Use of advanced flow cytometric and genomic methods to elucidate the pathophysiology of leukemias

Alexandre Bazinet
Division of Experimental Medicine
McGill University, Montreal

August 2021

A thesis submitted to McGill University in partial fulfillment of the requirements of the degree of Master of Science

TABLE OF CONTENTS

TABLE OF CONTENTS 2
LIST OF ABBREVIATIONS 4
ABSTRACT 6
RÉSUMÉ 7
ACKNOWLEDGEMENTS 9
CONTRIBUTION OF AUTHORS 10
INTRODUCTION 12
PROJECT 1: USE OF FLUORESCENCE-ACTIVATED CELL SORTING AND SINGLE-CELL TRANSCRIPTOMICS TO IDENTIFY VULNERABILITIES IN ACUTE MYELOID LEUKEMIA IN THE MINIMAL RESIDUAL DISEASE STATE 14
LITERATURE REVIEW 15
Acute Myeloid Leukemia 15
Minimal/Measurable Residual Disease in Acute Myeloid Leukemia 19
Mechanisms of Treatment Resistance in Acute Myeloid Leukemia 22
Single-Cell RNA Sequencing 23
METHODS 24
Overview 24
Patient samples 24
Design and optimization of a 17-color, single-tube, flow cytometry panel for AML MRD 25
Flow cytometry 26
Generation of a normal BM reference file 27
Flow cytometry data analysis 27
Patient survival analysis 29
Identification of AML driver mutations using NGS 29
Cell sorting and scRNA-seq 30
RESULTS 32
Application of the 17-color AML MRD panel to normal BM samples 32
Application of the 17-color AML MRD panel to diseased BM samples 36
Survival analysis 43
Use of the 17-color AML MRD panel for cell sorting and single-cell RNA sequencing 44
DISCUSSION 46

PROJECT 2: COMMON CLONAL ORIGIN OF CHRONIC MYELOMONOCYTIC LEUKEMIA AND B CELL ACUTE LYMPHOBLASTIC LEUKEMIA IN A PATIENT WITH A GERMLINE *CHEK2* VARIANT 50
LITERATURE REVIEW 51
METHODS 53
Flow cytometry and cell sorting 53
Whole-exome sequencing and analysis 54
Functional analysis of the *CHEK2* variant 55
RESULTS 58
Description of the patient case and family history 58
Identification of genomic variants in the buccal swab and CMML/B-ALL fractions 59
Functional analysis of the *CHEK2* c.475T>C, p.Y159H variant 62
DISCUSSION 65
CONCLUSIONS 68
REFERENCES 70
APPENDICES 80
Tables 80
Figures 82

LIST OF ABBREVIATIONS

ACMG	American College of Medical Genetics and Genomics
aHSCT	Allogeneic hematopoietic stem cell transplantation
AI	Allelic imbalance
ALL	Acute lymphoblastic leukemia
AML	Acute myeloid leukemia
ANOVA	Analysis of variance
APL	Acute promyelocytic leukemia
BM	Bone marrow
BSA	Bovine serum albumin
CGI	Cancer Genome Interpreter
CHIP	Clonal hematopoiesis of indeterminate potential
CLL	Chronic lymphocytic leukemia
CML	Chronic myeloid leukemia
CMML	Chronic myelomonocytic leukemia
CR	Complete remission
DfN	Different-from-normal
DFS	Disease-free survival
ECS	Error-corrected sequencing
ELN	European LeukemiaNet
EPO	Erythropoietin
ET	Essential thrombocytosis
FAB	French-American-British
FACS	Fluorescence-activated cell sorting
FBS	Fetal bovine serum
FCS	Flow Cytometry Standard
FHA	Forkhead-associated
FSC	Forward scatter
GSEA	Gene set enrichment analysis
HSC	Hematopoietic stem cell
HSPC	Hematopoietic stem and progenitor cell
IPSS	International Prognostic Scoring System
IPSS-R	Revised International Prognostic Scoring System
IR	Intermediate risk
KD	Kinase domain
LacO	Lac Operator
LacR	Lac Repressor
LAIP	Leukemia-associated immunophenotype
LOH	Loss of heterozygosity
LSC	Leukemia stem cell
MDS	Myelodysplastic syndrome
MDS-MLD	Myelodysplastic syndrome with multilineage dysplasia
MFC	Multiparameter flow cytometry

MFI	Mean fluorescence intensity
MNBM	Merged normal bone marrow
MPN	Myeloproliferative neoplasm
MRD	Minimal/measurable residual disease
MST	Minimum spanning tree
NGS	Next-generation sequencing
NHL	Non-Hodgkin's lymphoma
ORF	Open reading frame
OS	Overall survival
PBMC	Peripheral blood mononuclear cell
PBS	Phosphate-buffered solution
PCA	Principal component analysis
PCR	Polymerase chain reaction
PDX	Patient-derived xenograft
PON	Panel of normals
qPCR	Quantitative polymerase chain reaction
RBC	Red blood cell
RT	Reverse transcription
SCD	Serine-glutamine/threonine-glutamine cluster domain
scRNA-seq	Single-cell RNA sequencing
SI	Separation index
SNP	Single-nucleotide polymorphism
SNV	Single-nucleotide variant
SSC	Side scatter
VAF	Variant allele frequency
VUS	Variant of undetermined significance
WBC	White blood cell
WES	Whole-exome sequencing
WHO	World Health Organization
WT	Wild type

ABSTRACT

Recently, considerable advances have been made in the study of the pathophysiology of leukemias. However, significant knowledge gaps still exist in our understanding of the heterogeneity of these diseases at the molecular level and of the mechanisms of treatment resistance. Improvements in the fields of multiparameter flow cytometry (MFC) and genetics have the potential to answer some of these questions with regards to hematological malignancies. Such knowledge is essential for the development of rational therapeutic strategies. In this thesis, I present two projects related to this common theme.

In the first project, I designed a single-tube, 17-color flow cytometry strategy allowing for the identification of minimal/measurable residual disease (MRD) in acute myeloid leukemia (AML). In order to better evaluate this high-dimensional MFC data, I explored novel bioinformatic analysis approaches. I then demonstrated how this antibody panel can be used with fluorescence-activated cell sorting (FACS) to physically isolate AML cells that resist chemotherapy. Finally, in order to identify gene expression changes in these cells, I performed a single-cell RNA sequencing (scRNA-seq) protocol on the isolated MRD cells. This has the potential to lead to a better understanding of why these cells resist treatment. I obtained successful cDNA libraries in a small fraction of cases, demonstrating the feasibility of this approach. However, frequent RNA degradation in the single cells precluded the generation of full gene expression data. Further protocol modifications to address the issue of RNA degradation should be explored.

In the second project, I constructed a putative sequence of clonal evolution in a patient who sequentially developed myelodysplastic syndrome (MDS), chronic myelomonocytic leukemia (CMML), and B-cell acute lymphoblastic leukemia (B-ALL). This was achieved through FACS to isolate the different leukemias followed by whole-exome sequencing (WES) on these cell fractions and a buccal swab. I identified 11 potential driver mutations with various distributions between the clones. In this unusual case, I identified a germline variant in the tumor suppressor *CHEK2* gene (c.475T>C, p.Y159H) previously considered a variant of undetermined significance (VUS). I performed a protein binding assay demonstrating that this variant impairs binding to BRCA1 and may represent an inherited cancer predisposition. My findings support the re-classification of this variant as likely pathogenic.

RÉSUMÉ

Récemment, l'étude de la pathophysiologie des leucémies a subi des avancées considérables. Cependant, nous n'apprécions pas complètement l'hétérogénéité au niveau moléculaire ni les mécanismes de résistance au traitement de ces maladies. Les techniques de la cytométrie en flux et de la génétique sont très prometteuses pour répondre à ces questions sur les cancers du sang. Ces connaissances sont essentielles pour le développement de nouveaux traitements rationnels. Dans cette thèse, je présente deux projets liés à ce thème commun.

Au cours du premier projet, j'ai créé une stratégie de cytométrie de 17 couleurs permettant l'identification de la maladie résiduelle minimale (MRD) en leucémie myéloïde aiguë (LMA). Pour mieux évaluer ces données complexes de cytométrie, j'ai exploré une nouvelle méthode d'analyse bio-informatique. J'ai par la suite démontré comment ce panneau d'anticorps peut être utilisé pour isoler, à l'aide du tri cellulaire induit par fluorescence (FACS), les cellules de LMA qui survivent à la chimiothérapie. Finalement, dans le but d'identifier un profil d'expression génétique unique, j'ai appliqué un protocole de séquençage ARN de cellule unique sur ces cellules de LMA isolées. Ceci aurait le potentiel de mener à de nouveaux traitements qui exploitent un point faible de ces cellules. J'ai obtenu des librairies d'ADN complémentaire adéquates dans une fraction des cellules, démontrant que cette approche est réalisable. Cependant, une dégradation fréquente de l'ARN a empêché la génération de données d'expression génétique. Je propose des modifications au protocole dans le but de minimiser la dégradation de l'ARN.

Pour le deuxième projet, j'ai établi une séquence d'évolution clonale à partir d'une origine commune chez le cas rare d'un patient diagnostiqué successivement avec un syndrome myélodysplasique (SMD), une leucémie myélomonocytaire chronique (LMMC) et finalement une leucémie lymphoblastique aiguë à lymphocytes B (LLA-B). Pour ce faire, j'ai utilisé le FACS pour isoler ces différentes leucémies que j'ai ensuite soumises, ainsi qu'un frottis buccal, au séquençage de l'exome entier. J'ai identifié 11 mutations pilotes potentielles avec une distribution variée entre les différents clones. J'ai également identifié une variante dans la lignée germinale de ce patient affectant le gène suppresseur de tumeurs *CHEK2* (c.475T>C, p.Y159H). Cette variante étant auparavant considérée une variante de signification incertaine (VSI), j'ai exécuté un test de liaison protéique démontrant que cette variante entrave la liaison avec BRCA1 et représente possiblement

une prédisposition génétique au cancer. Ces trouvailles supportent la reclassification de *CHEK2* c.475T>C, p.Y159H comme étant probablement pathogène.

ACKNOWLEDGEMENTS

This work would not have been possible without the assistance of many individuals. I would like to thank my supervisors François Mercier and Sarit Assouline, as well as my thesis committee members John M. Storring, Anne Gatignol, and Jean-Sébastien Delisle for their guidance and support. Nathalie A. Johnson, Liliana Stoica, Ana Stirbu, Stephen Caplan, and the Jewish General Hospital Division of Hematology were indispensable in obtaining patient samples for these projects. I would especially like to thank Yoon Kow Young for his extensive teaching concerning all aspects of flow cytometry. I sincerely thank all of the following colleagues for providing technical assistance, intellectual contributions, fruitful discussions, and materials: Sylvain Gimmig, Emily Comyshyn, Yury Monczak, Zhen Shen, Kathleen Klein, Celia M. T. Greenwood, Kim Remans and the researchers at the European Molecular Biology Laboratory (EMBL), Zara Aryanpour**,** Gabriela Galicia-Vazquez, Samantha Worme, Maja Jankovic, Katharine Fooks, Wai Lam Poon, Alexandre Orthwein, John Heath, Estelle R. Simo-Cheyou, William D. Foulkes, Barbara Rivera Polo, and Anne-Sophie Chong. Importantly, I thank the Fonds de Recherche du Québec – Santé (FRQS) for providing me with the salary support that permitted this work. Additional funding for these projects was also provided by the Canadian Institute of Health Research (CIHR), the Cole Foundation, the Sir Mortimer B. Davis Foundation of the Jewish General Hospital, and the Richard and Edith Strauss Foundation.

I would not be the person I am today without the support of my family and friends. I sincerely thank them for all they have done for me throughout my years.

In addition, I thank all the patients and their families who consented to donate samples for research.

CONTRIBUTION OF AUTHORS

François E. Mercier provided supervision throughout the entirety of these projects.

Introduction: This chapter was written entirely by Alexandre Bazinet.

Project 1: Use of fluorescence-activated cell sorting and single-cell transcriptomics to identify vulnerabilities in acute myeloid leukemia in the minimal residual disease state

Literature Review: The literature review and writing of this chapter was performed entirely by Alexandre Bazinet.

Methods/Results: Nathalie A. Johnson, Liliana Stoica, and Ana Stirbu managed and processed the samples into the biobank. Sylvain Gimmig and Emily Comyshyn from BD Biosciences provided input into the design of the 17-color panel and reagent selection. Yoon Kow Young provided teaching and technical assistance concerning the BD FACSAria Fusion cell sorter. Kathleen Klein and Celia M. T. Greenwood helped design the statistical analysis applied to the FlowSOM free bioinformatic analysis. Katharine Fooks provided teaching and sample code for the generation of the Kaplan-Meier survival analysis and Cox proportional hazards model. Yury Monczak and Zhen Shen provided the raw FASTQ sequencing files for the myeloid gene panel. Wai Lam Poon provided teaching and technical assistance for the Target-Seq protocol. Maja Jankovic and Gabriela Galicia-Vazquez provided teaching and technical assistance regarding cell culture and bacterial culture. Kim Remans and the EMBL researchers provided the plasmids required for the generation of the Tn5 enzyme. Zara Aryanpour purified the Tn5 enzyme. Alexandre Bazinet performed all other work, including sample identification and selection, chart reviews, sample preparation for flow cytometry, operation of the cytometer for analysis and cell sorting, data analysis, coding for the bioinformatic analysis (NGS pipeline and FlowSOM), execution of the Target-Seq protocol, bacterial culture for Tn5 generation, writing, and generation of all figures and tables.

Discussion: This chapter was written entirely by Alexandre Bazinet.

Project 2: Common clonal origin of chronic myelomonocytic leukemia and B cell acute lymphoblastic leukemia in a patient with a germline *CHEK2* variant

This project has been published as a full-length article[1] in the journal Cold Spring Harbor Molecular Case Studies. Alexandre Bazinet is first author on this paper. The material presented in this thesis has been adapted and expanded from the published manuscript. The figures for project 2 were reproduced under the Creative Commons Attribution-NonCommercial 4.0 International License (CC-BY-NC), with no permission required given the nature of this license. These figures were first generated by Alexandre Bazinet for the publication of the original manuscript.

Literature Review: The literature review and writing of this chapter was performed entirely by Alexandre Bazinet.

Methods/Results: Nathalie A. Johnson, Liliana Stoica, and Ana Stirbu managed and processed the samples into the biobank. Yoon Kow Young provided teaching and technical assistance concerning the BD FACSAria Fusion cell sorter. Samantha Worme provided technical assistance for the DNA extraction. Anne-Sophie Chong, Barbara Rivera Polo, and William D. Foulkes provided guidance for the analysis of the whole-exome sequencing (WES) data. Anne-Sophie Chong performed the ExomeAI analysis. Alexandre Orthwein and Estelle R. Simo-Cheyou provided guidance, reagents, and technical expertise for the LacO/LacR assay. John Heath performed the transfection of the U2OS cells, confocal microscopy measurements, and statistical analysis for the CHK2-BRCA1 binding assay using the LacO/LacR system. Alexandre Bazinet performed all other work, including the chart review, sample preparation for flow cytometry, cell sorting, coding of the bioinformatic pipeline, analysis of the WES data, generation of the mutant mCherry-LacR-*CHEK2* constructs using site-directed mutagenesis/gateway cloning, confirmation of variants using Sanger sequencing, writing, and generation of all figures and tables with the exception of figure 15, panel C, which was generated by John Heath.

Discussion: This chapter was written entirely by Alexandre Bazinet.

Conclusion: This chapter was written entirely by Alexandre Bazinet.

INTRODUCTION

There is an increasing body of evidence suggesting that leukemias and related disorders exhibit considerable heterogeneity at the molecular level[2, 3, 4]. This concept applies not only between different patients, who may have leukemias exhibiting a variety of distinct driver mutations, but also within individual patients, where multiple clones with distinct functional and genetic properties may coexist. Understanding this heterogeneity is critical in the design of treatments able to eradicate subclones resistant to standard therapeutic approaches.

In recent years, major advances have been made in the field of multiparameter flow cytometry (MFC). Improvements in the selection of available fluorochromes and the performance of cytometers have allowed for an increasing quantity of parameters to be acquired simultaneously. This enables researchers to obtain very detailed immunophenotypes of cell populations of interest, which can then be purified using fluorescence-activated cell sorting (FACS). In parallel to the advances in MFC, increasing throughput and decreasing costs of next-generation sequencing (NGS) technologies now allow entire genomes/exomes to be sequenced rapidly and at relatively low cost. Furthermore, single-cell methodologies enable the detailed genomic and transcriptomic profiling of individual cells, allowing for the characterization of heterogeneous and/or rare populations. These flow cytometric and genomic methods are highly valuable and practical for the study of leukemias, where tumor cell suspensions can be readily obtained from the peripheral blood or bone marrow. In this thesis, I combined MFC and genetic methods with the goal of improving our understanding of the pathophysiology of different leukemias. Two projects are presented around this common theme.

My first project centers on acute myeloid leukemia (AML), an aggressive malignancy arising from hematopoietic stem and progenitor cells[5, 6]. In untreated AML, malignant myeloblasts rapidly expand, leading to death from bone marrow failure within weeks to months. Despite recent advances in AML therapeutics, the prognosis remains poor, and the majority of AML patients do not survive the disease[7]. Somewhat counterintuitively, a large proportion of AML patients are able to achieve complete remissions following induction chemotherapy[8]. Unfortunately, a small fraction of leukemic cells often survive treatment and subsequently expand, leading to relapse. This is the major cause of treatment failure and death in AML. With recent advances in flow cytometry and immunophenotyping, it is now possible to identify minute amounts of leukemia

cells in samples drawn from patients in complete remission. The identification of such cells is termed minimal/measurable residual disease (MRD) and is the focus of this project. These cells are of considerable interest as they are the putative cause of recurrent disease. A better understanding of why these cells selectively resist therapy could potentially lead to targeted treatments to eradicate them, with the potential to improve the cure rate in AML. The goal of this project was to identify, isolate and characterize AML minimal residual disease cells at the level of the transcriptome. Specifically, I studied paired diagnosis/remission samples of AML patients treated with intensive chemotherapy at the Jewish General Hospital. Using a custom-designed flow cytometry antibody panel, I identified and isolated AML MRD cells using fluorescence-activated cell sorting (FACS). I then attempted to perform single-cell RNA-sequencing on these cells to identify key up- or down-regulated cellular pathways. Adequate cDNA libraries were obtained in a small fraction of cells, suggesting this approach may be feasible. However, technical issues related to RNA degradation in single cells precluded the generation of the desired transcriptomic data. Protocol modifications, which are discussed in this thesis, may eventually allow for this powerful approach to link phenotypic and transcriptomic data.

My second project was inspired by the very rare case of a patient who had been sequentially diagnosed with myelodysplastic syndrome (MDS), chronic myelomonocytic leukemia (CMML), and B cell acute lymphoblastic leukemia (B-ALL). My goal was to clarify the genetic underpinnings of this complex case and to build a putative sequence of clonal evolution. I used FACS to isolate the concomitant CMML and B-ALL clones and processed the cell fractions using whole-exome sequencing (WES). I identified 11 potential driver mutations, including a possible inherited cancer predisposition in the *CHEK2* tumor suppressor gene. The effect of this *CHEK2* variant on protein function was not well established in the literature. I therefore performed a functional protein binding assay to confirm the negative effect of this variant on a downstream effector protein (BRCA1). The objective of this validation was to reclassify this *CHEK2* mutant as a likely pathogenic variant. I submitted my novel interpretation of this *CHEK2* variant on ClinVar and my findings were published in a peer-reviewed journal[1].

PROJECT 1:
USE OF FLUORESCENCE-ACTIVATED CELL SORTING AND SINGLE-CELL TRANSCRIPTOMICS TO IDENTIFY VULNERABILITIES IN ACUTE MYELOID LEUKEMIA IN THE MINIMAL RESIDUAL DISEASE STATE

LITERATURE REVIEW

Acute Myeloid Leukemia

Overview, epidemiology, and prognosis

Acute myeloid leukemia (AML) is an aggressive hematological malignancy. It is the most common form of acute leukemia affecting adults, with an incidence rate of 4.3 cases/100,000 population/year in the Unites States. AML is increasingly common in older individuals and the median age at diagnosis is 68 years[9]. Currently, the treatment of AML represents a substantial challenge in the field of malignant hematology. Standard therapeutic approaches consisting of intensive chemotherapy have changed relatively little since the 1970's and produce largely unsatisfactory outcomes. Most patients diagnosed with AML eventually die from the disease[7, 9]. In young, fit individuals, the development of relapsed and/or chemoresistant disease represents the major cause of treatment failure and death[10]. In elderly patients or those unfit for chemotherapy, treatment options have historically been limited. This group has long represented an unmet clinical need and has suffered from a dismal prognosis. Many of these patients do not receive any form of anti-leukemic therapy in a real-world setting[11]. Fortunately, recent years have seen improvement in low-intensity AML treatments and outcomes are likely to improve in older/unfit individuals[12].

Biology

Acute leukemias are characterized by uncontrolled proliferation and impaired differentiation of malignant blast cells[10]. It has become appreciated in recent years that many AML cases are preceded by a pre-leukemic state referred to as clonal hematopoiesis of indeterminate potential (CHIP). CHIP consists of the detection of cancer-associated mutations in the blood cells of apparently healthy individuals[13]. The three most common genes mutated in CHIP are *DNMT3A*, *TET2* and *ASXL1*, but many others have been described[14]. Three large studies performed whole-exome sequencing (WES) on the blood cells of many thousands of individuals unselected for hematologic phenotype and identified CHIP in over 5% of subjects over age 70[14, 15, 16]. This percentage continued to increase with advancing age. Many of the common CHIP mutations are considered weak cancer drivers and are insufficient to cause leukemia on their own. However, they promote clonal expansion or dominance within the affected pool of immature hematopoietic

cells[13]. This represents fertile ground for the acquisition of further driver mutations that may eventually lead to overt disease. Indeed, detectable CHIP is a strong risk factor for the subsequent development of hematological malignancies (hazard ratio 12.9) and also confers an increased risk of death (hazard ratio 1.4)[14]. Interestingly, CHIP has also been shown to be a risk factor for atherosclerotic cardiovascular disease, likely via aberrant macrophage-related inflammatory pathways[17].

Another concept of increasing importance in AML is that of the leukemia stem cell (LSC). The LSC hypothesis proposes that AML is organized hierarchically, with a pool of rare, self-renewing cells (LSCs) maintaining the bulk leukemic blasts, which are more mature and unable to propagate the leukemia on their own. The strict definition of an LSC requires demonstration of the cell's ability to propagate the leukemia in serial transplantation experiments in mice, as well as the capacity to give rise to non-LSC progeny (bulk leukemic blasts)[18]. Two landmark papers published in the 1990's identified LSCs from human AML samples using serial xenotransplantation in immunodeficient mice[19, 20]. Initially, LSCs appeared to be found exclusively in the very immature CD34+/CD38- compartment. More recent work has identified cells with LSC properties within the CD34+/CD38+ and even CD34- compartments[21, 22, 23]. However, it appears the frequency of LSCs is highest in the classical CD34+/CD38- compartment for most AML cases[23].

LSCs are characterized by proliferative and self-renewal properties. The cell of origin giving rise to the LSC is thought to be either a hematopoietic stem cell (HSC) or progenitor cell, possibly having already acquired a CHIP mutation. Non-malignant HSCs, which have naturally high self-renewal properties but low proliferative potential, could potentially be transformed into LSCs following the acquisition of mutations conferring increased proliferation. Conversely, normal progenitor cells, which are very proliferative but lack self-renewal capacity, could achieve LSC status following acquisition of self-renewal properties[5]. There is evidence to support both hypotheses regarding the cell of origin in AML[5, 6].

