



HAL
open science

La cartographie optique du génome, une alternative prometteuse aux approches cytogénétiques standards dans les leucémies aiguës lymphoblastiques : étude comparative en aveugle sur 10 patients

Valentin Lestringant

► To cite this version:

Valentin Lestringant. La cartographie optique du génome, une alternative prometteuse aux approches cytogénétiques standards dans les leucémies aiguës lymphoblastiques : étude comparative en aveugle sur 10 patients. Médecine humaine et pathologie. 2021. dumas-03212340

HAL Id: dumas-03212340

<https://dumas.ccsd.cnrs.fr/dumas-03212340>

Submitted on 29 Apr 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



UNIVERSITE DE PICARDIE JULES VERNE
FACULTE DE MEDECINE D'AMIENS
ANNEE 2021



THESE 2021-1

THESE POUR LE DIPLOME D'ETAT DE DOCTEUR EN MEDECINE
SPECIALITE BIOLOGIE MEDICALE

**LA CARTOGRAPHIE OPTIQUE DU GENOME,
UNE ALTERNATIVE PROMETTEUSE
AUX APPROCHES CYTOGENETIQUES STANDARDS
DANS LES LEUCEMIES AIGUES LYMPHOBLASTIQUES :
ETUDE COMPARATIVE EN AVEUGLE SUR 10 PATIENTS**

Présentée et soutenue publiquement

Le lundi 11 janvier 2021

Par

Valentin LESTRINGANT

PRESIDENT DU JURY :

Monsieur le Professeur Loïc GARÇON

MEMBRES DU JURY :

Monsieur le Professeur Jean-Pierre MAROLLEAU

Monsieur le Professeur Saïd KAMEL

Madame le Docteur Dominique PENTHER

Monsieur le Docteur Nicolas DUPLOYEZ

DIRECTEUR DE THESE :

Monsieur le Docteur Yann FERRET

REMERCIEMENTS

Monsieur le Professeur Loïc GARÇON

Professeur des Universités – Praticien Hospitalier (Hématologie – Transfusion)

Chef du service d'Hématologie Biologique

Chef du service de Génétique Constitutionnelle

Monsieur le Professeur,

Je vous remercie de me faire l'honneur de présider mon jury de thèse.

Merci pour l'intérêt que vous portez à ce travail. Merci pour votre écoute, votre disponibilité et votre soutien au quotidien.

Soyez assuré de mon plus profond respect et de ma sincère reconnaissance.

Monsieur le Professeur Jean-Pierre MAROLLEAU

Professeur des Universités-Praticien Hospitalier (Hématologie – Transfusion)

Chef du service d'hématologie clinique et de thérapie cellulaire

Chef du Pôle « Oncopôle »

Monsieur le Professeur,

Je vous remercie pour l'honneur que vous me faites en acceptant de participer à mon jury de thèse. Merci pour l'intérêt que vous portez à ce travail. J'espère qu'il saura retenir votre intérêt.

Soyez assuré de mon sincère respect et de ma profonde reconnaissance.

Monsieur le Professeur Saïd KAMEL

Professeur des Universités (Pharmacie)

Monsieur le Professeur,

Je vous remercie pour l'honneur que vous me faites en acceptant de participer à mon jury de thèse. Merci pour l'intérêt que vous portez à ce travail.

Soyez assuré de mon sincère respect et de ma profonde reconnaissance.

Madame le Docteur Dominique PENTHER

Praticien Hospitalier (Biologie Médicale)

Je vous remercie pour l'honneur que vous me faites en acceptant de participer à mon jury de thèse. Merci pour l'intérêt que vous portez à ce travail.

Merci également pour le soutien que vous m'apportez au quotidien, pour votre disponibilité et vos précieux conseils.

Soyez assuré de mon sincère respect et de ma profonde reconnaissance.

Monsieur le Docteur Nicolas DUPLOYEZ

Praticien Hospitalier (Biologie Médicale)

Je te remercie pour l'honneur que tu me fais en acceptant de participer à mon jury de thèse. Merci pour l'intérêt que tu portes à ce travail. Merci pour ton aide et ta disponibilité dans la réalisation de ce projet.

Sois assuré de mon sincère respect et de ma profonde reconnaissance.

Monsieur le Docteur Yann FERRET

Praticien Hospitalier (Biologie Médicale)

Je te remercie pour l'honneur que tu m'as fait d'avoir accepté d'encadrer ma thèse. Merci pour ton aide précieuse et ton implication dans ce travail.

Egalement un grand merci pour m'avoir transmis ta passion de la cytogénétique et pour tes enseignements. Merci pour tes nombreux conseils, ta bienveillance et ton soutien. J'espère être à la hauteur de tes attentes pour reprendre le flambeau.

Enfin merci pour ta bonne humeur et ton humour. J'ai eu beaucoup de plaisir à partager ton quotidien dans le service.

Sois assuré de mon sincère respect et de ma plus profonde reconnaissance.

L'équipe de cytogénétique onco-hématologique du CHU d'Amiens :

Merci pour votre bienveillance et votre accueil chaleureux. Merci de m'avoir montré toutes les ficelles et m'avoir transmis votre expérience avec tant de plaisir. J'admire votre implication, votre motivation et votre enthousiasme. Le service vous doit énormément.

Anne-Lise, merci pour ta douceur, ta gentillesse et merci également pour les maladresses et les *qui pro quo* qui viennent illuminer nos journées. **Mélanie**, merci pour ta joie de vivre, et pour animer le labo avec tant d'énergie. Tes talents de chanteuse sont sans égal. **Francis**, merci d'avoir en toutes circonstances le dicton, la citation ou le proverbe qui convient.

L'équipe de génétique constitutionnelle du CHU d'Amiens :

Monsieur le Professeur Henri Copin : Merci pour votre accueil chaleureux dans le service de cytogénétique. Merci pour votre bienveillance et vos bons mots ainsi que pour votre disponibilité durant mon internat. Je garderai également des souvenirs inoubliables de vos cours en première année de médecine. Soyez assuré de mon sincère respect.

Dr Marlène Gallet : Merci pour ta bonne humeur, pour ton aide et pour ta bienveillance au quotidien.

Dr Guillaume Jedraszak : Merci pour ton soutien et pour ton aide précieuse.

Dr Virginie Magry : Merci pour ta bonne humeur et ta sympathie.

Catherine, Florence, Fred, Nathalie : Merci de participer, chacun à votre manière, à rendre le quotidien du service si agréable.

