN_ANTIGEN_PRESENTATION_DN	1.772698	0.008406
REACTOME_PROCESSING_OF_CAPPED_INTRONLESS_PRE_MRNA	1.77239	0.008401
REACTOME_DEADENYLATION_DEPENDENT_MRNA_DECAY	1.771852	0.008434

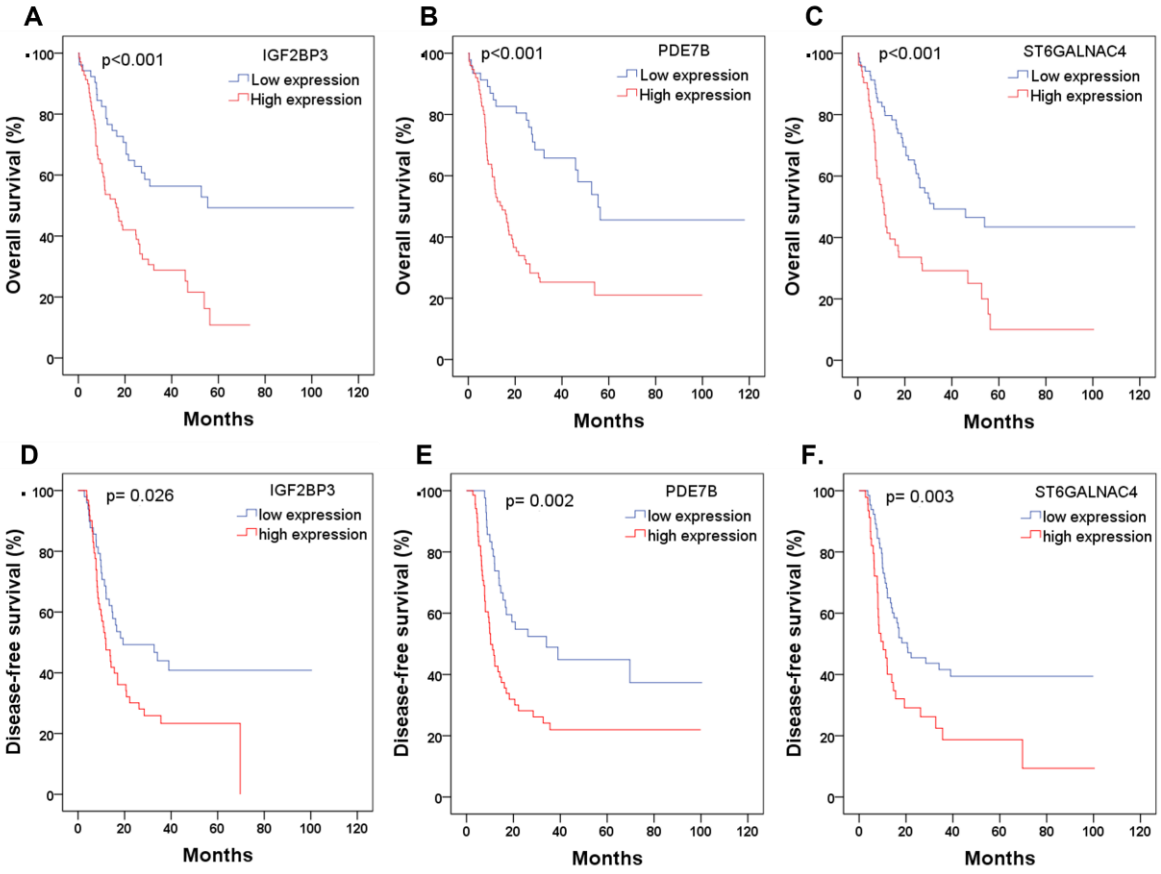
FERRANDO_T_ALL_WITH_MLL_ENL_FUSION_DN	1.77183	0.008412
WP_THE_EFFECT_OF_PROGERIN_ON_THE_INVOLVED_GENES_IN_HUTCHINSONGILFORD_PROGERIA_SYNDROME	1.768584	0.008715
VALK_AML_CLUSTER_8	1.766015	0.008952
LIN_MELANOMA_COPY_NUMBER_DN	1.765357	0.008993
REACTOME_MITOTIC_PROMETAPHASE	1.76418	0.009085
REACTOME_PROCESSIVE_SYNTHESIS_ON_THE_C_STRAND_OF_THE_TELOMERE	1.763908	0.009087
REACTOME_FORMATION_OF_INCISION_COMPLEX_IN_GG_NER	1.76317	0.009139
RICKMAN_METASTASIS_UP	1.762004	0.009221
REACTOME_FORMATION_OF_TC_NER_PRE_INCISION_COMPLEX	1.76129	0.00928
BOHN_PRIMARY_IMMUNODEFICIENCY_SYNDROM_UP	1.760732	0.009314
CHICAS_RB1_TARGETS_LOW_SERUM	1.759708	0.009397
MAYBURD_RESPONSE_TO_L663536_DN	1.758608	0.009476
DUTERTRE_ESTRADIOL_RESPONSE_6HR_UP	1.756467	0.009681
FLECHNER_BIOPSY_KIDNEY_TRANSPLANT_REJECTED_VS_OK_DN	1.754252	0.009895
FERRANDO_HOX11_NEIGHBORS	1.752647	0.01005
WP_ONE_CARBON_METABOLISM	1.751377	0.010185
KEGG_RNA_POLYMERASE	1.749048	0.010408
WELCSH_BRCA1_TARGETS_DN	1.747246	0.010562
BLALOCK_ALZHEIMERS_DISEASE_INCIPIENT_DN	1.745003	0.010808
KAAB_HEART_ATRIUM_VS_VENTRICLE_DN	1.741947	0.011123
REACTOME_HEDGEHOG_ON_STATE	1.738765	0.011462
SMID_BREAST_CANCER_LUMINAL_A_DN	1.737094	0.01165
SENGUPTA_NASOPHARYNGEAL_CARCINOMA_UP	1.736367	0.01172
WP_TCA_CYCLE_AKA_KREBS_OR_CITRIC_ACID_CYCLE	1.734924	0.011869
NAM_FXYD5_TARGETS_DN	1.73273	0.012091
KEGG_PROPANOATE_METABOLISM	1.729938	0.012447
REACTOME_RNA_POLYMERASE_II_TRANSCRIPTION_TERMINATION	1.728768	0.012546
ZHONG_RESPONSE_TO_AZACITIDINE_AND_TSA_DN	1.72865	0.01253

KEGG_BIOSYNTHESIS_OF_UNSATURATED_FATTY_ACIDS	1.726864	0.012721
SESTO_RESPONSE_TO_UV_C6	1.726299	0.012782
REACTOME_SUMOYLATION_OF_UBIQUITINYLATION_PROTEINS	1.725086	0.012928
WP_EUKARYOTIC_TRANSCRIPTION_INITIATION	1.724369	0.012983
JISON_SICKLE_CELL_DISEASE_DN	1.724145	0.012967
PID_E2F_PATHWAY	1.722546	0.013138
TAKAO_RESPONSE_TO_UVB_RADIATION_UP	1.722404	0.013118
GUTIERREZ_MULTIPLE_MYELOMA_UP	1.719478	0.013519
WHITFIELD_CELL_CYCLE_S	1.717124	0.013809
REACTOME_HSF1_ACTIVATION	1.714051	0.014226
REACTOME_PEROXISOMAL_PROTEIN_IMPORT	1.713109	0.014332
HALLMARK_FATTY_ACID_METABOLISM	1.712309	0.014418
WP_GENOTOXICITY_PATHWAY	1.709416	0.014775
MOOTHA_TCA	1.709186	0.014781
TOMLINS_PROSTATE_CANCER_UP	1.707793	0.014956
KEGG_PEROXISOME	1.701063	0.015915
KEGG_BUTANOATE_METABOLISM	1.700556	0.015968
WALLACE_PROSTATE_CANCER_UP	1.700497	0.01594
HOLLEMAN_ASPARAGINASE_RESISTANCE_B_ALL_UP	1.695943	0.016636
VANTVEER_BREAST_CANCER_METASTASIS_DN	1.694749	0.016787
WP_FLUOROPYRIMIDINE_ACTIVITY	1.694384	0.016811
ALONSO_METASTASIS_UP	1.688217	0.01777
JAEGER_METASTASIS_UP	1.68073	0.019184
OUELLET_CULTURED_OVARIAN_CANCER_INVASIVE_VS_LMP_UP	1.678897	0.019489
LOPES_METHYLATED_IN_COLON_CANCER_DN	1.678566	0.019514
PELLICCIOTTA_HDAC_IN_ANTIGEN_PRESENTATION_UP	1.676662	0.019825
KEGG_PYRUVATE_METABOLISM	1.6704	0.020987
REACTOME_CELLULAR_RESPONSE_TO_HEAT_STRESS	1.670296	0.020964

STEIN_ESRRA_TARGETS_UP	1.665591	0.021886
WP_UREA_CYCLE_AND_ASSOCIATED_PATHWAYS	1.664938	0.021967
REACTOME_APC_C_CDC20_MEDIATED_DEGRADATION_OF_CYCLIN_B	1.664363	0.022025
NAKAYAMA_SOFT_TISSUE_TUMORS_PCA2_UP	1.659278	0.023078
HAHTOLA_MYCOSIS_FUNGOIDES_CD4_DN	1.656093	0.023768
WP_GASTRIC_CANCER_NETWORK_1	1.654596	0.024008
REACTOME_RESOLUTION_OF_D_LOOP_STRUCTURES_THROUGH_SYNTHESIS_DEPENDENT_STRAND_ANNEALING_SDSA_	1.651457	0.024721
LY_AGING_PREMATURE_DN	1.649317	0.025219
REACTOME_REGULATION_OF_HSF1_MEDIATED_HEAT_SHOCK_RESPONSE	1.647309	0.025662
WONG_PROTEASOME_GENE_MODULE	1.64576	0.025954
REACTOME_MEIOSIS	1.644217	0.026263
KEGG_P53_SIGNALING_PATHWAY	1.641421	0.026926
VANHARANTA_UTERINE_FIBROID_WITH_7Q_DELETION_UP	1.640367	0.027098
VILLANUEVA_LIVER_CANCER_KRT19_UP	1.640182	0.027092
REACTOME_GENE_SILENCING_BY_RNA	1.636181	0.028031
WANG_SMARCE1_TARGETS_DN	1.635636	0.028131
REACTOME_FGFR2_ALTERNATIVE_SPLICING	1.632819	0.028802
REACTOME_PCP_CE_PATHWAY	1.632129	0.028924
REACTOME_RESOLUTION_OF_D_LOOP_STRUCTURES	1.63164	0.028987
REACTOME_ASSOCIATION_OF_TRIC_CCT_WITH_TARGET_PROTEINS_DURING_BIOSYNTHESIS	1.631171	0.02907
RODWELL_AGING_KIDNEY_NO_BLOOD_DN	1.630368	0.029187
REACTOME_SIGNALING_BY_FGFR2_IIIA_TM	1.62952	0.029354
REACTOME_RNA_POLYMERASE_III_TRANSCRIPTION_TERMINATION	1.626746	0.030039
WP_TRANSSULFURATION_AND_ONE_CARBON_METABOLISM	1.626741	0.029979
KEGG_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_ANCHOR_BIOSYNTHESIS	1.624439	0.030617
WP_DNA_IRDOUBLE_STRAND_BREAKS_DSBS_AND_CELLULAR_RESPONSE_VIA_ATM	1.623245	0.030893
STEIN_ESR1_TARGETS	1.623055	0.030888
REACTOME_NEDDYLATION	1.619289	0.031923

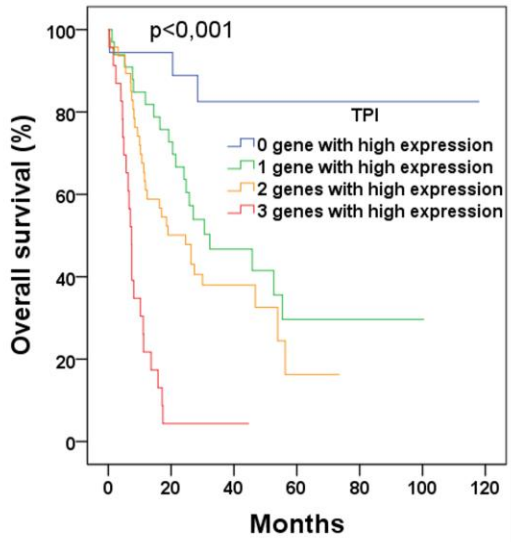
REACTOME_MITOCHONDRIAL_CALCIUM_ION_TRANSPORT	1.61846	0.032069
KIM_WT1_TARGETS_DN	1.618098	0.03211
REACTOME_HIV_TRANSCRIPTION_INITIATION	1.616934	0.032399
PID_DNA_PK_PATHWAY	1.616474	0.032446
REACTOME_SYNTHESIS_OF_GLYCOSYLPHOSPHATIDYLINOSITOL_GPI_	1.612905	0.033385
HAHTOLA_SEZARY_SYNDROM_DN	1.612339	0.033496
REACTOME_DEGRADATION_OF_BETA_CATENIN_BY_THE_DESTRUCTION_COMPLEX	1.611349	0.033709
REACTOME_PROTEIN_FOLDING	1.609486	0.034165
REACTOME_E3_UBIQUITIN_LIGASES_UBIQUITINATE_TARGET_PROTEINS	1.604155	0.035721
PYEON_CANCER_HEAD_AND_NECK_VS_CERVICAL_UP	1.603593	0.035842
KEGG_ARGININE_AND_PROLINE_METABOLISM	1.60314	0.035906
KIM_ALL_DISORDERS_DURATION_CORR_DN	1.602257	0.036091
REACTOME_CYCLIN_A_B1_B2_ASSOCIATED_EVENTS_DURING_G2_M_TRANSITION	1.602129	0.036045
WP_AMINO_ACID_METABOLISM	1.600829	0.036346
DAIRKEE_CANCER_PRONE_RESPONSE_BPA_E2	1.598578	0.037007
WILCOX_RESPONSE_TO_PROGESTERONE_UP	1.596629	0.037618
LEE_METASTASIS_AND_RNA_PROCESSING_UP	1.596191	0.037674
CHICAS_RB1_TARGETS_SENESCENT	1.594369	0.038193
PODAR_RESPONSE_TO_ADAPHOSTIN_DN	1.592978	0.038541
REACTOME_TRANSCRIPTIONAL_REGULATION_BY_RUNX1	1.592701	0.038559
REACTOME_HCMV_EARLY_EVENTS	1.591516	0.038885
REACTOME_REGULATION_OF_TP53_ACTIVITY	1.59021	0.039238
REACTOME_NUCLEAR_ENVELOPE_NE_REASSEMBLY	1.590015	0.039236
REACTOME_ANTIVIRAL_MECHANISM_BY_IFN_STIMULATED_GENES	1.588653	0.039623
WP_MITOCHONDRIAL_CIV_ASSEMBLY	1.58623	0.040354
SA_G1_AND_S_PHASES	1.585801	0.040397
REACTOME_RNA_POLYMERASE_III_CHAIN_ELONGATION	1.583825	0.041005
SU_TESTIS	1.581726	0.041681