Since LSCs are responsible for AML initiation and relapse following chemotherapy, their elimination is a highly attractive clinical goal[18]. Unfortunately, LSCs exhibit cell cycle and metabolic properties that may confer resistance to standard chemotherapy. LSCs have been shown to be relatively quiescent compared to the bulk leukemic blasts, thus promoting resistance to cell-cycle specific (S-phase) agents such as cytarabine[24]. These quiescent cells display lower metabolic

activity and reactive oxygen species, which offers protection against chemotherapy-induced oxidative stress[25]. These characteristics may explain why relapse is frequent in AML. Indeed, the expression of LSC-associated genes in AML cells has been associated with adverse outcomes in large patient cohorts[26]. Of note, some studies have contradicted the notion of LSC chemoresistance by showing that cytarabine could eradicate LSC populations[27, 28].

Leukemogenesis in AML can thus be conceptualized as a multistep process evolving through a variety of stages (CHIP, LSC generation, overt disease). During this process, genetic and epigenetic lesions accumulate within the immature hematopoietic compartment[29]. Single-cell and bulk sequencing studies have demonstrated complex patterns of clonal evolution in AML, which can be further altered once the selective pressures of anti-leukemic therapy are introduced[30, 31]. At the genetic level, AML is a highly heterogeneous disease. A wide variety of recurrent driver mutations have been identified and affect various cellular processes[32]. These include mutations in transcription factors, signaling proteins, epigenetic modifiers, splicing factors, and tumor suppressors. Larger-scale chromosomal abnormalities (translocations, aneuploidies) and epigenetic changes are also well-described in AML[8, 32]. These diverse genetic drivers both define prognosis[2, 8] and can be used to guide treatment[12] in the era of molecular targeted therapy.

Once the leukemia is initiated, malignant AML blasts accumulate in the blood and bone marrow. These cells proliferate rapidly and are unable to differentiate into functional blood cells. This process disrupts normal progenitors and leads to bone marrow failure with associated blood cell count abnormalities (anemia, thrombocytopenia, neutropenia). Therefore, the AML patient becomes highly vulnerable to infectious and bleeding complications. Without effective treatment, the disease is usually fatal within weeks to months[33].

Diagnosis and Treatment

AML patients will typically present with manifestations of anemia (fatigue, dyspnea), thrombocytopenia (easy bruising, petechiae, hemorrhage) and/or neutropenia (infections, fever). Occasionally, AML may also present with extramedullary disease such as tissue masses (myeloid sarcoma) or infiltration of the skin (leukemia cutis) and other organs[34]. Peripheral blood cytopenias are usually accompanied by an elevated white blood cell (WBC) count with abnormal morphology

(blasts). The diagnosis is confirmed by demonstrating ≥ 20% myeloid blasts in the blood or bone marrow[35]. In addition to the morphologic blast count which establishes the diagnosis, the bone marrow aspirate is sent for key ancillary studies including flow cytometry, cytogenetics, and molecular testing. Historically, AML was categorized using the French-American-British (FAB)[36] system based on morphological features and cytochemical staining. Modern AML classification systems such as the World Health Organization 2016[35] classification rely heavily on cytogenetic abnormalities. In addition, genetic information derived from targeted next-generation sequencing (NGS) panels is increasingly being used to define new AML categories based on genetic drivers. The categorization of AML based on genetics and cytogenetics is very clinically relevant as it allows for the risk stratification of patients into three distinct prognostic groups (favorable, intermediate, adverse) that are used to guide the intensity of therapy[8]. Furthermore, some genetic drivers can now be targeted using small molecule inhibitors (*FLT3*, *IDH1*, *IDH2*)[12].

The standard treatment for AML patients fit for intensive chemotherapy consists of 7 days of infusional cytarabine and 3 days of the anthracycline daunorubicin ("7+3" regimen)[10]. This regimen successfully induces complete remission (CR) in 60 to 80% of younger patients (age < 60) and 40 to 60% of older individuals (age ≥ 60)[8]. Select patient categories may benefit from the addition of a third agent such as the FLT3 inhibitor midostaurin[37] or CD33-directed gemtuzumab ozogamicin[38, 39]. Following induction, consolidation treatment is required to prevent relapse. For favorable risk AML, this usually consists of further courses of intermediate- or high-dose cytarabine. Intermediate and adverse risk AML patients are usually referred for consideration of allogeneic hematopoietic stem cell transplantation (aHSCT) assuming they are fit for this procedure and a suitable source of stem cells can be identified. Elderly and/or unfit AML patients typically do not have good outcomes with such intensive approaches due to increased treatment toxicities and lower response rates. The combination of the hypomethylating agent azacitidine with the BCL-2 inhibitor venetoclax has recently emerged as the new gold standard treatment to prolong survival in this group, although it is not curative[40].

In all patient categories, disease that is refractory to chemotherapy or that relapses after initial remission is a major clinical problem. There is no standard therapy for these cases. Further intensive chemotherapy can be attempted but responses are rarely durable without consolidation with aHSCT[41]. Patients with specific driver mutations can be treated with small molecule

inhibitors such as gilteritinib (*FLT3*-mutated AML), ivosidenib (*IDH1*-mutated AML) and enasidenib (*IDH2*-mutated AML)[12]. A full discussion of the treatment of relapsed-refractory AML is beyond the scope of this literature review but excellent articles are available on this topic[41, 42]. A clinical trial, if available, often represents a good option for these patients.

Minimal/Measurable Residual Disease in Acute Myeloid Leukemia

Concept and Techniques

Despite the fact that most AML patients are able to achieve CR with intensive chemotherapy, relapse is common. This implies that a small subset of leukemic cells selectively resist therapy and eventually re-expand, leading to recurrence of the disease. The detection of a small quantity of residual AML cells in remission is referred to as minimal/measurable residual disease (MRD). Such cells often only represent 0.1 to 0.01% of the total bone marrow WBCs[43]. Therefore, a highly sensitive laboratory assay is required to detect MRD.

Currently, two major techniques are widely applied to measure MRD in AML: multiparameter flow cytometry (MFC) and real-time quantitative polymerase chain reaction (qPCR)[43]. Most NGS-based gene panels do not have sufficient sensitivity to be used for MRD. However, modifications such as error-corrected sequencing (ECS), aimed at improving sensitivity to the 0.01 to 0.001% variant allele frequency (VAF) level, have led to the emergence of NGS as another valid MRD method[44]. With the exception of qPCR-based monitoring for patients with very specific genetic lesions (*RUNX1-RUNX1T1*, *CBFB-MYH11*, *PML-RARA*, *NPM1*)[45, 46], AML MRD assays have yet to be fully standardized or universally adopted[43]. These assays all rely on identifying leukemia-specific features (genetic or cell surface proteins) that distinguish malignant cells from their normal counterparts. Each method has its own advantages and limitations.

The detection of AML MRD by MFC relies on the principle that leukemic blasts have aberrant expression of cell surface markers compared to normal myeloid progenitors. These immunophenotypic aberrancies can be grouped into three major categories[47]. The first group consists of myeloid markers that are either over- or under-expressed on AML cells compared to normal myeloblasts. Another group consists of asynchronous antigen expression, which is described as the expression on immature blasts of markers which are normally only expressed at a

later maturational stage. Finally, cross-lineage antigen expression refers to the aberrant expression of lymphoid markers on myeloid cells. The abnormal immunophenotype found on the AML blasts is termed leukemia-associated immunophenotype (LAIP) and is highly variable from patient to patient. A wide variety of LAIPs are described in the literature[48, 49]. There are two major approaches to analysis for MFC-based MRD evaluation in AML. The first consists of identifying the LAIP present at diagnosis and attempting to track this same population in remission (LAIP-based approach)[43]. The major disadvantage of this method is that it may yield a false negative result if the LAIP changes between diagnosis and remission. This phenomenon, referred to as immunophenotypic shift, is well described in AML[50]. In addition, a sample drawn at the time of diagnosis must be analyzed with the same antibody panel that will be used for the MRD assay to establish the baseline LAIP. In contrast to the LAIP-based approach, the different-from-normal (DfN) approach seeks to identify aberrant cells forming a distinct population compared to the normal myeloid blasts. Effectively, the normal progenitors serve as internal controls from sample to sample and leukemic cells are identified based on their deviation from the normal pattern[51]. The DfN and LAIP-based approaches are increasingly being perceived as complementary rather than mutually exclusive[43]. Once a population is identified as being suspicious for MRD, it is quantified as a percentage of CD45+ WBCs in the bone marrow. The recommended threshold for MRD positivity in AML is 0.1%[43], although residual leukemia can be detected at the 0.01% level if enough cells (1 million) are acquired[51]. The interpretation of MFC-based AML MRD is complex and requires extensive knowledge of normal myeloid maturation patterns. There is also a degree of subjectivity in the identification of the abnormal population[51]. The assay is only available in large medical centers with access to expert operators. A major advantage of MFC for MRD is that this approach is applicable to over 90% of AML patients, irrespective of mutational status[52].

Molecular MRD methods include qPCR and NGS (ECS), which rely on the longitudinal tracking of AML driver mutations between diagnostic and remission samples. These techniques are highly sensitive and prone to less subjectivity. The major challenge in this case is the identification of mutations suitable for MRD tracking, which is not applicable to all AML patients. Common CHIP mutations such as *DNMT3A*, *TET2* and *ASXL1* are known to persist in pre-leukemic clones during remission and do not necessarily represent residual leukemia[53]. Such mutations are unsuitable for MRD tracking. On the other hand, mutations such as *NPM1*, *FLT3*, *NRAS/KRAS* and *KIT* are expected to be cleared with successful treatment[54]. Their persistence

strongly suggests MRD is present. Unfortunately, for many mutations, suitability for MRD is not yet known. For AML cases driven by specific chromosomal translocations (*RUNX1-RUNX1T1*, *CBFB-MYH11*, *PML-RARA*) and *NPM1* mutations, qPCR is widely used to monitor for MRD[43, 45, 46].

Applications

The most well-established clinical application for MRD in AML is prognostication. Numerous studies have demonstrated that detectable MRD is associated in multivariate analyses with an increased risk of relapse and, in many cases, lower overall survival (OS) depending on the method used and AML subtype[43, 46, 55, 56, 57, 58, 59, 60]. This includes MRD assessed using a variety of methods (MFC, qPCR and NGS) and at various time points (after induction chemotherapy, at the end of treatment, in follow-up and pre-transplant). Importantly, it appears that this increased risk cannot be entirely mitigated with aHSCT[61].

There is considerable interest in using MRD to tailor the treatment of AML patients, and this represents a very active field of research[62]. Theoretically, this could allow for de-escalation of therapy in MRD-negative patients (who are known to have good outcomes) with the purpose of minimizing treatment-related toxicity. Conversely, MRD-positive patients, who have poor outcomes, could potentially benefit from intensification of treatment with the goal of eradicating MRD prior to overt relapse. The GIMEMA AML1310 trial explored such a strategy by assigning different transplantation strategies (autologous versus allogeneic) based on MRD status in intermediate risk (IR) AML patients in first remission[63]. Using this approach, the IR patients had disease-free survival (DFS) comparable to favorable risk patients, with no noted difference in OS and DFS between the MRD-positive and MRD-negative IR patients. The AML05 study used MRD to stratify AML patients with t(8;21) into high and low risk groups[64]. The high risk (MRD-positive) group was assigned to a more intensive consolidation regimen (aHSCT) and this appeared to reduce the risk of relapse and improve survival. In the RELAZA2 study, preemptive azacitidine administration was able to delay/prevent relapse in MRD-positive AML patients in remission[65]. Although MRD-directed therapy is still considered experimental, these studies indicate the strategy holds promise in the treatment of AML. Many immunotherapies are currently being investigated for their potential to eradicate MRD[66].

From a basic/translational research point of view, the AML cells responsible for MRD persistence are of great interest. A better understanding of why these cells selectively resist treatment and lead to relapse has the potential to lead to powerful new therapeutics.

Mechanisms of Treatment Resistance in Acute Myeloid Leukemia

It is not entirely understood why a subset of AML cells are able to survive chemotherapy and subsequently lead to relapse. At the genetic level, it has been shown that the founding clone or subclones may acquire additional mutations and expand at relapse[31]. In addition, a growing body of evidence supports the concept that a cell population with LSC-like properties already exists at diagnosis, is selected for by cytotoxic chemotherapy, and subsequently leads to relapse[67, 68]. As previously discussed in this review, LSC's appear to possess properties that make them relatively chemoresistant. While a considerable amount of work has been done to study AML in the relapsed state, relatively few studies have attempted to characterize the MRD state. This is likely related to the challenge in identifying and isolating this rare cell population. Nonetheless, promising work has been done in this field. A study using a patient-derived xenograft (PDX) model revealed that the AML cells responsible for regenerating the leukemia following cytarabine exposure have a distinct gene expression signature enriched for G-protein-coupled receptor signaling[28]. A similar PDX-based study showed that chemoresistant AML cells upregulate genes involved in oxidative phosphorylation and fatty acid metabolism, as well as the cell surface marker CD36 (fatty acid translocase)[27]. These features appeared to confer resistance to cytarabine. Taken together, these studies suggest that residual AML cells surviving chemotherapy exhibit unique functional and metabolic changes. These features represent potential vulnerabilities that could be targeted therapeutically. As proof of principle, an inhibitor of *DRD2*, a gene found to be upregulated in chemoresistant AML cells, was shown to be able to delay relapse in treated mice[28]. This drug has since translated to phase I clinical trials in humans[69].

In human AML patients, bulk RNA-sequencing performed on paired diagnostic and sorted MRD samples identified specific gene expression changes following chemotherapy[70]. This included upregulation of genes related to Myc, E2F, cell cycle checkpoints, DNA repair and oxidative phosphorylation. In addition, immune-related genes were downregulated in LSCs, potentially promoting immune evasion. Recently, a senescence-like phenotype has been described

in chemoresistant AML cells following treatment[71]. These pathways deserve further study as potential mechanisms of resistance to treatment and represent promising therapeutic targets.

Single-Cell RNA Sequencing

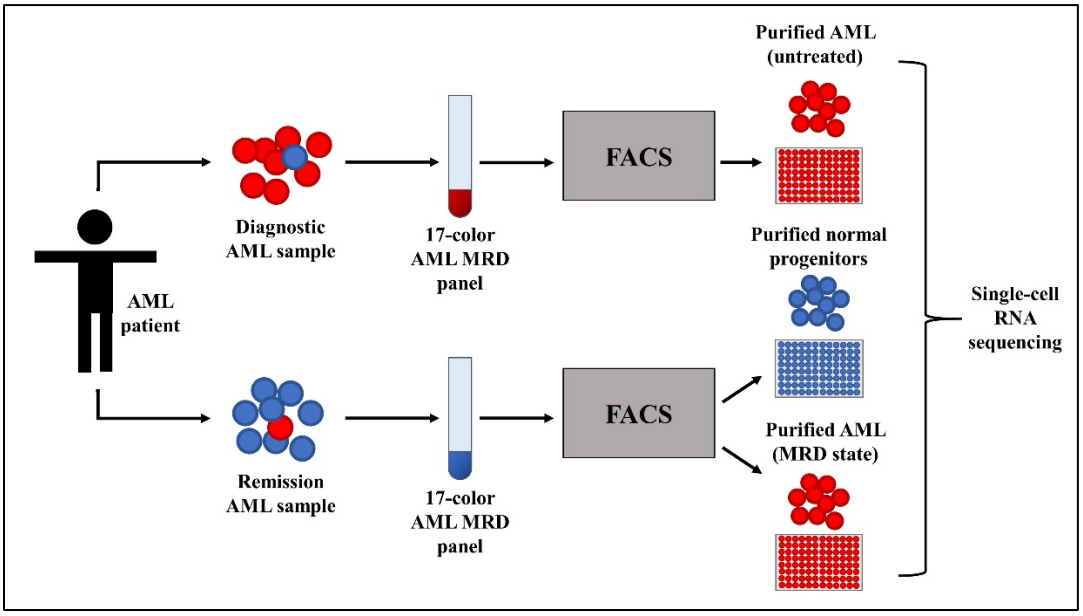
Single-cell RNA sequencing (scRNA-seq) represents an ideal tool to study drug resistance in cancer. Whereas traditional bulk RNA-sequencing technologies will provide only the average gene expression profile for a tissue, single-cell technologies allow for the transcriptomic characterization of heterogeneous and/or rare cell types[72]. A variety of scRNA-seq protocols have been developed over the past decade. These exhibit considerable variation in terms of single-cell isolation methods, library preparation, throughput and read length generated (3'/5'-biased versus full-length)[72]. Microfluidics-based methods such as the Chromium 10X protocol allow for a high throughput (thousands of cells) at relatively low cost per cell[73]. For very rare cells (such as MRD cells), it is desirable to purify the target population with fluorescence-activated cell sorting (FACS) into microwell plates prior to proceeding with scRNA-seq. The plate-based Smart-Seq2[74] approach also has the advantage of generating full-length reads (as opposed to 3'-biased reads with the Chromium 10X approach). With certain protocol adjustments, Smart-Seq2 can allow for the concomitant generation of somatic variant and transcriptomic data. These protocol modifications consist of adding mutation-specific primers at key steps in the protocol to selectively amplify regions of interest and enable sensitive detection of somatic variants[75]. This modified protocol is referred to as Target-Seq[75] and represents a powerful tool to study drug resistance in cancer cells. As an example, this method has allowed for the identification of a rare population of therapy-resistant stem cells in chronic myeloid leukemia (CML) displaying upregulation of genes associated with quiescence[3]. To the best of our knowledge, application of Target-Seq to FACS-sorted AML MRD cells has not been performed. Such an assay has the potential to reveal heterogeneity and targetable vulnerabilities within the MRD compartment.

METHODS

Overview

In this project, we obtained primary AML samples from patients drawn at diagnosis and in remission. Using a custom 17-color AML MRD flow cytometry panel, we isolated untreated AML cells, residual AML cells persisting in remission (MRD), and normal progenitors in remission. These three key populations were then submitted to scRNA-seq with the goal of identifying differentially expressed gene pathways between the populations. An overview of the experimental design is shown in **Figure 1**.

Figure 1: Experimental design overview



Patient samples

We obtained bone marrow (BM) samples from patients diagnosed with AML at the Montreal Jewish General Hospital (JGH) between June 2014 and September 2020. Acute promyelocytic leukemia (APL) cases were excluded. These samples were drawn after the patient had achieved CR following intensive chemotherapy. When possible, a paired diagnostic sample was also studied. Samples to be used as normal references were obtained from non-Hodgkin's lymphoma patients undergoing staging pre-therapy (and confirmed by pathology to have

unaffected BM). All samples were purified using Ficoll gradient centrifugation and cryopreserved in a biobank at -80°C until recovery for the study. All patients provided written, informed consent for the collection of their samples for research. This project was approved by the local Research Ethics Board (protocol 11-047 for the collection of samples, protocol 2020-1925 for use of these samples to study MRD).

Design and optimization of a 17-color, single-tube, flow cytometry panel for AML MRD

The European LeukemiaNet (ELN) MRD Working Party published guidelines in 2018 in an attempt to harmonize AML MRD measurement[43]. We designed a custom 17-color AML MRD panel based on the cell surface markers recommended in this guideline document by 7 major European and North American MRD centers. This included markers from all 3 categories of immunophenotypic aberrancies described in AML[47] (under/overexpression, asynchronous, cross-lineage). When possible, we selected well-validated (EuroFlow) antibody-fluorochrome combinations[76]. This panel was designed to be compatible with a BD FACSAria Fusion cell sorter. We favored a single-tube design to maximize the number of events acquired and sensitivity of the panel. The final panel design is shown in **Table 1**. We carefully tested and titrated all antibodies on single-stained normal BM samples. For each antibody-fluorochrome, we selected the titration with the highest separation index (SI)[77] (**Appendix Figure A1**).

Table 1: 17-color AML MRD flow cytometry panel

Laser	Fluorochrome	Filter	Marker	Clone
Ultraviolet 375nm	BUV395	379/28	CD11b	D12
	BUV496	515/30	CD14	MφP9
	BUV737	740/35	CD123	9F5
Violet 405nm	V450	450/40	HLA-DR	L243
	V500	525/50	CD45	2D1
	BV605	610/20	CD19	SJ25C1
	BV650	660/20	CD56	NCAM16.2
	BV711	710/50	CD2	RPA-2.10
	BV786	780/60	CD64	10.1
Blue 488nm	FITC	530/30	CD15	MMA
	BB700	695/40	CD34	8G12
Yellow Green 561nm	PE	582/15	CD13	L138
	PE-CF594	610/20	CD7	M-T701
	PE-Cy7	780/60	CD117	104D2
Red 640nm	APC	670/30	CD33	P67.6
	APC-R700	730/45	LIVE/DEAD	
	APC-H7	780/60	CD38	HB7

Flow cytometry

Patient samples were thawed in a 37°C water bath and washed once with phosphate-buffered solution (PBS). To exclude dead cells from analysis, we stained the samples with BD Horizon Fixable Viability Stain 700 (1:100 000, cat. #564997) for 10 minutes, followed by washing with PBS + 2% fetal bovine serum (FBS). We then added 5uL of BD Fc Block (cat. #564220) to minimize non-specific antibody binding (10-minute incubation). The samples were stained with the 17-color antibody cocktail in BD Brilliant Stain Buffer (cat. #563794) for 45 minutes. This buffer is required when multiple Brilliant Violet dyes are combined within the same tube. Finally, the sample was washed with PBS +2% FBS and resuspended in 500uL prior to proceeding to the cytometer. All handling and incubation steps were performed on ice and protected from light. Vortex was avoided to minimize cellular stress. We used UltraComp eBeads (Invitrogen, cat. #01-2222-42) compensation beads to prepare fresh compensation controls on each day the cytometer was used. To minimize day-to-day instrument variation in mean fluorescence

intensity (MFI), we used BD 8-peak Rainbow Calibration Beads (cat. #559123) to adjust the voltages to obtain a consistent MFI in each channel prior to every experiment.

The data was acquired on a BD FACSAria Fusion cell sorter. We used BD FACSDiva (version 8.0.2) to calculate the compensation matrix. We aimed to acquire a minimum of 500 000 to 1 000 000 leukocyte events per sample as recommended[43]. When diagnostic and remission samples were evaluated on the same day, the remission samples were always acquired first to eliminate any risk of contamination from the gross leukemic samples. A line backflush was performed for 30 seconds between each sample.

Generation of a normal BM reference file

We evaluated 10 normal BM samples with the 17-color MRD panel. These samples serve as a valuable reference for later comparison with the diseased samples. We performed a quality control on these flow cytometry standard (FCS) files using the flowAI[78] package in the R programming language[79]. This package detects and removes low-quality events based on flow rate instability, signal acquisition instability and dynamic range margin/outlier status[78]. Using the flowCore[80] package, we then randomly selected 100 000 high-quality events from each normal bone marrow file to create a merged normal bone marrow (MNBM)[81] master reference FCS file (1 000 000 leukocytes) for use in later analyses. Debris, doublets, and dead cells are excluded from this selection using flowCore tools.

Flow cytometry data analysis

Standard gating approach for AML MRD

We first applied a standard gating approach using Infinicyt software (version 1.7) to identify AML MRD in the remission samples. This approach first consists of identifying the immature myeloid compartment ($CD45_{dim}$, SSC_{low}, CD34+, CD19-), followed by careful examination for immunophenotypic aberrancies in relation to the normal progenitors (DfN approach)[47, 51, 52]. When a paired diagnostic sample was available, we also attempted to track the diagnostic LAIP within the remission sample (LAIP-based approach). We thus used a combination

of the DfN and LAIP approaches as recommended by the ELN MRD Working Party[43]. A remission sample was considered positive for MRD if it contained an aberrant population suspicious for residual AML consisting of ≥0.01% of the CD45+ events with a minimum of 20 cells[47, 52]. There is evidence to suggest MRD detected at this level is associated with worse outcomes[55, 61]. An example of the gating strategy to identify the immature myeloid compartment is shown in **Appendix Figure A2**.