L'équipe du secrétariat de BDR et Cytogénétique : **Nathalie, Laure** et **Audrey**, merci pour votre bonne humeur communicative.

A mes co-internes de biologie médicale : Merci pour ces années formidables et tous les petits moments du quotidien. La plupart d'entre vous ne sont déjà plus internes mais je tenais à vous remercier : **Baptiste, Cédric, Sophie, Quentin, Laurine, Arthur, Ben, Elodie, Samy, Julien, Manon, Geoffrey ...** et tous les autres.

Julie : Merci d'être une oreille attentive avec qui l'on peut ronchonner de tout et de rien.

Mathilde : Je ne le dirais qu'une seule et unique fois : tu es formidable !

Yaël : Je ne serais évidemment pas arrivé jusqu'ici sans ton aide. Alors merci de m'avoir supporté et d'être présente à mes côtés après toutes ces années.

Merci à ma famille et particulièrement à mes parents **Thierry** et **Valérie**, à mon frère **Tristan** et à ma grand-mère **Lucette**. Je vous dois évidemment beaucoup alors un immense merci pour votre soutien inconditionnel et pour tout le bonheur que vous m'apportez.

Enfin une pensée particulière pour mon grand-père, **Bernard** : Merci d'avoir su tant éveiller ma curiosité. Tu es pour beaucoup responsable de la personne que je suis aujourd'hui et j'espère que tu en serais fier.

TABLE DES MATIERES

TABLES DES ILLUSTRATIONS	9
RESUME	10
ABSTRACT	11
LISTE DES ABREVIATIONS	12
INTRODUCTION	13
1. Leucémies aiguës lymphoblastiques : généralités	13
1.1 Physiopathologie	13
1.2 Epidémiologie	13
1.3 Diagnostic	14
1.3.1 Signes cliniques.....	14
1.3.2 Examens biologiques.....	14
1.4 Facteurs pronostiques	17
1.5 Principes thérapeutiques	17
2 Place de la cytogénétique dans les LAL	18
2.1 Anomalies cytogénétiques dans les LAL-B.....	18
2.2 Anomalies cytogénétiques dans les LAL-T.....	23
2.3 Apport et limites des différentes techniques de cytogénétique.....	25
2.3.1 Caryotype conventionnel	25
2.3.2 FISH	25
2.3.3 SNP-Array	26
2.3.4 RT-MLPA.....	27
3 Cartographie optique du génome.....	28
3.1 Principes	28
3.2 Apport en cytogénétique onco-hématologique	30
OBJECTIFS	32
MATERIEL ET METHODES	33
1 Sélection des échantillons	33
2 Caryotype	33
3 FISH.....	33
4 SNP-Array.....	34
5 RT-MLPA	34
6 Cartographie optique du génome.....	34
6.1 Extraction de l'ADN et collection des données	35

6.2	Analyse par le Rare Variant Pipeline et sélection des variants de structure.....	35
6.3	Comparaison des données.....	36
7	Confirmation des anomalies additionnelles	36
RESULTATS		37
1	Caractéristiques des patients	37
2	Données qualité de la cartographie optique du génome	37
3	Identification des anomalies d'intérêt clinique par cartographie optique du génome	38
3.1	Discordances entre les techniques.....	41
3.2	Discordances entre les opérateurs	41
4	Identification d'anomalies chromosomiques non détectées par les techniques standards.....	42
5	Corrélation entre les coordonnées chromosomiques des anomalies décrites en COG et en SNP-Array.....	47
6	Détection des pertes d'hétérozygotie en cartographie optique du génome.....	47
DISCUSSION		48
1	Efficacité de la COG pour la détection des anomalies chromosomiques des LAL.....	48
2	Accessibilité de la technologie de COG	49
3	Identification par COG d'anomalies non détectées par les techniques standards	49
4	Description de nouvelles anomalies potentiellement impliquées dans la leucémogénèse.....	50
5	Capacité de la COG pour la détection des CN-LOH.....	50
6	Interférences pré-analytiques et analytiques de la COG.....	51
CONCLUSION		52
REFERENCES BIBLIOGRAPHIQUES.....		53
ANNEXE 1 : Liste des 186 gènes d'intérêts utilisés pour la sélection des anomalies.....		58
ANNEXE 2 : Analyse de l'échantillon 3 en cartographie optique par l'algorithme CNV		59
ANNEXE 3 : Analyse de l'échantillon 4 en cartographie optique par l'algorithme CNV		59
ANNEXE 4 : Caractérisation par cartographie optique du génome d'un réarrangement des chromosomes 7 et 8 impliquant les gènes <i>MYC</i> et <i>TRB</i>		60
ANNEXE 5 : Anomalies détectées en cartographie optique, ne répondant pas aux critères d'inclusion mais impliquant des gènes de la base de données COSMIC.....		62
ANNEXE 6 : Manuscrit de l'article de thèse, en préparation pour soumission		63
ANNEXE 7 : Serment d'Hippocrate		80

TABLES DES ILLUSTRATIONS

TABLEAUX

Tableau 1 : Classification EGIL des LAL-B	16
Tableau 2 : Classification EGIL des LAL-T.....	16
Tableau 3 : Classification OMS 2016 des LAL-B	19
Tableau 4 : Caractéristiques des patients.....	37
Tableau 5 : Anomalies détectées par les techniques standards et correctement identifiées par cartographie optique du génome	40
Tableau 6 : Anomalies détectées par les techniques standards et manquées en cartographie optique du génome	42
Tableau 7 : Anomalies décrites par cartographie optique et non détectées par les techniques standards	43

FIGURES

Figure 1: Fréquence des anomalies chromosomiques primaires chez les enfants et adultes atteints de LAL-B, d'après Moorman <i>et al.</i>	23
Figure 2 : Principes de l'hybridation fluorescente <i>in situ</i>	26
Figure 3 : Etapes de la SNP-Array (sur Cytoscan HD par Affimetrix).....	27
Figure 4 : Etapes de la RT-MLPA (exemple d'un transcrit <i>BCR-ABL1</i>).....	28
Figure 5 : Etapes de la cartographie optique du génome (adapté de Bionano).....	29
Figure 6 : Représentation la reconstruction du génome par chevauchement des molécules d'ADN marquées (d'après Bionano).....	30
Figure 7 : Diagramme de flux des anomalies détectées par les techniques standards et par cartographie optique du génome	39
Figure 8 : Caractérisation par cartographie optique du génome d'un réarrangement complexe des chromosomes 10, 11 et 14 et impliquant les gènes <i>TLX1</i> , <i>LMO2</i> et <i>TRA</i>	44

RESUME

Introduction : Les leucémies aiguës lymphoblastiques (LAL) sont caractérisées par une grande diversité d'anomalies génétiques. Leur détection en routine requiert l'emploi de nombreuses techniques complémentaires : caryotype, hybridation fluorescente *in situ* (FISH), Single nucleotide polymorphism Array (SNP-Array) et recherche de transcrits de fusion. La cartographie optique du génome (COG) est une technologie permettant l'étude de l'ensemble du génome à haute résolution ainsi que la détection des anomalies de nombre et de structure, y compris les translocations équilibrées. L'objectif était d'évaluer les performances de cette nouvelle technologie face aux techniques standards.