ACOSTA_PROLIFERATION_INDEPENDENT_MYC_TARGETS_UP	1.581524	0.041677
MULLIGHAN_NPM1_MUTATED_SIGNATURE_2_DN	1.579238	0.042433
REACTOME_KINESINS	1.577695	0.042958
RAHMAN_TP53_TARGETS_PHOSPHORYLATED	1.577403	0.042961
REACTOME_TRANSCRIPTIONAL_REGULATION_BY_TP53	1.576636	0.043157
WIERENGA_PML_INTERACTOME	1.576071	0.043285
REACTOME_METABOLISM_OF_COFACTORS	1.57597	0.043239
WARTERS_RESPONSE_TO_IR_SKIN	1.571474	0.044918
RICKMAN_TUMOR_DIFFERENTIATED_WELL_VS_POORLY_UP	1.570771	0.045091
WP_PYRIMIDINE_METABOLISM_AND_RELATED_DISEASES	1.570225	0.045217
REACTOME_PROTEIN_UBIQUITINATION	1.569457	0.04543
ZHAN_MULTIPLE_MYELOMA_CD2_DN	1.567928	0.045911
STEIN_ESRRA_TARGETS	1.565927	0.046666
LUI_THYROID_CANCER_PAX8_PPARG_DN	1.565523	0.046736
PUIFFE_INVASION_INHIBITED_BY_ASCITES_UP	1.564012	0.047284
REACTOME_COPI_DEPENDENT_GOLGI_TO_ER_RETROGRADE_TRAFFIC	1.56346	0.047417
KEGG_HISTIDINE_METABOLISM	1.562869	0.047558
REACTOME_MRNA_CAPPING	1.560687	0.048334
REACTOME_RNA_POLYMERASE_III_TRANSCRIPTION	1.558453	0.049213



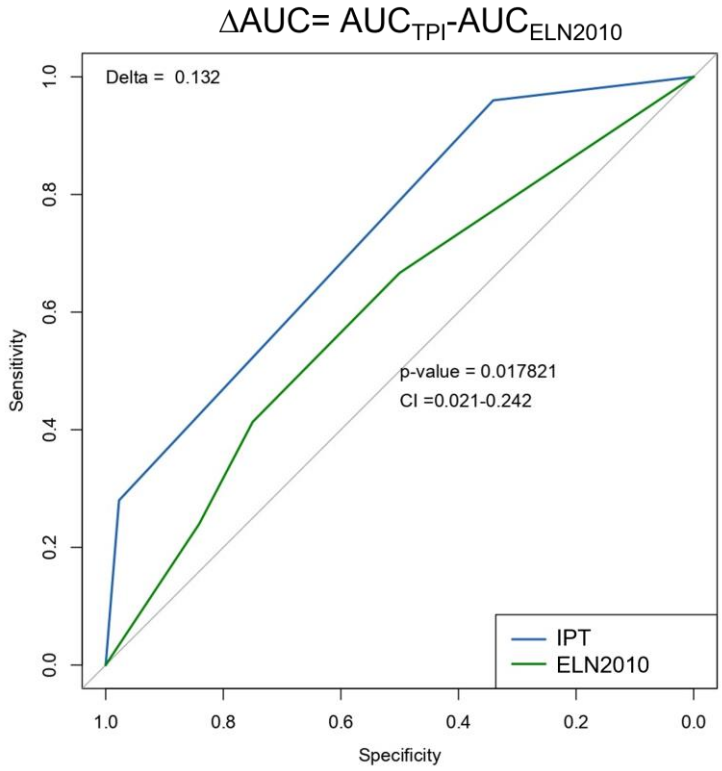
Gene	OS%3y	Median OS	95%CI	DFS%3y	Median DFS	95%CI
IGF2BP3						
Low expression	56.3	55.4	not calculated	43.9	19.3	0.0-41.15
High expression	28.8	16.3	9.21-23.39	23.3	11.9	8.28-15.52
PDE7B						
Low expression	65.7	55.4	not calculated	48.9	34.1	9.22-58.98
High expression	25.2	14.5	9.15-19.85	21.9	10.3	7.69-12.9
ST6GALNAC4						
Low expression	49.3	32.3	6.08-58.5	41.6	20.6	7.91-33.29
High expression	29.2	11.1	9.02-13.2	18.7	10.3	6.08-14.5

Supplemental figure 1: Comparison of OS (A) and DFS (D) curves between patients with high and low expression of the *IGF2BP3* gene. Comparison of OS (B) and DFS (E) curves between patients with high and low expression of the *PDE7B* gene. Comparison of OS (C) and DFS (F) curves between patients with high and low expression of the *ST6GALNAC4* gene.

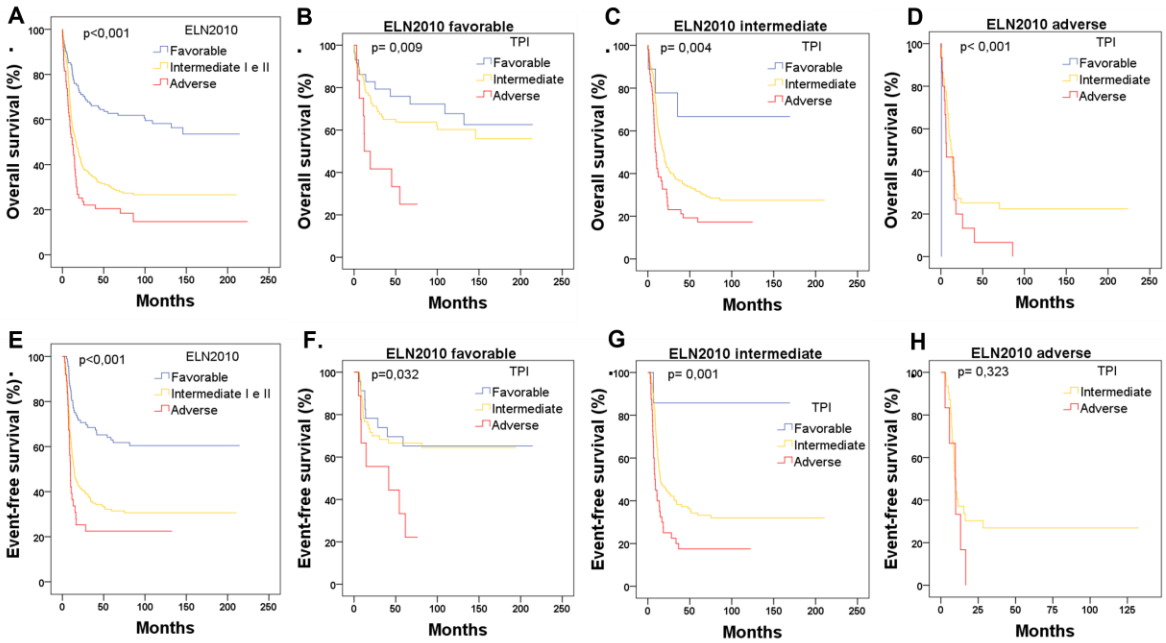


Curve comparison	*p-value*
0 x 1 gene with high expression	0,004
0 x 2 genes with high expression	<0,001
0 x 3 genes with high expression	<0,001
1 x 2 genes with high expression	0,205
1 x 3 genes with high expression	<0,001
2 x 3 genes with high expression	<0,001

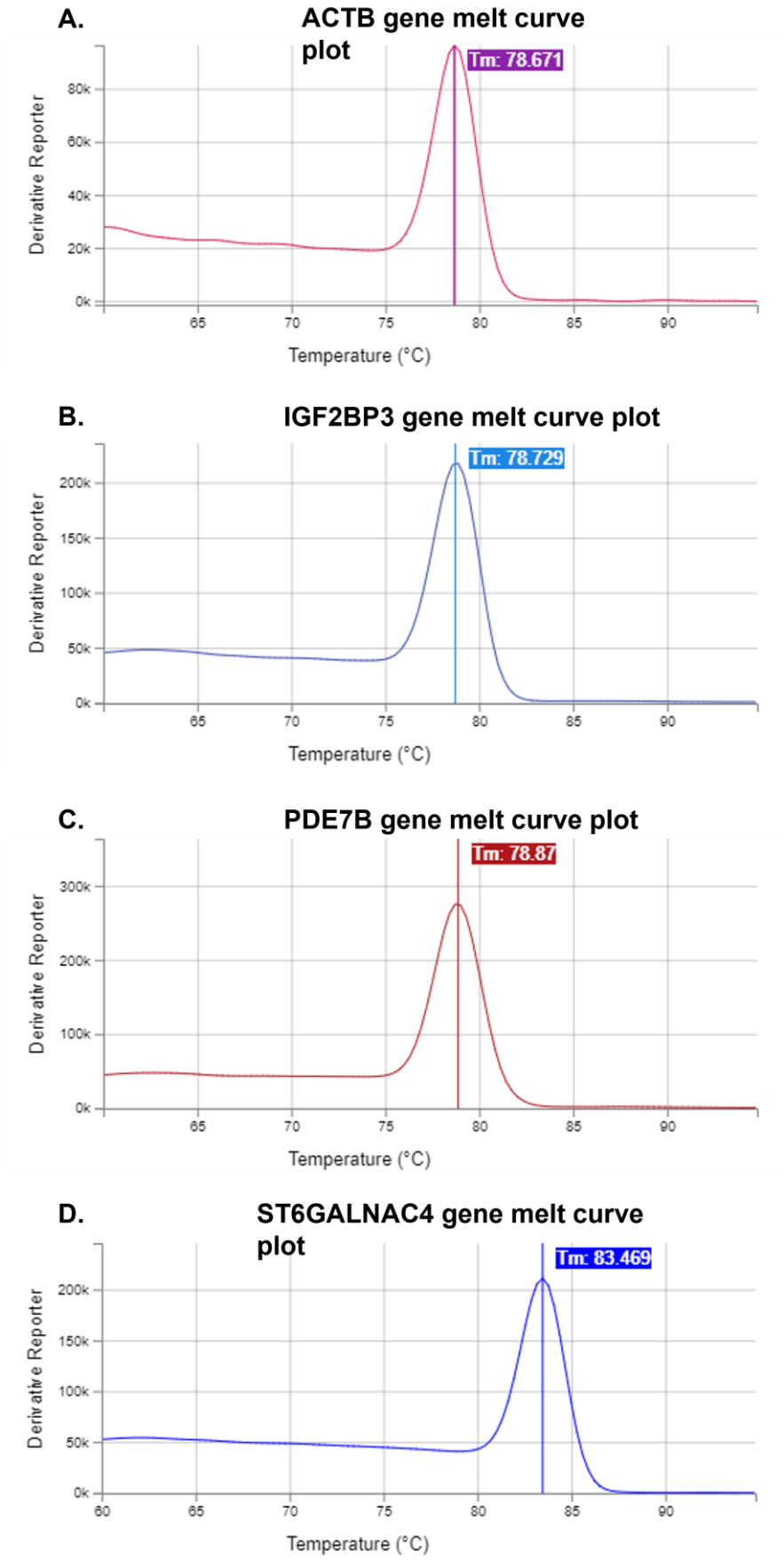
Supplemental figure 2: Inspection of the Kaplan-Meyer curve for categorization of risk groups by the transcriptome-based prognostic index.



Supplemental figure 3: ROC curves for the ELN2010 and TPI risk stratification.



Supplemental figure 4: Comparison between the OS (A) and EFS (E) curves of the favorable, intermediate I/II and adverse ELN2010 risk groups of the GSE6891 external validation cohort. Comparison of OS and DFS curves between favorable, intermediate and adverse TPI risk groups of patients in the favorable (B and F, respectively), intermediate I/II (C and G, respectively) and adverse (D and H, respectively) ELN2010 risk group of the GSE6891 external validation cohort. Due to the small number of patients, we merged intermediate ELN2010 risk groups I and II into just one intermediate group.



Supplemental figure 5: Melting curve plot for amplification of *ACTB* (A), *IGF2BP3* (B), *PDE7B* (C) and *ST6GALNAC4* (D) genes.

APÊNDICE F

UNIVERSIDADE FEDERAL DE PERNAMBUCO
CENTRO DE BIOCIÊNCIAS
LABORATÓRIO CENTRAL

TERMO DE CONSENTIMENTO LIVRE E ESCLARECIDO

(PARA MAIORES DE 18 ANOS OU EMANCIPADOS - Resolução 466/12)

Convidamos o (a) Sr.(a) para participar como voluntário(a) da pesquisa "Impacto prognostico de alterações moleculares em pacientes com leucemia mieloide aguda do adulto", que está sob a responsabilidade do pesquisador Antonio Roberto Lucena de Araujo, residente a Rua Charles Darwin, número 183, Boa Viagem CEP: 51021-520, telefone: (81) 999089382, email: araujoarl@hotmail.com. Também participam desta pesquisa os pesquisadores: Marcos André Cavalcanti Bezerra (81-998008105), Aleide Santos de Melo Lima (81-988358596), Matheus Filgueira Bezerra (81-998734248), Marianny Fernanda Teixeira Pacheco (81-999444491), Guilhermy Victor Sousa de Araújo (81-996569414), Elaine Cristina Fontinele Martins (81-99179-8040) e Fábio Rodrigo Barbosa Dutra Nascimento (81-988612131).