Bioinformatic approach for AML MRD

We explored two major bioinformatic approaches to analyze our flow cytometry data. The first method was to leverage the principal component analysis (PCA) tool built into Infinicyt. Alternatively, we applied the FlowSOM[82] R package with the standard settings except for a larger grid size (15 x 15, 225 nodes) to facilitate the identification of rare populations. For the FlowSOM frozen[81] method, the minimal spanning tree (MST) was generated using the MNBM reference FCS file. The cellular identity of each node was determined by examining its contents in traditional 2D dot plots and comparing the immunophenotype with that of known BM populations[83]. Patient sample FCS files were then mapped back onto the "fixed" MST generated from the MNBM file as previously described[81]. For the FlowSOM free method[81], the MST was individually generated for each patient case using the MNBM, diagnosis, and remission FCS files. In this case, the cellular identity of each node was established by identifying its closest normal counterpart in the FlowSOM frozen MNBM MST using Euclidean distance. To complete the FlowSOM free protocol, each file (MNBM, diagnosis, remission) was mapped back onto the MST for the patient in question to establish a portrait of disease evolution over time in relation to a normal reference[81].

To evaluate for statistical significance, we used a combination of resampling and bootstrapping methods in R to attribute a p-value to each node. Briefly, the 10 normal BM samples were used to gauge variability in normal samples. One of the 10 normal samples was randomly selected to represent a diagnostic "leukemic" sample and one was selected to represent a "remission" sample. One million events were randomly selected from the "leukemic", "remission" and 8 remaining normal BM samples. These events were then used to run the FlowSOM free protocol and generate a matrix containing the event count for each node by sample. This process was repeated to generate 500 bootstrapped matrices. When FlowSOM free was then applied to

actual patient remission samples, each node was queried for the proportion of bootstrapped matrices containing at least one node with equal or more extreme variability compared to the MNBM sample. This proportion represents the p-value for each node.

Patient survival analysis

We assessed differences in survival between the MRD-positive and MRD-negative AML patients using the Kaplan-Meier estimate[84]. For the univariate analysis, statistical significance was determined using the log-rank test[84]. We also performed a multivariate survival analysis with a Cox proportional hazards model[84] considering MRD status, age, and ELN risk category as explanatory variables.

Identification of AML driver mutations using NGS

Our clinical molecular laboratory has access to a myeloid NGS panel targeting 35 genes recurrently mutated in hematological malignancies (**Appendix Table A1**). This panel covers the full-length sequence of each gene. Each genomic sequence is amplified and sequenced (paired-end reads) on an Illumina instrument (MiSeq). An average coverage of 100X is obtained with this method (minimum coverage 50X, 30X for CEPBA). This panel is routinely performed for AML patients at diagnosis in our institution.

We obtained the raw sequencing files (FASTQ) generated by this panel from the clinical laboratory. For analysis, we created a custom bioinformatics pipeline based on the Broad Institute GATK best practices workflow for somatic short variant discovery[85]. Adaptor and duplicate sequences were removed. BWA-MEM[86] was used to align the reads to the reference genome (hg19). Variants were called using Mutect2[87] with a panel of normals (GATK, Broad Institute, Cambridge, MA). The calls were then filtered using FilterMutectCalls (GATK, Broad Institute, Cambridge, MA). We used all built-in filters with the exception of the germline and clustered events filters. We chose to ignore these two filters because we found the germline filter was excluding CHIP variants present at relatively high frequency in the general population and the clustered events filter was excluding clearly pathogenic mutations. We used ANNOVAR[88] to

annotate the filtered variants. This list was further refined by excluding synonymous single-nucleotide variants, non-coding variants, variants known to be benign in the ClinVar database, and variants present in the general population at a rate greater than 0.01% (as per the 1000G, ExAC, ESP6500si, and gnomAD databases). We only retained variants present above a 5% variant allele frequency (VAF) cutoff. At this stage, a final curation step was performed. A variant was only retained if it could be confidently expected to be a leukemic driver in AML. This could consist of a gain-of-function variant that is well described in the literature. As an example, *NRAS* G12 or *IDH1* R132 variants would fall into this category. Alternatively, a damaging variant in a gene where loss of function is associated with leukemia would also be retained. A truncating *TET2* or *ASXL1* mutation would meet this criterion. A variant was defined as damaging if it resulted in protein truncation (nonsense, frameshift) or if it was predicted as being pathogenic by a majority of 6 *in silico* tools (SIFT, Polyphen2, FATHMM, PROVEAN, M-CAP, CADD) as per previously established methodology[89].

Cell sorting and scRNA-seq

For scRNA-seq, we used the Target-Seq protocol[75]. Three cell populations of interest (AML blasts at diagnosis, AML blasts in the MRD state, and normal remission myeloid blasts) were identified using the AML MRD flow cytometry panel. These were sorted using a BD FACSAria Fusion cell sorter into 96-well plates containing the cell lysis solution. The plates were incubated 5 minutes at room temperature for protease digestion, briefly centrifuged, and frozen at -80°C for short-term storage.

For the next phase of the protocol, the plates were thawed and incubated at 72°C for 15 minutes to inactivate the protease. We then performed reverse transcription (RT) using SMARTScribe reverse transcriptase (Clontech, cat. #639538). This was followed by a polymerase chain reaction (PCR) step (24 cycles) using SeqAMP DNA polymerase (Clontech, cat. #638509). RNAse inhibitor was added at both the protease digestion and RT steps to minimize RNA degradation. All steps prior to PCR were performed in a dedicated pre-amplification hood treated to eliminate environmental DNA and RNAses.

Target-Seq requires mutation specific primers to improve the sensitivity of somatic mutation detection in parallel to single-cell transcriptomic data. These primers amplify specific loci from both gDNA and mRNA. The gDNA primers were designed to bind within the introns upstream and downstream from the exon containing the mutation of interest. The mRNA primers were designed to bind the exons preceding and following the mutation. These mutation-specific primers were added at both the RT and PCR steps of Target-Seq[75].

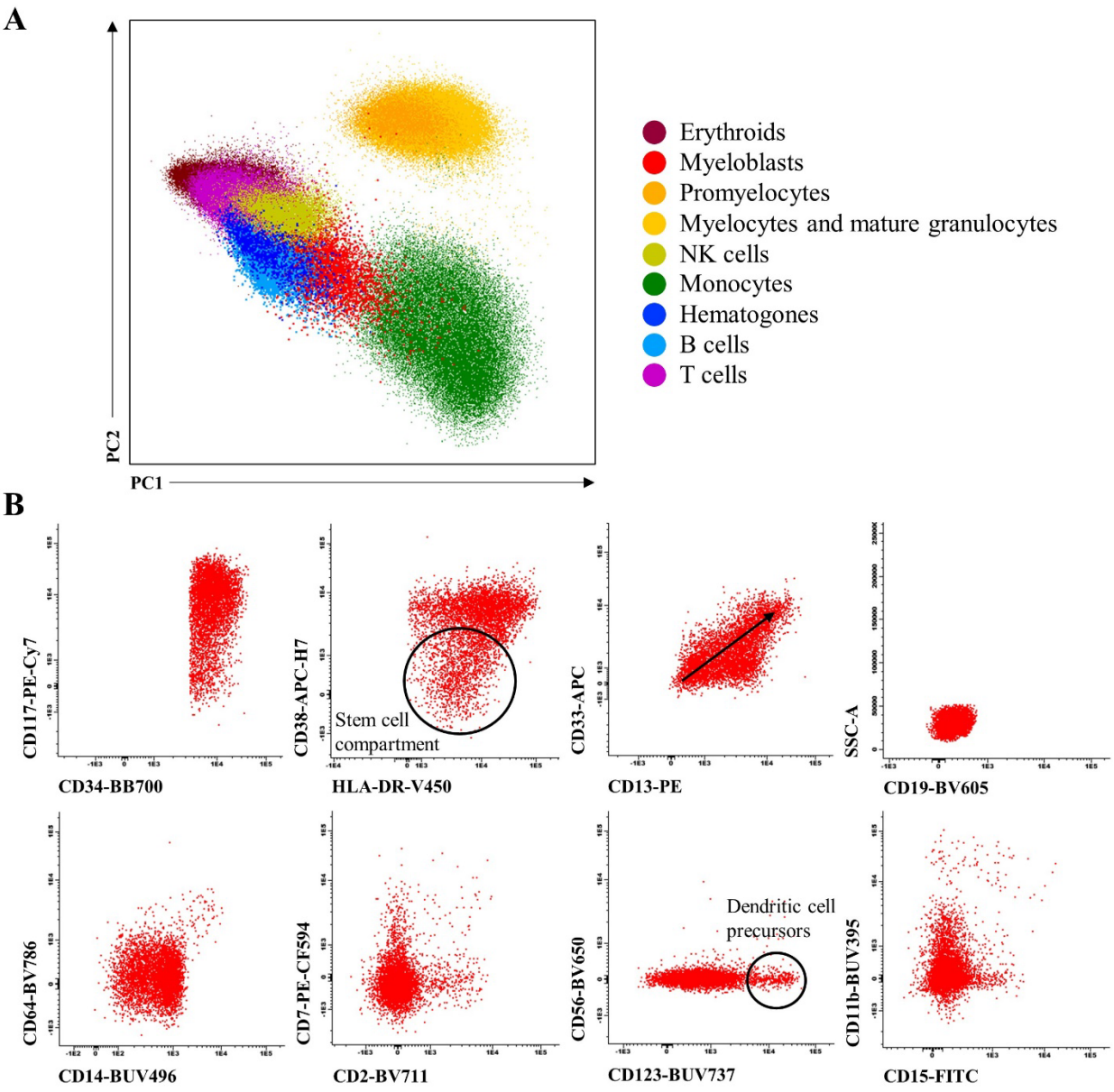
Following RT and PCR amplification, we purified the cDNA/amplicon mix using Ampure XP Beads (Beckman Coulter, cat. #A63881) as per the manufacturer recommendations. The quality of the cDNA libraries was assessed using a Bioanalyzer DNA High Sensitivity chip (Agilent). We quantified the cDNA libraries using the AccuClear Ultra High Sensitivity dsDNA Quantification Kit (Biotium, cat. #31028). Tagmentation was performed using Tn5 transposase (prepared in house) to generate sequencing-ready whole transcriptome libraries. This Tn5 transposase was prepared using a pETM11-Sumo3-Tn5 construct, a generous gift from Dr. Kim Remans and the researchers at the European Molecular Biology Laboratory (EMBL)[90]. Tagmentation was followed by PCR amplification, bead-purification, and sequencing using an Illumina platform. We planned to analyze this data using the Seurat[91] R package for dimensionality reduction and clustering, followed by gene set enrichment analysis (GSEA)[92] to identify up or down-regulated pathways.

RESULTS

Application of the 17-color AML MRD panel to normal BM samples

We first evaluated 10 normal BM samples stained with the 17-color AML MRD panel. We analyzed these FCS files in Infinicyt and found our panel could reliably identify most major BM cell lineages based on well-described immunophenotypic patterns[83]. These populations included myeloblasts (CD34+, CD19-, CD117+), hematogones (CD34+, CD19+), promyelocytes (SSC_{high}, CD33+, CD15+, CD11b-), myelocytes to mature granulocytes (SSC_{high}, CD33+, CD15+, CD11b+), monocytes (CD14+, CD64+, $CD33_{bright}$), B cells (CD19+), T cells (CD2+ and/or CD7+), NK cells (CD56+), and erythroid precursors (CD45-, CD19-). These immunophenotypes are simplified in this text for concision. All 17 markers were considered during the gating strategy. Using the PCA function in Infinicyt, we could visualize all these key populations on a single 2-dimensional plot (**Figure 2, panel A**). When gating on the immature progenitor compartment ($CD45_{dim}$, SSC_{low}, CD34+), we obtained a detailed immunophenotype of normal myeloblasts (CD19-) and hematogones (CD19+) (**Figure 2, panel B**). Normal myeloblasts expressed strong CD117 and HLA-DR and were negative for CD2, CD7, CD11b, CD14, CD15, CD56, and CD64. When CD13 and CD33 were plotted against each other, a characteristic diagonal pattern was observed as described in the literature[52]. Interestingly, we also identified the rare CD34+/CD38- stem cell compartment[83] and a CD34+/$CD123_{bright}$ population, likely representing common dendritic cell progenitors[93] (**Figure 2, panel B**). Collectively, these data served as a valuable reference when the panel was later applied to diseased samples.

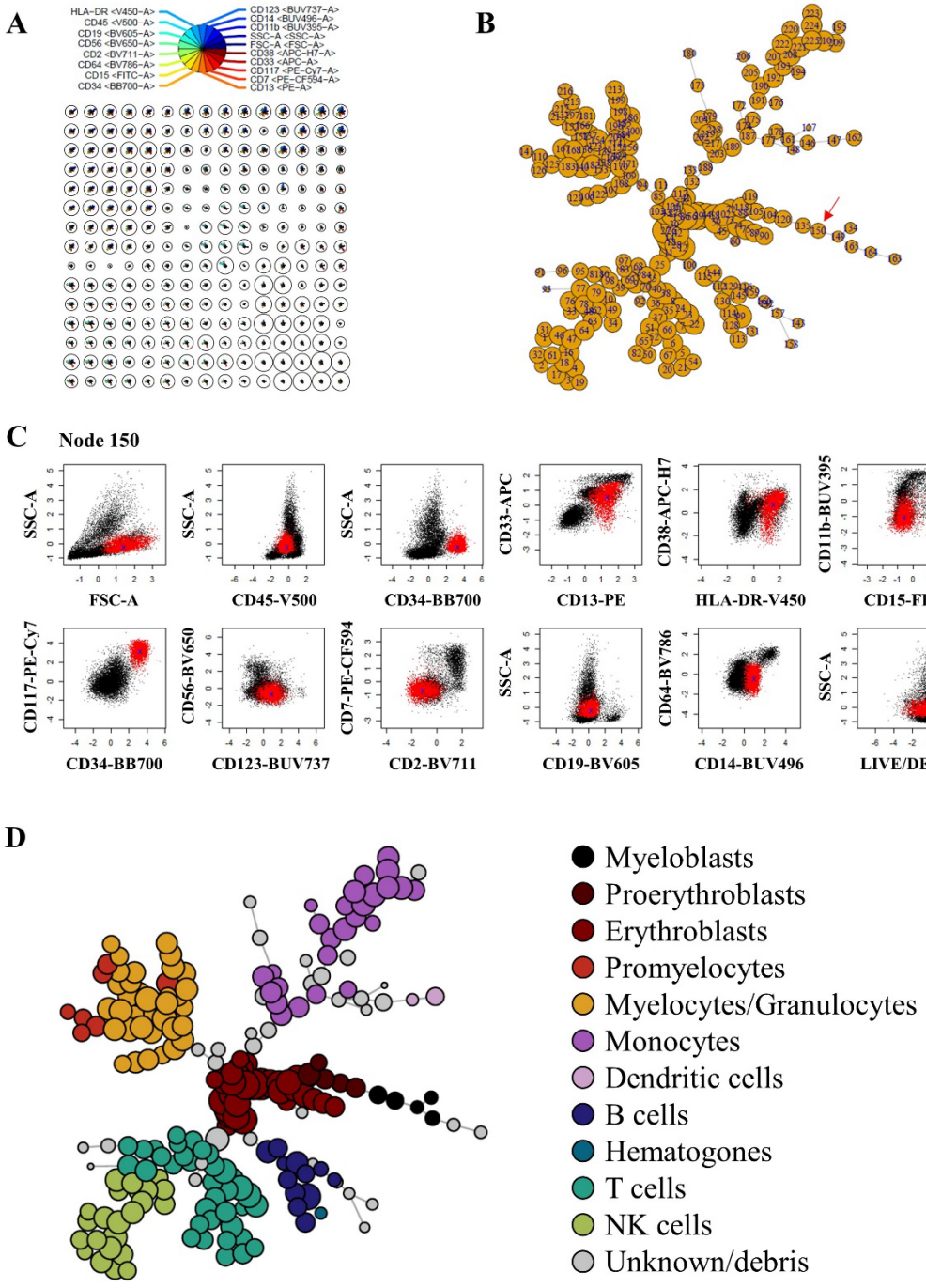
Figure 2: 17-color AML MRD panel applied to a normal bone marrow sample



A, Normal bone marrow hematopoietic compartments identified with the 17-color panel. This plot represents a principal component analysis in Infinicyt software. This dimensionality reduction tool that can automatically cluster cell populations considering all 17 markers simultaneously on a single plot. **B**, Detailed immunophenotype of the myeloblast compartment ($CD45_{dim}$, SSC_{low}, CD34+, CD19-). A typical diagonal pattern was observed when plotting CD13 versus CD33 (black arrow). The myeloblast compartment also includes dendritic cell precursors ($CD123_{bright}$), and the very immature stem cell compartment (CD34+/CD38-).

The analytical complexity of multiparameter flow cytometry data was recognized as early as the 1980s[94]. With increasing parameters, it becomes increasingly difficult to fully evaluate the data using traditional 2D gating dot plots. For this reason, we explored if the bioinformatic package FlowSOM[82] could facilitate interpretation of our data. We first used the flowCore[80] R package to randomly select 100 000 high-quality events from each normal BM sample to generate the merged normal bone marrow (MNBM) reference (1 million cells). We then analyzed the MNBM file using FlowSOM. This algorithm first performs clustering using a grid of nodes (**Figure 3, panel A**), where each node is situated in proximity to other similar nodes based on marker expression. This grid is initialized using randomly selected events from the dataset and updated as each new event is assigned to the node it resembles most. Finally, these nodes are organized into a minimal spanning tree (MST)[82], a branching structure in which the sum of the distances between the nodes is minimized (**Figure 3, panel B**). Each node can then be queried, and its contents visualized on standard 2D dot plots to identify cellular identity (**Figure 3, panel C**). The end result is a detailed graphical representation of the different cell populations present in the MNBM sample (**Figure 3, panel D**). All cell surface markers and light scatter parameters (FSC-A, SSC-A) were used for this clustering.

Figure 3: FlowSOM analysis applied to the merged normal bone marrow (MNBM) file



A, Events in the MNBM file are clustered into 225 nodes (15 x 15) grid. Each node is in proximity to other similar nodes. **B**, The minimal spanning tree (MST) is built to minimize the total weight of the connections between the nodes. This results in similar nodes being grouped together in the tree. **C**, Each node is interrogated on standard 2D dot plots to identify cellular identity. In this example, node #150 (red arrow in panel **B**) is shown to contain myeloblasts (red population: $CD45_{dim}$, CD34+, CD19-, CD117+). **D**, The MST is recolored based on cellular identity.

Application of the 17-color AML MRD panel to diseased BM samples

For MRD evaluation, we identified 22 samples in our biobank drawn from AML patients after achieving CR with intensive chemotherapy. For 14 of these patients, a paired diagnostic sample was also available. The characteristics of this patient cohort are detailed in **Table 2**.

Table 2: AML patient cohort characteristics

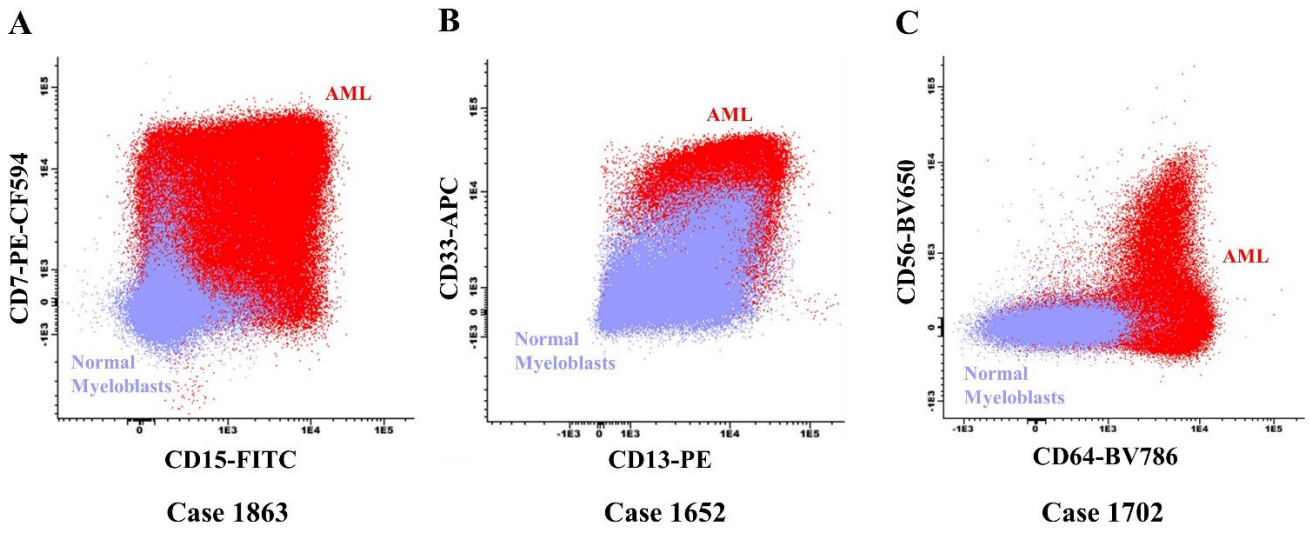
Patient	Age	Sex	FAB	Karyotype	Molecular	ELN risk category	Diagnostic sample
344	52	F	n/a	normal	FLT3-/NPM1-	Int	no
767	62	F	M1	t(3;10), der(10)	n/a	Int	yes
819	69	F	M1	normal	FLT3-/NPM1+/CEPBA+	Fav	yes
1024	59	M	M4	inv(16)	FLT3-/NPM1-	Fav	no
1395	65	M	M1	+8, +13	FLT3-/NPM1-/CEPBA-	Int	yes
1533	71	M	M2	-7p	FLT3-/NPM1-/CEPBA+	Fav	no
1550	24	M	M2	normal	FLT3+/NPM1-/CEPBA-	Adv	no
1551	66	M	M2	complex	FLT3-/NPM1-/CEPBA-	Adv	no
1582	53	M	M2	+8	FLT3-/NPM1-/CEPBA-	Int	yes
1602	54	M	M5	complex	FLT3+/NPM1-/CEPBA-	Adv	yes
1652	44	F	M2	t(6;9)	FLT3-/NPM1-/CEPBA-	Adv	yes
1683	39	F	M4	normal	FLT3-/NPM1+/CEPBA-	Fav	yes
1702	50	F	M5	normal	FLT3+/NPM1+/CEPBA-	Int	yes
1763	66	F	M2	t(8;21), -X	FLT3-/NPM1-/CEPBA-	Fav	no
1821	22	M	M5	complex	FLT3-/NPM1-/CEPBA-	Adv	yes
1863	18	M	M1	normal	FLT3+/NPM1-/CEPBA-	Adv	yes
1887	58	M	M4	inv(16)	FLT3-/NPM1-/CEPBA-	Fav	yes
1974	30	F	M2	normal	FLT3-/NPM1+/CEPBA-	Fav	no
1989	61	F	n/a	complex	FLT3-/NPM1-/CEPBA-	Adv	yes
2006	62	M	M1	+8	FLT3-/NPM1-/CEPBA-	Int	no
2061	55	M	M1	normal	FLT3-/NPM1+/CEPBA-	Fav	yes
2143	65	M	M5	normal	FLT3-/NPM1+/CEPBA-	Fav	yes

Our cohort was representative of the diversity seen in AML patients. The median age of the cohort was 56.5 years (range 18-71) with 59% males and 41% females. We had cases

representative of four different FAB subtypes (6 M1, 7 M2, 3 M4, 4 M5). No cases of M0, M6 or M7 could be obtained, likely due to their rarity. We specifically excluded M3 (APL) cases given well-established molecular MRD methods (*PML-RARA* qPCR) and lack of relevance of MFC-based MRD in these cases. The cases exhibited a wide variety of karyotypic and molecular abnormalities (**Table 2**). Following ELN 2017 risk categories[8], we had 9 (41%) favorable, 6 (27%) intermediate, and 7 (32%) adverse risk patients. Since we only selected patients treated with intensive chemotherapy and successfully achieving CR, our cohort was biased towards younger and lower-risk patients compared to what would be expected in the AML population in general.

We first validated that our 17-color MRD panel could reliably identify various aberrant immunophenotypes in diagnostic AML samples. Representative examples are shown in **Figure 4**.