Matériel et méthodes : Une COG Bionano a été réalisée pour 10 patients atteints de LAL-B ou T au diagnostic. Les données étaient analysées en aveugle par deux opérateurs indépendants puis comparées à celles obtenues par caryotype, FISH, SNP-Array et Reverse Transcription Multiplex Ligation-dependent Probe Amplification (RT-MLPA).

Résultats : Sur les 80 anomalies décrites par les techniques standards, 73 étaient retrouvées par les opérateurs en COG.

Discussion : Les 10 anomalies non retrouvées en COG étaient dues à des problèmes de filtre des algorithmes ou à un défaut de détection de la technologie. Douze nouvelles anomalies étaient décrites en COG, dont 7 ont pu être confirmées par une autre technique, certaines impliquant des gènes d'intérêt. La COG permettait également la résolution d'évènements chromosomiques complexes.

Conclusion : La COG semble être une alternative innovante, rapide et rentable aux techniques de cytogénétique standards pour la caractérisation génétique des LAL.

Mots-clefs : Leucémie aiguë lymphoblastique, cytogénétique, cartographie chromosomique, hémopathies malignes, hématologie, génétique, biologie

ABSTRACT

Introduction: Acute lymphoblastic leukemia (ALL) is characterized by a large number of genetic abnormalities. Their routine detection requires the use of many complementary techniques: karyotyping, fluorescent *in situ* hybridization (FISH), Single nucleotide polymorphism Array (SNP-Array) and the search for fusion transcripts. Optical genome mapping (OGM) is a technology that allows a high-resolution genome-wide analysis as well as the detection of copy number anomalies and structure anomalies, including balanced translocations. The objective was to evaluate the performance of this new technology compared to standard techniques.

Material and methods: A Bionano OGM was performed for 10 patients with B or T-ALL at diagnosis. The data were blindly analyzed by two independent operators and compared to those obtained by karyotype, FISH, SNP-Array and Reverse Transcription Multiplex Ligation-dependent Probe Amplification (RT-MLPA).

Results: Among the 80 abnormalities described by standard techniques, 73 were found by operators in OGM.

Discussion: The 10 anomalies not found in OGM were due to algorithms' filter issues or to a detection defect in the technology. Twelve anomalies were newly described in OGM, 7 of which could be confirmed by another technique and some involving genes of interest. OGM also allowed the resolution of complex chromosomal events.

Conclusion: OGM appears to be an innovative, rapid and cost-effective alternative to standard cytogenetic techniques for the genetic characterization of ALL.

Keywords: Acute lymphoblastic leukemia, cytogenetics, chromosome mapping, hematological malignancies, hematology, genetics, biology

LISTE DES ABREVIATIONS

ACPA : Analyse Chromosomique sur Puce à ADN

ADN THPM : ADN de Très Haut Poids Moléculaire

CGH : Comparative Genomic Hybridization

CN-LOH : Copy Neutral Loss Of Heterozygosity

CNV : Copy Number Variant

COG : Cartographie Optique du Génome

COSMIC : Catalogue Of Somatic Mutations In Cancer

DMSO : Diméthylsulfoxyde

EGIL : European Group for the Immunological characterization of Leukemias

FISH : Fluorescent *in situ* Hybridization

ISCN : International System for human Cytogenomic Nomenclature

LAL : Leucémie aigüe lymphoblastique

LD-RTPCR : Ligation Dependent RTPCR (ou RT-MLPA)

NGS : Next Generation Sequencing

NLV : Negative Label Variance

OMS : Organisation Mondiale de la Santé

PAR : PseudoAutosomal Region

Pb : paire de bases

PCR : Polymerase Chain Reaction

RT-MLPA : Reverse Transcription Multiplex Ligation-dependent Probe Amplification (ou LD-RTPCR)

RT-PCR : Reverse Transcription PCR

RVP : Rare Variant Pipeline

SNP : Single Nucleotide Polymorphism

SV : Structural Variant

TCR : T-Cell Receptor

INTRODUCTION

1. Leucémies aiguës lymphoblastiques : généralités

1.1 Physiopathologie

Les leucémies aiguës lymphoblastiques (LAL) sont des hémopathies lymphoïdes, qui résultent d'un blocage de la maturation des progéniteurs hématopoïétiques lymphoïdes au stade de lymphoblaste. On distingue les LAL de lignée B (LAL-B) et de lignée T (LAL-T), selon la lignée dans laquelle est engagé le progéniteur impliqué. Le potentiel de prolifération important des lymphoblastes va entraîner un envahissement de la moelle osseuse puis généralement du sang et des organes lymphoïdes, et plus rarement d'autres localisations (système nerveux central, séreuses). Dans certains cas, la prolifération initiale des lymphoblastes intervient au niveau ganglionnaire ou extra-ganglionnaire, provoquant la formation d'une masse tumorale ; on parle alors de lymphome lymphoblastique. L'accumulation clonale de lymphoblastes est la conséquence d'un ensemble de processus hématopoïétiques aberrants résultants de mutations, de réarrangements chromosomiques et de modifications épigénétiques. Schématiquement, la survenue d'anomalies au niveau de facteurs de transcription de l'hématopoïèse, aboutissant au blocage de la maturation des progéniteurs, s'associe à des anomalies activatrices favorisant la prolifération. Des facteurs épigénétiques viennent s'ajouter, amplifiant cette altération du fonctionnement de la cellule leucémique.

1.2 Epidémiologie

Les LAL sont des pathologies rares dans la population générale, avec une nette prédominance masculine, représentant moins de 5% de l'ensemble des hémopathies lymphoïdes. Le taux d'incidence standardisé à la population mondiale est de 2 cas/an/100 000 habitants chez les hommes et de 1,5 chez les femmes (sex ratio : 1,35)[1]. Il existe cependant une grande disparité de l'incidence de cette pathologie en fonction de l'âge. En effet, 45% des patients atteints ont moins de 15 ans, avec un pic d'incidence entre 0 et 4 ans. Il s'agit ainsi des cancers les plus fréquents chez l'enfant. Chez l'adulte, l'incidence augmente avec l'âge, la maladie survenant dans la plupart des cas après 50 ans.