Caso este Termo de Consentimento contenha informações que não lhe sejam compreensíveis, as dúvidas podem ser tiradas com a pessoa que está lhe entrevistando e apenas ao final, quando todos os esclarecimentos forem dados, caso concorde com a realização do estudo pedimos que rubrique as folhas e assine ao final deste documento, que está em duas vias, uma via lhe será entregue e a outra ficará com o pesquisador responsável.

Caso não concorde, não haverá penalização, bem como será possível retirar o consentimento a qualquer momento, também sem nenhuma penalidade.

INFORMAÇÕES SOBRE A PESQUISA:

Essa pesquisa tem como objetivo "avaliar o panorama genético de pacientes adultos com leucemia mieloide aguda (LMA), bem como seu impacto prognóstico nos principais desfechos clínicos da doença (remissão, recaída e morte)". Serão incluídos na pesquisa pacientes adultos (acima de 18 anos) com diagnóstico de LMA, atendidos e acompanhados no Hospital Hemope, no Hospital do Câncer de Pernambuco (HCP) e no Instituto de Medicina Integral Professor Fernando Figueira (IMIP). O material a ser analisado na pesquisa será de medula óssea (tutano) e/ou sangue periférico. Essas amostras são coletadas de rotina no momento do diagnóstico e só utilizaremos na pesquisa o material que sobrar dessas análises, não sendo necessária nenhuma coleta adicional, evitando transtornos ao paciente. A pesquisa não pretende modificar o tratamento ou acompanhamento já preconizado para o paciente.

Essa pesquisa não traz riscos adicionais para o paciente, além dos riscos que são intrínsecos aos procedimentos que já seriam realizados na rotina de diagnóstico, pois o seu material biológico será o mesmo utilizado na rotina laboratorial do HCP, IMIP e Hospital Hemope e será encaminhado para a pesquisa somente após seu uso na rotina, sendo utilizado apenas o material que sobrar. Em nenhuma ocasião será coletado material excedente para uso da pesquisa.

Com essa pesquisa pretendemos, futuramente, melhorar o entendimento sobre o desenvolvimento da LMA e o refinamento da análise de prognóstico (classificação de risco) e, assim, poderá haver uma maior adequação do protocolo de tratamento a ser seguido para cada paciente, podendo aumentar as chances de cura.

Apesar de a pesquisa de algumas destas mutações serem preconizadas pela Organização Mundial da Saúde, o alto custo destes exames muitas vezes não permite sua realização nas instituições públicas ou por custeio próprio do paciente. Nesse contexto, esse projeto se propõe a realizar os testes em questão e fornecer os resultados à equipe dos hospitais envolvidos e aos pacientes.

Todas as informações desta pesquisa serão confidenciais e serão divulgadas apenas em eventos ou publicações científicas, não havendo identificação dos voluntários, a não ser entre os responsáveis pelo estudo, sendo assegurado o sigilo sobre a sua participação. Os dados de prontuário coletados nesta pesquisa, ficarão armazenados em computador pessoal, sob a responsabilidade do pesquisador principal, no endereço acima informado, pelo período mínimo de 5 anos.

Nada lhe será pago e nem será cobrado para participar desta pesquisa, pois a aceitação é voluntária, mas fica também garantida a indenização em casos de danos, comprovadamente decorrentes da participação na pesquisa, conforme decisão judicial ou extra-judicial. Se houver necessidade, as despesas para a sua participação serão assumidas pelos pesquisadores (ressarcimento de transporte e alimentação).

Em caso de dúvidas relacionadas aos aspectos éticos deste estudo, você poderá consultar o Comitê de Ética em Pesquisa Envolvendo Seres Humanos da UFPE no endereço: **(Avenida da Engenharia s/n – 1º Andar, sala 4 - Cidade Universitária, Recife-PE, CEP: 50740-600, Tel.: (81) 2126.8588 – e-mail: cepccs@ufpe.br).**

_____

Prof. Dr. Antônio Roberto Lucena de Araújo
Pesquisador responsável

CONSENTIMENTO DA PARTICIPAÇÃO DA PESSOA COMO VOLUNTÁRIO (A)

Eu, _____, CPF _____, abaixo assinado, após a leitura (ou a escuta da leitura) deste documento e de ter tido a oportunidade de conversar e ter esclarecido as minhas dúvidas com o pesquisador responsável, concordo em participar do estudo “Impacto prognostico de alterações moleculares em pacientes com leucemia mieloide aguda do adulto”, como voluntário (a). Fui devidamente informado (a) e esclarecido (a) pelo(a) pesquisador (a) sobre a pesquisa, os procedimentos nela envolvidos, assim como os possíveis riscos e benefícios decorrentes de minha participação. Foi-me garantido que posso retirar o meu consentimento a qualquer momento, sem que isto leve a qualquer penalidade.

.

Impressão digital (opcional)

Local e data _____

Assinatura do participante: _____

Presenciamos a solicitação de consentimento, esclarecimentos sobre a pesquisa e o aceite do voluntário em participar. (02 testemunhas não ligadas à equipe de pesquisadores):

Nome:	Nome:
Assinatura:	Assinatura:

ANEXO A

Normas da revista *Blood*

Journal Scope

Blood**, the flagship journal of the American Society of Hematology**, published online and in print, provides an international forum for the publication of original articles describing basic laboratory, translational, and clinical investigations in hematology. Acceptance of manuscripts is based on the originality and importance of the observations or investigations, the quality of the work and validity of the evidence, the clarity of presentation, and the relevance to our readership and field. All articles are expected to be concise, well-organized and clearly written. Authors submit a manuscript with the understanding that the manuscript (or its essential substance) has not been published other than as an abstract in any language or format and is not currently submitted elsewhere for print or electronic publication.

Primary research articles will be published under the following scientific categories: ***Clinical Trials and Observations***; Gene Therapy; Hematopoiesis and Stem Cells; ***Immunobiology and Immunotherapy***; Myeloid Neoplasia; Lymphoid Neoplasia; Phagocytes, Granulocytes and Myelopoiesis; Platelets and Thrombopoiesis; Red Cells, Iron and Erythropoiesis; Thrombosis and Hemostasis; Transfusion Medicine; Transplantation; and Vascular Biology. Papers can be listed under more than one category as appropriate. Authors are invited to submit a presubmission inquiry if they are uncertain whether their work falls within the general scope of the journal.

Blood has formulated more precise scopes for immunobiology and immunotherapy, vascular biology, HIV/HTLV, and clinical trial submissions. The journal is very interested in submissions as reflected in the definitions linked below.

Immunobiology and Immunotherapy scope

- ***Vascular Biology scope***
- ***HIV/HTLV scope***
- ***Clinical Trials scope***

Blood welcomes submission of manuscripts reporting on clinical trials whether phase 1, 2, 3 or 4. Reports should include a full description of the study design, patient population, methodology and conduct, and statistical plan. Immunobiology and Immunotherapy encompass a wide spectrum of research, but *Blood* can accommodate only papers that have clear and important implications for hematology. Preference is given to papers focusing on human immunobiology, immunotherapy, and immune pharmacology approaches in the human setting and those with significant implications for understanding of normal or malignant hematologic processes. Papers on tumor immunology, tumor vaccines and cancer immunotherapy development may be appropriate if the target cells or antigens are relevant to hematologic malignancies, but generally *Blood* cannot accommodate tumor immunology papers focusing solely on non-hematologic tumor types.
Papers investigating autoimmunity and utilizing non-hematologic models are not

within the scope of *Blood*. Vascular biology papers that focus primarily on atherosclerosis are outside the scope of *Blood* and instead should be considered for alternative journals. Papers felt to be outside the scope of *Blood* will be returned to the author without full peer review.

Original research articles

Regular Articles

Maximum length for a Regular Article is 4,000 words of text - counting only the Introduction, Methods, Results, and Discussion. Submissions are limited to a total of 7 figures, and digital images are required. We recommend a limit of 100 references. The sections of a Regular Article should be ordered as follows:

- Abstract
- Introduction
- Methods
  (must include sufficient information to allow readers to understand the article content)
- Results
- Discussion
- Acknowledgements
- Authorship Contributions
- Disclosure of Conflicts of Interest
- References
- Tables
- Figure Legends
- Figures

Supplemental data - to be published online only - may include additional information regarding methodology, supplemental figures or tables, or primary data sets; it must be submitted with the original manuscript submission so it can be peer reviewed. (***See "Supplemental data"***)

Any involvement of medical writers/researchers, particularly those employed or supported by the pharmaceutical industry, in the writing of an article must be clearly defined and disclosed in the Authorship and/or the Acknowledgements section(s) as appropriate. This type of involvement must also be disclosed to the Editor-in-Chief in the Cover Letter. For more information, see the journal **Conflict of interest disclosure** and the ***editorial policies for authors.***

Definitive original research articles of exceptional scientific importance may be considered for designation as Plenary Papers. The decision to highlight an article as a Plenary Paper rests entirely with the Editors.

Systematic Review and Meta-analysis

The journal welcomes submission of systematic reviews with meta-analysis, which should be submitted as a Regular Article. A literature review should be performed and summarized. Data (where available) should be systematically extracted, and summative statistics should be presented where possible.

Brief Reports

Short manuscripts definitively documenting either experimental results or informative clinical observations will be considered for publication in this category. Single-case reports or case series cannot be accommodated unless they elucidate very novel and important disease biology or approaches to therapy. Brief Reports are not intended to allow publication of incomplete or preliminary findings. The

review process is equally rigorous as for Regular Articles and the acceptance rate is lower. Brief Reports may not exceed 1,200 words of text -counting only the Introduction, Methods, Results, and Discussion. Abstracts must not exceed 200 words and should be a single paragraph with no subheadings. Only 2 figures/tables and 25 references may be included. The sections of a Brief Report should be ordered as follows:

- Abstract
- Introduction
- Methods
  (must include sufficient information to allow readers to allow reproduction of the data)
- A combined Results and Discussion section
- Acknowledgements
- Authorship Contributions
- Disclosure of Conflicts of Interest
- References
- Tables
- Figure Legends
- Figures

e-Blood

Some years ago, we created the e-Blood article type for manuscripts that presented novel datasets or methodologies with high relevance for hematology research. Over time, it has become increasingly common for a wide range of research papers to include significant supplemental datasets, and as a result we are now folding e-Blood into our Regular Article designation. Blood remains interested in these articles, and under the new article type designation they will be published both online and in print.

Other article types

Review Articles

Review articles are welcomed by the Journal and are generally solicited by the Editor-in-Chief; authors wishing to submit an unsolicited Review Article are invited to contact the Editor-in-Chief prior to submission in order to screen the proposed topic for relevance and priority, given other review articles that may already be in preparation. Review articles should focus on recent scientific or clinical advances in an area of broad interest to those in the field of hematology. Such articles must be concise and critical and should include appropriate references to the literature. All Review Articles, including those solicited by the Editors, are rigorously peer reviewed before a final publication decision is made.

Review articles should not exceed 4,000 words in length; the abstract must not exceed 250 words; we recommend a limit of 100 references. We wish for Review Articles to be written by experts who are personally committed to writing the manuscript, and therefore limit authorship to a maximum of 3 authors. The use of tables and color figures to summarize critical points is encouraged; the Journal offers assistance with preparation or improvement of figures by professional illustrators, once the article is accepted.

Pharmaceutical or medical device company employees and medical writers supported by a pharmaceutical or medical device company are not permitted to have any role in writing Review, Perspective, How I Treat, Blood Spotlight, or Evidence-based Focused Review articles. The use of ***editing services for non-***

English speakers is permissible, but it must be disclosed. Please direct any questions regarding this policy to the ***Editor-in-Chief*** prior to submission.

Special Reports

Special Reports encompass manuscripts that are neither reviews nor original reports of primary research. They may include consensus statements, guidelines, statements from task forces, or recommendations. Pharmaceutical or medical device company employees cannot act as authors, and medical writers supported by a pharmaceutical or medical device company are not permitted to have any role in writing Special Reports.

These articles should not exceed 4,000 words of body text and 250 words for the abstract; a limit of 100 references is recommended, although this can be flexible. Supplemental files are permitted. The use of tables and figures to summarize critical points is encouraged; the journal offers aid with preparation or improvement of figures by professional illustrators, once the article is accepted. The journal may also consider interactive links associated with the online article.

Perspectives

Perspectives are articles discussing significant topics and controversies relevant to hematology, generally from a more personal or opinion-based standpoint than a Review Article. Interested authors should ***correspond with the Editor-in-Chief*** prior to submission to discuss the suitability of the proposed subject matter. The length of the article should not exceed 4,000 words; the abstract must not exceed 250 words; we recommend a limit of 100 references. We wish for Perspectives to be written by experts who are personally committed to writing the manuscript, and therefore limit authorship to a maximum of 3 authors. Typically, Perspectives should state the topic and background information concisely, discuss opposing viewpoints, and make recommendations for further investigations or actions.

Pharmaceutical or medical device company employees and medical writers supported by a pharmaceutical or medical device company are not permitted to have any role in writing Review, Perspective, How I Treat, **Blood** Spotlight, or Evidence-based Focused Review articles. The use of ***editing services for non-English speakers*** is permissible, but it must be disclosed. Please direct any questions regarding this policy to the ***Editor-in-Chief*** prior to submission.