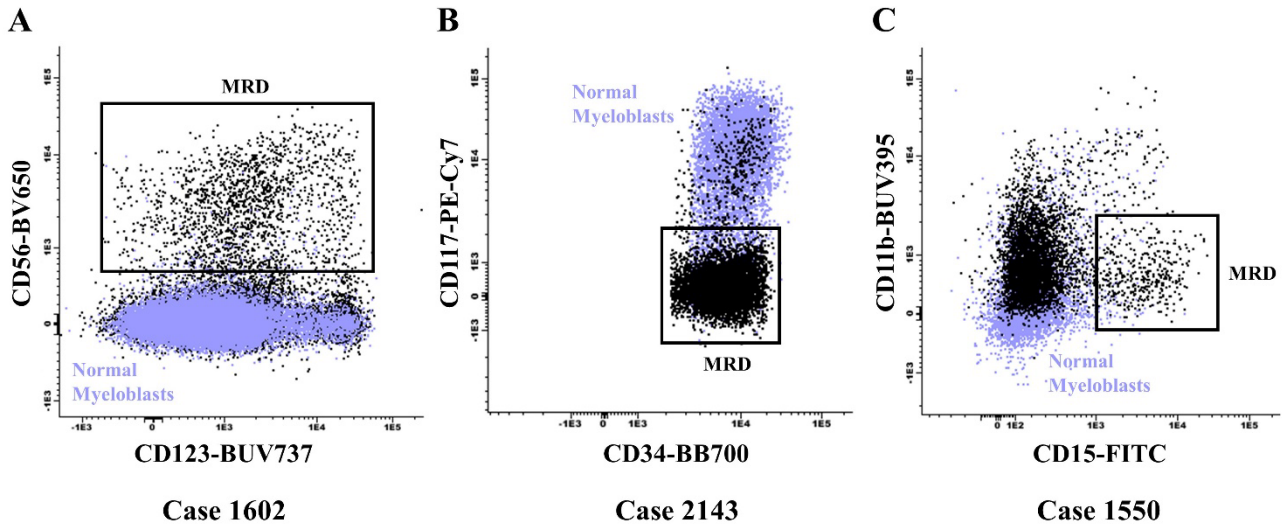
Figure 4: 17-color AML MRD panel applied to diagnostic AML samples



A-B, These panels represent two cases of CD34+ AML at diagnosis. The malignant blasts are shown in red. The reference merged normal myeloblasts are shown in light blue. Case **A** shows aberrant CD15 and CD7 expression on the AML blasts. In case **B**, CD33 is overexpressed on the AML compared to the normal myeloblasts. **C**, This is a case of CD34- AML with monocytic differentiation (FAB M5b). A large quantity of CD64+, CD56+ blasts are seen.

We then applied a LAIP-based DfN[43] approach to identify MRD in AML patient remission samples using a standard gating strategy[47, 52] in Infinicyt. Representative MRD populations are shown in **Figure 5**. Using this method, we identified 6 cases harboring a population suspicious for MRD in the remission sample (**Table 3**).

Figure 5: 17-color AML MRD panel applied to remission samples with suspected MRD



Examples of AML remission samples positive for MRD. The reference merged normal myeloblasts are shown in light blue. The remission sample is shown in black, with the population suspicious for MRD framed in the black rectangle. These cases represent examples of cross-lineage marker expression (CD56+ myeloblasts, **panel A**), under-expression of a normal myeloid marker (CD117- myeloblasts, **panel B**), and asynchronous antigen expression (CD15+ myeloblasts, **panel C**). These LAIPs are not observed in the normal reference.

Table 3: MRD-positive AML cases

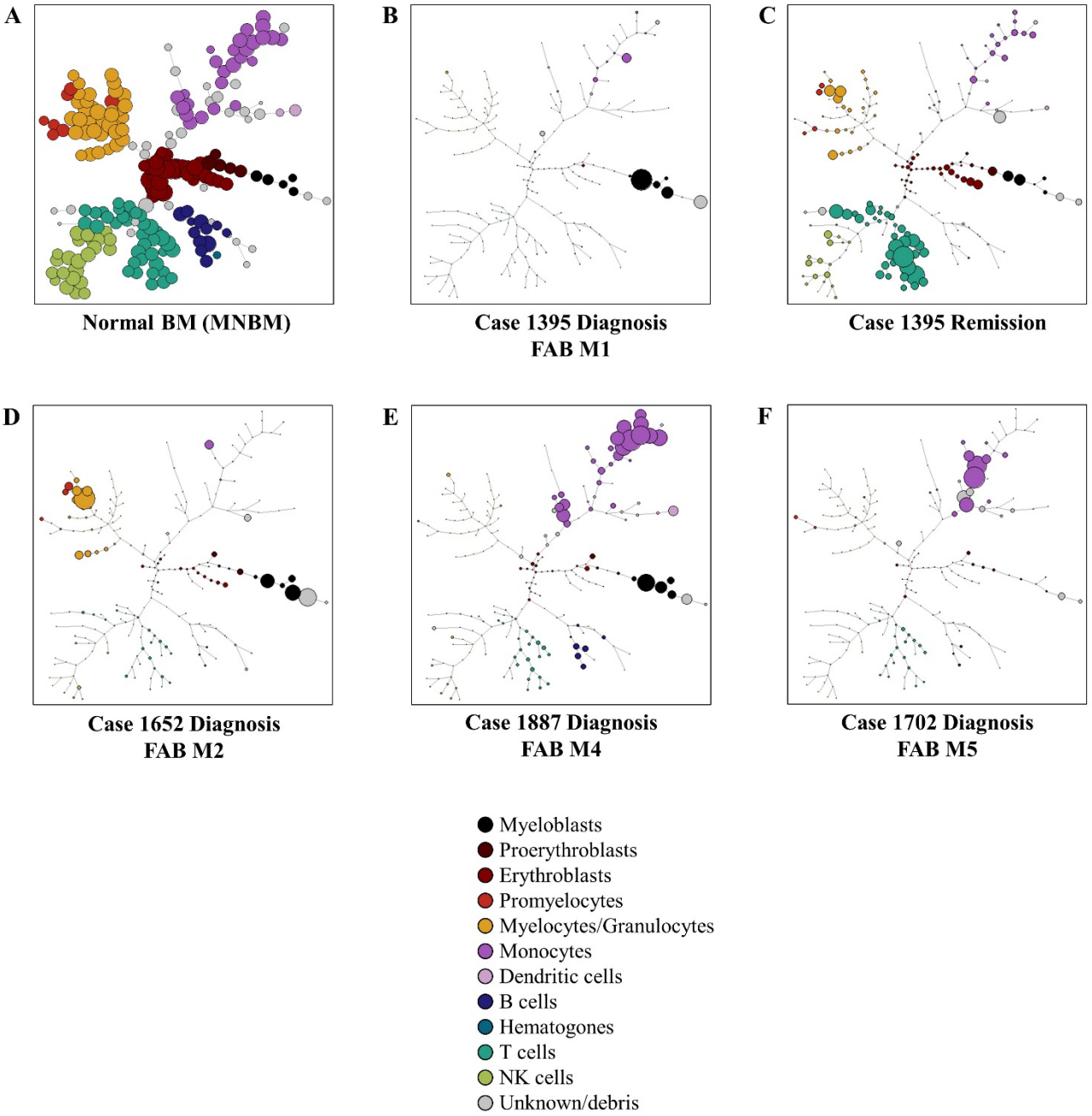
Patient	Leukemia-associated immunophenotype (LAIP)	MRD quantification (as % of CD45+ leukocytes)	Diagnostic sample available	Same LAIP at diagnosis	Selected driver mutations
344	CD7+ myeloblasts	1.10%	no	n/a	n/a
767	CD56+ myeloblasts	0.18%	yes	no*	n/a
1395	CD56+ myeloblasts	0.10%	yes	yes	FLT3(TKD) EZH2 ZRSR2 RUNX1
1550	CD15+ myeloblasts	0.22%	no	n/a	WT1
1602	CD56+ myeloblasts	1.04%	yes	no	FLT3(ITD) U2AF1
2143	CD34+/CD117-/HLA-DR-/CD11b+ myeloblasts	4.77%	yes	yes	NPM1 IDH2

*The majority of diagnostic blasts were CD34- with aberrant CD56 expression. In remission, the CD34-/CD56+ blasts were eliminated, but a residual CD34+/CD56+ population was identified.

Populations suspicious for MRD were identified at levels ranging from 0.10 to 4.77% of the total CD45+ leukocytes. For two cases (1395, 2143), the LAIP identified in remission was identical to the diagnostic LAIP. In two other cases (767, 1602), an immunophenotypic shift was seen following chemotherapy. For the cases identified as MRD-positive, we obtained (if possible) the raw NGS sequencing data performed at diagnosis (panel of 35 myeloid genes, **Appendix Table A1**). Using a custom bioinformatic analysis pipeline, we identified key AML driver mutations in these samples (**Table 3**) for later use in the Target-Seq protocol.

After identifying populations suspicious for MRD using a traditional gating approach, we sought to explore if bioinformatic flow cytometry tools could provide additional information. We applied two variations of the FlowSOM[82] algorithm to the diseased samples. The first method is referred to as FlowSOM frozen[81] and consists of mapping AML diagnostic or remission samples to the MST previously generated using the MNBM reference samples. This allows for a very rapid assessment of the major bone marrow populations present in the diseased samples. We observed expanded myeloblast and/or monocytic populations in diagnostic AML samples according to FAB subtypes. This was accompanied by loss of normal bone marrow populations. Following chemotherapy and remission, normal hematopoiesis was restored (**Figure 6**).

Figure 6: FlowSOM frozen applied to diagnostic and remission AML samples.



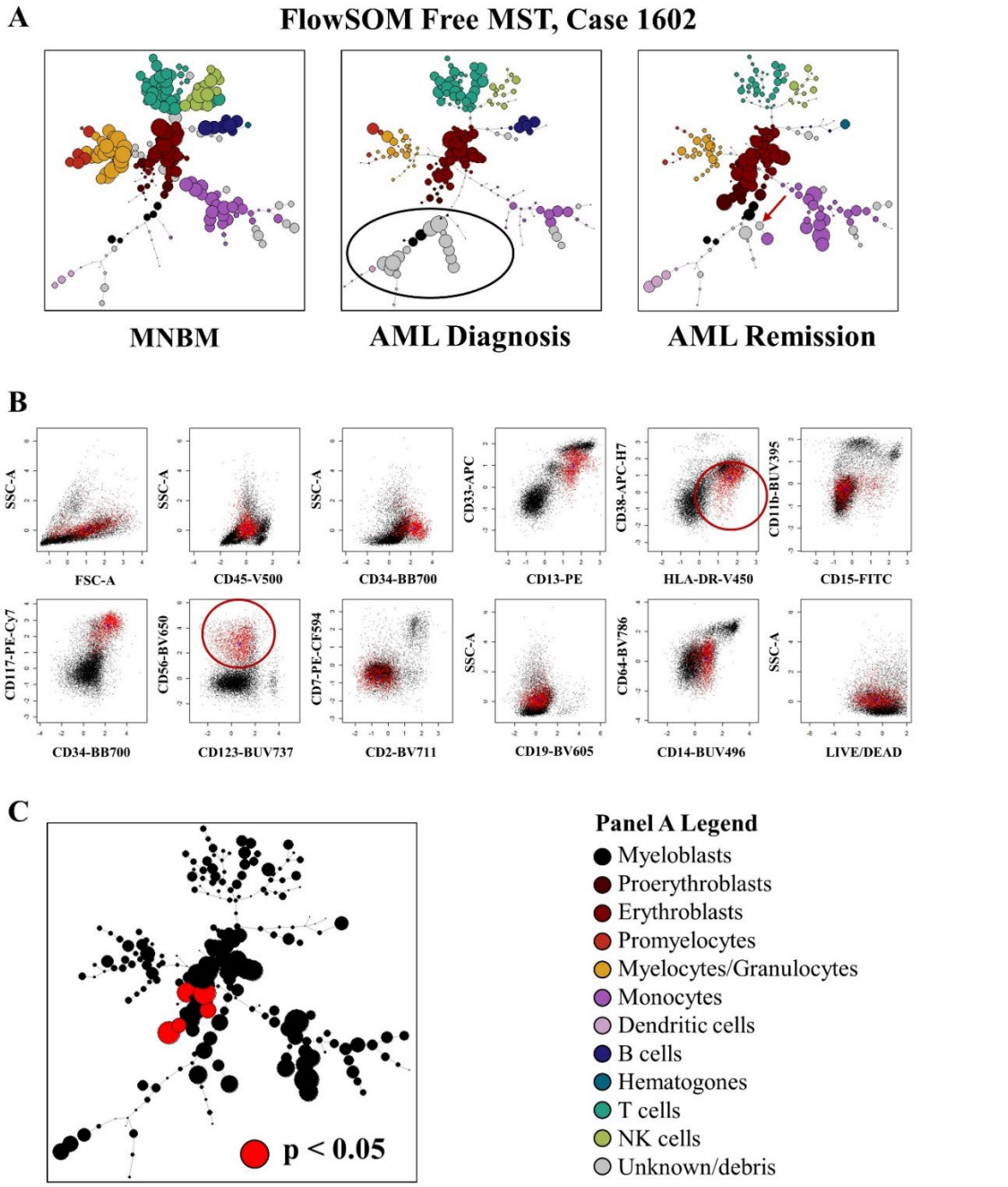
A, FlowSOM MST derived from the MNBM reference file. **B-C**, Longitudinal patient samples for case 1395. At diagnosis (**B**), most of the bone marrow is replaced by myeloblasts as would be expected from a FAB M1 morphology. Following chemotherapy (**C**), the myeloblast population is reduced and normal hematopoiesis restored. **D**, FAB M2 AML sample showing a myeloblast population accompanied by maturing granulocytes. **E**, FAB M4 AML sample containing a mixed myeloblast/monocytic infiltrate. **F**, In M5 AML, the majority of the sample consists of cells with monocytic differentiation.

Although FlowSOM frozen allows for the rapid assessment of 17-color flow cytometry data, it is not suited for MRD as aberrant myeloblasts in remission samples will be assigned to their closest normal counterpart during clustering (usually this represents the normal myeloblast nodes). Therefore, subtle immunophenotypic differences between leukemic and normal myeloblasts will be lost in the graphical representation. FlowSOM free is a variation on FlowSOM frozen that is better suited for MRD[81]. In FlowSOM free, the MST is built using a triplet of samples consisting of the MNBM normal reference, the diagnostic AML sample for the patient, and the remission sample for the same patient. The resulting MST will thus contain nodes representing LAIPs. Each sample is then mapped back to the common MST. This allows for the identification of nodes present in the remission sample but absent from the normal reference, ideally occupying LAIP nodes. Such populations potentially represent MRD.

A certain degree of variability is expected in bone marrow populations, even between normal patients. For this reason, we applied a statistical analysis to the FlowSOM free remission MST. This allowed us to identify remission-emergent nodes containing a new population (compared to the normal reference) of a size larger than could be explained by normal inter-sample variation. We used a combination of resampling and bootstrapping techniques to assign a p-value to each node in the remission MSTs (see methods). This highlights populations that are absent or extremely rare in samples drawn from normal patients.

We applied FlowSOM free to our entire cohort. For cases in which the diagnostic sample was unavailable, the MST was built using only the MNBM and AML remission samples. Although many of the MRD populations we identified using the traditional gating approach were seen in the FlowSOM free remission MSTs, none were statistically significant at the $p < 0.05$ level. This highlights the similarity between these MRD populations and normal myeloblasts, often only being distinguishable by 1 or 2 markers. However, we observed that immature erythroid cells (CD117+ erythroblasts) and monocytes expressing CD123 or CD56 were often significantly increased ($p < 0.05$) in the remission samples. This was a trend throughout the cohort. The increase in immature erythroid progenitors is expected in BM recovering from chemotherapy. Likewise, CD56 expression on monocytes has been described as a chemotherapy-induced phenomenon[95, 96]. A representative example of the FlowSOM free analysis is shown in **Figure 7**.

Figure 7: FlowSOM free applied to a diagnostic-remission AML sample pair

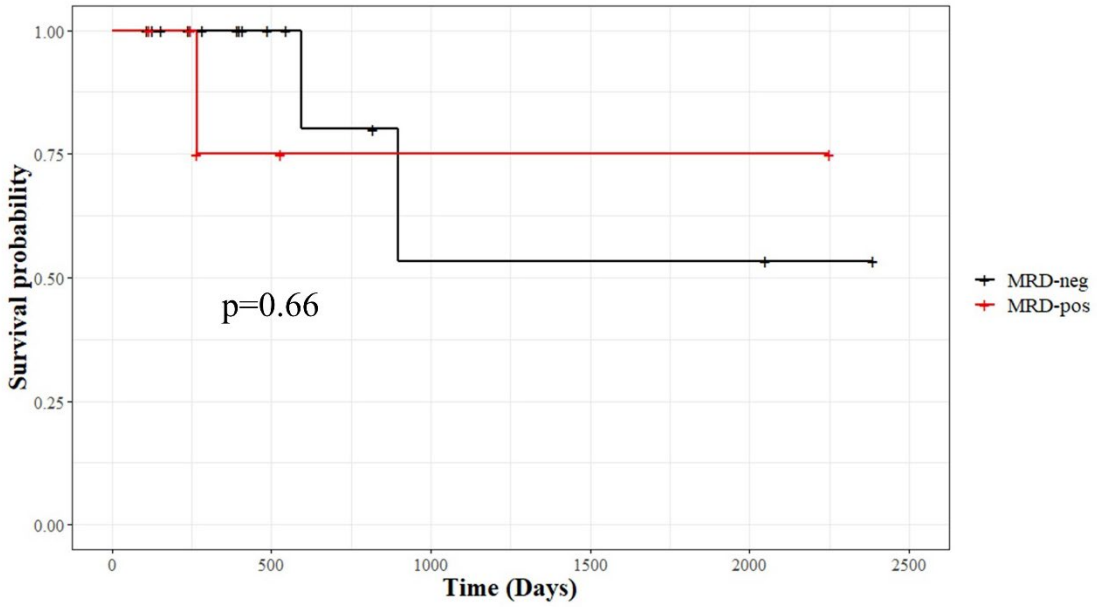


A, FlowSOM free performed on case 1602. A large cluster containing the AML blasts is seen in the diagnostic sample (black oval). In remission, a new node (red arrow), which is empty in the MNBM normal reference, appears close to the AML cluster. **B**, When interrogated, the node identified by the red arrow in panel A contains a CD56+/$CD38_{dim}$ myeloblast population, highly suspicious for MRD. This is the same population identified using the manual gating in **Figure 5A**. **C**, Remission sample nodes colored by statistical significance. The MRD node did not reach statistical significance at the $p < 0.05$ level. However, the proerythroblast population is significantly increased in the remission sample compared to the MNBM.

Survival analysis

After having identified MRD-positive cases within our patient cohort, we sought to evaluate if these patients had a worse OS compared to the MRD-negative cases. We first performed a univariate analysis (MRD + versus MRD -) using the Kaplan-Meier estimate method with the log-rank test to assess for statistical significance[84]. Unfortunately, over half of the cohort (12/22 patients) were transferred to another institution for aHSCT during the observation period. We could not obtain ethical approval for chart review in these other institutions, resulting in a large proportion of patients being lost to follow-up and censored from the Kaplan-Meier analysis. The resulting survival curves (**Figure 8**) did not achieve statistically significant separation by the log-rank test (p=0.61). In addition, a key assumption of the Kaplan-Meier method is that censored and uncensored patients have an equal chance of dying[97]. In our case, this assumption is clearly not met as only the higher-risk patients referred for aHSCT were lost to follow-up. No definitive conclusions can be drawn from this survival analysis. Likewise, in the multivariate analysis using a Cox proportional hazards model considering age, sex, MRD status, and ELN risk category as potential explanatory variables, no statistically significant associations with survival were identified. This is also likely related to the small sample size and the many patients lost to follow-up.

Figure 8: Kaplan-Meier survival analysis stratified by MRD status



Use of the 17-color AML MRD panel for cell sorting and single-cell RNA sequencing

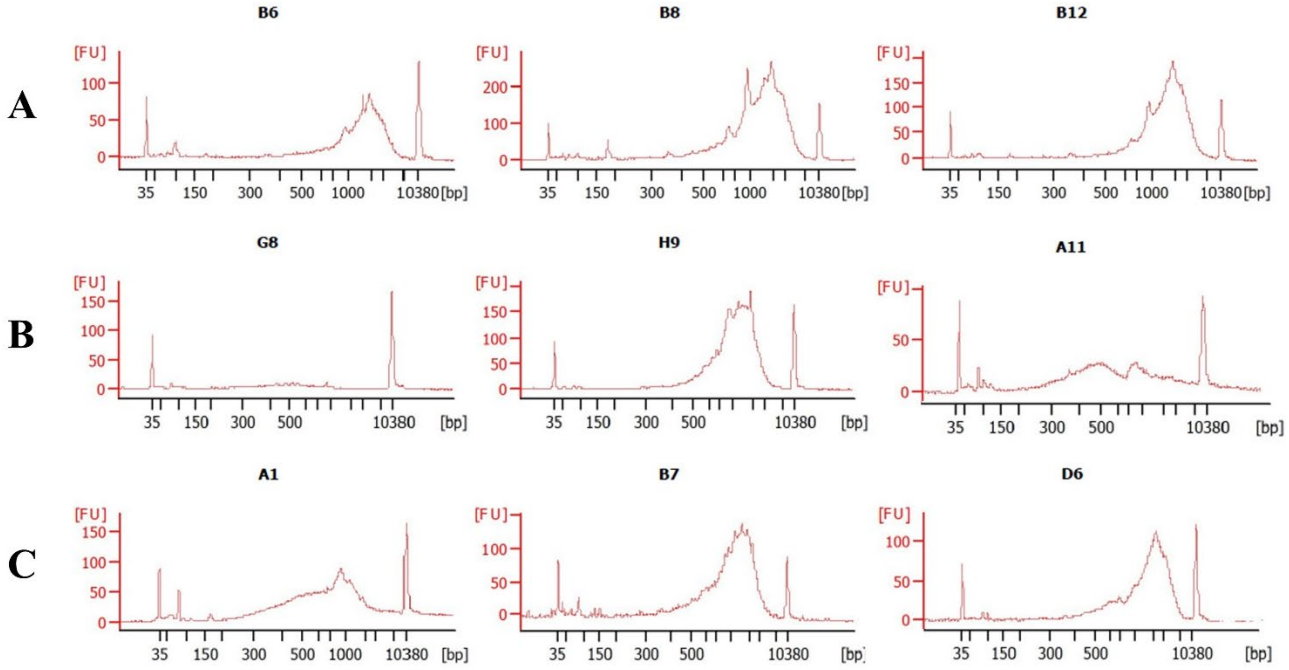
After identifying 6 samples (**Table 3**) containing populations suspicious for MRD, we used the 17-color panel to sort these MRD cells into 96-well plates. As a comparator, we also sorted myeloblasts with normal immunophenotype from the remission samples (presumed non-leukemic recovering myeloid progenitors). When available, untreated AML blasts from the diagnostic sample were also sorted. We then sought to process these three key populations (MRD blasts, normal remission blasts, and untreated AML) using Target-Seq[75]. This scRNA-seq protocol allows for the generation of transcriptomic data with parallel high-sensitivity somatic mutation identification. We used custom primers to track the mutations listed in **Table 3**. The goal of tracking these mutations was to validate that the populations labelled as MRD based on aberrant immunophenotype were enriched in AML driver mutations. We regarded this validation as important given the operator-dependence of flow-based MRD in AML.