En l'absence de traitement, l'issue est systématiquement fatale. Cependant l'utilisation de protocoles de polychimiothérapies depuis les années 1960, aujourd'hui accompagnés par l'allogreffe de moelle osseuse et les thérapies ciblées, ont permis d'améliorer drastiquement le pronostic de ces pathologies [2,3]. Actuellement, une rémission complète est atteinte chez près de 85% des enfants avec une survie globale de 90% à 5 ans [4]. Mais le pronostic décline avec

l'âge. Ainsi la survie globale à 5 ans chez l'adulte est comprise entre 30 et 40%, malgré l'obtention d'une rémission complète dans 90% des cas [5]. Cette disparité s'explique en partie par des différences dans le profil génétique des LAL en fonction de l'âge.

1.3 Diagnostic

1.3.1 Signes cliniques

Les manifestations cliniques des LAL, secondaires à la prolifération des cellules leucémiques, surviennent généralement avec une présentation aiguë et une altération de l'état général. Différents signes sont observables selon la localisation de cette prolifération.

Le syndrome d'insuffisance médullaire, provoqué par l'envahissement de la moelle osseuse par les lymphoblastes, est le témoin d'un retentissement de la maladie sur l'hématopoïèse. Il combine syndrome anémique, syndrome infectieux et syndrome hémorragique, en fonction des lignées cellulaires atteintes.

Il existe fréquemment un syndrome tumoral associé, conséquence de l'envahissement des organes lymphoïdes mis en évidence par une splénomégalie et la présence d'adénopathies.

Le syndrome de leucostase survient, quant à lui, dans les formes avec une forte hyperleucocytose (au-delà de 100 G/l). Il se traduit par des manifestations respiratoires (hypoxie, détresse respiratoire, épanchement pleural) et des troubles de la conscience.

D'autres localisations sont également possibles. L'envahissement neuro-méningé est plus fréquent dans les leucémies aiguës lymphoïdes que myéloïdes. Il a pour conséquence des céphalées voire un syndrome d'hypertension intracrânienne, avec une atteinte des nerfs crâniens et est de pronostic péjoratif. Des douleurs osseuses et articulaires peuvent survenir secondairement à des localisations osseuses, plus volontiers chez l'enfant que chez l'adulte. Chez les individus masculins, l'envahissement des testicules est possible, bien que rare.

1.3.2 Examens biologiques

Hémogramme

L'hémogramme est l'examen d'orientation de première intention. Une anémie est présente dans 90% des cas, classiquement normocytaire, normochrome, arégénérative. La numération leucocytaire est variable, allant de formes leucopénique (< 4 G/l, jusqu'à moins de 1 G/l, avec parfois une absence de blastes) jusqu'aux formes hyperleucocytaires (> 50 G/l, avec une blastose majeure). Une neutropénie (< 1,5 G/l) est fréquemment associée, pouvant aller jusqu'à

l'agranulocytose (< 0,5 G/l). Une thrombopénie est également présente dans la majorité des cas au diagnostic. Le frottis sanguin permet de mettre en évidence la présence, fréquente mais inconstante, de lymphoblastes circulants.

Myélogramme

Indispensable pour le diagnostic, le myélogramme permet, par la mise en évidence d'au moins 20% de blastes sur le frottis médullaire, de poser le diagnostic de leucémie aiguë. Une analyse morphologique, éventuellement complétée par une analyse cytochimique (réaction à la myéloperoxydase négative), permet d'évoquer l'origine lymphoïde des blastes.

Immunophénotypage

Réalisé préférentiellement sur un prélèvement médullaire, l'immunophénotypage des blastes par cytométrie en flux permet de distinguer les blastes myéloïdes des blastes lymphoïdes. Il détermine également l'appartenance des lymphoblastes à la lignée B ou T, ainsi que leur stade de maturation, selon la classification de l'EGIL (European Group for the Immunological characterization of Leukemias, Tableau 1 et Tableau 2). Enfin, l'immunophénotypage permet la mise en évidence de marqueurs aberrants, utiles pour le suivi de la maladie résiduelle.

Cytogénétique

Le caryotype des cellules leucémiques, obtenu après mise en culture de la moelle osseuse, permet de mettre en évidence les différentes anomalies chromosomiques, de nombre et de structure. Il est généralement complété par une analyse par hybridation fluorescente *in situ* (FISH).

Ces examens sont généralement complétés par une analyse sur puce à ADN (ACPA), plus résolutive, permettant de mettre en évidence les anomalies chromosomiques de nombre avec une taille plus réduite.

Tableau 1 : Classification EGIL des LAL-B

LAL de type B	Antigènes de la lignée B : CD19, CD22 (s/c), cCD79a	CD10	cμ	sIg
B-I (Pro-B)	+	-	-	-
B-II (B commune)	+	+	-	-
B-III (Pré-B)	+	±	+	-
B-IV (B mature)	+	±	±	+

Tableau 2 : Classification EGIL des LAL-T

LAL de type T	cCD3, CD7	CD2, CD5, CD4 et/ou CD8	CD1a	sCD3
T-I (Pro-T)	+	-	-	-
T-II (Pré-T)	+	+	-	-
T-III (T corticale)	+	+	+	±
T-IV (T mature)	+	+	-	+

Biologie moléculaire

Cet ensemble de techniques s'intéresse aux conséquences moléculaires de la leucémogénèse, à l'échelle de l'acide nucléique. La PCR (Polymerase Chain Reaction) permet la détection de délétions ou d'amplifications de régions codantes de l'ADN. La RT-PCR (Reverse Transcription PCR) et la RT-MLPA (Reverse Transcription Multiplex Ligation-dependent Probe Amplification) peuvent mettre en évidence les transcrits de fusions, en complément des examens de cytogénétique. Les mutations de gènes d'intérêts sont recherchées par NGS (Next Generation Sequencing). Enfin, il est possible de mettre en évidence la clonalité des cellules leucémiques et de caractériser les clones par l'étude du réarrangement des gènes des immunoglobulines (LAL-B) ou du TCR (T-Cell Receptor) (LAL-T) en PCR suivie d'un séquençage Sanger ou en NGS. Le suivi de la maladie résiduelle pourra ensuite être effectué par quantification en PCR quantitative (voire en NGS) des réarrangements séquencés lors de cette étude de clonalité.