Blood **Spotlight**

Blood Spotlights are articles that focus on emerging scientific and clinical developments or a recent burst of advances on a particular theme in a circumscript area and are generally solicited by the Editor-in-Chief; authors wishing to submit an unsolicited **Blood** Spotlight are invited to contact the Editor-in-Chief prior to submission to allow screening of the proposed topic for relevance and priority in relation to other Spotlight articles that may already be in preparation. Spotlights deal with scientific or clinical topics that typically have surfaced in the last 1-3 years and that fomented a significant interest in the field of hematology. Spotlights must offer critical discussions in a highly condensed and succinct format. They should include appropriate references to the literature. All **Blood** Spotlights, including those solicited by the Editors, are rigorously peer reviewed before a final publication decision is made.

Blood Spotlights should not exceed 1,500-2,000 words in length, should include an abstract of no more than 80 words, and we recommend a limit of 80 references. We wish for **Blood** Spotlight articles to be written by experts who are personally committed to writing the manuscript, and therefore limit authorship to a maximum

of 3 authors. Article titles should be concise. The inclusion of no more than 2 tables and/or figures to highlight and summarize critical points is encouraged; the Journal offers assistance with preparation or improvement of figures by professional illustrators, once the article is accepted.
Pharmaceutical or medical device company employees and medical writers supported by a pharmaceutical or medical device company are not permitted to have any role in writing Review, Perspective, How I Treat, **Blood** Spotlight, or Evidence-based Focused Review articles. The use of ***editing services for non-English speakers*** is permissible, but it must be disclosed. Please direct any questions regarding this policy to the ***Editor-in-Chief*** prior to submission.

How I Treat

The Journal welcomes articles written by expert clinicians offering up-to-date information and guidance regarding diagnosis and treatment of hematological diseases and clinical situations based on longstanding clinical experience. Each How I Treat article focuses on a single disease for which new information has recently emerged. Because many hematologic diseases are rare, their clinical management cannot always be based on large clinical trials. This increases the value of having articles authored by leading experts with in depth experience in the chosen disease.
How I Treat articles need to allow the clinical reader to gain greater insight into the biology of the disease and how it shapes clinical decision making through the synthesis of an expert clinician. Furthermore, these articles provide directions in the therapeutic management of common or complex clinical situations of the disease. The How I Treat articles contain illustrative clinical cases which should enhance their practical usefulness among the readers. The exemplary clinical cases in the text focus on the daily reality of clinical practice and present the information in an easily digestible format. Thus, the clinical cases need to be functionally integrated into the text for a full understanding of the issues and offer a link to the scientific background and diagnostic and therapeutic considerations in a practically meaningful way.

For instance, a How I Treat article could comprise: I. Introduction; II. Two to six cases each highlighting distinct clinically relevant issues and/or management dilemmas which are linked to the discussion that is presented; and III. Conclusion. Alternatively, one or two cases are presented that set the stage and are referred throughout the discussion. Selected supporting figures and tables are recommended. For examples see eg AM Vannucchi, How I treat polycythemia vera, ***Blood (2014) 124 (22): 3212-3220***; G. Ossenkoppele and B. Löwenberg, How I treat the older patient with acute myeloid leukemia, ***Blood (2015) 125 (5): 767-774***.

These pieces are generally solicited by the Editor-in-Chief, though any interested author should correspond with the Editor-in-Chief prior to submission to discuss the suitability of the proposed subject matter. We wish for How I Treat articles to be written by experts who are personally committed to writing the manuscript, and therefore limit authorship to a maximum of 2 authors. The length should not exceed 4,000 words; the abstract must not exceed 200 words; we recommend a limit of 100 references.
Pharmaceutical or medical device company employees and medical writers supported by a pharmaceutical or medical device company are not permitted to

have any role in writing Review, Perspective, How I Treat, Blood Spotlight, or Evidence-based Focused Review articles. The use of *editing services for non-English speakers* is permissible, but it must be disclosed. Please direct any questions regarding this policy to the *Editor-in-Chief* prior to submission.

Blood Work

Blood welcomes submissions of photomicrographs and brief case descriptions to emphasize the value that the microscope adds to the history and physical exam. Blood Work places importance on the peripheral smear. The figure must clearly demonstrate the feature(s) being described and the discussion should emphasize this teaching point(s). The objective used (e.g. ×100 objective) should be stated within the text. Blood Work is an educational instrument to physicians and hematology students. Studies or case studies will not be accepted.

Each submission must contain a single, high-resolution figure (may be a composite) formatted as a TIFF (minimum 300 dpi) and a discussion of no more than 200 words describing the clinical case linked to the image and summary teaching point. Each piece should have a maximum of two authors and should not contain references. All other policies governing submissions to the Journal apply to Blood Work. There is no submission fee for Blood Work. For questions about submission to this section, please contact the Journal's Editorial Office at *editorial@hematology.org*.

Accepted material will be submitted to the ***ASH Image Bank*** for the Editor-in Chief's consideration.

Letters to Blood

The journal accepts original communications that bring out a focused but novel and important message on basic or clinical topics in hematology. All clinical submissions must have been approved by an ethics committee or institutional review board. Consideration for publication will be based on priority and interest to readership as determined by Editorial evaluation and peer review. In general, Letters to Blood report primary investigations that provide novel and important insights into hematologic biology or pathobiology, or the therapy of hematologic disease. Submissions are not intended to allow publication of incomplete or preliminary findings.

Single-case reports or small case series are considered in the Journal only if they offer truly important data elucidating disease biology or therapy.

Letters to Blood include no more than 1,200 words of text, 25 references, and 2 figures or tables. Letters have no abstract. A clear title is required. Letters to Blood are posted to First Edition, indexed by Medline, and appear in PubMed. Submission fees and page charges do not apply to Letters.

Blood Commentary

The Editors invite experts in the field to write brief commentaries introducing and placing into context selected primary research articles included in each issue of Blood.

e-Letters

Short, relevant comments that would expand reader understanding of Blood articles can be submitted as e-Letters. These contributions undergo a rapid editorial review and are posted shortly after acceptance. Further requirements and policies:

- e-Letters must have no more than 10 authors
- no more than 300 words of text
- No more than 5 references

- Tables and figures cannot be submitted
- Any conflicts of interest (see Conflict of interest disclosure)
- must accompany e-Letters

e-Letters that are not directly related to a published article, that duplicate points similar to those of already posted comments, or that are characterized by profanity, personal attacks, unprofessional tone or content, or offensive, abusive or libelous language will not be posted. Authors should pay special attention to spelling and grammar.

Authors of published articles are encouraged to respond with their own comments to e-Letters that have been posted in reference to their articles. e-Letters are not indexed by Medline.

Manuscript Preparation

Before submitting your manuscript online at *eJournalPress*, please read and carefully follow the guidelines below. Any deviations could result in significant delay in the submission and review process.

Please note that Blood adheres to the criteria of the *International Committee of Medical Journal Editors*, which has established Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals.

Manuscript length restrictions

Please adhere to the length, figure/table and reference number restrictions described on the previous page for each article type. On the title page, include a text word count, abstract word count, the number of figures and tables, and the number of references. Submissions exceeding these limits will not be considered for review and will be returned to the author.

Manuscript organization

Organize the content of your manuscript file as follows: Title Page, Key Points, Abstract, Introduction, Methods, Results, Discussion, Acknowledgments, Authorship Contributions, Conflict of Interest Disclosures, References, Tables, Figure Legends, and Figures. For Brief Reports, Results and Discussion must be combined. Your text document must include page numbers, meaning there must be page numbers inserted into the header or the footer of your document.

Article title and title page

The title should succinctly and effectively convey to non-specialists the content of the article with no more than 120 characters, including spaces. Titles should be in active rather than passive voice, without the use of punctuation or abbreviations. If commonly-understood abbreviations are included in the title, they must be defined in the abstract. If the article reports on results utilizing solely non-human model systems, the species must be indicated in the title.

Title page must contain the following: article title; short title for the running head (not to exceed 50 characters, including spaces between words); full and accurate names of all authors (as you want them to appear in online searches and citations); affiliations of institutions where the research was done, reflecting the order of authorship by using superscripted numbers; corresponding author's full name, address, e-mail address, and phone and fax numbers; word counts for text and abstract, figure/table count and reference count.

Regular Articles and Brief Reports should also include on the title page an appropriate scientific category chosen during submission.

Key Points

Blood now publishes 1 to 2 Key Point summaries of research papers - specifically, Regular Articles, Brief Reports. The purpose of these short, bullet-pointed statements is to identify the most relevant outcomes of the paper and to provide a synopsis encapsulating the significance of the research and its implications for readers.

Key Points should be written clearly and succinctly. Avoid using scientific jargon whenever possible. Each Key Point should be no more than 140 characters, including spaces. Key Points are required upon manuscript submission, immediately preceding the Abstract in both the submission form metadata and the text document, and they will be reviewed by the assigned Editor.

Key Points are published online, in First Edition, and in print immediately preceding the Abstract and will be freely available upon publication. They will not be indexed by PubMed, but will be searchable via Google and other search engines.

Abstract

The abstract should contain 250 words or fewer (200 words or fewer for Brief Reports; check the word count limit in the description for other article types) and succinctly, in a logical progression state the rationale/hypothesis, objectives, findings/results, and conclusions of the study. Abstracts should be a continuous narrative and not broken up into subheadings, and should not contain references. Authors need to ensure that abstracts are easily readable and understandable to a broad readership. The abstract should accurately reflect the content of the article, be written in plain and succinct language and, as much as possible, avoid jargon and acronyms.

The abstract of a research paper should preferably contain the following elements (per ICMJE recommendations):

- The context or background for the study. The authors should consider that a vast majority of readers have either no or limited knowledge of the article context: one or two plain-language sentences should clearly describe this background.
- The study's purpose, i.e., why the study was done. The objectives of the research should be explicitly provided, rather than in general statements.
- Methods/procedures (selection of study participants, settings, measurements, analytical methods).
- Main findings, giving specific effect sizes and their statistical and clinical significance, if possible.
- Main conclusions and interpretation of findings with emphasis on new and important aspects of the study and/or observations.

Methods

The materials and methods section should be detailed enough to provide clear information on what was done experimentally, including all major experimental plans and procedures. The Journal will not consider manuscripts that include significant portions of the methods section as supplemental data.

Methods: clinical trials or human subjects research

- See policies below regarding reporting of investigations involving human subjects and clinical trial registration.Include in the Methods section as appropriate:

- A statement that the research was approved by the relevant institutional review boards or ethics committees and that all human participants gave written informed consent.
- A statement regarding the identity of those who analyzed the data and confirming access of all authors to primary clinical trial data.
- The clinical trial registration number and approved registry name for all clinical trials.

For phase 3 randomized clinical trials, we request that the authors provide a flow diagram in CONSORT format and include all of the information required by the CONSORT checklist within the body of the manuscript. When restrictions on length prevent the inclusion of some of this information in the manuscript, it may be provided instead as supplemental data. The CONSORT statement, checklist, and flow diagram are available at *http://www.consort-statement.org***.**

For all clinical trials that report on a parenteral or high-intensity treatment regimen, information required for actual administration of the treatment regimen in practice should be included as a separate supplemental file. The following components should be included:

- Drug name (chemical, generic, and brand name or names)
- Dose (along with any modifications made for BMI, hematologic parameters, renal function, or other factors)
- Route (if parenteral, is central venous access required?)
- Type and volume of diluent if drug is not given IV push direct from vial; rate of administration
- Cycle length and number of cycles, or criteria for discontinuation
- Premedications and concurrent medications (including hydration, anti-emetics, growth factors, or any other relevant supportive medications)
- Patient-monitoring parameters (frequency of visits and blood draws during therapy)

Methods: high-throughput studies

The Journal requires that authors deposit their high-throughput data, including mRNA, miRNA, proteomic, and genomic DNA (arrayCGH, ChIP-chip, and SNP) arrays into a public database, such as Gene Expression Omnibus (GEO) or Array Express, or provide open access to their own Web-based data repository. An accession number or Web site link, with valid access codes, active at the time of submission for access by Editors and reviewers, must be supplied in the Methods section of the text.

Methods: animal studies

Blood recommends that authors follow the ARRIVE guidelines when reporting in vivo experiments in animal research. (Kilkenny C, Browne WJ, Cuthill IC, et al. *Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research*. PloS Biol 2010; 8(6): e1000412).

Methods: data sharing statement

Per Blood's *data sharing policy*, authors must make datasets and protocols available to other investigators without unreasonable restrictions.

All original research submissions must include a subsection titled "Data Sharing Statement" at the end of the Methods section. This subsection should include information regarding how to access any data that have been deposited to publicly accessible databases as well as a person to contact for data that are not publicly accessible.

Example statements:

- “For original data, please contact name@example.org.”
- “Microarray data are available at GEO under accession number XXXXXXX.”
- “X data may be found in a data supplement available with the online version of this article. Y data have been deposited to www.example.org.”

Methods: Clinical trial data sharing

International Committee of Medical Journal Editors *(ICMJE)* guidelines require that clinical trial reports address the authors’ intent to share deidentified individual participant data. The Data Sharing Statement for clinical trial reports should include the following information:

- Will deidentified individual participant data be shared
- Which particular data will be shared
- Any additional documents that will be shared
- The method through which data will be shared
- The timeframe during which data will be accessible
- The conditions, if any, required to access to the data
- For more information regarding what should be stated please see the ICMJE’s recommendation page, which includes a table of detailed example statements that was the basis for the examples below.

Example statements

- “Deidentified individual participant data are available indefinitely at www.example.org. The study protocol, analytic code…[etc.] are also available at the same website.”
- “Deidentified individual participant data that underlie the reported results will be made available 3 months after publication for a period of 5 years after the publication date at www.example.org. Proposals for access should be sent to name@example.org. The study protocol is included as a data supplement available with the online version of this article.”
- “Individual participant data will not be shared.”