We first tested the Target-Seq protocol on the K562 cell line to confirm that we were able to obtain good quality cDNA libraries. These K562 cells were taken directly from culture, stained with a viability dye, sorted into 96-well plates (live cells only), and frozen at -80°C. The plates were processed the following day using Target-Seq. As K562 cells are derived from chronic myeloid leukemia (CML) blast crisis, we added previously validated[3] primers specific for the BCR-ABL transcript at the RT and PCR steps. We evaluated the quality of the single-cell cDNA libraries obtained using a Bioanalyzer device and found that 73% (8/11) of the wells tested displayed good quality libraries (**Figure 9A**). Of note, due to the high cost of the Bioanalyzer service, only a random subset of the wells in each plate were evaluated for quality. Single cells yielded on average 49.1ng (range 5.0ng to 133.3ng) of dsDNA after the PCR step. In addition, the BCR-ABL mutant transcript could be detected by Sanger sequencing in 100% (4/4) of single-cell libraries, indicating successful amplification of this sequence.

We then proceeded to apply Target-Seq to the 96-well plates containing the sorted patient-derived populations (cryopreserved). We encountered significant challenges in obtaining good quality cDNA libraries from these cells. During our first attempts (patient 1550), only 14% (3/22) of cells generated acceptable libraries. Many libraries were either devoid of any product or displayed evidence of RNA degradation (**Figure 9B**). This prompted us to adopt optimization measures with the goal of improving the rate of successful library generation. These included

performing all cell handling steps on ice, shorter pipetting times, strict measures to eliminate RNAses, and an increase in the number of PCR cycles. When we first applied these measures to sorted CD34+/CD117+/CD19- myeloblasts from cryopreserved normal BM samples, we obtained an improved rate of successful library generation (32%, 7/22 of wells tested) (**Figure 9C**). Unfortunately, this improvement did not translate to the populations sorted from the remainder of the patient samples, which yielded only 11% good-quality libraries (5/44 of wells tested). This suggests persistent issues with RNA degradation in the MRD cells undergoing a lengthy preparation and stressful sorting procedure.

Figure 9: Bioanalyzer traces for cDNA libraries obtained with Target-Seq



A, High quality cDNA libraries obtained from the cell line K562. Ideally, the whole transcriptome amplification product average is expected to be about 2000bp. **B,** cDNA libraries generated from cryopreserved cells prior to optimization. Some cells show lack of any product (**G8**) and others show evidence of RNA degradation (**A11**). **C,** Improvements in library generation from single cryopreserved cells following protocol optimization.

DISCUSSION

Most patients diagnosed with AML and treated with intensive chemotherapy successfully achieve remission. Despite this, relapse is common. This implies that a subset of AML cells selectively resists therapy and regenerates the leukemia. Using advanced flow cytometry and genomic techniques, we sought to study these MRD cells at the level of single-cell gene expression. The overarching goal was to clarify why these cells resist treatment, and to identify weaknesses that could be targeted therapeutically (with small-molecule inhibitors for example). In this project, we successfully identified human remission AML samples containing populations suspicious for MRD based on aberrant immunophenotype. In addition, we used FACS to isolate these cells for processing with a scRNA-seq protocol. Unfortunately, the protocol was only successful in generating adequate cDNA libraries in a small fraction of cells, precluding a gene set enrichment analysis. It should be noted that, to the best of our knowledge, performing scRNA-seq on sorted primary human AML MRD cells has not been previously performed. From a proof-of-concept point of view, our results demonstrate that this experimental design is possible as we did obtain acceptable cDNA libraries at a low rate. However, further optimizations are required for this protocol to generate meaningful transcriptomic data.

In terms of MFC-based MRD, we expanded on traditional approaches by creating a single-tube, 17-color panel. Most AML MRD panels currently in use are 8 to 10 colors and split across multiple tubes[47, 51]. These 8-10 color panels are compatible with common clinical cytometers and are simpler from a compensation point of view. However, a single-tube design has the advantages of maximizing the number of cells acquired as the sample does not need to be split. This has the potential to improve the capacity to detect very rare populations such as MRD. In addition, the single-tube approach does not rely on backbone markers and can assess any combination of markers on each cell. In our cohort, we identified 6 samples containing LAIPs that are well-described in the literature[48]. MRD positivity was seen in roughly one quarter of the cases. This is in line with the expected positivity rate published by large MRD centers[51]. We felt it was important to attempt to validate the populations we sorted truly represented leukemic blasts. We first sought to do this by comparing survival in the MRD-positive versus MRD-negative patients. Unfortunately, barriers related to patient consent did not allow us to obtain up-to-date survival data for over half the cohort because they were transferred to another institution for aHSCT. No

significant conclusions could be drawn from the resulting Kaplan-Meier survival analysis. We then tried to evaluate if the sorted MRD fractions were enriched for AML driver mutations. For 4 out of the 6 MRD-positive samples, we could obtain raw NGS sequencing data performed at diagnosis to identify such mutations. We attempted to track these mutations in the sorted cell fractions using Target-Seq. However, the sorted cryopreserved cells had evidence of severe RNA degradation and did not generate acceptable libraries for sequencing. It was therefore not possible to fully validate that the MRD populations identified and sorted with the 17-color AML MRD panel truly represented residual leukemia. Nonetheless, the aberrant immunophenotypes observed and positivity rate were consistent with previously published literature[47, 48, 51].

Lack of standardization and operator-dependence are major challenges facing MFC-based MRD in AML. Recognition of aberrant immunophenotypic patterns requires considerable experience. In addition, these aberrancies can be subtle and vary widely from patient to patient. For these reasons, standardization of MFC-based MRD in AML has lagged compared to other leukemias such as chronic lymphocytic leukemia (CLL)[98, 99] and acute lymphoblastic leukemia (ALL)[100]. Bioinformatic flow cytometry analysis approaches have to potential to remedy some of these difficulties with AML MRD. As previously described by Lacombe et al.[81], we found the FlowSOM tool in R to be particularly useful in analyzing our high-parameter data. The FlowSOM frozen protocol could rapidly identify major bone marrow populations, whereas the FlowSOM free modification allowed for the identification of emergent populations in remission samples compared to a normal reference. We took this method further by applying a custom statistical analysis to highlight the most significant new nodes in remission. We believe FlowSOM is a very promising tool to study MRD with more reproducibility and less operator-dependence. In our project, we used bone marrow obtained from untreated NHL patients as a normal control as this is what we had available in the biobank. However, a more appropriate control would be bone marrow obtained from patients recovering from chemotherapy for non-myeloid malignancies. This would allow us to control for non-malignant, chemotherapy-induced populations such as the regenerating erythroid progenitors and CD56+ monocytes we observed in the AML remission samples. Considerable hurdles remain prior to the application of a full bioinformatic analysis strategy in the clinic. More extensive testing of different AML MRD panels would be required to identify optimal reagent combinations. Such panels should be validated with known positive and negative controls (established by a molecular MRD method such as *NPM1* qPCR for example) to confirm the

bioinformatic analysis can reliably identify MRD populations. Finally, these tools, which are currently code-based and require programming skills to use, would need to be developed into a graphical user interface that is usable by the average physician or pathologist.

A major limitation of this project was the very low rate of successful cDNA libraries generated using the sorted cryopreserved patient samples. This is likely the result of RNA degradation related to the cryopreservation process followed by thawing, a lengthy sample preparation, and the stress of the cell sorting procedure. This hypothesis is supported by the fact that we obtained much better cDNA libraries from the K562 cells compared to the patient samples. These K562 cells were taken directly from culture, stained only with a viability dye (< 30 minutes of preparation time), and immediately sorted. Cells in culture are also larger and more metabolically active, thus containing more starting RNA material than cryopreserved cells. We could not significantly improve the rate of successful library generation in the patient samples despite multiple optimization procedures including temperature control (all steps on ice), shortening the preparation time, taking great care to keep the environment free of RNAses, and increasing the number of PCR cycles. The fact that we did, on occasion (~10%), observe successful cDNA libraries with the cryopreserved patient samples argues against a problem with the Target-Seq protocol itself or the reagents used and favors the RNA degradation hypothesis. RNA sequencing has previously been successful on sorted MRD cells[70]. However, these were bulk-sorted (not single-cell), making the issue of viability and RNA quality less critical as many cells are processed together. It is also unclear if cryopreservation was used in this abstract[70]. Our protocol could be further optimized for cellular viability. This should probably involve avoiding cryopreservation and performing testing immediately after fresh sample collection. However, this would greatly increase the logistical complexity of the assay. Close coordination between the clinical team and research laboratory would be required. In addition, MRD would need to be identified immediately on the cell sorter without the benefit of advanced analysis tools such as Infinicyt or bioinformatics. A large proportion of cases (~75%) would be expected to be MRD-negative[51] and have no residual leukemia cells to sort, representing a waste of the Target-Seq plate prepared right before the sort. Nonetheless, if these logistical hurdles could be overcome, we believe the approach we have developed would represent a powerful tool to study chemoresistance in AML. Some investigators have explored an alternative approach consisting of applying a higher-throughput scRNA-seq protocol (Chromium 10x) to unsorted AML remission samples,

instead relying on bioinformatic techniques to identify residual leukemia[101]. It does not appear RNA degradation was a major barrier using this protocol.

In summary, we developed a high-dimensional (17-color) flow cytometry AML MRD panel. We showed this panel was compatible with traditional (gating) and bioinformatic analysis methods. We also demonstrated how such a panel has the potential to be used to isolate rare cell populations of interest for scRNA-seq. Although further optimization is required, the combination of advanced flow cytometry and genomic methods has the potential to yield valuable insight into treatment failure in AML and lead to novel therapies.

PROJECT 2:
COMMON CLONAL ORIGIN OF CHRONIC MYELOMONOCYTIC LEUKEMIA AND B CELL ACUTE LYMPHOBLASTIC LEUKEMIA IN A PATIENT WITH A GERMLINE *CHEK2* VARIANT

LITERATURE REVIEW

Myelodysplastic syndromes (MDS) consist of a group of hematological malignancies that are increasingly frequent as people age and represent a growing disease burden[102]. Clinically, these disorders are characterized by ineffective hematopoiesis, chronic cytopenias, and a variable risk of leukemic transformation. The bone marrow aspirate typically reveals dysplastic morphological changes, with or without increased blasts[103]. MDS patients suffer from complications related to cytopenias, such as fatigue, dyspnea, infections, bleeding, and reduced quality of life[104, 105, 106]. MDS is associated with a wide range of chromosomal and genetic abnormalities. These typically affect cellular pathways such as epigenetic regulation, RNA splicing, transcription, signaling, cell cycle control, and DNA repair[4].

Patients with MDS can be stratified into risk categories based on the number/degree of affected bone marrow lineages, blast percentage, and genetic abnormalities. This is usually done using tools such as the International Prognostic Scoring System (IPSS)[107] or its revised version (IPSS-R)[108]. There is thus considerable variation in disease severity from patient to patient. Transformation to AML is a well-described component of the natural history of MDS and occurs on average in 30% of patients at some point during the disease course[109]. This transformation is more common in higher-risk IPSS categories.

Transformation of MDS to other leukemias besides AML is considered exceedingly rare. To date, only a few dozen cases of MDS to ALL transformation have been described in the literature[110, 111, 112]. From a disease biology point-of-view, this phenomenon is very interesting as it represents a switch from the myeloid (MDS) to lymphoid (ALL) lineage. With the exception of lymphoid blast crisis in CML[35], lineage switch is rare throughout hematological malignancies. The underlying mechanisms underlying MDS to ALL transformation are still poorly understood. One plausible hypothesis is that the self-renewing multipotent hematopoietic stem or early progenitor cell responsible for the MDS clone acquires further genetic or epigenetic changes leading to lymphoid transformation[110, 111]. It has been appreciated for some time that MDS-specific cytogenetic changes (such as -5q or -7) can be detected within the lymphocytes of a subset of MDS patients[113, 114]. However, it is unclear how this low-level lymphoid output occasionally evolves to overt ALL. Alternatively, different hematological malignancies may originate from a common preleukemic HSC pool (CHIP). Such preleukemic cells can be detected in the blood of individuals

without a hematological cancer[115], AML patients[116], and lymphoma patients[117]. It is thus possible that branching clonal evolution may evolve from this pre-malignant reservoir.

MDS to chronic myelomonocytic leukemia (CMML) transformation also appears to be quite rare, with only a handful of cases described in the literature[118, 119, 120, 121]. These two entities are more closely related than MDS and ALL. Indeed, CMML was previously included as a type of MDS in the FAB classification. It now occupies the category of myelodysplastic/myeloproliferative (MDS/MPN) overlap syndromes[35]. MDS and CMML share a number of features including bone marrow dysplasia, a propensity towards elevated bast counts, and typical cytogenetic/molecular abnormalities[35]. Predominantly dysplastic versus proliferative forms of CMML are described based on total WBC count[35]. Often, the only distinguishing feature between dysplastic type CMML and MDS is peripheral blood monocytosis ($\geq 1 \times 10^9$/L) persisting for 3 months. This has been criticized as a somewhat arbitrary distinction by some authors[121]. Overall, given the similarity between MDS and CMML, and the fact that both are thought to originate from a transformed hematopoietic stem or progenitor cell, it is not surprising that one entity may transform into the other. However, the underlying mechanisms remain to be clarified. CMML appears to have a different mutational landscape compared to MDS, with a higher frequency of *ASXL1*, *NRAS/KRAS*, *CBL*, and *SRSF2* mutations[122, 123, 124]. The acquisition of such mutations may lead to differences in proliferative and differentiation potential and underly transformation from MDS to CMML.

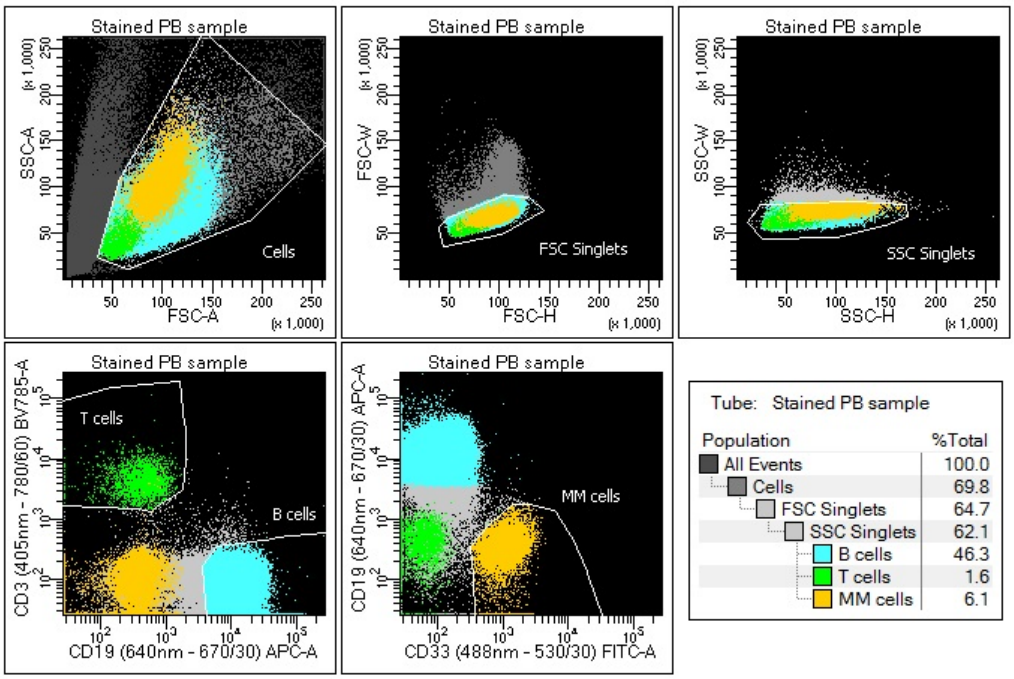
Overall, given the rarity of MDS transformation to CMML and ALL, we believe such cases deserve to be aggressively investigated. In this project, we studied such a case with FACS and whole-exome sequencing (WES) to elucidate the underlying pathophysiology at the genetic level.

METHODS

Flow cytometry and cell sorting

The patient provided written informed consent for the collection of a peripheral blood sample and buccal swab for this project (Research Ethics Board protocols 11-047 and 12-052). We purified peripheral blood mononuclear cells (PBMCs) from the sample using Ficoll density gradient centrifugation. These cells were cryopreserved at -80°C until processing. We then thawed and stained this sample using CD19-APC (eBioscience, clone HIB19, 17-0199-41) and CD33-FITC (eBioscience, clone WM53, 11-0338-41). We sorted the B-lymphoid (CD19+/CD33-) and myelomonocytic (CD19-/CD33+) fractions using a BD FACSAria Fusion cell sorter. To isolate normal T cells for *CHEK2* variant detection, we added CD3-BV786 (BD Biosciences, clone SK7, 563800) to this antibody panel and sorted CD19-/CD33-/CD3+ cells. The cell sorting strategy is illustrated in **Figure 10**. Post-sorting purity check showed that the sorted fractions were >99% pure for the target population.

Figure 10: Cell sorting strategy



Abbreviations: APC, allophycocyanin; BV785, Brilliant Violet 785; FITC, fluorescein isothiocyanate; FSC, forward scatter; MM, myelomonocytic; PB, peripheral blood; SSC, side scatter.

Whole-exome sequencing and analysis

We extracted genomic DNA from the sorted cell fractions and the concomitantly collected buccal swab using the DNeasy Blood and Tissue Kit (QIAGEN, 69506). We sent the B-lymphoid, myelomonocytic and buccal swab gDNA for WES (Novogene Co., Beijing, China). The Agilent SureSelect Human All Exon kit (Agilent Technologies) was used to generate the sequencing libraries. Sequencing was performed on an Illumina platform (150-bp paired-end reads).

We performed analysis of the WES data on a Compute Canada Cluster (Beluga). We used BWA-MEM[86] to align the clean reads to the GRCh37 (hg19) reference genome. To call the germline variants from the buccal swab, we used HaplotypeCaller (GATK, Broad Institute). We used the following default hard filters on the germline variants: QD < 2.0, FS > 60.0, MQ < 40.0, HaplotypeScore > 13.0, MappingQualityRankSum < -12.5, ReadPosRankSum < -8.0. Because HaplotypeCaller is not well suited for variant calling on tumor samples (somatic variant detection), we opted to use the variant caller Mutect2 (GATK, Broad Institute) with a panel of normals (PON) to identify the variants in the tumor samples (B-lymphoid, myelomonocytic). Mutect2 is specifically designed for this purpose. We used FilterMutectCalls (GATK, Broad Institute) to filter the tumor sample variants. We performed variant annotation using ANNOVAR[88]. We excluded synonymous single-nucleotide variants (SNVs), non-coding variants, variants identified at a rate over 0.01% in population databases (1000G, ExAC, ESP6500si, gnomAD), and variants known to be benign on the ClinVar database. Furthermore, only variants with at least 10X coverage in at least one sample were retained. These first curation steps allow for the exclusion of artifacts, single-nucleotide polymorphisms (SNPs), and variants very unlikely to be pathogenic. However, many variants will remain at this stage that may not be implicated in cancer biology. To further refine our list of variants, we used the web-based application Cancer Genome Interpreter (CGI) to identify probable cancer drivers[125]. In parallel to the CGI analysis, we used a previously described method[89] to assess pathogenicity in the retained variants. Frameshift and nonsense mutations are considered pathogenic, whereas SNVs are assigned a category (benign, variant of uncertain significance, likely pathogenic, or pathogenic) based on a combination of 6 *in silic*o tools (SIFT, PolyPhen2, M-CAP, FATHMM, PROVEAN, CADD). No benign variants were retained in the final mutation list. All mutations identified by NGS were confirmed using Sanger sequencing.

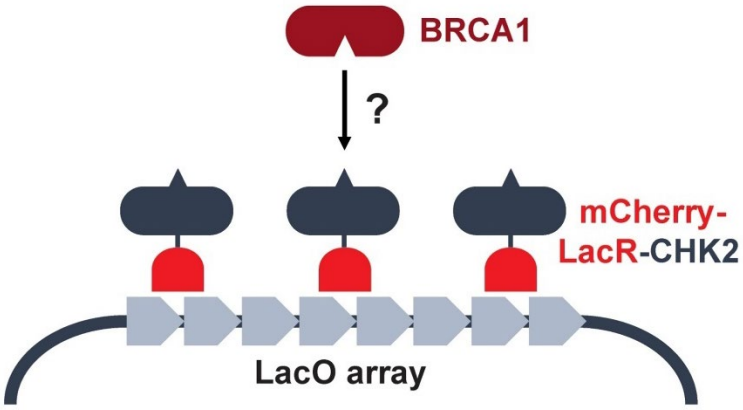
ExomeAI[126] is a web-based bioinformatic tool (http://genomequebec.mcgill.ca/exomeai) that allows for the identification of allelic imbalances (AIs) in NGS data. AIs can include large chromosomal deletions/gains or areas of loss of heterozygosity (LOH). This AI analysis complements the standard analysis that can only detect point mutations and small insertions/deletions. We used ExomeAI with the default settings to identify such larger scale abnormalities in our sequencing data.

Functional analysis of the *CHEK2* variant

The germline variant we identified in the tumor suppressor *CHEK2* (c.475T>C, p.Y159H) was of great interest to us for its possible role as an inherited cancer predisposition. Therefore, we performed a functional assay to evaluate the effect of this variant on protein function. The CHK2 Y159H variant affects the forkhead-associated (FHA) domain that is critical to bind downstream effectors such as BRCA1[127]. We thus leveraged a LacO/LacR protein binding assay to evaluate the effect of Y159H on the CHK2-BRCA1 protein interaction.

The LacO/LacR system[128, 129] takes advantage of the binding between the lac operator (LacO) repeat DNA sequence and the lac repressor (LacR) protein. A protein of interest (in this case CHK2) is tagged with LacR and a fluorescent protein (mCherry). By expressing this protein in a cell line containing a stable integration of a repeat LacO array, CHK2 is tethered at a specific point within the cell that can be identified by the mCherry signal. The interaction between CHK2 and a second protein of interest (in this case BRCA1) can be evaluated using a second fluorochrome bound to BRCA1 and measured at the LacO array. A schematic representation of this assay is shown in **Figure 11**.

Figure 11: LacO/LacR system to assess the CHK2-BRCA1 interaction



To generate variant CHK2 protein, we obtained the CHEK2 open reading frame (ORF) vector from the ORFeome Collaboration provided by the McGill Platform for Cellular Perturbation of the Goodman Cancer Research Centre and Biochemistry at McGill University. We used the QuikChange II Site-Directed Mutagenesis kit (Agilent Technologies) to generate the Y159H mutant as well as the known dysfunctional variants I157T and H143Y[130] as positive controls. The oligonucleotides used for the QuikChange reaction are detailed in **Appendix Table A2**. We then used Gateway cloning technology (Invitrogen) to insert the variant and wild type (WT) *CHEK2* vectors into a pDEST-mCherry-LacR vector. Successful cloning of the different variants into the final constructs was confirmed using Sanger sequencing.

For the single-cell assay, we used U2OS cells containing a stable integration of the LacO repeat sequence. These cells were plated onto #1.5 coverslips in a 12-well plate. We transfected the U2OS-LacO cells at 60% confluency with 1 µg of each mCherry-LacR-*CHEK2* construct. The coverslips were then rinsed with PBS, fixed with PBS +2% paraformaldehyde (Thermofisher), washed again with PBS, and permeabilized with PBS + 0.3% Triton-100 solution for 20 minutes. After blocking the coverslips with PBS + 5% BSA + 0.1% Triton-100 (PBSA-T), we performed a one-hour incubation with 1:400 rabbit anti-BRCA1 primary antibody (EMD Millipore, 07-434). This was followed by two washes and incubation with goat anti-rabbit Alexa Fluor 647 secondary antibody (Thermofisher, A-21244). Finally, the coverslips were washed twice in PBS, once in ddH2O, and mounted with Fluoromount-G (Thermofisher). The images were acquired using a LSM 800 confocal microscope (Carl Zeiss AG) and analyzed as previously described[131]. The

colocalization of BRCA1 (AF647 signal) with WT and variant CHK2 was measured at the LacO repeat array and quantified by calculating the mean fluorescence intensity (MFI) of AF647 at the mCherry focus divided by the MFI of AF647 in the background region of the nucleus (relative fluorescence). We performed two biological replicates for each variant and a minimum of 25 cells were evaluated per replicate. We assessed statistical significance using ordinary one-way ANOVA with Dunnett's multiple comparisons test using WT CHK2 as the control (GraphPad Prism software, version 9.0.0).