1.4 Facteurs pronostiques

La définition du pronostic de la maladie est un point central de la prise en charge des LAL. Elle définit l'intensité des traitements et son éventuelle adaptation. Ces facteurs, souvent associés, sont dépendants du patient, de la maladie, mais également de la réponse au traitement.

L'âge est le principal facteur pronostique lié au patient. Comme évoqué en 1.2, le taux de survie chute drastiquement avec l'âge, entre autres car la fréquence des anomalies génétiques est différente selon l'âge. Ainsi, chez l'adulte un âge au-delà de 50 ans est considéré comme un facteur péjoratif. Chez l'enfant, les LAL survenant avant 1 an sont de pronostic défavorable en raison de la surreprésentation des réarrangements de *KMT2A* [6].

L'hyperleucocytose (> 50 G/l chez l'enfant et > 30 G/l chez l'adulte), l'envahissement du système nerveux central et le phénotype T sont des marqueurs de mauvais pronostic [7]. Les anomalies génétiques sont étroitement liées au pronostic de la maladie et seront développées dans le chapitre 2.

Enfin, la corticorésistance et la chimiorésistance sont deux facteurs pronostiques majeurs de réponse au traitement. La corticorésistance se définit par un nombre de blastes sanguins supérieur à 1 G/l après 7 jours de corticothérapie et une injection de méthotrexate intra-thécale. Elle n'est évaluable que si la blastose sanguine est supérieure à ce seuil au diagnostic. La chimiorésistance est quant à elle définie par un taux de blastes supérieur à 5 % sur le frottis médullaire à 21 jours de traitement. La quantification de la maladie résiduelle, par PCR quantitative en temps réel ou par cytométrie en flux, est également un outil important de la prise en charge. En effet, une intensification du traitement ou le recours à une greffe de moelle osseuse serait bénéfique pour les patients ayant une maladie résiduelle supérieure à 0,01 % à la fin de la phase d'induction [8]. Selon le protocole CALLF01 (protocole français pour le traitement des LAL de l'enfant, en cours d'inclusion actuellement) par exemple, la détection d'une maladie résiduelle supérieure à 0,1 % à J35 ou $> 0,01$ % à J65 de traitement entraîne le reclassement des patients dans un groupe pronostique plus défavorable, menant à l'intensification du protocole de chimiothérapie voire à une greffe de moelle osseuse.

1.5 Principes thérapeutiques

Le traitement des LAL débute classiquement par une préphase, consistant en l'administration de corticoïdes à fortes doses pendant une semaine et dont le but est de diminuer la quantité de cellules tumorales. Elle est suivie de plusieurs séquences de chimiothérapie. L'induction, qui dure généralement 4 à 6 semaines et dont le but est l'éradication du clone leucémique, combine

une anthracycline, un alcaloïde de la pervenche, de la L-asparaginase et un corticoïde. Elle est suivie par la phase de consolidation, pendant 2 à 3 mois, dont l'objectif est d'éviter la sélection d'un clone résistant et combine méthotrexate, cytarabine et L-asparaginase. Une allogreffe de moelle osseuse peut être réalisée en fonction du pronostic, de l'âge et de l'état général du patient. Le traitement se poursuit par une phase d'entretien de longue durée (de l'ordre de 2 ans), généralement à base de méthotrexate et de 6-mercaptopurine, qui permet la prévention de la rechute. Dans le cas des LAL avec fusion *BCR-ABL1*, un inhibiteur de tyrosine kinase comme l'imatinib peut être utilisé.

L'évolution des protocoles de traitement a permis l'amélioration de la survie des patients mais pose désormais le problème des conséquences à long terme de ces traitements lourds, en particulier chez l'enfant (retards de croissance et de puberté, cancers secondaires). L'enjeu actuel est donc d'obtenir le meilleur rapport bénéfices/risques. L'objectif est ainsi l'obtention d'une rémission complète tout en limitant la toxicité des chimiothérapies. Dans ce contexte, la stratification du risque en fonction du pronostic individuel, est essentielle. Ce pronostic est conditionné par de nombreux facteurs, cliniques et biologiques, parmi lesquels le profil génétique des cellules leucémiques apparaît comme étant le plus impactant [9,10]. De plus, l'essor des thérapies ciblées est un argument supplémentaire en faveur d'un typage génétique le plus exhaustif possible à la recherche de cibles thérapeutiques actuelles mais également futures. La détermination des anomalies génétiques des LAL est donc nécessaire pour une meilleure prise en charge des patients.

2 Place de la cytogénétique dans les LAL

2.1 Anomalies cytogénétiques dans les LAL-B

L'Organisation Mondiale de la Santé (OMS) a identifié un certain nombre d'anomalies récurrentes dans les LAL-B [11]. Ces LAL avec anomalies récurrentes sont distinguées des autres, désignées comme « not otherwise specified » (LAL-B NOS ou « other »). Il existe des anomalies de nombre (gain ou perte de chromosomes entiers) ainsi que des anomalies de structure, qui peuvent être équilibrées, c'est-à-dire sans perte de matériel chromosomique (translocations, inversions) ou déséquilibrées (délétions, répétitions, amplifications etc.). La classification OMS actuelle de 2016 décrit 9 entités avec anomalies récurrentes (Tableau 3).

Tableau 3 : Classification OMS 2016 des LAL-B

Leucémie aiguë / lymphome lymphoblastique B, NOS

Leucémie aiguë / lymphome lymphoblastique B avec t(9;22)(q34;q11.2) ; *BCR-ABL1*

Leucémie aiguë / lymphome lymphoblastique B avec t(v;11q23) ; *KMT2A* réarrangé

Leucémie aiguë / lymphome lymphoblastique B avec t(12;21)(p13;q22) ; *ETV6-RUNX1*

Leucémie aiguë / lymphome lymphoblastique B avec hyperdiploïdie

Leucémie aiguë / lymphome lymphoblastique B avec hypodiploïdie

Leucémie aiguë / lymphome lymphoblastique B avec t(5;14)(q31;q32) ; *IL3-IGH*

Leucémie aiguë / lymphome lymphoblastique B avec t(1;19)(q23;p13.3) ; *TCF3-PBX1*

Entité provisoire : Leucémie aiguë / lymphome lymphoblastique B, BCR-ABL like

Entité provisoire : Leucémie aiguë / lymphome lymphoblastique B avec *iAMP21*

La translocation t(9;22)(q34;q11) entraîne la fusion des gènes *BCR* et *ABL1*. Initialement décrite dans les leucémies myéloïdes chroniques, cette anomalie est également récurrente dans les LAL-B. Le gène de fusion ainsi obtenu est à l'origine d'une dérégulation de l'activité tyrosine kinase d'*ABL1*, qui devient constitutive. Il s'agit d'une anomalie de mauvais pronostic, même si la détection de cette translocation autorise l'adjonction d'inhibiteurs de tyrosine kinase à la chimiothérapie classique. Une délétion du gène *IKZF1* est associée dans 80% des cas. Cette translocation est fréquente chez l'adulte, où elle représente près d'un quart des LAL-B, et plus rare chez l'enfant (3 %)[12].