While these guidelines do not mandate that individual participant data must be shared, refusal to share underlying data must be documented in the Data Sharing Statement as in the final example. For any questions regarding this policy, please contact Joanna Robertson, Senior Manager, Data Integrity–Publications, at *jrobertson@hematology.org***.**

Acknowledgments

Support received from individuals, organizations, grants, corporations, and/or any other sources must be acknowledged. For work involving a biomedical product or potential product partially or wholly supported by corporate funding, a note must be included stating: This study was supported (in part) by research funding from [company name] to [author’s or authors’ initials]**.** Grant support, if received, needs to be stated and the specific granting institution(s) name(s) and grant numbers provided when applicable. Any individuals involved in the writing/editing/researching of the paper not named as authors should be identified, their role specified, and their funding source specified; for example, “Joseph Smith, a medical writer supported by funding from [company name], provided drafts and editorial assistance to the authors during preparation of this

manuscript." Prior to submission of the manuscript, we recommend that authors notify all individuals being included in the acknowledgments section to ensure their names and roles are being identified accurately.

Authorship and conflict-of-interest statements

For each author, include in this section his or her category of contribution and list any potential conflicts of interest. These statements will be printed and posted online in the First Edition and in the final version in the Authorship section.

If the author(s) declare no competing financial interests, this must be explicitly stated and will be included in all versions of the article. Contributions and COI must appear both in the metadata and in the manuscript text.

References

Include references in numerical order at the end of the article according to the order of citation in the manuscript text. Text citations of reference should consist of superscript numbers. Format references per the instructions of the Blood Style Guide. If you use citation software, check it carefully to ensure that it formats your references according to the current Blood style.

Authors can now have Medline links in their HTML references for citations that have only been published via prepublication in Blood First Edition or in other prepublished articles. Since prepublished articles have PubMed records and a PubMed ID (PMID) is listed at the bottom of every PubMed record as the citation identifier, an author can include the PMID within his or her manuscript references to link the prepublication citation to its PubMed record. Citation of a paper prepublished in First Edition must also include its DOI number, as shown in the prepublished article.

Footnotes and abbreviations

Do not use footnotes; instead, sparingly use parenthetical statements within text. Abbreviations should be defined at first mention and thereafter applied consistently throughout the article. Do not use nonstandard abbreviations or abbreviate terms appearing fewer than three times. Give the chemical name of a compound after the first use of the common name. The common name may be used throughout the article. Abbreviate units of measure only when used with numbers. See the Blood Style Guide for more information.

Figures

When submitting a manuscript for review, image file formats accepted for uploading include: GIF, JPEG (.jpg), PDF, TIFF, and EPS. PowerPoint (.ppt) files are acceptable but are strongly discouraged due to conversion issues and poor resolution in the published article.

High-resolution image files are not preferred for initial submission as the file sizes may be too large. The total file size of the PDF for peer review should not exceed 5 MB. However, high-resolution figures are required for accepted articles entering into prepublication and print production. To prepare print-quality figures, see Figure preparation and sizing for the final print publication. Detailed instructions for submitting digital artwork can be found at *Digital artwork for production in* Blood.

Important guidelines for image preparation

(This set of instructions is adapted with permission from the *Journal of Cell Biology* instructions to authors.)

Note that no specific feature within an image may be enhanced, obscured, moved, removed, or introduced. If groupings of images from different parts of the same gel or microscopic field, or from different gels, fields, or exposures are used, they must

be made explicit by the arrangement of the figure (i.e., by inserting black dividing lines) and in the text of the figure legend, explaining what steps were taken to produce the final image and for what reason. Adjustments of brightness, contrast, or color balance are acceptable if they are applied to the whole image and as long as they do not obscure, eliminate, or misrepresent any information present in the original, including backgrounds. Without background information, it is not possible to evaluate how much of the original gel is actually shown. Nonlinear adjustments (e.g., changes to gamma settings) must be disclosed in the figure legend. The use of special software tools (e.g., erasing, cloning) available in popular image-editing software is strongly discouraged unless absolutely necessary, and any such manipulations must be explained in the figure legend.

All images in Figures and Supplemental information from manuscripts accepted for publication are examined for any indication of improper manipulation or editing. Questions raised by Blood staff will be referred to the Editors, who may then request the original data from the author(s) for comparison with the submitted figures. Such manuscripts will be put on hold and will not be prepublished in Blood First Edition until the matter is satisfactorily resolved. If the original data cannot be produced, the acceptance of the manuscript may be revoked.

Cases of deliberate misrepresentation of data will result in revocation of acceptance and will be reported to the corresponding author's home institution or funding agency.

Figure legends

All legends must begin with a short, descriptive sentence that summarizes the intent and content of the figure. This sentence should be in boldfaced font. A more detailed explanation of the data contained in the figure and/or its parts should follow in standard (non-boldfaced) font.

Whenever possible, the following information should be provided in figure legends regarding the acquisition and processing of images:

- Make and model of microscope
- Type, magnification, and numerical aperture of the objective lenses
- Temperature
- Imaging medium
- Fluorochromes
- Camera make and model
- Acquisition software
- Any subsequent software used for image processing, with details about types of operations involved (e.g., type of deconvolution, 3D reconstructions, surface or volume rendering, gamma adjustments, etc.).

If some of the information is not available, an explanation must be provided.

Tables

Each table should have a brief, specific, descriptive title, giving sufficient explanation to make the data intelligible without reference to the text. Number all tables and cite in numerical order in the text, using Arabic numerals.

Display of sequences

Prepare sequences as figures (images), not tables. This will ensure that proper alignment within a sequence is preserved.

Supplemental data

The Journal encourages the submission of supplemental data linked to primary research articles, including videos and short movies, that enhance the understanding of the science discussed in the manuscript. Supplemental data

must be included during the initial submission of the parent manuscript. All supplemental data, other than videos, must be contained in a single PDF or Microsoft Word (.doc or .docx) file — not as separate files for each individual component. Do not include any supplemental data in the main manuscript text document, including appendices (e.g., lists of contributors to a consortium), methods, tables, figures, and legends of any kind. The Editors will review the supplemental material along with the manuscript, but acceptance of the manuscript does not guarantee ultimate acceptance of the supplement.

Supplemental data may or may not appear alongside an accepted article at the time of its publication in First Edition, depending on the time needed to process the supplemental material. Blood instituted a publication fee of $105 for each standard data supplement accompanying an accepted paper. Any supplement exceeding 5 MB will incur an additional $105 (USD) fee; exceptions are possible for certain video files at the Editor's discretion. The fee is waived for Review Articles, How I Treat, Perspectives. For more information, please see Supplemental data in Blood. Any information necessary for a reader to fully evaluate and understand an article must be included in the main text of a paper — not included solely in supplemental data.

Editing services for non-English speakers

Blood has partnered with Enago to provide language editing services for authors. It is available to any author. Pricing and timescale options vary and authors can choose the level of assistance that meet their requirements. Enago are experts in the field of technical and language editing as well as academic and technical translation.

To support our authors, Blood has negotiated an exclusive discount for authors with the editing services provider, Enago. Click here (*www.enago.com/pub/blood*) to create an account or to order an editing service

Please note that having your work language edited by Enago will not in any way guarantee publication. The edited manuscript will still be subject to the same rigorous editorial assessments and checks at peer review; the editorial decision is based exclusively on the merits of the manuscript.

Supplemental Data

The main manuscript should contain adequate information to permit understanding and interpretation. Supplemental data may include additional figures, tables, materials and methods, or other items that add value to the manuscript, but are not necessary to understand the underlying research. Supplemental data should not be used to circumvent the word count or figure limits of a submission. Files that are not printable, such as videos or long gene sequences, must be submitted as supplemental files.

Supplemental data are evaluated as part of peer review; current, clearly labeled files must be included at every stage of review. Please do not include any supplemental material in your manuscript files. Supplemental files are uploaded separately from manuscript files.

When uploading supplemental data to the Blood *submission system*, please upload all supplemental data as a single PDF when possible, with content appearing in the following order:

- Methods
- Tables (with legends)
- Figures (with legends)

- Legends for videos or other separate supplemental files
- Appendices (e.g., list of participants in a multi-center study)

Tables and figures in the supplemental PDF must be labeled as “supplemental”, as in “Supplemental Table 1” and ”Supplemental Figure 1.” Videos must be labeled as “Video 1;” videos should not be labeled as “movies.”
For files in formats that may be difficult or impossible to combine into PDF (e.g., Excel or video), we ask that you upload those files separately and include the legends or descriptions for those supplements in the PDF.

Publication of Supplements
For each supplemental file that was uploaded during submission with an answer of “Yes” regarding whether or not to publish the file online, the most recent version of that file will be published as it was uploaded or combined into a PDF if applicable. Supplemental data will be published alongside the article in Blood *First Edition* if possible, but delays may occur for articles that include multiple supplemental files.

Blood requires a publication fee of $105 for each standard data supplement accompanying an accepted paper. The fee is waived for invited articles.

Title Page

Titles
The maximum length of an article title in Blood is 120 characters, including spaces. The title may have one subtitle (which counts toward the 120-character limit), separated from the main title by a colon. Avoid using more than one subtitle. It is preferable that the title be a phrase, but if necessary, it may be a declarative sentence or question.
If a drug name is mentioned in the title, the generic name should be used unless several proprietary versions of the drug are being compared or the article is commenting on a specific proprietary version of the drug.
matinib (not Gleevec)

alemtuzumab (not Campath)

Abbreviations are generally allowed in titles. If an abbreviation is used in a title, it must be expanded (spelled out) in the abstract.

A study group name may appear in the title and/or the byline; see Bylines (below) for rules for including it in the byline. An abbreviation of the group name may be used alone in the title if it is expanded in the abstract.
If a study group is part of the title, the group's full membership should be listed in an appendix; include a title page note that refers to that appendix.
A complete list of the members of the German Chronic Lymphocytic Leukemia Study Group appears in the "Appendix."
If a study group is not part of the title or byline but is involved in the article, any membership list should be listed in an appendix, but put the statement referring to the group in "Acknowledgments" rather than in a title page note.

Bylines
Each author name should consist of a full given name and last name; initials and patronymics may be included. Separate names with commas. Use and before the

last author's name.

Each author name should be followed by at least 1 superscript number keyed to that author's affiliation(s); see Affiliation lines below.
Stuart T. Fraser Jr,[1] Joan Isern,[2] and Margaret H. Baron[3]
If an author has more than 2 affiliations, use a hyphen.
Stuart T. Fraser Jr,1 Joan Isern,[2] and Margaret H. Baron[3–5]
Bylines should not include titles or academic degrees. Symbols, other than the ones that match authors to affiliations, may be used only if they are keyed to title page footnotes indicating equal contributions by specific authors.
Examples of study groups as authors are shown below.

Dieter Huhn, Christoph von Schilling, Martin Wilhelm, Anthony D. Ho, Michael Hallek, Rolf Kuse, Wolfgang Knauf, Ute Riedel, Axel Hinke, Stefanie Srock, Stefan Serke, Christian Peschel, and Bertold Emmerich, for the German Chronic Lymphocytic Leukemia Study Group

Dieter Huhn, Christoph von Schilling, Martin Wilhelm, Anthony D. Ho, Michael Hallek, Rolf Kuse, Wolfgang Knauf, Ute Riedel, Axel Hinke, Stefanie Srock, Stefan Serke, Christian Peschel, Bertold Emmerich, and the German Chronic Lymphocytic Leukemia Study Group
In rare circumstances, a study group may be listed as the only author.
An abbreviation of a study group name may be used alone in the byline if it is expanded in the abstract.

If a study group is part of the byline, the group's full membership should be listed in an appendix; include a title page note that refers to that appendix.

A complete list of the members of the German Chronic Lymphocytic Leukemia Study Group appears in the "Appendix."
If a study group is not part of the title or byline but is involved in the article, any membership list should be listed in an appendix but put a statement in "Acknowledgments" rather than in a title page note.

Affiliation lines

The affiliation line should be placed below the byline, and each affiliation should be preceded by a superscript number (no space) that is keyed to an author on the byline. Affiliations should be in the order of byline names' number keys. It may be necessary to treat different departments at the same institution as separate affiliations.
[1]Department of Oncological Sciences, University of Turin Medical School, Turin, Italy; [2]Laboratory Division of Clinical Oncology, Institute for Cancer Research and Treatment, Turin, Italy; and [3]Laboratory of Gene Therapy, Institute for Cancer Research and Treatment, Turin, Italy
Do not abbreviate anything in the affiliation lines except for names of US and Australian states and Canadian provinces; use official postal versions of these abbreviations.
If an author's affiliation has changed since the work reported in the article was done, a current-affiliation statement may be added to the Authorship section.

Abstracts

All articles except Editorials, **Blood** Forums, **Blood** Consults, and Letters to **Blood** must have an abstract.

The abstract must be a single paragraph. The maximum length of an an abstract is 250 words for Regular Articles, e-**Blood** articles, Review Articles, and Perspectives; 200 words for Brief Reports and How I Treat articles; and 80 words for **Blood** Spotlights and Evidence-Based Focused Reviews.

Use of abbreviations in abstracts is allowed; see ***Abbreviations*** for details.

Parts of an article (eg, references, tables, figures, appendixes) should not be cited in the abstract.

The guidelines for drug names in Titles apply to abstracts too.

Avoid the use of manufacturer names in the abstract unless a specific brand of equipment is being investigated.

Any clinical trial registry information should be given as the last sentence of the abstract.

...We conclude that specific anti–B-cell therapy with rituximab may be beneficial for patients with steroid-refractory chronic GVHD. This trial was registered at www.clinicaltrials.gov as #NCT00136396.

Title page notes

The following kinds of footnotes are allowed on the title page, in order: statement of equal author contributions, statement of article series, and statement of prior presentation. Other footnotes, such as copyright statements and submitted-accepted-prepublished date notes, will be added by the printer.

Statement of equal author contribution

The following is an example of a statement of equal author contributions.

H.K.K. and M.D.L.L.S. contributed equally to this study.