RESULTS

Description of the patient case and family history

A 56-year-old male was referred to the hematology clinic for an incidental finding of severe neutropenia. His family history was positive for essential thrombocytosis (ET) in his mother, colorectal cancer in his father, and an unspecified malignancy in his brother. The patient himself did not have any previous history of cancer or hematological disorders. A bone marrow aspirate and biopsy was performed and was compatible with a diagnosis of myelodysplastic syndrome with multilineage dysplasia (MDS-MLD) by modern World Health Organization (WHO) 2016 criteria[35]. G-band karyotyping showed -Y in all metaphases. His IPSS[107] risk score was low. Given the low-risk nature of his disease and the asymptomatic neutropenia, he was initially managed with observation alone for 10 years. Eventually, he developed worsening anemia which prompted the initiation of recombinant erythropoietin (EPO). Although he did initially respond to EPO, the anemia eventually became refractory. This was accompanied by progressive neutrophilia and monocytosis. Repeat diagnostic workup including bone marrow examination and flow cytometry was now consistent with transformation to CMML without elevated blasts (CMML-0 by WHO 2016[35]). His karyotype remained unchanged (-Y) and IPSS risk score was now intermediate-1. He was treated with lenalidomide 10 mg daily. This helped control the leukocytosis, but he remained dependent on red blood cell (RBC) transfusions. After a two-year period of stability, he experienced an abrupt onset of leukocytosis with abundant blasts on the peripheral blood smear. Morphology and flow cytometry performed on the bone marrow now showed a B-ALL clone in addition to the background CMML. His karyotype remained -Y. The patient was admitted to the hospital and treated with combination chemotherapy (prednisone, vincristine, daunorubicin, cyclophosphamide, and L-asparaginase). His treatment was complicated by an invasive fungal infection (fusariosis) and liver toxicity. This resulted in prolonged hospital admission after which he was no longer considered fit for intensive chemotherapy. The B-ALL and CMML both relapsed 8 months after the induction regimen. The sample we evaluated in this project was drawn at this time point (with both B-ALL and CMML clones in the peripheral blood). The patient's physician attempted to control the disease with sequential vincristine monotherapy, mitoxantrone monotherapy, and finally inotuzumab ozogamicin. Ultimately, these agents were ineffective in

controlling the leukemias. A decision was made to pursue comfort care and the patient died 14 years after the initial MDS diagnosis.

Identification of genomic variants in the buccal swab and CMML/B-ALL fractions

We processed the patient sample by using FACS to isolate the B-ALL (CD19+/CD33-) and CMML (CD19-/CD33+) clones. We then extracted the DNA from these two malignant fractions and a concomitantly collected buccal swab. We sent these three samples for WES. The sequencing metrics are shown in **Appendix Table A3**. We identified a total of 11 different variants meeting retention criteria (see methods) distributed between the three samples (**Table 4**). These were comprised of 6 nonsynonymous single-nucleotide variants (SNVs), 4 insertions or deletions, and 1 stop-gain (nonsense) variant. All variants reported in **Table 4** had their presence or absence confirmed using Sanger sequencing. This was especially important for the buccal swab sample where the DNA quantity and sequencing coverage (12.72X) were lower.

By evaluating which variants were unique or shared between the various samples, we built a putative clonal evolution sequence for this patient case (**Figure 12**). Specifically, variants shared between all samples (buccal swab, CMML, B-ALL) were considered to have a germline origin. Variants absent from the buccal swab but shared between the CMML and B-ALL clones are likely to have been acquired at the MDS or CMML disease stages, prior to the emergence of the B-ALL. Finally, mutations unique to the CMML or B-ALL clones may represent later events in the disease course.

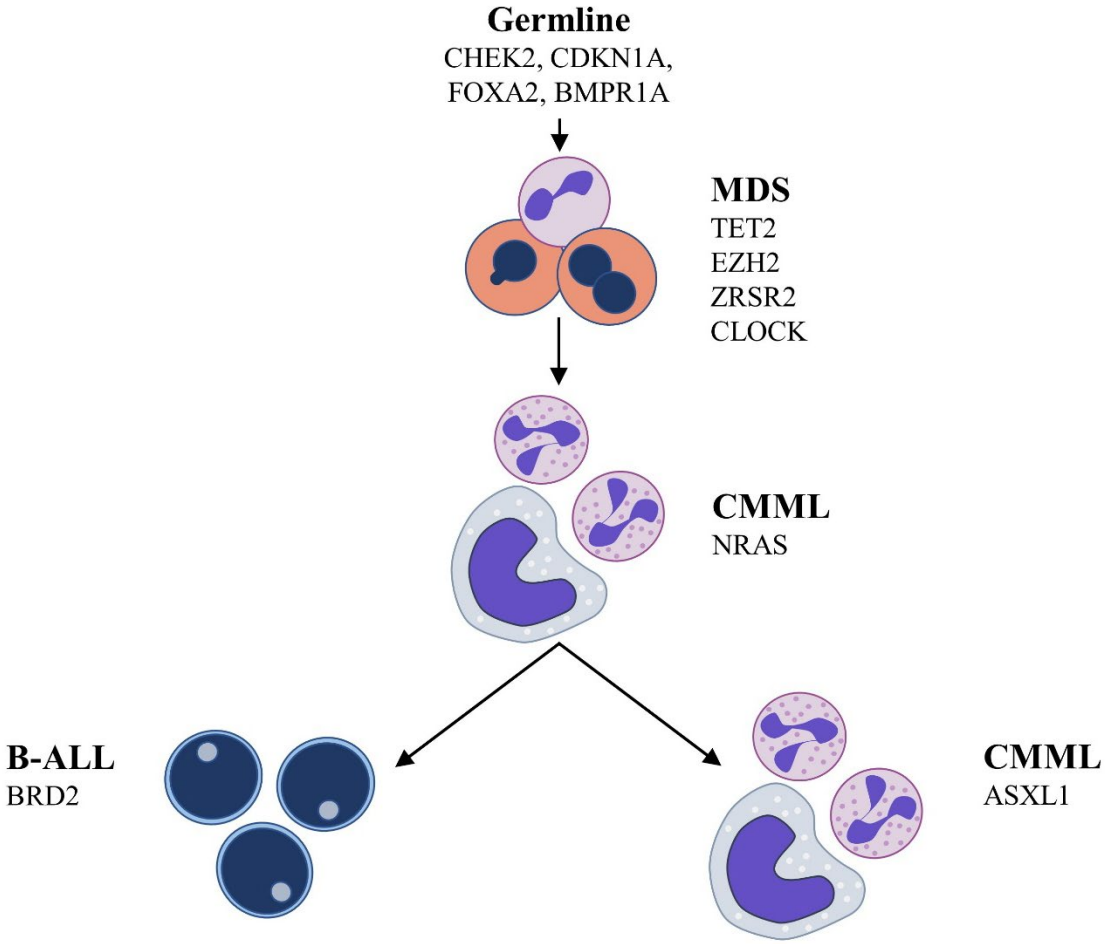
We used the web-based application ExomeAI[126] to evaluate for larger scale allelic imbalances (AIs). This analysis showed a large AI in the long arm of chromosome 4 (chr4q) in the CMML and B-ALL samples. This is consistent with the *TET2* frameshift deletion (located on chr4q) with a variant allele frequency (VAF) of 1.0 identified in both tumor samples. We also identified a large AI on chromosome 16q unique to the B-ALL sample.

Table 4: Variants identified in the buccal swab sample and sorted cell populations using whole-exome sequencing

Gene	Chromosome and GRCh37 position	HGVS DNA Reference	HGVS Protein Reference	Variant Type	*In Silico* Pathogenicity Assessment (see methods)	ClinVar status (if available)	Buccal Swab	Myeloid	Lymphoid
							Genotype	Variant Allele Frequency	
NRAS	Chr 1:115258747	NM_002524:c.35G>A	p.G12D	Nonsynonymous SNV	VUS*	Pathogenic	-	0.47	0.39
CLOCK	Chr 4:56322165	NM_001267843:c.881C>T	p.P294L	Nonsynonymous SNV	Likely pathogenic	Not reported	-	0.41	0.38
TET2	Chr 4:106158398_106158399	NM_001127208:c.3299_3300del2	p.V1100Afs*2	Frameshift deletion	Pathogenic	Not reported	-	1.00	1.00
BRD2	Chr 6:32943865_32943865	NM_001113182:c.529_530insTGA	p.V177_A178insI	Nonframeshift insertion	Unknown	Not reported	-	-	0.41
CDKN1A	Chr 6:36652068	NM_001291549:c.292G>A	p.A98T	Nonsynonymous SNV	VUS	Not reported	Heterozygous	0.47	0.47
EZH2	Chr 7:148507467_148507468	NM_004456:c.1986_1987del2	p.Y663fs*0	Frameshift deletion	Pathogenic	Not reported	-	0.54	0.39
BMPR1A	Chr 10:88683261	NM_004329:c.1471G>A	p.E491K	Nonsynonymous SNV	VUS	VUS	Heterozygous	0.39	0.64
FOXA2	Chr 20:22562771	NM_021784:c.1109T>G	p.L370R	Nonsynonymous SNV	Pathogenic	Not reported	Heterozygous	0.50	0.58
ASXL1	Chr 20:31022922_31022922	NM_015338:c.2407delC	p.Q803Kfs*14	Frameshift deletion	Pathogenic	Not reported	-	0.41	-
CHEK2	Chr 22:29121082	NM_007194:c.475T>C	p.Y159H	Nonsynonymous SNV	Pathogenic	VUS	Heterozygous	0.40	0.38
ZRSR2	Chr X:15809088	NM_005089:c.73G>T	p.E25X	Stopgain	Pathogenic	Not reported	-	1.00	0.75

*Although this variant was classified as a VUS using the combination of *in silico* tools, it is a well validated pathogenic variant in ClinVar and we considered it as such.

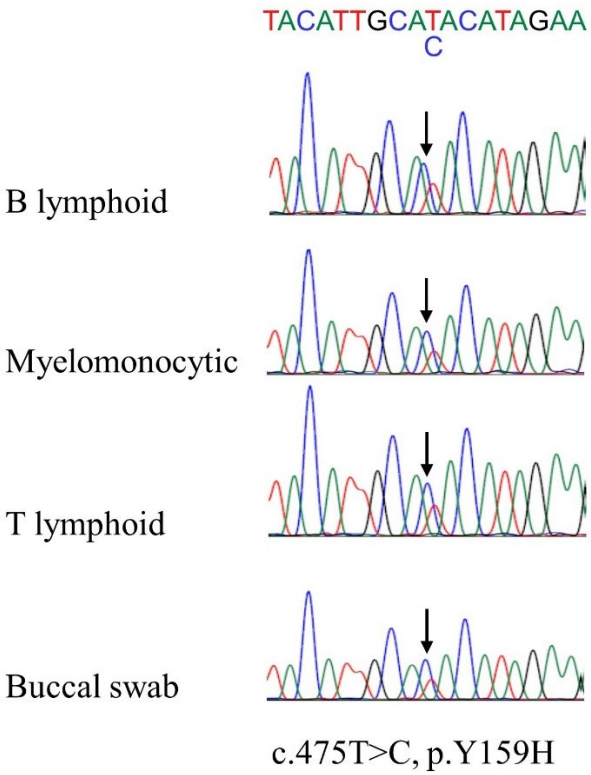
Figure 12: Putative sequence of clonal evolution in the presented case



Functional analysis of the *CHEK2* c.475T>C, p.Y159H variant

The variant we identified in the tumor suppressor *CHEK2* represents a potential inherited cancer predisposition. We confirmed the presence of this variant by Sanger sequencing in the buccal swab, myelomonocytic, and lymphoid samples. In addition, the variant was also observed in non-malignant T cells sorted from this patient (**Figure 13**). This strongly suggests a germline origin for *CHEK2* c.475T>C.

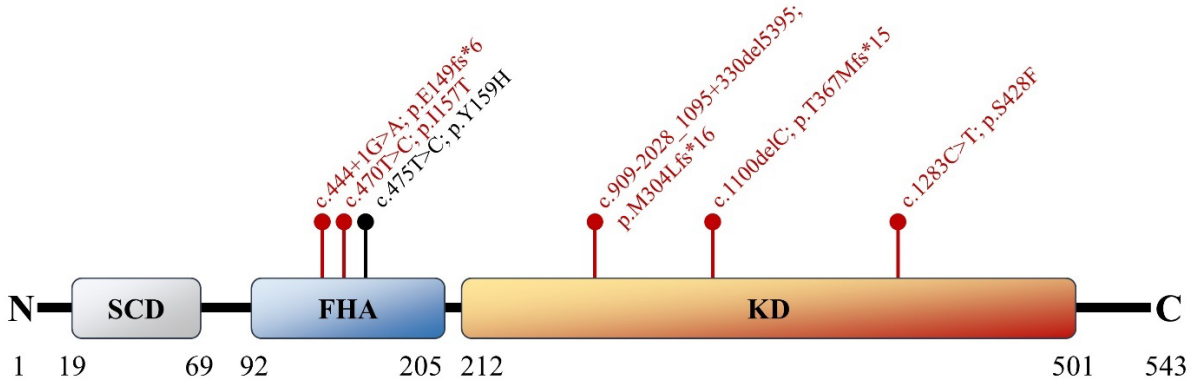
Figure 13: *CHEK2* c.475T>C chromatograms



The *CHEK2* c.475T>C SNV leads to a tyrosine to histidine substitution at codon 159 (Y159H) of the CHK2 protein. This change occurs in the forkhead-associated (FHA) domain, which binds downstream effectors including BRCA1, p53, and CDC25 proteins[127]. The Y159H variant is located only two amino acid positions away from the I157T variant, which is known to disrupt the FHA domain (**Figure 14**)[127]. In addition, the Y159H variant was predicted as damaging/pathogenic by all six *in silico* prediction tools we used. Therefore, we hypothesized that the Y159H variant could have a deleterious effect comparable to the known damaging I157T

variant. Currently, the Y159H variant is considered a variant of uncertain significance (VUS) in the ClinVar database. This is in part due to limited functional studies showing a lack of growth perturbation in yeast with this variant[132]. The American College of Medical Genetics and Genomics (ACMG) provides criteria for the classification of genetic variants. Prior to performing our functional assay, we evaluated the Y159H variant as meeting the ACMG criteria PM1 (located in a critical protein domain), PM2 (absent/rare in normal control databases), and PP3 (*in silico* prediction), giving an overall score of VUS (2 moderate, 1 supporting criteria) similarly to the ClinVar assessment[133].

Figure 14: Graphical representation of the CHK2 protein and domains

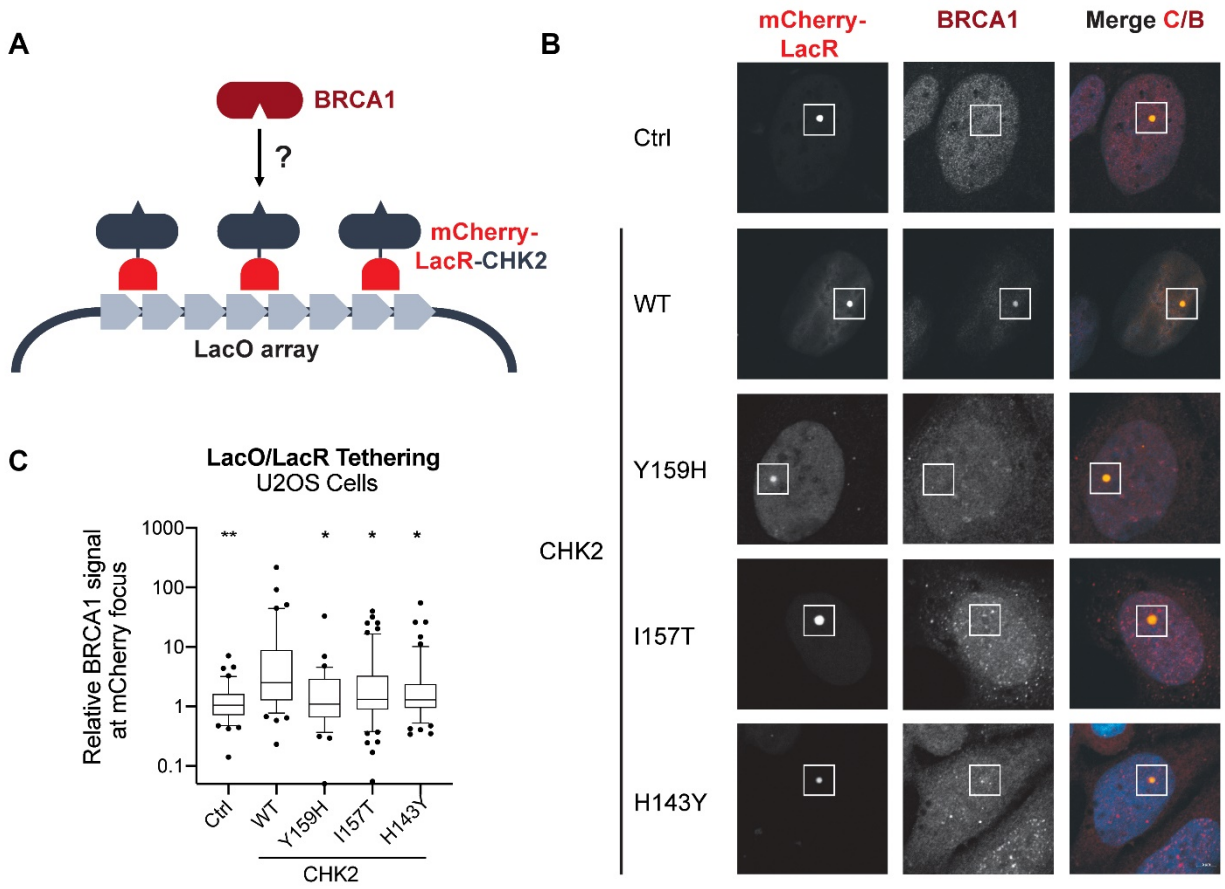


The CHK2 protein and its domains are illustrated. A selection of five pathogenic mutations described in the literature are shown in red. The Y159H variant identified in this case report is shown in black, highlighting its proximity to I157T. Figure redrawn using data from Cai et al. 2009[134] and Stolarova et al. 2020[135].

To test our hypothesis that CHK2 Y159H impairs binding to BRCA1, we used a LacO/LacR system. This assay allows for the evaluation of the association of two proteins at an integrated LacO array by tethering an mCherry-LacR tagged version of the protein of interest (CHK2). A detailed description of the technique is provided in the previous chapter (Methods). We first confirmed that our mCherry-LacR-tagged version of WT CHK2 could recruit endogenous BRCA1 to the LacO array. We then compared the ability of the Y159H variant to bind BRCA1 using the I157T and H143Y as positive controls, and WT CHK2 as a negative control. I157T is known to disrupt the FHA domain and H143Y produces an unstable protein product[127, 130]. Our results showed that the Y159H variant strongly impairs CHK2 association to BRCA1 compared to WT CHK2 (**Figure 15**). This impaired binding was similar in degree to the known pathogenic

variants (I157T, H143Y). This strongly suggests Y159H is disruptive to the FHA domain. Our data supports the reclassification of CHK2 Y159H as likely pathogenic based on this additional functional data. Our novel interpretation of this clinical variant was deposited to the ClinVar database under accession number (SCV001593269).

Figure 15: Effect of CHK2 variants on BRCA1 binding



A, Graphical representation of the LacO/LacR system. **B,** Confocal microscopy images. The LacO array position is identified by the accumulation of mCherry signal. Colocalization of BRCA1 and CHK2 is assessed at this mCherry focus (white square) by measuring the AF-647 signal (bound to endogenous BRCA1). The control (Ctrl) condition represents an mCherry-LacR construct lacking CHK2. **C,** Relative AF-647 signal (representing BRCA1) at the mCherry focus for WT and variant CHK2. Statistical significance was assessed using one-way ANOVA (Dunnett's multiple comparisons test) using WT CHK2 as the comparator (* = $p \leq 0.05$, ** = $p \leq 0.01$).

Abbreviations: AF-647, Alexa Fluor 647; BRCA1, breast cancer type 1 susceptibility protein; CHK2, checkpoint kinase 2; WT, wild-type.

DISCUSSION

Sequential transformation of MDS to CMML to B-ALL is exceedingly rare. We sought to combine FACS and WES to elucidate the genetic underpinning of such a case. Using this approach, we identified 11 suspected driver mutations with a varied distribution between the sorted leukemic populations and a buccal swab. In addition, we identified a previously poorly characterized *CHEK2* variant and performed functional validation to evaluate its pathogenicity. Finally, we proposed a putative sequence of clonal evolution applicable to this patient case (**Figure 12**).

Germline variants affecting genes implicated in cancer are of interest to researchers as they may represent inherited predispositions to a wide range of malignancies. This has implications regarding genetic counselling and cancer screening. In our patient case, we identified four germline variants meeting retention criteria. Three of the affected genes (*CDKN1A*, *BMPR1A*, and *FOXA2*) have been implicated in human malignancies to a certain degree[136, 137, 138]. However, we could not find any evidence that the specific variants identified in this case were pathogenic beyond *in silico* prediction tools. We therefore chose to focus our study on the *CHEK2* variant because this gene is considered a multiorgan cancer susceptibility gene, albeit with a low to moderate penetrance[135, 139]. *CHEK2* encodes the protein CHK2, which binds and activates BRCA1 to promote the DNA damage response (homologous recombination repair pathway). Other functions that have been associated with CHK2 include cell cycle arrest, apoptosis, and senescence[139]. We hypothesized *CHEK2* c.475T>C, p.Y159H could be a damaging mutation based on *in silico* tools and its proximity to the I157T variant that is known to disrupt CHK2 binding to BRCA1, p53, and CDC25[127]. Case control studies have shown that the I157T variant is a risk factor for hematological malignancies including MDS and ET[140, 141]. Furthermore, recently published functional studies demonstrated that I157T leads to hematopoietic stem and progenitor cell (HSPC) expansion, representing a mechanism by which this variant may promote hematological malignancies[142]. In this project, we generated functional data using a LacO/LacR system showing that Y159H impairs binding to BRCA1 similarly to I157T. Although we did not evaluate the entire spectrum of CHK2 function, our findings suggest Y159H likely interferes with the DNA damage response and may represent a cancer susceptibility allele. The Y159H variant is currently considered a VUS based on ACMG criteria. Our functional assay results would qualify as an additional supporting criterion for pathogenicity (PS3_supporting)[143]. Assuming our findings can be reproduced by other

laboratories, this could reclassify Y159H as likely pathogenic (2 moderate, 2 supporting ACMG criteria)[133]. Further work is required to fully elucidate the effects of Y159H.

We considered variants shared between the CMML and B-ALL clones, but absent in the buccal swab, to likely represent somatic mutations acquired earlier in the course of the disease (prior to the emergence of the B-ALL clone). The variants in *TET2*, *EZH2*, and *ZRSR2* are highly typical of MDS. These genes are involved in cellular processes that are frequently dysregulated in MDS[4] (DNA methylation, histone modification, and RNA splicing, respectively). The finding of a variant in the circadian regulator *CLOCK* was intriguing. Little is known of the role of this gene in hematological malignancies. However, there is increasing evidence to suggest that the circadian clock acts as a tumor suppressor via modulation of the immune system, apoptosis, and cellular proliferation[144]. An increased risk of cancer has been identified in individuals who experience altered circadian rhythms from night-shift work[145]. *NRAS* mutations are rare in low-risk MDS but frequent in CMML[4, 122]. The NRAS G12D variant we identified is known to promote aberrant cellular proliferation and survival[146]. We believe this variant was acquired at the MDS to CMML transition in the patient's history, when he transformed from a low-risk, indolent MDS to the more proliferative CMML. This remains speculative however as we did not have access to samples drawn earlier in the disease course to confirm this hypothesis.