Les translocations impliquant le gène *KMT2A* (anciennement *MLL*) en 11q23 sont particulièrement fréquentes chez le nourrisson (80 % des cas de LAL-B avant 1 an). Elles restent relativement fréquentes chez l'enfant (8 %) et chez l'adulte (10 %). Actuellement près de 135 partenaires de translocation ont été identifiés [6]. Le plus fréquent dans les LAL-B est le gène *AFF1* en 4q21, retrouvé dans plus de la moitié des cas. Les gènes *MLLT3* en 9p21 et *MLLT1* en 19p13.3 sont également des partenaires fréquents de *KMT2A*. Ces 3 partenaires représentent ensemble 88 % des remaniements de *KMT2A* dans les LAL-B. Le pronostic de ces

anomalies est majoritairement défavorable quel que soit l'âge, mais peut être corrélé au gène partenaire, un remaniement avec *AFF1* étant par exemple très défavorable.

La translocation t(12;21)(p13;q22) est à l'origine d'une fusion des gènes *ETV6* et *RUNX1*. Elle est présente principalement chez l'enfant (25 % des LAL-B) et est bien plus rare chez l'adulte (1 %), survenant alors essentiellement avant 25 ans. Il s'agit d'une anomalie cryptique (indétectable au caryotype) qui nécessite donc la réalisation d'une FISH ou d'une RT-PCR pour être mise en évidence. Le pronostic est favorable [13].

La translocation t(5;14)(q31;q32), provoque la fusion des gènes *IL3* et *IGH*. Bien qu'il s'agisse d'une entité très rare, la présentation en est particulière. La soumission d'*IL3*, facteur de croissance de la lignée éosinophile, à l'activateur du gène *IGH* est à l'origine d'une hyperéosinophilie majeure, pouvant masquer le diagnostic de LAL-B.

Plus généralement, les translocations avec le locus *IGH* sont retrouvées dans 5 % des LAL-B, principalement chez l'enfant et l'adulte jeune, et sont associées à un pronostic moins favorable, sans être un facteur de risque indépendant [14].

La translocation t(1;19)(q23;p13) entraîne la fusion des gènes *TCF3* et *PBX1*. Elle est présente dans 5 % des LAL-B (enfant et adulte). Deux formes coexistent pour ce remaniement. L'une, équilibrée, avec les 2 dérivés de translocation. L'autre, plus fréquente et déséquilibrée, avec la seule présence du dérivé 19, les deux chromosomes 1 étant alors normaux. Initialement considérée comme une anomalie de mauvais pronostic, l'adaptation du protocole de chimiothérapie permettrait une amélioration significative de la survie [10]. Ainsi si le pronostic de cette anomalie reste discuté, il est actuellement considéré comme intermédiaire. Plus rarement le gène *TCF3* fusionne avec le gène *HLF* par une translocation t(17;19)(q22;p13) s'associant à une hypercalcémie et une coagulation intravasculaire disséminée, et dont le pronostic est très défavorable [15].

Nouvelle entité introduite par l'édition 2016 de la classification de l'OMS, l'amplification intrachromosomique du chromosome 21 (iAMP21) est mise en évidence par FISH, avec une sonde *RUNX1*. Elle se définit par l'existence d'au moins 5 copies de ce gène par cellule, ou plus de 3 copies sur un même chromosome 21 en FISH métaphasique. Bien qu'étant le gène utilisé pour mettre en évidence l'anomalie, il s'agit en réalité d'un réarrangement bien plus complexe du chromosome 21 par des mécanismes de cassure-fusion-pont entraînant un véritable chromotripsis (multiples réarrangements d'un même chromosome). Il s'agit d'une anomalie de mauvais pronostic rencontrée essentiellement chez l'enfant (2 % des LAL-B)[16].

L'hyperdiploïdie répertoriée par l'OMS correspond à un nombre de chromosomes compris entre 51 et 65. Une hyperdiploïdie entre 47 et 50 chromosomes est possible mais généralement associée à des anomalies de structure et n'a donc pas d'impact pronostique propre. L'hyperdiploïdie élevée, à 51-65 chromosomes est fréquente chez l'enfant (25 % des LAL-B) et plus rare chez l'adulte (7 %). Les gains chromosomiques observés sont récurrents et concernent préférentiellement les chromosomes X, 4, 6, 10, 14, 17, 18 et 21. Son pronostic est très favorable [17].

L'hypodiploïdie telle que décrite par l'OMS, concerne les LAL avec 43 chromosomes ou moins, associée à un pronostic très péjoratif [18]. En effet, de la même façon que les hyperdiploïdies, l'impact pronostique des hypodiploïdies à 44-45 chromosomes n'est pas indépendant. Cette entité est rare, concernant 1 à 5 % des LAL-B de l'enfant et de l'adulte. Deux formes se distinguent : la quasi-haploïdie (23 à 29 chromosomes) et l'hypodiploïdie sévère (30 à 39 chromosomes). Encore une fois, le profil des pertes chromosomiques n'est pas aléatoire. Dans l'hypodiploïdie sévère, les chromosomes 1 et 6 sont généralement conservés tandis que dans la quasi-haploïdie, ce sont les chromosomes X, Y, 14, 18 et 21 qui sont généralement conservés. La monosomie 7 est, elle, constante dans les deux groupes. Ce profil hypodiploïde peut être masqué par une endoréplication du clone, conduisant à des cellules quasi-triploïdes (60-78 chromosomes) pour les hypodiploïdies sévères et à une hyperdiploïdie (50-58 chromosomes) pour les quasi-haploïdes. Le clone dupliqué peut en effet être le seul mis en évidence par le caryotype, posant alors le problème du diagnostic différentiel avec un clone hyperdiploïde. L'observation du profil des gains et pertes chromosomiques est alors primordial pour distinguer ces deux entités.