Statement of article series

If the article is part of a series of articles, a note to that effect may be included and should cite all the previous articles.

This article is a continuation of a previous report.[15]

Statement of prior presentation

Begin this note with “Presented ...,” if at all possible.

Presented in abstract form at the 57th annual meeting of the American Society of Hematology, Orlando, FL, 7 December 2015.[15]

Main Text

Text headings

All headings should be capitalized sentence style, except that scientific terms should be capitalized or lowercased for accuracy.

Main (first-level) headings

All Regular Articles and **e-Blood** articles should have an introduction, a methods section, a results section, and a discussion section; they may also have acknowledgments and/or an appendix.

- Introduction
- Materials and methods (or Methods, Patients and methods, etc, if appropriate)
- Results
- Discussion

Brief Reports should have abbreviated sections:

- Introduction
- Study design
- Results and discussion

Authors revising their papers to be Brief Reports should be sure to combine their results and discussion sections.
All other items with special section designations may use whatever main headings the authors and editors deem appropriate. Inside Blood Commentaries and Letters to Blood should avoid the use of text headings.

Subheadings

Main sections may be subdivided, and the resulting subheadings are not predefined.
Headings should be constructed in parallel; there should never be one subheading in a given section. If you use subheadings, be sure to use a consistent style for each subheading level.
See *Abbreviations* for information on the use of abbreviations in headings.
Parts of an article (eg, references, tables, figures, appendixes) should not be cited in a heading.

Notes added in proof

Any information added (or deleted) after acceptance is subject to approval by the accepting editor.
Information added between acceptance and page proof composition will be added into the manuscript's existing text by the Blood Production Office.
Information added in page proofs may be added either to existing text or as a note added in proof. Any note added in proof will appear as the last paragraph of "Discussion."

Acknowledgments

This optional section should appear right after "Discussion." List all organizational support, including fellowships, chairs, and grants, in this section. Also include all recognition of nonauthor individual contributions. Begin grant support statements “This work was supported ... .” Designation of which author received which support may be indicated by inserting parenthetical sets of initials.
This work was supported by grants from the Swiss National Science Foundation (81BS-52825) (B.U.M.) and the National Institutes of Health (grants CA41456 and CA72009) (D.G.T.).

Names of acknowledged persons should be presented consistently within each article. Their professional affiliations are optional. Do not include salutations (Dr., Prof., Ms., Mr., etc.). Dedications are not allowed.

Include all organizational support in the grant support note, including fellowships and grants. Include in the Acknowledgments all recognition of nonauthor individual contributions.

If applicable, acknowledgments should include, but are not limited to, the following:

1. Any statement regarding degree candidacy and the relationship of the work to the degree.

A.B.C. and D.E.F. are PhD candidates at Any University. This work is submitted in partial fulfillment of the requirement for the PhD.

2. Any statement summarizing the role or views of sponsors, government, or employers.

The sponsors of this study are public or nonprofit organizations that support science in general. They had no role in gathering, analyzing, or interpreting the data.

The content of this publication does not necessarily reflect the views or polices of the Department of Health and Human Services, nor does mention of trade names, commercial products, or organizations imply endorsement by the US government.

3. Any statement of manuscript number for a particular institution.

This is article no. 0000-MEM from the Scripps Research Institute.

4. Any declaration of no financial interest.

The International Safety Monitoring Committee that commissioned this study is an independent body of scientists including Elias Schwartz (chair, US), Samuel Charache (US), Chaim Hershko (Israel), Stuart MacLeod (Canada), and Giuseppe Masera (Italy). This committee was convened by Apotex Inc in accordance with section 5.5.2 of the International Conference on Harmonization Good Clinical Practice guidelines. The members of the Safety Monitoring Committee and the authors have no financial interest in the development of deferiprone.

5. Any reference to an appendix listing the membership of a study group (see ***Titles and Bylines*** for exceptions).

Thanks to the German Chronic Lymphocytic Leukemia Study Group for its participation in this study; a complete membership list appears in "Appendix."

Authorship

This required section should appear after the article text and after any acknowledgments. It must contain an authorship statement, a conflict-of-interest statement (positive or negative), and contact information for the corresponding author, and it may contain other information.

The authorship contribution statement should be the first paragraph, introduced by “Contribution:”. The positive or negative conflict-of-interest statement should be the second paragraph, introduced by “Conflict-of-interest disclosure:”. The corresponding author’s contact information should constitute the final paragraph, introduced by “Correspondence:”

Contribution: H.K.K., M.D.L.L.S., and A.V.G. performed experiments; C.K.K. analyzed results and made the figures; G.T. and H.K.K. designed the research and wrote the paper.

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

Correspondence: Giovanna Tosato, Basic Research Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892; e-mail: tosatog@mail.nih.gov.

More information about conflict-of-interest disclosures is in the ***Author Guide***.

The corresponding author's contact information should consist of the name, mailing address, and e-mail address of the author prepared to handle all official correspondence. Avoid academic degrees and phone and fax numbers. If necessary, more than one author's contact information may be listed.
Correspondence: Pier Giuseppe Pelicci, Department of Experimental Oncology, European Institute of Oncology, Via Ripamonti 435, 20141 Milan, Italy; e-mail: pgpelicci@ieo.it; and Saverio Minucci, Department of Experimental Oncology, European Institute of Oncology, Via Ripamonti 435, 20141 Milan, Italy; e-mail: sminucci@ieo.it.

Other information that may be put in "Authorship" includes any statement of author death, any current author affiliations that have changed since the study was done, and any notice of a study group membership list as an online supplement. All this information should be placed together in the next-to-last paragraph (ie, between the disclosure and the author contact information).
Contribution: H.K.K., M.D.L.L.S., and A.V.G. performed experiments; C.K.K. analyzed results and made the figures; G.T. and H.K.K. designed the research and wrote the paper.
Conflict-of-interest disclosure: The authors declare no competing financial interests.
The current affiliation for R.F. is Division of Haematology/Oncology, Hospital for Sick Children, Toronto, Ontario, Canada.
John M. Jones died on 2 August 2016.
A complete list of the members of the German Chronic Lymphocytic Leukemia Study Group appears as a data supplement to the online version of this article.
Correspondence: Giovanna Tosato, Basic Research Laboratory, Center for Cancer Research, National Cancer Institute, National Institutes of Health, Bethesda, MD 20892; e-mail: tosatog@mail.nih.gov.

In notes that mention authors (other than the death statement and the correspondence information), refer to authors with complete sets of initials, with periods and closed up. If 2 or more authors have the same initials, write out last names.
H.D.K.
J.-M.F.
J. M. Faint and J. M. Fletcher

Appendixes
All appendixes appear immediately before "References."
If an appendix contains a figure or table that is referenced anywhere in text, give it a number in succession with any other figures or tables in the article. If the figure or table is not referenced by name anywhere in the text, it need not have a number. If an appendix consists largely of a figure or table, it should not have a number or a caption at all; however, it may have a title and/or legend.

List study group members in an appendix. See *Titles, Bylines,* and Acknowledgments (above) for instructions on how to refer to such an appendix.

Only if an article has more than 1 appendix should an appendix be numbered; multiple appendixes should be numbered with arabic numerals. If a substantive

name for an appendix would be helpful, include it after a colon.

Appendix
Appendix: study group members
Appendix 1, Appendix 2
Appendix 1: study group members

Addenda

The journal does not publish addenda; other options are notes added in proof or appendixes, if appropriate.

Text references to article elements

When referring to other parts of an article, be as specific as possible; avoid general statements such as "see above," "see below," or "see text."

To refer to an article element that does not have a heading, use a lowercase general description of the item, without quotation marks or other punctuation or formatting.

the abstract
the title

To refer to an element of an article that has a printed, unnumbered heading, use quotation marks around the heading exactly as it appears. Do not precede the heading with "the," and do not follow it with "section."

"Introduction"
"Materials and methods" (not "the Materials and methods section")
"Chemotaxis assay"
"Results"
"CCL19-induced dendritic extension of SPDCs"
"References"
"Appendix"

But refer to a numbered element of an article without using quotation marks.

Figure 1
Table 3
Appendix 2

Figure and table callouts

Figures and tables should appear in the order in which they are introduced in the main text. (The main text mention of a figure or table after which the item appears is called its callout.) If necessary, however, figures and tables may be mentioned in a preliminary fashion out of order, as long as this is done prior to "Results," but in such situations the actual figure or table will not appear in the article until its official callout, in order.

Parenthetical mentions of a figure or table should not include "see" unless necessary.

(Figure 1)
(see Figure 1 for a description of the process)

When more than 1 of the same type of element is cited, separate consecutive items with a hyphen and nonconsecutive items with "and."

Figures 1–3
Tables 1 and 3
Figures 2–4 and 6

Tables 2, 4, and 6

Also do this when multiple elements not of the same type are cited in running text.
Tables 2 and 3 and Figures 1 and 4–6 summarize ...
But when multiple elements not of the same type are cited in parentheses, separate them with a semicolon.
(Table 2; Figure 1)
(Figures 1–3; Tables 2 and 4)
Text references to figures may refer directly to panels and subpanels by appending them to the figure name. Separate consecutive panel labels with hyphens and nonconsecutive panel labels with commas followed by no space.
Figure 1A
Figure 2B–C
Figure 3B,D
Figure 4Aiii
Figure 5Ci–iii
Figure 6Bi,iii
Refer to parts of a figure that do not have panel labels or to parts of a table in general terms; avoid punctuation or possessives if at all possible.
Figure 2B inset
Figure 3 filled squares
Table 1 footnotes
the rightmost column of Table 2
"Congenital anomalies" in Table 2
Data presented in a table or figure should not be entirely repeated in the text. When the text does repeat a piece of information from a table or figure, care should be taken to cross-check accuracy.

Text references to companion articles

Companion articles (closely related articles that appear side-by-side in the same issue) are designated as such by the journal editors upon acceptance, if not before. Authors who want their article to be a companion to another should contact the journal editors during the review process or upon acceptance.

If an article references a designated companion article, use the standard citation-reference combination.

(see accompanying article by DiMartino et al,29)

Lists

Structure

A list of 3 or more items may be presented in the text or as a table. Lists of 2 items should always be presented in the text.
If a list consists of words or phrases and is in the main text, it should be presented as run-in text punctuated by commas or semicolons.

The response of Fanca bone marrow to in vitro stimulation is characterized by (1) an accelerated depletion in CFU-GM progenitors, (2) an evident granulocyte/macrophage differentiation disbalance, and (3) a marked susceptibility

of the expanded population to enter into apoptosis.

If a list consists of complete sentences and is in the main text, it should be presented as run-in text or as a set of paragraphs, but always punctuated by periods.

(1) IVIG (black IgGs) and 7E3 (white IgGs) are taken into the cell by pinocytosis. (2) At physiologic pH, IgG has low affinity for the FcRn receptor. (3) Bound IgG molecules are protected from release into the lysosome.

Enumeration

A list need not be numbered, but numbering is encouraged to promote clarity. Number items in a table list or a list of paragraphs using arabic numerals (1, 2, 3, etc), followed by a period, a space, and the list item.

1. IVIG (black IgGs) and 7E3 (white IgGs) are taken into the cell by pinocytosis.
2. At physiologic pH, IgG has low affinity for the FcRn receptor.
3. Bound IgG molecules are protected from release into the lysosome.

In a list that is run in as part of a sentence, number items using arabic numerals (preferred) or lowercase letters surrounded by parentheses and followed by a space and the list item.

Supplemental data

Supplemental data (not part of the main manuscript) may be accepted by the editors for publication in the online journal. Examples include videos and spreadsheets. Supplemental data should be cited at least once in the main text.

Figures

Image size and layout

Images should be laid out as compactly as is consistent with conveying the relevant data. Images will be sized to fit the smallest possible space while retaining all relevant detail. If an article is accepted for publication, the figures may be altered by *Blood's* publication management vendor to conform to *Blood* style, which includes, but is not limited to, standard colors and fonts.

In order to prevent radical changes in figure content, authors should prepare the figures 8.4 cm wide (1-column width) or, if necessary, 12.7 cm wide (1½ column width). The maximum width is 17.7 cm (2-column width).

Tabular material should almost always appear as a table that is not part of a figure. However, if the tabular material is graphically related to parts of a figure, it can appear within the figure.

Text legend

The figure legend must contain a boldface (a) name ("Figure" + arabic figure number) and (b) substantive title. Do not refer to figure panels, other figure parts, or any other part of an article in a figure title. A nonboldface description of the figure usually follows, run in after the title, describing each panel, subpanel, inset, or other part of the figure.

Figure 4. Clusters of genes categorized by the expression patterns in purified stem and progenitor cells. The vertical axis represents the normalized gene expression values. (A) Representative genes that are predominantly expressed in HSCs and downregulated in MPPs, CLPs, and CMPs. (B) Representative genes that were upregulated in MPPs. (C) Representative genes

that are highly expressed in CLPs. (D) Representative genes that are highly expressed in CMPs.

Symbol and text labels

Figure images may contain symbol and text labels that are necessary to convey relevant detail. Avoid any text labels that might be confused with panel and subpanel labels. In the example below, “control” should be abbreviated “Ctrl,” not “C,” to avoid confusion with panel C of the figure.

... (C) Plasma levels from 10 healthy control (Ctrl) donors.

Any abbreviations used in the figure image but not explained in the article's main text should be defined in the legend.

Symbols should be explained in the text legend if possible, including the symbols themselves if possible.

Symbols represent IVIG treatment groups (n = 4 rats/group): saline (●), 0.4 g/kg (■), 1 g/kg (▲), and 2 g/kg (♦).

If necessary, words may be used to describe the symbols.

Splenocytes from DO11.10 mice were incubated with 0, 1, 10, 100, or 1000 ng of MIP-1a per mL (triangles) or MIP-1b (squares) in plates containing 0 (open symbols) or 500 (filled symbols) µg of OVA per ml.

If necessary, symbols may be explained in a key within the figure image. In that case, the key should be placed into the figure image as compactly as possible.