Only two variants were found to be unique to a single clone. It is reasonable to assume these variants emerged later in the disease course, after divergence into distinct CMML and B-ALL clones. The B-ALL clone harbored a unique 3-bp insertion in the gene *BRD2*, resulting in the in-frame addition of a single amino acid. *BRD2* encodes the BET protein family member BRD2 that acts as a transcriptional regulator[147]. Experiments performed in murine models have shown that overexpression of BRD2 induces B cell malignancies[148], but the role of this protein in human cancers is poorly defined. We believe this *BRD2* variant is unlikely to be solely responsible for the lineage shift that was observed in this patient. It is possible other driver mutations unique to the B-ALL clone were missed due to the imperfect coverage obtained with WES. Alternatively, the myeloid-to-lymphoid shift could be explained by additional karyotypic abnormalities beyond -Y that were missed with G-band karyotyping. However, the more likely explanation, in our view, is that the lineage shift may have been induced by epigenetic alterations. Epigenetics are increasingly recognized as critical in the pathobiology of cancer[149]. Normal hematopoietic progenitors undergo

progressive epigenetic changes (including DNA methylation and histone modifications) as they progress through lineage determination and differentiation[150], and it is likely similar processes are active in malignant hematopoiesis. Our case notably demonstrated variants in two key epigenetic regulators (*TET2* and *EZH2*), which may have contributed to epigenetic alterations leading to lineage switch. Unfortunately, our study design could not directly evaluate epigenetic state in the various cell populations. The CMML clone contained one unique variant in the *ASXL1* gene. Truncating mutations in exon 12 of *ASXL1*, such as the one we identified in this case, are very common in CMML (roughly 45% of cases)[151]. This variant has also been described as cooperating with the NRAS G12D variant (which was also observed in our patient) to promote accelerated leukemogenesis[152]. Mechanistically, loss of ASXL1 protein function leads to aberrant chromatin remodelling and activation of oncogenes such as *HOXA* and the PI3K/AKT pathway[152, 153]. Because *ASXL1* mutations are generally considered early events in leukemogenesis[154], we expected this variant would be shared between both the CMML and B-ALL clones. It is nonetheless possible that the subclone responsible for the emergence of the B-ALL clone simply did not carry the *ASXL1* variant, or that this variant was lost upon transformation to B-ALL due to lack of selective advantage.

Overall, the methodology we employed in this project allowed us to elucidate the genetic underpinnings of a case of complex and unusual hematological disease. We identified multiple potential driver mutations and proposed a model of branching clonal evolution from a common origin. We also identified a *CHEK2* variant (Y159H) that possibly represented an inherited cancer predisposition in this patient. The functional data we generated concerning this variant has the potential to reclassify it as likely pathogenic and represents a novel contribution in the field of cancer genetics and DNA repair.

CONCLUSIONS

Understanding molecular heterogeneity and mechanisms of disease resistance in leukemias represents a major challenge in the field of hematology. Multiparameter flow cytometry, FACS, scRNA-seq, and WES are powerful technologies to tackle such questions.

In the first project, we created a 17-color AML MRD flow cytometry panel. This represents the acquisition of considerably more parameters than most panels currently in use for this purpose, which are usually 8 to 10 colors[47, 51]. The 17-color panel has the advantages of providing a highly detailed immunophenotype while maximizing the number of events that can be acquired for each sample. As clinical cytometers gain the ability to detect over 10 colors, it is likely larger panels such as ours will make their way into the clinic. With the increasing complexity of such panels, bioinformatic analysis tools will become important adjuncts in data interpretation. We expanded upon the FlowSOM approach described by Lacombe et al.[81] by applying it to a diverse cohort of AML remission samples and adding a statistical analysis method allowing for the rapid identification of the most relevant remission-emergent nodes. Although a major goal of our project was to analyze the transcriptome of AML MRD blasts at the single-cell level, we encountered significant challenges in obtaining adequate cDNA libraries from the cryopreserved sorted cells. This was likely due to RNA degradation related to the stress of freeze-thaw cycles, lengthy sample preparation, and cell sorting. However, a small percentage of cells did yield adequate libraries, suggesting our approach may be feasible with protocol modification. This would likely require the use of fresh (as opposed to cryopreserved) samples. This would be challenging from a logistical point of view but nonetheless possible in our opinion. We believe this is worth pursuing as the generation of single-cell transcriptomic data from AML MRD cells has the potential to identify targetable vulnerabilities that may lead to novel therapeutics.

In the second project, we combined FACS and WES to establish a complex pattern of clonal evolution underlying a case of MDS sequentially transforming into CMML and B-ALL. We identified many known pathogenic variants driving leukemogenesis in this case (*NRAS*, *TET2*, *EZH2*, *ZRSR2*, and *ASXL1*). Although such mutations are not currently directly actionable[155], this type of case-level knowledge may eventually be used for the selection of targeted treatments, fulfilling the goal of personalized medicine. The most notable contribution of this project to the medical literature was the identification of a germline *CHEK2* variant (c.475T>C, p.Y159H).

Although this variant had been identified previously in cancer patients, its biological effects were largely unknown beyond some preliminary yeast studies, and it was considered a VUS by most laboratories (ClinVar). The functional data we generated demonstrates impaired binding of this variant protein to one of its principal downstream effectors (BRCA1). Assuming our results can be reproduced by other laboratories, this finding has the potential to reclassify the Y159H variant as likely pathogenic by ACMG criteria. Further functional and case control studies will be required to fully establish this variant as an inherited cancer predisposition allele.

REFERENCES

1. Bazinet A, Heath J, Chong AS, Simo-Cheyou ER, Worme S, Rivera Polo B, et al. Common clonal origin of chronic myelomonocytic leukemia and B cell acute lymphoblastic leukemia in a patient with a germline CHEK2 variant. Cold Spring Harb Mol Case Stud. 2021.
2. Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, et al. Genomic Classification and Prognosis in Acute Myeloid Leukemia. N Engl J Med. 2016;374(23):2209-21.
3. Giustacchini A, Thongjuea S, Barkas N, Woll PS, Povinelli BJ, Booth CAG, et al. Single-cell transcriptomics uncovers distinct molecular signatures of stem cells in chronic myeloid leukemia. Nat Med. 2017;23(6):692-702.
4. Ogawa S. Genetics of MDS. Blood. 2019;133(10):1049-59.
5. Chopra M, Bohlander SK. The cell of origin and the leukemia stem cell in acute myeloid leukemia. Genes Chromosomes Cancer. 2019;58(12):850-8.
6. Goardon N, Marchi E, Atzberger A, Quek L, Schuh A, Soneji S, et al. Coexistence of LMPP-like and GMP-like leukemia stem cells in acute myeloid leukemia. Cancer Cell. 2011;19(1):138-52.
7. Boddu PC, Kadia TM, Garcia-Manero G, Cortes J, Alfayez M, Borthakur G, et al. Validation of the 2017 European LeukemiaNet classification for acute myeloid leukemia with NPM1 and FLT3-internal tandem duplication genotypes. Cancer. 2019;125(7):1091-100.
8. Dohner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Buchner T, et al. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood. 2017;129(4):424-47.
9. Howlader N, Noone AM, Krapcho M, Miller D, Brest A, Yu M, et al. SEER Cancer Statistics Review, 1975-2017. Bethesda, MD. 2020 [cited 2021 Mar 31]. Available from: https://seer.cancer.gov/csr/1975_2017/.
10. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. N Engl J Med. 2015;373(12):1136-52.
11. Medeiros BC, Satram-Hoang S, Hurst D, Hoang KQ, Momin F, Reyes C. Big data analysis of treatment patterns and outcomes among elderly acute myeloid leukemia patients in the United States. Ann Hematol. 2015;94(7):1127-38.
12. Bazinet A, Assouline S. A review of FDA-approved acute myeloid leukemia therapies beyond '7 + 3'. Expert Rev Hematol. 2021;14(2):185-97.
13. Steensma DP. Clinical consequences of clonal hematopoiesis of indeterminate potential. Blood Adv. 2018;2(22):3404-10.
14. Genovese G, Kahler AK, Handsaker RE, Lindberg J, Rose SA, Bakhoum SF, et al. Clonal hematopoiesis and blood-cancer risk inferred from blood DNA sequence. N Engl J Med. 2014;371(26):2477-87.
15. Jaiswal S, Fontanillas P, Flannick J, Manning A, Grauman PV, Mar BG, et al. Age-related clonal hematopoiesis associated with adverse outcomes. N Engl J Med. 2014;371(26):2488-98.
16. Xie M, Lu C, Wang J, McLellan MD, Johnson KJ, Wendl MC, et al. Age-related mutations associated with clonal hematopoietic expansion and malignancies. Nat Med. 2014;20(12):1472-8.

17. Jaiswal S, Natarajan P, Silver AJ, Gibson CJ, Bick AG, Shvartz E, et al. Clonal Hematopoiesis and Risk of Atherosclerotic Cardiovascular Disease. N Engl J Med. 2017;377(2):111-21.
18. Thomas D, Majeti R. Biology and relevance of human acute myeloid leukemia stem cells. Blood. 2017;129(12):1577-85.
19. Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, et al. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature. 1994;367(6464):645-8.
20. Bonnet D, Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med. 1997;3(7):730-7.
21. Taussig DC, Miraki-Moud F, Anjos-Afonso F, Pearce DJ, Allen K, Ridler C, et al. Anti-CD38 antibody-mediated clearance of human repopulating cells masks the heterogeneity of leukemia-initiating cells. Blood. 2008;112(3):568-75.
22. Taussig DC, Vargaftig J, Miraki-Moud F, Griessinger E, Sharrock K, Luke T, et al. Leukemia-initiating cells from some acute myeloid leukemia patients with mutated nucleophosmin reside in the CD34(-) fraction. Blood. 2010;115(10):1976-84.
23. Eppert K, Takenaka K, Lechman ER, Waldron L, Nilsson B, van Galen P, et al. Stem cell gene expression programs influence clinical outcome in human leukemia. Nat Med. 2011;17(9):1086-93.
24. Ishikawa F, Yoshida S, Saito Y, Hijikata A, Kitamura H, Tanaka S, et al. Chemotherapy-resistant human AML stem cells home to and engraft within the bone-marrow endosteal region. Nat Biotechnol. 2007;25(11):1315-21.
25. Lagadinou ED, Sach A, Callahan K, Rossi RM, Neering SJ, Minhajuddin M, et al. BCL-2 inhibition targets oxidative phosphorylation and selectively eradicates quiescent human leukemia stem cells. Cell Stem Cell. 2013;12(3):329-41.
26. Ng SW, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, et al. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature. 2016;540(7633):433-7.
27. Farge T, Saland E, de Toni F, Aroua N, Hosseini M, Perry R, 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):716-35.
28. Boyd AL, Aslostovar L, Reid J, Ye W, Tanasijevic B, Porras DP, et al. Identification of Chemotherapy-Induced Leukemic-Regenerating Cells Reveals a Transient Vulnerability of Human AML Recurrence. Cancer Cell. 2018;34(3):483-98 e5.
29. Grove CS, Vassiliou GS. Acute myeloid leukaemia: a paradigm for the clonal evolution of cancer? Dis Model Mech. 2014;7(8):941-51.
30. Ediriwickrema A, Aleshin A, Reiter JG, Corces MR, Kohnke T, Stafford M, et al. Single-cell mutational profiling enhances the clinical evaluation of AML MRD. Blood Adv. 2020;4(5):943-52.
31. Ding L, Ley TJ, Larson DE, Miller CA, Koboldt DC, Welch JS, et al. Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing. Nature. 2012;481(7382):506-10.
32. Cancer Genome Atlas Research N, Ley TJ, Miller C, Ding L, Raphael BJ, Mungall AJ, et al. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. N Engl J Med. 2013;368(22):2059-74.

33. Khwaja A, Bjorkholm M, Gale RE, Levine RL, Jordan CT, Ehninger G, et al. Acute myeloid leukaemia. Nat Rev Dis Primers. 2016;2:16010.
34. Bakst RL, Tallman MS, Douer D, Yahalom J. How I treat extramedullary acute myeloid leukemia. Blood. 2011;118(14):3785-93.
35. Swerdlow SH, Campo E, Lee Harris N, Jaffe ES, Pileri SA, Stein H, et al. WHO Classification of Tumours of Haematopoietic and Lymphoid Tissues. Lyon: International Agency for Research on Cancer; 2017. 585 p.
36. Bennett JM, Catovsky D, Daniel MT, Flandrin G, Galton DA, Gralnick HR, et al. Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group. Br J Haematol. 1976;33(4):451-8.
37. Stone RM, Mandrekar SJ, Sanford BL, Laumann K, Geyer S, Bloomfield CD, et al. Midostaurin plus Chemotherapy for Acute Myeloid Leukemia with a FLT3 Mutation. N Engl J Med. 2017;377(5):454-64.
38. Lambert J, Pautas C, Terre C, Raffoux E, Turlure P, Caillot D, et al. Gemtuzumab ozogamicin for de novo acute myeloid leukemia: final efficacy and safety updates from the open-label, phase III ALFA-0701 trial. Haematologica. 2019;104(1):113-9.
39. Hills RK, Castaigne S, Appelbaum FR, Delaunay J, Petersdorf S, Othus M, et al. Addition of gemtuzumab ozogamicin to induction chemotherapy in adult patients with acute myeloid leukaemia: a meta-analysis of individual patient data from randomised controlled trials. Lancet Oncol. 2014;15(9):986-96.
40. DiNardo CD, Jonas BA, Pullarkat V, Thirman MJ, Garcia JS, Wei AH, et al. Azacitidine and Venetoclax in Previously Untreated Acute Myeloid Leukemia. N Engl J Med. 2020;383(7):617-29.
41. Thol F, Ganser A. Treatment of Relapsed Acute Myeloid Leukemia. Curr Treat Options Oncol. 2020;21(8):66.
42. Koenig K, Mims A. Relapsed or primary refractory AML: moving past MEC and FLAG-ida. Curr Opin Hematol. 2020;27(2):108-14.
43. Schuurhuis GJ, Heuser M, Freeman S, Bene MC, Buccisano F, Cloos J, et al. Minimal/measurable residual disease in AML: a consensus document from the European LeukemiaNet MRD Working Party. Blood. 2018;131(12):1275-91.
44. Young AL, Wong TN, Hughes AE, Heath SE, Ley TJ, Link DC, et al. Quantifying ultra-rare pre-leukemic clones via targeted error-corrected sequencing. Leukemia. 2015;29(7):1608-11.
45. Gabert J, Beillard E, van der Velden VH, Bi W, Grimwade D, Pallisgaard N, et al. Standardization and quality control studies of 'real-time' quantitative reverse transcriptase polymerase chain reaction of fusion gene transcripts for residual disease detection in leukemia - a Europe Against Cancer program. Leukemia. 2003;17(12):2318-57.
46. Ivey A, Hills RK, Simpson MA, Jovanovic JV, Gilkes A, Grech A, et al. Assessment of Minimal Residual Disease in Standard-Risk AML. N Engl J Med. 2016;374(5):422-33.
47. Xu J, Jorgensen JL, Wang SA. How Do We Use Multicolor Flow Cytometry to Detect Minimal Residual Disease in Acute Myeloid Leukemia? Clin Lab Med. 2017;37(4):787-802.
48. Al-Mawali A, Gillis D, Hissaria P, Lewis I. Incidence, sensitivity, and specificity of leukemia-associated phenotypes in acute myeloid leukemia using specific five-color multiparameter flow cytometry. Am J Clin Pathol. 2008;129(6):934-45.

49. Rossi G, Giambra V, Minervini MM, De Waure C, Mancinelli S, Ciavarella M, et al. Leukemia-associated immunophenotypes subdivided in "categories of specificity" improve the sensitivity of minimal residual disease in predicting relapse in acute myeloid leukemia. Cytometry B Clin Cytom. 2020;98(3):216-25.
50. Baer MR, Stewart CC, Dodge RK, Leget G, Sule N, Mrozek K, et al. High frequency of immunophenotype changes in acute myeloid leukemia at relapse: implications for residual disease detection (Cancer and Leukemia Group B Study 8361). Blood. 2001;97(11):3574-80.
51. Wood BL. Acute Myeloid Leukemia Minimal Residual Disease Detection: The Difference from Normal Approach. Curr Protoc Cytom. 2020;93(1):e73.
52. Jaso JM, Wang SA, Jorgensen JL, Lin P. Multi-color flow cytometric immunophenotyping for detection of minimal residual disease in AML: past, present and future. Bone Marrow Transplant. 2014;49(9):1129-38.
53. Jongen-Lavrencic M, Grob T, Hanekamp D, Kavelaars FG, Al Hinai A, Zeilemaker A, et al. Molecular Minimal Residual Disease in Acute Myeloid Leukemia. N Engl J Med. 2018;378(13):1189-99.
54. Hasserjian RP, Steensma DP, Graubert TA, Ebert BL. Clonal hematopoiesis and measurable residual disease assessment in acute myeloid leukemia. Blood. 2020;135(20):1729-38.
55. Terwijn M, van Putten WL, Kelder A, van der Velden VH, Brooimans RA, Pabst T, et al. High prognostic impact of flow cytometric minimal residual disease detection in acute myeloid leukemia: data from the HOVON/SAKK AML 42A study. J Clin Oncol. 2013;31(31):3889-97.
56. Freeman SD, Virgo P, Couzens S, Grimwade D, Russell N, Hills RK, et al. Prognostic relevance of treatment response measured by flow cytometric residual disease detection in older patients with acute myeloid leukemia. J Clin Oncol. 2013;31(32):4123-31.
57. Grimwade D, Jovanovic JV, Hills RK, Nugent EA, Patel Y, Flora R, et al. Prospective minimal residual disease monitoring to predict relapse of acute promyelocytic leukemia and to direct pre-emptive arsenic trioxide therapy. J Clin Oncol. 2009;27(22):3650-8.
58. Yin JA, O'Brien MA, Hills RK, Daly SB, Wheatley K, Burnett AK. Minimal residual disease monitoring by quantitative RT-PCR in core binding factor AML allows risk stratification and predicts relapse: results of the United Kingdom MRC AML-15 trial. Blood. 2012;120(14):2826-35.
59. Press RD, Eickelberg G, Froman A, Yang F, Stentz A, Flatley EM, et al. Next-generation sequencing-defined minimal residual disease before stem cell transplantation predicts acute myeloid leukemia relapse. Am J Hematol. 2019;94(8):902-12.
60. Thol F, Gabdoulline R, Liebich A, Klement P, Schiller J, Kandziora C, et al. Measurable residual disease monitoring by NGS before allogeneic hematopoietic cell transplantation in AML. Blood. 2018;132(16):1703-13.
61. Walter RB, Buckley SA, Pagel JM, Wood BL, Storer BE, Sandmaier BM, et al. Significance of minimal residual disease before myeloablative allogeneic hematopoietic cell transplantation for AML in first and second complete remission. Blood. 2013;122(10):1813-21.
62. Ngai LL, Kelder A, Janssen J, Ossenkoppele GJ, Cloos J. MRD Tailored Therapy in AML: What We Have Learned So Far. Front Oncol. 2020;10:603636.

63. Venditti A, Piciocchi A, Candoni A, Melillo L, Calafiore V, Cairoli R, et al. GIMEMA AML1310 trial of risk-adapted, MRD-directed therapy for young adults with newly diagnosed acute myeloid leukemia. Blood. 2019;134(12):935-45.
64. Zhu HH, Zhang XH, Qin YZ, Liu DH, Jiang H, Chen H, et al. MRD-directed risk stratification treatment may improve outcomes of t(8;21) AML in the first complete remission: results from the AML05 multicenter trial. Blood. 2013;121(20):4056-62.
65. Platzbecker U, Middeke JM, Sockel K, Herbst R, Wolf D, Baldus CD, et al. Measurable residual disease-guided treatment with azacitidine to prevent haematological relapse in patients with myelodysplastic syndrome and acute myeloid leukaemia (RELAZA2): an open-label, multicentre, phase 2 trial. Lancet Oncol. 2018;19(12):1668-79.
66. Short NJ, Ravandi F. How close are we to incorporating measurable residual disease into clinical practice for acute myeloid leukemia? Haematologica. 2019;104(8):1532-41.
67. Shlush LI, Mitchell A, Heisler L, Abelson S, Ng SWK, Trotman-Grant A, et al. Tracing the origins of relapse in acute myeloid leukaemia to stem cells. Nature. 2017;547(7661):104-8.
68. Ho TC, LaMere M, Stevens BM, Ashton JM, Myers JR, O'Dwyer KM, et al. Evolution of acute myelogenous leukemia stem cell properties after treatment and progression. Blood. 2016;128(13):1671-8.
69. Aslostovar L, Boyd AL, Almakadi M, Collins TJ, Leong DP, Tirona RG, et al. A phase 1 trial evaluating thioridazine in combination with cytarabine in patients with acute myeloid leukemia. Blood Adv. 2018;2(15):1935-45.
70. Issa GC, Benton CB, Mohanty V, Shen Y, Alaniz Z, Wang F, et al. Identification of Gene Expression Signatures in Leukemia Stem Cells and Minimal Residual Disease Following Treatment of Adverse Risk Acute Myeloid Leukemia. Blood. 2019;134(Supplement_1):2717-.
71. Duy C, Li M, Teater M, Meydan C, Garrett-Bakelman FE, Lee TC, et al. Chemotherapy induces senescence-like resilient cells capable of initiating AML recurrence. Cancer Discov. 2021.
72. Hwang B, Lee JH, Bang D. Single-cell RNA sequencing technologies and bioinformatics pipelines. Exp Mol Med. 2018;50(8):1-14.
73. Choi JR, Yong KW, Choi JY, Cowie AC. Single-Cell RNA Sequencing and Its Combination with Protein and DNA Analyses. Cells. 2020;9(5).
74. Picelli S, Faridani OR, Bjorklund AK, Winberg G, Sagasser S, Sandberg R. Full-length RNA-seq from single cells using Smart-seq2. Nat Protoc. 2014;9(1):171-81.
75. Rodriguez-Meira A, Buck G, Clark SA, Povinelli BJ, Alcolea V, Louka E, et al. Unravelling Intratumoral Heterogeneity through High-Sensitivity Single-Cell Mutational Analysis and Parallel RNA Sequencing. Mol Cell. 2019;73(6):1292-305 e8.
76. Euroflow Consortium. List of reference and alternative antibodies for hematological malignancy panels Version 1.11. Leiden, Netherlands. 2019 [cited 2021 Apr 22]. Available from: https://www.euroflow.org/usr/pub/protocols.php.
77. Telford WG, Babin SA, Khorev SV, Rowe SH. Green fiber lasers: an alternative to traditional DPSS green lasers for flow cytometry. Cytometry A. 2009;75(12):1031-9.
78. Monaco G, Chen H, Poidinger M, Chen J, de Magalhaes JP, Larbi A. flowAI: automatic and interactive anomaly discerning tools for flow cytometry data. Bioinformatics. 2016;32(16):2473-80.