Une autre entité introduite par l'édition 2016 de la classification de l'OMS est la LAL *BCR-ABL1*-like. Cette entité est caractérisée par un profil transcriptomique semblable aux LAL avec fusion *BCR-ABL1*, sans pour autant être porteuse de ce remaniement mais partage le mauvais pronostic qui lui est associé. Elle représente de 10 à 15 % des LAL-B de l'enfant, 20 % de l'adolescent et 27 % de l'adulte jeune [19]. Cette entité forme en réalité un groupe hétérogène ayant pour point commun une activation de gènes à activité tyrosine kinase, par réarrangement ou par mutation. On distingue ainsi plusieurs groupes. Les réarrangements de *CRLF2* sont les plus fréquemment rencontrés et s'accompagnent de mutations *JAK1* ou *JAK2*. La fusion se fait soit avec le gène *IGH* par translocation t(X;14)(p22;q32) ou t(Y;14)(p11;q32), soit avec *P2RY8* après une microdélétion sur un chromosome sexuel. Ces deux remaniements provoquent une expression aberrante du récepteur de cytokines *CRLF2*. Les réarrangements peuvent également survenir au niveau des gènes des tyrosines kinases de la classe *ABL* (*ABL1*, *ABL2*, *PDGFRB* et

CSF1R), du gène *EPOR* (récepteur à l'EPO) ou de *JAK2*. La dérégulation peut également survenir par des mutations activatrices de la voie RAS, impliquant les gènes *NRAS*, *KRAS*, *PTPN11* et *NF1* ou sur d'autres gènes avec activité tyrosine kinase comme *NTRK3*.

D'autres anomalies, non présentées dans l'OMS 2016, doivent être mentionnées. Les délétions d'*IKZF1* en 7p12 surviennent dans 15 % des LAL-B pédiatriques et 30 % des LAL-B de l'adulte. Elles sont fréquemment associées aux LAL avec fusion *BCR-ABL1*. Le pronostic de cette anomalie, initialement considéré comme défavorable est discuté, et ne semble pas affecter le pronostic favorable d'autres anomalies comme les délétions d'*ERG*, auxquelles les délétions *IKZF1* sont fréquemment associées, ou encore les réarrangements *ETV6-RUNX1* [20–22]. Les délétions du gène *ERG* en 21q22 sont rencontrées dans 2 à 5 % des LAL-B de l'enfant et sont associées à un excellent pronostic, avec une survie à 5 ans de plus de 90 % [21]. Retrouvée dans 5 % des cas, la complexité du caryotype, définie pour les LAL-B par la présence d'au moins 5 anomalies, est un facteur indépendant de mauvais pronostic [23]. La fréquence relative et l'impact pronostique des principales anomalies chromosomiques sont résumés dans la Figure 1.

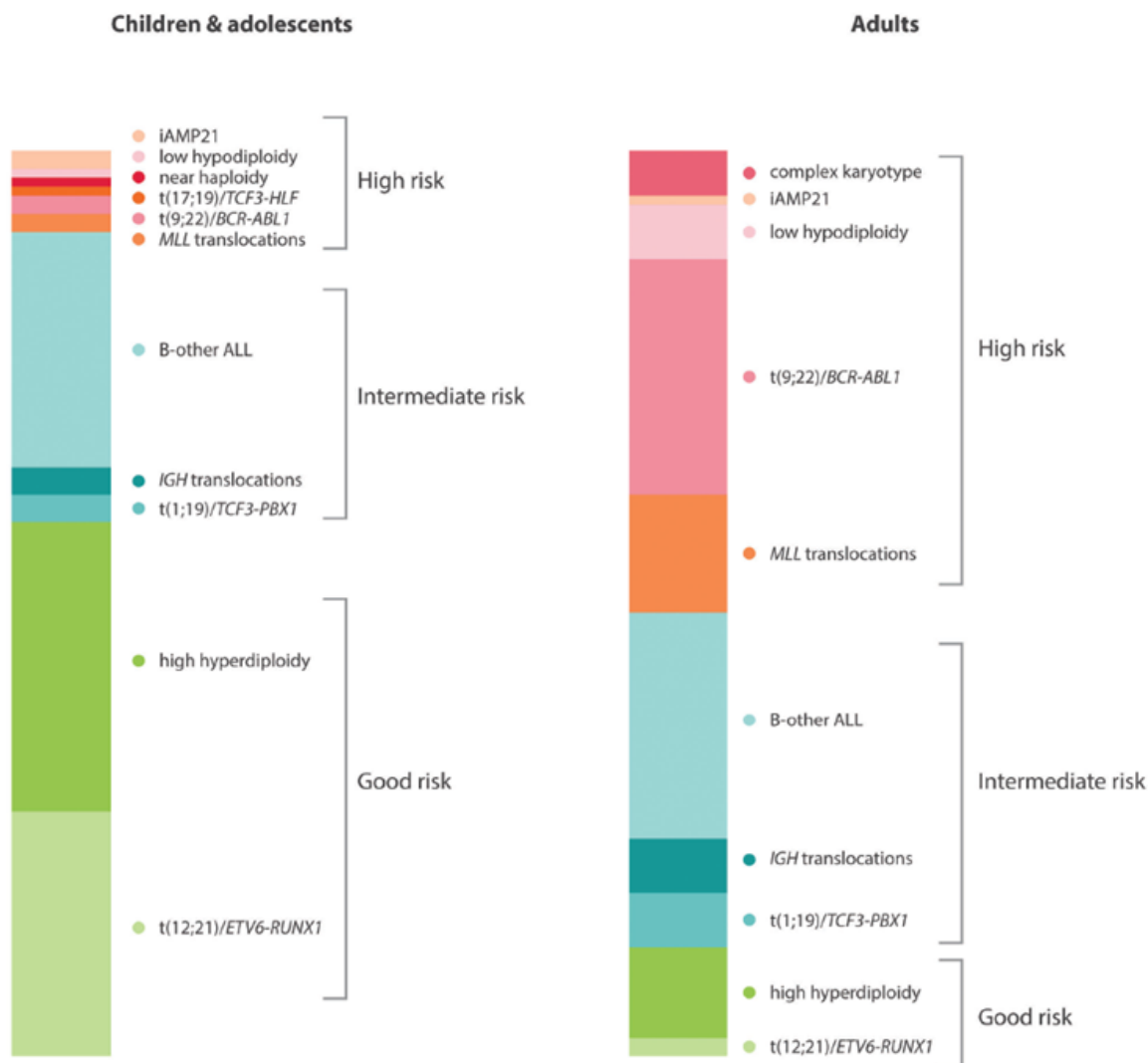


Figure 1: Fréquence des anomalies chromosomiques primaires chez les enfants et adultes atteints de LAL-B, d'après Moorman *et al.*

2.2 Anomalies cytogénétiques dans les LAL-T

Contrairement aux LAL-B, la classification 2016 de l'OMS n'individualise pas d'anomalies génétiques récurrentes. Le pronostic des LAL-T est par ailleurs davantage impacté par la survenue de mutation de gènes [24]. Cependant, même si elles sont plus rares que pour les LAL-B, les LAL-T ne sont pas exemptes de remaniements chromosomiques.