Insets

Insets are generally explained in the figure legend, unless their meaning is obvious from the image.

Reproductions and adaptations

If a figure is reproduced or adapted from previously published material, written permission must be obtained by the author from the copyright holder (usually the publisher). The author must send the *Blood* editorial office a copy of this written permission before the article can be published, so it is advisable to request permission early.

Acknowledge the original source by creating the appropriate reference and citing it in the legend, making reference to the acquired permission.

Adapted from Foster and Hunt[47] with permission.

Reprinted from Begley et al[4] with permission.

Tables

Structure

Tables should generally consist of tabular material (at least 2 rows × 2 columns, including row heads but not column heads). But nontabular lists of 3 or more items may be presented as tables; lists of 1 or 2 items should be incorporated into the main text. Data of the same kind generally should read downward in columns, not across in rows.

Multipart tables (tables with two or more separate sets of column heads) are not allowed and must be split into separate numbered tables. No column should contain more than 1 column head. Place any title of a table section to the left as a row head, being sure to indent existing row heads as row subheads.

Title
Every table should begin with “Table” and the table number, followed by a period and a brief, substantive title.
Table 1. Human inhibitory NK cell receptors

Data fields
Organize tables as compactly and logically as possible. Minimize the number of cells with no data, and avoid empty cells. Use row subheads to minimize the number of columns. If all entries in a column or row are identical, remove that column or row and put that information in a footnote. If all entries in a column are percentages, use the percent symbol in the column head, not after each numeric entry in the column.

Column and row heads
All columns except the stub (the leftmost column) must have a column head. It is preferred but not required that the stub have a column head.

Row heads should not cut into any columns beyond the first column, nor should any head be centered across the table between rows.

In row or column heads, specify the unit for that row or column; separate the head and the unit with a comma.

Hemoglobin level, g/L

Parentheses should be used in row or column heads only to explain data entries that contain parentheses.

Average hemoglobin level, g/L (mean)

Column and row heads should be accurate (including singular or plural) for the data presented in the respective column or row.

P (not P value)
Capitalization
Capitalize table titles, column heads, row heads, and text entries in the data field sentence style, unless capitalizing the first word would change its meaning.

Abbreviations and arithmetic symbols
Use abbreviations where possible in the column and row heads and in the data field; expand in a footnote any abbreviations not expanded in the main text.

Units and measurements
Scientific units should be given according to the SI system. If non-SI units are deemed useful, give a conversion factor in a footnote; do not give both the SI and the non-SI units in tables.
Use NA (not applicable), ND (not determined), and other similar abbreviations where appropriate and define them in the abbreviation footnote. If absolutely

necessary, indicate unavailable data with an em (long) dash but explain its meaning with a footnote.

Footnotes
Footnotes should occur in the following order: (1) note applying to entire table, (2) abbreviations note, and (3) all notes with callout symbols.

1. Any note that applies to the entire table should not have a note callout symbol.

2. Any abbreviations should be given in a separate note without a note callout symbol.

BM indicates bone marrow; SCT, stem cell transplantation; and VWF, von Willebrand factor.

Abbreviations should be defined in alphabetical order. (Subsequent tables using the same abbreviations should reference the previous abbreviation footnote.)

Abbreviations are explained in Table 1.

3. All footnotes that apply to columns, rows, or individual data entries use symbols in this order: ∗, †, ‡, §, ||, ¶, #, ∗∗, ††, and ‡‡. If more than 10 such notes are necessary, do not use symbols but use superscripted lowercase letters in alphabetical order instead. The symbol should not be part of the grammar of the footnote. Treat each note as a sentence, including capitalization and a period, but do not capitalize the the first word if doing so would change the scientific meaning.
*p53 sequencing was performed on patient MM samples, as previously described.
Callout symbols should proceed left to right, starting with the table title, then the column heads, and then reading across each successive row.
References

Responsibility for citation accuracy
Authors are responsible for the accuracy of their reference and citation information. Accuracy of journal titles, volume numbers, and page numbers is particularly important: References in the online journal will be accompanied by links to Medline records and other online content, but these links will work only if citation information is completely accurate.

If citation software is used, it should be checked carefully to ensure that it formats references according to current AMA Manual of Style, 10th edition, style.
The Blood *Production Office* can answer any questions about current style, but contact your citation software's maker if the software does not generate references correctly.
Titles of articles, books, and so forth should not be edited for style; they should be accurately reproduced as they were published.

Reference citation or in-text citation?
A resource should be referenced only if it is a publication. This excludes manuscripts not accepted for publication, among other resources; such resources

are carefully delineated below.

Principles of reference citation

Reference list
References are listed at the end of the article, before the figure legends and tables.

Numeric order
References should be listed in order of their first citation in text. Authors should take care to renumber references when adding, deleting, or moving citations during the peer-review process.

Reference-citation agreement
Every citation must have a reference, and every reference must have a citation. Authors should take care to ensure this when adding, deleting, or moving citations during the peer-review process.

Location of citation numbers
References may be cited only in main text, title-page notes, figure legends, and tables. Do not cite references in abstracts, figure titles, or table titles.

Reference citation formats
Basic format
Cite references in text by inserting the superscripted reference number.
During the last decade, high-dose therapy with autologous stem cell support has become a common treatment in younger patients with myeloma.[1]

For clarity, superscripted citations may appear within a sentence or clause.
During the last decade, high-dose therapy with autologous stem cell support[1] has become a common treatment in younger patients with myeloma. The largest lesion found in the first study[2] was 10 cm.

Punctuate multiple cited references with hyphens (3 or more consecutive references) and/or commas (nonconsecutive references or 2 consecutive references).

Rh is a highly complex red cell blood group system with 52 antigens and numerous phenotypes.[1,2,5]
Virtually all patients eventually die of progressive disease, with a median survival of approximately 3 years.[1-3]

Mention of referenced work in text

If the text needs to refer directly to the referenced work, use the author’s surname.

Holmgren[4] raised the question of whether DNA can be transferred from one cell to another via the phagocytosis of apoptotic bodies.

If the reference has 2 authors, use both surnames.
Holmgren and Smith[5] raised the question of whether DNA can be transferred from one cell to another via the phagocytosis of apoptotic bodies.

If the reference has 3 or more authors, use only the first surname, plus "et al."

Holmgren et al[6] raised the question of whether DNA can be transferred from one cell to another via the phagocytosis of apoptotic bodies.

A superscript number should never be used alone as part of the grammar of the text.

For a review, see Zupanska.[8] (Not: For a review, see [8].)

If necessary, another author involved may be mentioned in the text, but for clarity it is preferable to also follow the author name formulas above**.**

Jones's group (Holmgren et al[6]) raised the question of whether DNA can be transferred from one cell to another via the phagocytosis of apoptotic bodies. [6-8]

Also if necessary, the referenced work may be referred to by its title.

The full-text TEC evidence report,Use of Epoetin for Anemia in Oncology**.**[7] should be consulted by those interested in a more detailed treatment of the state of the evidence supporting the use of epoetin in clinical oncology practice than the information provided in this guideline.

Citation of parts of referenced work

Cite a figure or a table in the referenced work by including in the superscript a parenthetical immediately after the reference number; use "Fig" or "Tab" with the item number closed up afterward.

We have shown, however, that 48 hours after IVIG injection, platelet-associated antiplatelet IgG remained at the same levels as seen in mice receiving antiplatelet antibody alone.[2(Fig1)]

In this respect, we remark that patient 2 described in our series as alive with sMDS[1(Tab2)] subsequently developed sAML and died recently of resistant leukemia.

Cite a page of another work in the same way, using "p" or "pp" with the number(s) closed up afterward.

We have shown, however, that 48 hours after IVIG injection, platelet-associated antiplatelet IgG remained at the same levels as seen in mice receiving antiplatelet antibody alone.[2(pp1945-1946)]

Always cite direct quotations with a page number. Authors are responsible for the accuracy of quotations.

Based upon their work, the authors state that "decreases in plasma antibody levels would lead to decreases in the degree of platelet opsonization."[1(p2092)]

Reference formats

Author list

Begin each reference with the author list. List each author's surname followed by a space and a set of initials closed up without periods; include a hyphen if applicable. Separate each author name with a comma; do not use and between any entries. If there are 1 to 6 authors, list all authors. If there are more than 6 authors, list the first 3 authors followed by et al.

Arlin ZA.

Sallan SE, Weinstein HJ.

Kim S-C, Hahn J-S, Min Y-H, Yoo N-C, Ko Y-W, Lee W-J.

Wetzler M, Dodge RK, Mrózek K, et al.Lenhoff S, Hjorth M, Holmberg E, et al.

Journal articles

Include the article title (capitalized sentence style and ending with a period), journal name (italicized and abbreviated according to the *National Library of Medicine*), year (followed by a semicolon), volume number, issue number in parentheses (followed by a colon), and inclusive page numbers (followed by a period).

Lenhoff S, Hjorth M, Holmberg E, et al. Impact on survival of high-dose therapy with autologous stem cell support in patients younger than 60 years with newly diagnosed multiple myeloma: a population-based study. Blood. 2000;95(1):7-11.

Prepublished journal articles

Articles accepted for publication and prepublished (aka published ahead of print) should be referenced like a journal article, except that the DOI (digital object identifier) and the date of prepublication should supplant the year, volume number, and page numbers.

Hou TZ, Verma N, Wanders J, et al. Identifying functional defects in patients with immune dysregulation due to LRBA and CTLA-4 mutations [published online ahead of print 3 February 2017]. Blood. doi:10.1182/blood-2016-10-745174.

Journal articles in press

Journal articles that are in press but are not prepublished should be referenced like other journal articles, except instead of year, volume number, and page numbers, include the phrase “In press.”

Lenhoff S, Hjorth M, Holmberg E, et al. Impact on survival of high-dose therapy with autologous stem cell support in patients younger than 60 years with newly diagnosed multiple myeloma: a population-based study. Blood. In press.

Only manuscripts that have been accepted should be referenced this way. If a manuscript has not yet been accepted, it should be cited per *Citations in text only***.**

Books and chapters thereof

Include the book title (capitalized title style and ending with a period), place of publication (followed by a colon), publisher (followed by a semicolon and a space), and the year of publication (followed by a period).

McGarry MP, Protheroe CA, Lee JJ. Mouse Hematology: a Laboratory Manual. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press; 2010.

If referencing a chapter of a book, include the chapter title (capitalized sentence style and ending with a period), "In:" (followed by a space), any book editors (listed in author-list style and ending with "ed." or "eds."), the book title, the volume and/or edition if any, and the place, publisher, and year of publication, followed by a colon and the chapter’s page numbers.

Wilson ML, Weinstein MP, Reller LB. Laboratory detection of bacteremia and fungemia. In: Jorgensen JH, Pfaller MA, Carroll KC, et al, eds. Manual of Clinical Microbiology**.** Vol 1. 11th ed. Washington, DC: ASM Press; 2015:15-28.

Papers and abstracts of papers

When referencing a paper presented at a meeting with a published abstract, reference the published abstract. Always give an abstract number; the page number is optional.

Mueller MC, Gattermann N, Lahaye T, et al. Dynamics of BCR-ABL mRNA transcript expression in newly diagnosed CML patients treated with imatinib or interferon alpha [abstract]. **Blood**. 2002;100(11):365a. Abstract 1413.

When referencing a paper presented at a meeting but not published in any form, include the title, the meeting name, the date of presentation, and the geographic location.

Eisenberg J. Market forces and physician workforce reform: why they may not work. Paper presented at Annual Meeting of the Association of American Medical Colleges. 28 October 1995. Washington, DC.

Editorials, commentaries, introductions, and letters

When a reference is made to one of these article types, include the type of item in brackets at the end of any title.

Cazzola M. Introduction to a review series on myeloproliferative neoplasms [editorial]. **Blood**. 2017;129(6):659.

Besson-Fournier C, Gineste A, Latour C, et al. Hepcidin upregulation by inflammation is independent of Smad1/5/8 signaling by activin B [letter]. **Blood**. 2017;129(4):533-536.

If referencing an Inside **Blood** Commentary, refer to it as a commentary.

Roberts I, de la Fuente J. Sickle cell disease: the price of cure [commentary]. **Blood** 2016;128(21):2486-2488.

Errata and retractions

Cite retracted articles and articles with errata in full, with the information about the erratum or retraction in square brackets after the article title.

Baraff LJ, Bass JW, Fleisher GR, et al. Practice guideline for the management of infants and children 0 to 36 months of age with fever without source [published correction appears in **Ann Emerg Med**. 1993;22(9):1490]. **Ann Emerg Med**. 1993;22(7):1198-1210.

Yu Y, Cao F, Ran Q, Sun X. Regulatory T cells exhibit neuroprotective effect in a mouse model of traumatic brain injury [retracted in **Mol Med Rep.** 2017;15(5):2897]. **Mol Med Rep.** 2016;14(6):5556-5566.

Official reports and guidelines

National Research Council. Guide for the Care and Use of Laboratory Animals. 8th ed. Washington, DC: National Academy Press; 2011.

Software manuals

Cite software manuals by using the book reference format.

SAS user's guide: statistics. Cary, NC: SAS Institute; 1999.

Online documents

If a publication or document is available in online form only, then reference it providing author name (or any institutional author), document title, the URL, and the date on which it was accessed.

Gilmore T. NF-κB transcription factors. ***http://www.bu.edu/nf-kb***. Accessed 10 February 2017.

European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Version 7.1, valid from 2017-03-10. ***http://www.eucast.org/fileadmin/src/media/PDFs/EUCAST_files/Breakpoint_tables/v_7.1_Breakpoint_Tables.pdf.*** Accessed 18 May 2017.

Online databases

Reference an online database with its author name (or any institutional author), database name, its URL, and the date on which it was accessed.

McKusick-Nathans Institute of Genetic Medicine, Johns Hopkins University. OMIM: online Mendelian inheritance in man. ***http://www.omim.org***. Accessed 10 February 2017.