79. R Core Team. R: A language and environment for statistical computing. Vienna, Austria. 2020 [cited 2021 Mar 23]. Available from: https://www.R-project.org/.
80. Hahne F, LeMeur N, Brinkman RR, Ellis B, Haaland P, Sarkar D, et al. flowCore: a Bioconductor package for high throughput flow cytometry. BMC Bioinformatics. 2009;10:106.
81. Lacombe F, Lechevalier N, Vial JP, Bene MC. An R-Derived FlowSOM Process to Analyze Unsupervised Clustering of Normal and Malignant Human Bone Marrow Classical Flow Cytometry Data. Cytometry A. 2019;95(11):1191-7.
82. Van Gassen S, Callebaut B, Van Helden MJ, Lambrecht BN, Demeester P, Dhaene T, et al. FlowSOM: Using self-organizing maps for visualization and interpretation of cytometry data. Cytometry A. 2015;87(7):636-45.
83. Wood B. Multicolor immunophenotyping: human immune system hematopoiesis. Methods Cell Biol. 2004;75:559-76.
84. Bewick V, Cheek L, Ball J. Statistics review 12: survival analysis. Crit Care. 2004;8(5):389-94.
85. Broad Institute. GATK Best Practices Workflows: Somatic short variant discovery. Cambridge, MA. 2020 [cited 2020 May 14]. Available from: https://gatk.broadinstitute.org/hc/en-us/articles/360035894731-Somatic-short-variant-discovery-SNVs-Indels-.
86. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25(14):1754-60.
87. Cibulskis K, Lawrence MS, Carter SL, Sivachenko A, Jaffe D, Sougnez C, et al. Sensitive detection of somatic point mutations in impure and heterogeneous cancer samples. Nat Biotechnol. 2013;31(3):213-9.
88. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010;38(16):e164.
89. de Kock L, Wu MK, Foulkes WD. Ten years of DICER1 mutations: Provenance, distribution, and associated phenotypes. Hum Mutat. 2019;40(11):1939-53.
90. Hennig BP, Velten L, Racke I, Tu CS, Thoms M, Rybin V, et al. Large-Scale Low-Cost NGS Library Preparation Using a Robust Tn5 Purification and Tagmentation Protocol. G3 (Bethesda). 2018;8(1):79-89.
91. Stuart T, Satija R. Integrative single-cell analysis. Nat Rev Genet. 2019;20(5):257-72.
92. Subramanian A, Tamayo P, Mootha VK, Mukherjee S, Ebert BL, Gillette MA, 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):15545-50.
93. Breton G, Lee J, Liu K, Nussenzweig MC. Defining human dendritic cell progenitors by multiparametric flow cytometry. Nat Protoc. 2015;10(9):1407-22.
94. Murphy RF. Automated identification of subpopulations in flow cytometric list mode data using cluster analysis. Cytometry. 1985;6(4):302-9.
95. Gong P, Metrebian F, Dulau-Florea A, Wang Z-X, Bajaj R, Gulati G, et al. Aberrant expression of CD56 on granulocytes and monocytes in myeloproliferative neoplasm. Journal of Hematopathology. 2013;6(3):127-34.
96. Loken MR, van de Loosdrecht A, Ogata K, Orfao A, Wells DA. Flow cytometry in myelodysplastic syndromes: report from a working conference. Leuk Res. 2008;32(1):5-17.

97. Bland JM, Altman DG. Survival probabilities (the Kaplan-Meier method). BMJ. 1998;317(7172):1572.
98. Rawstron AC, Fazi C, Agathangelidis A, Villamor N, Letestu R, Nomdedeu J, et al. A complementary role of multiparameter flow cytometry and high-throughput sequencing for minimal residual disease detection in chronic lymphocytic leukemia: an European Research Initiative on CLL study. Leukemia. 2016;30(4):929-36.
99. Rawstron AC, Villamor N, Ritgen M, Bottcher S, Ghia P, Zehnder JL, et al. International standardized approach for flow cytometric residual disease monitoring in chronic lymphocytic leukaemia. Leukemia. 2007;21(5):956-64.
100. Theunissen P, Mejstrikova E, Sedek L, van der Sluijs-Gelling AJ, Gaipa G, Bartels M, et al. Standardized flow cytometry for highly sensitive MRD measurements in B-cell acute lymphoblastic leukemia. Blood. 2017;129(3):347-57.
101. Jiang S, He F, Gao L, Chen A, Hu Y, Fan L, et al. Identification of Chemo-Resistant Residual Cell Population in Pediatric AML of Complete Remission By Single Cell RNA Sequencing. 62nd ASH Annual Meeting and Exposition; Dec 5-8 2020; Online.
102. Cogle CR. Incidence and Burden of the Myelodysplastic Syndromes. Curr Hematol Malig Rep. 2015;10(3):272-81.
103. Cazzola M. Myelodysplastic Syndromes. N Engl J Med. 2020;383(14):1358-74.
104. Goldberg SL, Chen E, Corral M, Guo A, Mody-Patel N, Pecora AL, et al. Incidence and clinical complications of myelodysplastic syndromes among United States Medicare beneficiaries. J Clin Oncol. 2010;28(17):2847-52.
105. Jansen AJ, Essink-Bot ML, Beckers EA, Hop WC, Schipperus MR, Van Rhenen DJ. Quality of life measurement in patients with transfusion-dependent myelodysplastic syndromes. Br J Haematol. 2003;121(2):270-4.
106. Kantarjian H, Giles F, List A, Lyons R, Sekeres MA, Pierce S, et al. The incidence and impact of thrombocytopenia in myelodysplastic syndromes. Cancer. 2007;109(9):1705-14.
107. Greenberg P, Cox C, LeBeau MM, Fenaux P, Morel P, Sanz G, et al. International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood. 1997;89(6):2079-88.
108. Greenberg PL, Tuechler H, Schanz J, Sanz G, Garcia-Manero G, Sole F, et al. Revised international prognostic scoring system for myelodysplastic syndromes. Blood. 2012;120(12):2454-65.
109. Porwit A, Saft L. The AML–MDS interface—leukemic transformation in myelodysplastic syndromes. Journal of Hematopathology. 2011;4(2):69-79.
110. Martins F, Kruszewski M, Scarpelli I, Schoumans J, Spertini O, Lubbert M, et al. Characterization of myelodysplastic syndromes progressing to acute lymphoblastic leukemia. Ann Hematol. 2021;100(1):63-78.
111. Disperati P, Ichim CV, Tkachuk D, Chun K, Schuh AC, Wells RA. Progression of myelodysplasia to acute lymphoblastic leukaemia: implications for disease biology. Leuk Res. 2006;30(2):233-9.
112. Guo ZP, Tan YH, Li JL, Xu ZF, Chen XH, Xu LR. Acute pro-B-Cell lymphoblastic leukemia transformed from myelodysplastic syndrome with an ASXL1 missense mutation: A case report with literature review. Oncol Lett. 2018;15(6):9745-50.
113. Nilsson L, Astrand-Grundstrom I, Arvidsson I, Jacobsson B, Hellstrom-Lindberg E, Hast R, et al. Isolation and characterization of hematopoietic progenitor/stem cells in 5q-deleted

myelodysplastic syndromes: evidence for involvement at the hematopoietic stem cell level. Blood. 2000;96(6):2012-21.
114. van Lom K, Hagemeijer A, Smit E, Hählen K, Groeneveld K, Löwenberg B. Cytogenetic clonality analysis in myelodysplastic syndrome: monosomy 7 can be demonstrated in the myeloid and in the lymphoid lineage. Leukemia. 1995;9(11):1818-21.
115. Zink F, Stacey SN, Norddahl GL, Frigge ML, Magnusson OT, Jonsdottir I, et al. Clonal hematopoiesis, with and without candidate driver mutations, is common in the elderly. Blood. 2017;130(6):742-52.
116. Jan M, Snyder TM, Corces-Zimmerman MR, Vyas P, Weissman IL, Quake SR, et al. Clonal evolution of preleukemic hematopoietic stem cells precedes human acute myeloid leukemia. Sci Transl Med. 2012;4(149):149ra18.
117. Quivoron C, Couronne L, Della Valle V, Lopez CK, Plo I, Wagner-Ballon O, et al. TET2 inactivation results in pleiotropic hematopoietic abnormalities in mouse and is a recurrent event during human lymphomagenesis. Cancer Cell. 2011;20(1):25-38.
118. Niscola P, Tendas A, Scaramucci L, Giovannini M, Fratoni S, de Fabritiis P. Evolution of chronic myelomonocytic leukemia from refractory anemia: the unusual course of a myelodysplastic syndrome. Blood Res. 2013;48(2):152-3.
119. Hasegawa Y, Sakai N, Toyama M, Ninomiya H, Abe T. [Chronic myelomonocytic leukemia transformed from refractory anemia with ring sideroblasts with a rare abnormal chromosome, inv (12)]. Rinsho Ketsueki. 1990;31(1):75-9.
120. Bursztyn B, Douer D, Ramot B. Chronic myelomonocytic leukemia following refractory anemia with sideroblasts: report of two cases. Eur J Haematol. 1987;38(2):197-9.
121. Breccia M, Cannella L, Frustaci A, Stefanizzi C, D'Elia GM, Alimena G. Chronic myelomonocytic leukemia with antecedent refractory anemia with excess blasts: further evidence for the arbitrary nature of current classification systems. Leuk Lymphoma. 2008;49(7):1292-6.
122. Meggendorfer M, Jeromin S, Haferlach C, Kern W, Haferlach T. The mutational landscape of 18 investigated genes clearly separates four subtypes of myelodysplastic/myeloproliferative neoplasms. Haematologica. 2018;103(5):e192-e5.
123. Haferlach T, Nagata Y, Grossmann V, Okuno Y, Bacher U, Nagae G, et al. Landscape of genetic lesions in 944 patients with myelodysplastic syndromes. Leukemia. 2014;28(2):241-7.
124. Yasinski E. Ash Clinical News: CMML: A Unique Overlap Syndrome Receiving Increased Attention. 2019 [cited May 5, 2021]. Available from: https://www.ashclinicalnews.org/chronic-leukemia/cmml-unique-overlap-syndrome-receiving-increased-attention/.
125. Tamborero D, Rubio-Perez C, Deu-Pons J, Schroeder MP, Vivancos A, Rovira A, et al. Cancer Genome Interpreter annotates the biological and clinical relevance of tumor alterations. Genome Med. 2018;10(1):25.
126. Nadaf J, Majewski J, Fahiminiya S. ExomeAI: detection of recurrent allelic imbalance in tumors using whole-exome sequencing data. Bioinformatics. 2015;31(3):429-31.
127. Li J, Williams BL, Haire LF, Goldberg M, Wilker E, Durocher D, et al. Structural and functional versatility of the FHA domain in DNA-damage signaling by the tumor suppressor kinase Chk2. Mol Cell. 2002;9(5):1045-54.

128. Orthwein A, Noordermeer SM, Wilson MD, Landry S, Enchev RI, Sherker A, et al. A mechanism for the suppression of homologous recombination in G1 cells. Nature. 2015;528(7582):422-6.
129. Tang J, Cho NW, Cui G, Manion EM, Shanbhag NM, Botuyan MV, et al. Acetylation limits 53BP1 association with damaged chromatin to promote homologous recombination. Nat Struct Mol Biol. 2013;20(3):317-25.
130. Bell DW, Kim SH, Godwin AK, Schiripo TA, Harris PL, Haserlat SM, et al. Genetic and functional analysis of CHEK2 (CHK2) variants in multiethnic cohorts. Int J Cancer. 2007;121(12):2661-7.
131. Findlay S, Heath J, Luo VM, Malina A, Morin T, Coulombe Y, et al. SHLD2/FAM35A co-operates with REV7 to coordinate DNA double-strand break repair pathway choice. EMBO J. 2018;37(18).
132. Delimitsou A, Fostira F, Kalfakakou D, Apostolou P, Konstantopoulou I, Kroupis C, et al. Functional characterization of CHEK2 variants in a Saccharomyces cerevisiae system. Hum Mutat. 2019;40(5):631-48.
133. Richards S, Aziz N, Bale S, Bick D, Das S, Gastier-Foster J, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genet Med. 2015;17(5):405-24.
134. Cai Z, Chehab NH, Pavletich NP. Structure and activation mechanism of the CHK2 DNA damage checkpoint kinase. Mol Cell. 2009;35(6):818-29.
135. Stolarova L, Kleiblova P, Janatova M, Soukupova J, Zemankova P, Macurek L, et al. CHEK2 Germline Variants in Cancer Predisposition: Stalemate Rather than Checkmate. Cells. 2020;9(12).
136. Calva-Cerqueira D, Chinnathambi S, Pechman B, Bair J, Larsen-Haidle J, Howe JR. The rate of germline mutations and large deletions of SMAD4 and BMPR1A in juvenile polyposis. Clin Genet. 2009;75(1):79-85.
137. Halmos B, Basseres DS, Monti S, D'Alo F, Dayaram T, Ferenczi K, et al. A transcriptional profiling study of CCAAT/enhancer binding protein targets identifies hepatocyte nuclear factor 3 beta as a novel tumor suppressor in lung cancer. Cancer Res. 2004;64(12):4137-47.
138. Kreis NN, Louwen F, Yuan J. The Multifaceted p21 (Cip1/Waf1/CDKN1A) in Cell Differentiation, Migration and Cancer Therapy. Cancers (Basel). 2019;11(9).
139. Antoni L, Sodha N, Collins I, Garrett MD. CHK2 kinase: cancer susceptibility and cancer therapy - two sides of the same coin? Nat Rev Cancer. 2007;7(12):925-36.
140. Janiszewska H, Bak A, Pilarska M, Heise M, Junkiert-Czarnecka A, Kuliszkiewicz-Janus M, et al. A risk of essential thrombocythemia in carriers of constitutional CHEK2 gene mutations. Haematologica. 2012;97(3):366-70.
141. Janiszewska H, Bak A, Skonieczka K, Jaskowiec A, Kielbinski M, Jachalska A, et al. Constitutional mutations of the CHEK2 gene are a risk factor for MDS, but not for de novo AML. Leuk Res. 2018;70:74-8.
142. Bao EL, Nandakumar SK, Liao X, Bick AG, Karjalainen J, Tabaka M, et al. Inherited myeloproliferative neoplasm risk affects haematopoietic stem cells. Nature. 2020;586(7831):769-75.

143. Brnich SE, Abou Tayoun AN, Couch FJ, Cutting GR, Greenblatt MS, Heinen CD, et al. Recommendations for application of the functional evidence PS3/BS3 criterion using the ACMG/AMP sequence variant interpretation framework. Genome Med. 2019;12(1):3.
144. Fu L, Lee CC. The circadian clock: pacemaker and tumour suppressor. Nat Rev Cancer. 2003;3(5):350-61.
145. Fu L, Kettner NM. The circadian clock in cancer development and therapy. Prog Mol Biol Transl Sci. 2013;119:221-82.
146. Prior IA, Lewis PD, Mattos C. A comprehensive survey of Ras mutations in cancer. Cancer Res. 2012;72(10):2457-67.
147. French CA. Small-Molecule Targeting of BET Proteins in Cancer. Adv Cancer Res. 2016;131:21-58.
148. Greenwald RJ, Tumang JR, Sinha A, Currier N, Cardiff RD, Rothstein TL, et al. E mu-BRD2 transgenic mice develop B-cell lymphoma and leukemia. Blood. 2004;103(4):1475-84.
149. Sharma S, Kelly TK, Jones PA. Epigenetics in cancer. Carcinogenesis. 2010;31(1):27-36.
150. Cedar H, Bergman Y. Epigenetics of haematopoietic cell development. Nat Rev Immunol. 2011;11(7):478-88.
151. Gelsi-Boyer V, Brecqueville M, Devillier R, Murati A, Mozziconacci MJ, Birnbaum D. Mutations in ASXL1 are associated with poor prognosis across the spectrum of malignant myeloid diseases. J Hematol Oncol. 2012;5:12.
152. Abdel-Wahab O, Adli M, LaFave LM, Gao J, Hricik T, Shih AH, et al. ASXL1 mutations promote myeloid transformation through loss of PRC2-mediated gene repression. Cancer Cell. 2012;22(2):180-93.
153. Cao L, Xia X, Kong Y, Jia F, Yuan B, Li R, et al. Deregulation of tumor suppressive ASXL1-PTEN/AKT axis in myeloid malignancies. J Mol Cell Biol. 2020;12(9):688-99.
154. Fujino T, Kitamura T. ASXL1 mutation in clonal hematopoiesis. Exp Hematol. 2020;83:74-84.
155. Carr TH, McEwen R, Dougherty B, Johnson JH, Dry JR, Lai Z, et al. Defining actionable mutations for oncology therapeutic development. Nat Rev Cancer. 2016;16(5):319-29.

APPENDICES

Tables

Appendix Table A1: Next-generation sequencing myeloid gene panel*

ABL1	MPL
ASXL1	MYD88
BRAF	NOTCH1
CALR	NPM1
CBL	NRAS
CEBPA	PDGFRA
CSF3R	PTEN
CXCR4	PTPN11
DNMT3A	RUNX1
EZH2	SETBP1
FLT3	SF3B1
HRAS	SRSF2
IDH1	TET2
IDH2	TP53
JAK2	U2AF1
KIT	WT1
KRAS	ZRSR2
KMT2D/MLL2	

*Courtesy of Dr. Yury Monczak

Appendix Table A2: QuikChange oligonucleotides used to generate the CHK2 variants

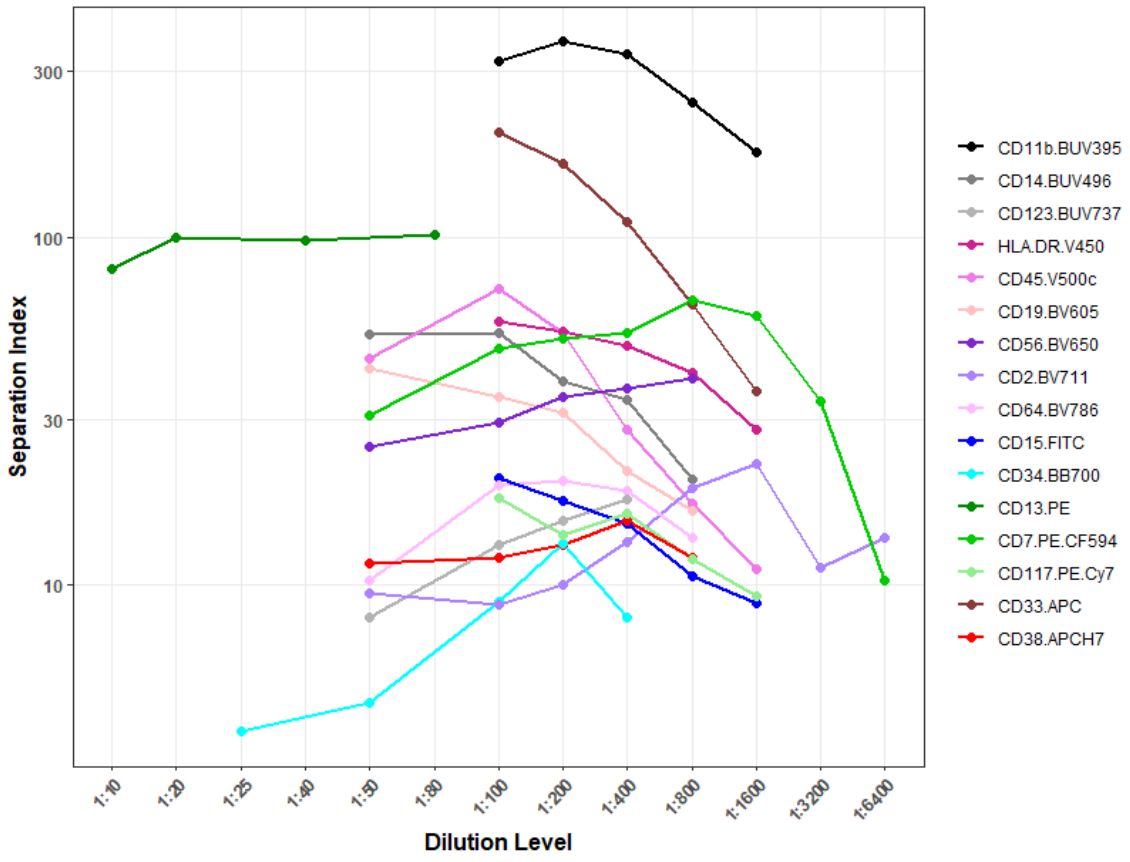
Variant	**Oligonucleotide pair**
Y159H	**F:** CCACTGTGATCTTCTATGTGTGCAATGTAAGAGTTTTTAGGAC **R:** GTCCTAAAAACTCTTACATTGCACACATAGAAGATCACAGTGG
I157T	**F:** CTTCTATGTATGCAGTGTAAGAGTTTTTAGGACCCACTTCCC **R:** GGGAAGTGGGTCCTAAAAACTCTTACACTGCATACATAGAAG
H143Y	**F:** TCCCTGAAAATCCGAAAATATTTCTTGCTGTATGTTCGGTATTTATCTGTT **R:** AACAGATAAATACCGAACATACAGCAAGAAATATTTTCGGATTTTCAGGGA

Appendix Table A3: Whole-exome sequencing metrics

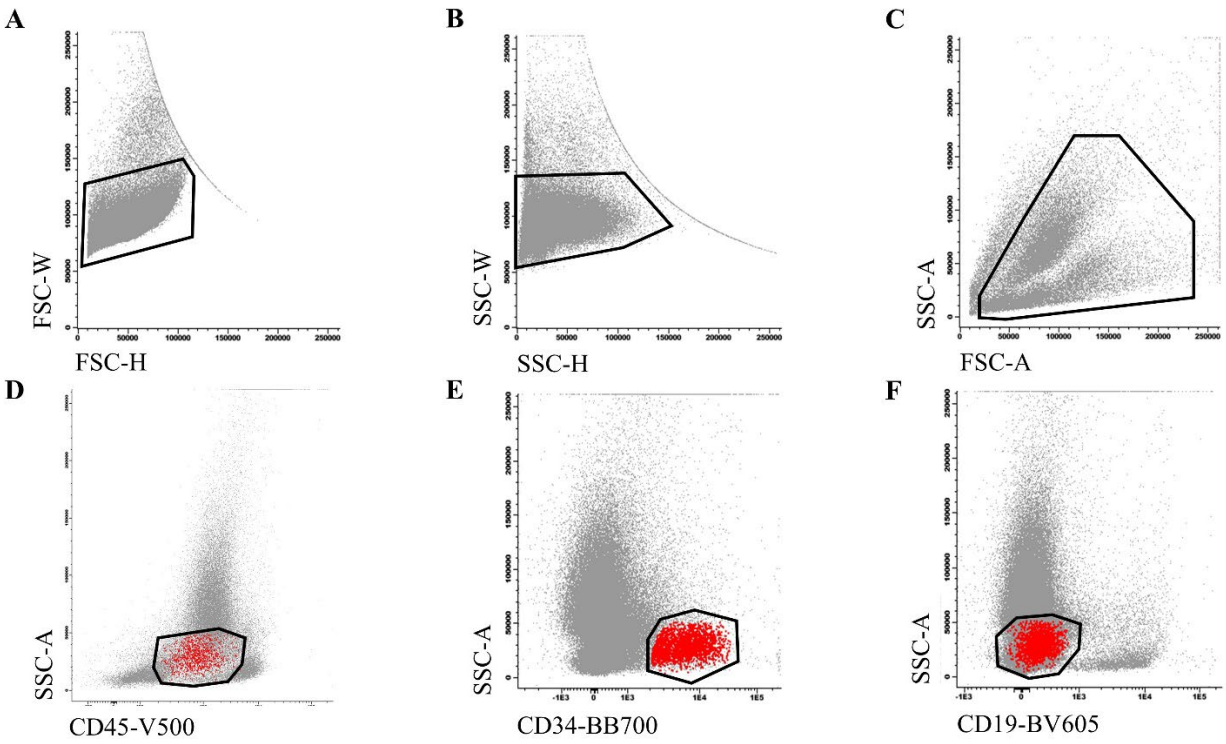
Sample	Reads	% Duplicate	% Mapped	Mean coverage on target	% of target covered with at least 10X
B Lymphoid	47,361,768	25.10	99.88	79.17X	97.0
Myeloid	63,717,732	25.45	99.92	97.19X	97.8
Buccal Swab	6,849,050	23.71	99.89	12.72X	49.8

Figures

Appendix Figure A1: Antibody titration curves for 17-color AML MRD panel



Appendix Figure A2: Myeloblast gating strategy



Doublets are excluded using FSC-H vs FSC-W and SSC-H vs SSC-W as is customary on a BD FACSAria Fusion instrument (**A-B**). Debris are excluded using a FSC-A vs SSC-A plot (**C**). A $CD45_{dim}$, SSC_{low} gate is drawn (**D**). This gate will contain myeloblasts, hematogones, and erythroid progenitors. Myeloblasts are identified within this gate as CD34+, CD19- (**E-F**). CD19 is used to exclude hematogone (CD19+). The final myeloblast population (red) is shown backgated on plots **D** to **F**.