Citations in text only

Some items should not be referenced and should be cited in text only in parentheses.

Personal communications

List the person's name, the affiliation, the method of communication, and the date.

Mice deficient in fibrinogen are born at the expected Mendelian frequency but can exhibit perinatal hemorrhage and death at frequencies that vary from 10% to 70% depending upon genetic background (Jane A. Smith, Harvard University, email, 19 July 2003).

Authors are responsible for acquiring written permission when citing a personal communication; the journal may request a copy of this permission.

Unpublished observations, data, or procedures

List all persons responsible for the material, a brief description, and a date. If a person is an author of the current article, use that person's initials.

Not all the protein in the standards binds to the well (L.X. and K.P.P., unpublished data, 14 May 2017).

If a nonauthor is responsible for the material, list that person's full name if at all possible.

Other studies have clearly demonstrated that HSCs can contribute to the brain (Christopher R. Cogle, Anthony T. Yachnis, Eric D. Laywell, unpublished observations, 21 August 1999)

Not all the protein in the standards binds to the well (L.X. and Eric D. Laywell, unpublished data, 14 May 2017).

Manuscripts in preparation

List all persons responsible for the material. If a person is an author of the current article, use that person's initials.

These phenotypes can be conferred to wild-type mice by bone marrow reconstitution with PAR4-deficient marrow and rescued in PAR4 null mice by reconstitution with wild-type marrow (J.R.H. and S.R.C., manuscript in preparation).

If a nonauthor is responsible for the material, list that person's full name if at all possible.

These phenotypes can be conferred to wild-type mice by bone marrow reconstitution with PAR4-deficient marrow and rescued in PAR4 null mice by reconstitution with wild-type marrow (Justin R. Hamilton and Shaun R. Coughlin, manuscript in preparation).

Manuscripts submitted but not accepted

List all persons responsible for the material and the date of submission. If a person is an author of the current article, use that person's initials; otherwise, use the person's full name.

We analyzed the potential of these cells to support HIV-1 infection and found that productive infection is dependent on the presence of exogenously administered rhGM-CSF (F.S. and Antonio Bernad, manuscript submitted January 2016).

If a nonauthor is responsible for the material, list that persons's full name if at all possible.

We analyzed the potential of these immature human dendritic cells to support HIV-1 infection, and found that productive infection is dependent on the presence of exogenously administered rhGM-CSF (Fernando Serrano and Antonio Bernad, manuscript submitted January 2004).

Software programs

Cite software programs like materials used, including the manufacturer and the manufacturer's location.

Toxicities were graded according to the Common Toxicity Criteria (CTC) version 2.0 (National Cancer Institute, Bethesda, MD).

The following primers were designed using Primer Express 1.0 software (Applied Biosystems, Foster City, CA): ...

Alternatively, a website may be given instead of a location.

The following primers were designed using Primer Express 1.0 software (Applied Biosystems, http://www.appliedbiosystems.com): ...

ANEXO B

Parecer consubstanciado do Comitê de Ética da Universidade Federal de Pernambuco

UNIVERSIDADE FEDERAL DE PERNAMBUCO CENTRO DE CIÊNCIAS DA SAÚDE / UFPE-

PARECER CONSUBSTANCIADO DO CEP

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: IMPACTO PROGNÓSTICO DE ALTERAÇÕES MOLECULARES EM PACIENTES COM LEUCEMIA MIELOIDE AGUDA DO ADULTO

Pesquisador: Antonio Roberto Lucena de Araujo

Área Temática:

Versão: 2

CAAE: 77527717.4.0000.5208

Instituição Proponente: Universidade Federal de Pernambuco - UFPE

Patrocinador Principal: Financiamento Próprio
MINISTERIO DA CIENCIA, TECNOLOGIA E INOVACAO

DADOS DO PARECER

Número do Parecer: 2.412.470

Apresentação do Projeto:

Trata-se de um projeto de pesquisa do Professor ANTONIO ROBERTO LUCENA DE ARAUJO, vinculado ao Departamento de Biofísica e Radiobiologia da Universidade Federal de Pernambuco. O projeto será desenvolvido no Laboratório Central do Centro de Biociências da UFPE, Centro de Pesquisa Aggeu Magalhães, Hospital de Cancer de Pernambuco, Instituto de Medicina Integral Prof. Fernando Figueira, Hospital das Clínicas da Universidade Federal de Pernambuco (HC-UFPE), Fundação de Hematologia e Hemoterapia de Pernambuco (HEMOPE). A proposta consiste basicamente na análise da hipótese: A caracterização molecular dos pacientes com LMA poderá ser de valor clínico significativo, no que diz respeito ao prognóstico. Alterações moleculares específicas poderão ainda, estar associadas a subtipos ou características clinico-laboratoriais da LMA. Para este fim realizar-se-á um estudo do tipo coorte retrospectivo e prospectivo.

O estudo será do tipo coorte retrospectivo e prospectivo. Serão incluídos 300 pacientes adultos com LMA, diagnosticados de março de 2010 a dezembro de 2020. Esses pacientes serão diagnosticados e acompanhados no Hospital de Hematologia e Hemoterapia da Fundação Hemope, no Hospital do Câncer de Pernambuco (HCP) e no Instituto de Medicina Integral Professor Fernando Figueira (IMIP), seguindo os critérios preconizados pela Organização Mundial de Saúde (OMS). Os pacientes serão maiores de 18 anos. Em todos os pacientes serão coletadas amostras de médula óssea e/ou sangue periférico no

Endereço: Av. da Engenharia s/nº - 1º andar, sala 4, Prédio do Centro de Ciências da Saúde
Bairro: Cidade Universitária **CEP:** 50.740-600
UF: PE **Município:** RECIFE
Telefone: (81)2126-8588 **E-mail:** cepccs@ufpe.br

ANEXO C

Parecer consubstanciado do Comitê de Ética da Fundação Hemope

FUNDAÇÃO DE
HEMATOLOGIA E
HEMOTERAPIA DO ESTADO

PARECER CONSUBSTANCIADO DO CEP

Elaborado pela Instituição Coparticipante

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: IMPACTO PROGNÓSTICO DE ALTERAÇÕES MOLECULARES EM PACIENTES COM LEUCEMIA MIELOIDE AGUDA DO ADULTO

Pesquisador: Antonio Roberto Lucena de Araujo

Área Temática:

Versão: 2

CAAE: 77527717.4.3001.5195

Instituição Proponente: Fundação de Hematologia e Hemoterapia do Estado de Pernambuco -

Patrocinador Principal: Financiamento Próprio
MINISTERIO DA CIENCIA, TECNOLOGIA E INOVACAO

DADOS DO PARECER

Número do Parecer: 2.644.746

Apresentação do Projeto:

Esse projeto tem como objetivo avaliar o perfil genético de pacientes adultos com leucemia mieloide aguda, incluindo mutações/polimorfismos de genes (NPM1, FLT3, CEBPA, WT1, MLL, IDH1, IDH2, DNMT3A, ASXL1, TP53), expressões aberrantes de genes (BAALC, ERG, EVI1, KMT2E, ID1 e MN1) e alterações no número de cópias de DNA mitocondrial (mtDNA), bem como o impacto prognóstico dessas alterações nos principais
desfechos clínicos da doença (remissão, recaída e morte). Para isso, pretende-se incluir um total de 300 pacientes diagnosticados e acompanhados no Hospital de Hematologia e Hemoterapia da Fundação Hemope, no Hospital do Câncer de Pernambuco e no Instituto de Medicina Integral Professor Fernando Figueira. Para as análises de expressão gênica e do número de cópias de mtDNA, serão utilizadas, como controle
saudável, amostras de sangue periférico e medula óssea de 20 indivíduos saudáveis. Entende-se por indivíduo saudável aquele que não apresentar nenhuma doença hematológica ou oncológica e, para essa análise, será necessário avaliar os prontuários dos pacientes selecionados. Esses indivíduos serão pacientes acompanhados no ambulatório do Serviço de Ortopedia e Traumatologia do Hospital das Clínicas de Pernambuco (HCPE) e que serão submetidos à cirurgia ortopédica em que a crista ilíaca esteja sendo abordada.

Endereço: Rua Joaquim Nabuco, 171
Bairro: Graças **CEP:** 52.011-000
UF: PE **Município:** RECIFE
Telefone: (81)3182-4771 **Fax:** (81)3182-4660 **E-mail:** cep.hemope@gmail.com

ANEXO D

Parecer consubstanciado do Comitê de Ética do Hospital do Câncer de Pernambuco

SOCIEDADE PERNAMBUCANA DE COMBATE AO CÂNCER-SPCC

Plataforma Brasil

PARECER CONSUBSTANCIADO DO CEP

Elaborado pela Instituição Coparticipante

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: IMPACTO PROGNÓSTICO DE ALTERAÇÕES MOLECULARES EM PACIENTES COM LEUCEMIA MIELOIDE AGUDA DO ADULTO

Pesquisador: Antonio Roberto Lucena de Araujo

Área Temática:

Versão: 3

CAAE: 77527717.4.3003.5205

Instituição Proponente: SOCIEDADE PERNAMBUCANA DE COMBATE AO CÂNCER -SPCC

Patrocinador Principal: Financiamento Próprio
MINISTERIO DA CIENCIA, TECNOLOGIA E INOVACAO

DADOS DO PARECER

Número do Parecer: 2.656.851

Apresentação do Projeto:

O projeto intitulado:IMPACTO PROGNÓSTICO DE ALTERAÇÕES MOLECULARES EM PACIENTES COM LEUCEMIA MIELOIDE AGUDA DO ADULTO, esta sendo apresentado na sua terceira versão 3

Objetivo da Pesquisa:

Objetivo Primário: Avaliar o panorama genético de pacientes adultos com leucemia mieloide aguda (LMA), bem como seu impacto prognóstico nos principais desfechos clínicos da doença (remissão, recaída e morte).

Objetivo Secundário: • Determinar a frequência de mutações/polimorfismos nos genes NPM1, FLT3, CEBPA, WT1, MLL, IDH1, IDH2, DNMT3A, ASXL1 e TP53 e determinar o significado prognóstico desses achados, utilizando a sobrevida global como desfecho;• Determinar a expressão gênica relativa dos genes BAALC, ERG, EVI1, KMT2E, ID1 e MN1, escolher um ponto de corte ideal (cutt off) para esses genes e determinar o significado prognóstico desses achados, utilizando a sobrevida global como desfecho;• Determinar o número de cópias de mtDNA e o significado prognóstico desse achado, utilizando a sobrevida global como desfecho;• Determinar o valor aditivo desses achados com base na proporção de risco de cada marcador molecular (índice prognóstico integrado, IPI) na predição dos principais desfechos clínicos da doença (remissão, recaída e morte);

Endereço: Av. Cruz Cabugá, 1597
Bairro: Santo Amaro **CEP:** 50.040-000
UF: PE **Município:** RECIFE
Telefone: (81)3217-8005 **Fax:** (81)3217-8005 **E-mail:** cep@hcp.org.br

ANEXO E

Parecer consubstanciado do Comitê de Ética do Instituto de Medicina Integral Professor Fernando Fiqueira - IMIP

INSTITUTO DE MEDICINA INTEGRAL PROFESSOR FERNANDO FIGUEIRA -

PARECER CONSUBSTANCIADO DO CEP

Elaborado pela Instituição Coparticipante

DADOS DO PROJETO DE PESQUISA

Título da Pesquisa: IMPACTO PROGNÓSTICO DE ALTERAÇÕES MOLECULARES EM PACIENTES COM LEUCEMIA MIELOIDE AGUDA DO ADULTO

Pesquisador: Antonio Roberto Lucena de Araujo

Área Temática:

Versão: 1

CAAE: 77527717.4.3002.5201

Instituição Proponente: Instituto de Medicina Integral Professor Fernando Figueira - IMIP/PE

Patrocinador Principal: Financiamento Próprio
MINISTERIO DA CIENCIA, TECNOLOGIA E INOVACAO

DADOS DO PARECER

Número do Parecer: 3.195.996

Apresentação do Projeto:

O estudo será do tipo coorte retrospectivo e prospectivo. Serão incluídos, neste estudo, pacientes adultos com LMA, diagnosticados de março de 2010 a dezembro de 2020, pretendendo incluir um total de 300 pacientes.Esses pacientes serão diagnosticados e acompanhados no Hospital de Hematologia e Hemoterapia da Fundação Hemope, no Hospital do Câncer de Pernambuco (HCP) e no Instituto de Medicina Integral Professor Fernando Figueira (IMIP), seguindo os critérios preconizados pela Organização Mundial de Saúde (OMS).Os pacientes serão maiores de 18 anos e submetidos ao esquema de tratamento 3+7 (LIMA et al., 2015).O seguimento dos pacientes será censurado/atualizado em março de 2021 e para aqueles em que o seguimento foi perdido serão censurados na última data em que sabidamente estavam vivos.Para as análises moleculares dos pacientes com LMA, serão utilizadas amostras de medula óssea (MO) e/ou sangue periférico (SP), desde que este último apresente mais de 80% de blastos leucêmicos circulantes.Essas amostras são coletadas no momento do diagnóstico e somente o material excedente será utilizado na pesquisa.Para as análises de expressão gênica e do número de cópias de mtDNA, serão utilizadas, como controle saudável, amostras de 4mL de SP e 5mL de MO de 20 indivíduos saudáveis.Entende-se por indivíduo saudável aquele que não apresentar nenhuma doença hematológica ou oncológica.O processamento das amostras será realizado no Laboratório Central da UFPE e no Centro de Pesquisas Aggeu Magalhães.Para obtenção dos dados clínicos, serão utilizados os

Endereço: Rua dos Coelhos, 300
Bairro: Boa Vista **CEP:** 50.070-902
UF: PE **Município:** RECIFE
Telefone: (81)2122-4756 **Fax:** (81)2122-4782 **E-mail:** comitedeetica@imip.org.br