La majorité des réarrangements des LAL-T impliquent les locus du TCR (T-Cell Receptor) : *TRA/D* en 14q11 ou *TRB* en 7q34 et sont la conséquence d'accidents de recombinaison VDJ. La soumission au promoteur de ces locus du TCR, entraîne ainsi une expression aberrante des différents gènes partenaires.

- *TLX1* en 10q24, 7 % des LAL-T de l'enfant et 20 % de l'adulte, d'impact pronostique favorable [24,25].
- *LMO1* en 11p15 et *LMO2* en 11p13, évènements rares, sans valeur pronostique, et fréquemment associés à une activation de *TALI*. L'existence de délétions en 11p13 conduisant également à la surexpression de *LMO2* porte la fréquence des remaniements de ce gène à 20 % des LAL-T de l'enfant [26].
- *MYB* en 6q23, évènement rare présent dans 3 % des LAL-T, principalement avant l'âge de 2 ans. Une duplication cryptique de ce gène est également décrite dans 10 % des LAL-T [27].
- *MYC* en 8q24, 3 % des LAL-T, associé à un risque accru de rechute [28].
- *TALI* en 1p32, évènement rare, présent dans 1 à 3 % des LAL-T, sans valeur pronostique. L'expression aberrante de ce gène peut également être obtenue par une délétion de petite taille sur le bras court du chromosome 1, rapprochant *TALI* du promoteur du gène *STIL*, retrouvée dans 15 à 30 % des LAL-T [24,25].
- *NOTCH1* en 9q34. Ce gène est également muté dans 50 % des LAL-T ce qui est associé à un pronostic favorable [29].

On décrit également l'inversion (7)(p15q34) et la translocation (7;7) (p15;q34) qui placent l'enhancer de *TRB* à proximité des gènes de la famille *HOXA* en 7p15, entraînant leur surexpression. Présentes dans 5 % des LAL-T, elles n'ont pas d'impact pronostique avéré [30]. De manière similaire aux réarrangements impliquant les locus du TCR, la translocation t(5;14)(q35;q32), entraîne la surexpression de *TLX3* en 5q35 par soumission à la région régulatrice située en 3' du gène *BCL11B* en 14q32. Elle est présente dans 20 % des LAL-T de l'enfant et 10 % de l'adulte, sans valeur pronostique établie [24].

Par ailleurs un certain nombre d'autres réarrangements sont à l'origine de gènes de fusion. En dehors d'exceptionnelles LAL-T avec fusion *BCR-ABL1*, d'autres mécanismes peuvent conduire à la surexpression d'*ABL1* [31]. Présente dans 6 % des LAL-T, la fusion *NUP214-ABL1* est la forme la plus fréquente, retrouvée en tant qu'anomalie secondaire dans les LAL-T avec surexpression de *TLX1* ou *TLX3*. Des délétions cryptiques en 9q34, à l'origine d'une fusion *SET-NUP214* qui provoque la surexpression de *HOXA*, sont présentes dans 6 % des LAL-T de l'adulte [32,33]. Les réarrangements de *KMT2A* en 11q23 sont décrits dans 8 % des LAL-T, où le pronostic est alors intermédiaire. Les partenaires les plus fréquents sont *MLLT1* en 19p13.3 et *MLLT4* en 6q27 [6]. La translocation t(10;11)(p13;q14) *PICALM-MLLT10* est présente dans 4 à 8 % des LAL-T de l'enfant et 10 % de l'adulte, avec une valeur pronostique défavorable [24].

2.3 Apport et limites des différentes techniques de cytogénétique

Les LAL sont des pathologies d'une grande diversité génétique qui requièrent systématiquement l'utilisation de plusieurs techniques de cytogénétique conventionnelle et moléculaire pour obtenir le profil génétique le plus exhaustif possible, nécessaire pour établir un pronostic précis. Cette caractérisation génétique est généralement établie par une combinaison de caryotype conventionnel, d'analyses FISH, CGH-Array (Comparative Genomic Hybridization-Array) ou SNP-Array (Single Nucleotide Polymorphism-Array), et d'une étude de l'expression des gènes par exemple par RT-MLPA ou transcriptome [34,35]. Toutes ces techniques sont ainsi complémentaires.

2.3.1 Caryotype conventionnel

Le caryotype est obtenu après culture cellulaire d'un prélèvement sanguin ou idéalement médullaire. Dans le cas des LAL, il s'agit d'une culture courte de 16 à 24 heures, sans mitogènes ni facteurs de croissance [36]. L'ajout de colchicine, puissant poison du fuseau mitotique, à la fin de la culture, permet d'obtenir des cellules bloquées en métaphases. Après étalement des cellules bloquées en métaphases, les mitoses sont photographiées pour en classer les chromosomes. C'est une analyse pangénomique qui permet de détecter des anomalies de structure équilibrées et déséquilibrées, mais sa résolution est limitée (de l'ordre de 10 à 20 Mpb) et son succès dépend de la phase de culture cellulaire, ainsi que de la qualité des métaphases obtenues. Dans les LAL en particulier, le faible indice prolifératif des blastes *in vitro* est la cause de fréquentes défaillances du caryotype.

2.3.2 FISH

La FISH consiste en l'incorporation, par dénaturation thermique, de sondes d'ADN complémentaire marquées par un fluorochrome au niveau d'une séquence d'intérêt (Figure 2). Il est possible d'analyser les métaphases obtenues après culture cellulaire (FISH métaphasique) ou les noyaux en interphase (FISH interphasique) avec ou sans étape de culture préalable. La FISH permet la détection de nombreuses anomalies différentes avec une résolution d'environ 200 kpb selon les sondes utilisées, mais il s'agit d'une analyse ciblée et limitée en débit qui nécessite souvent, en particulier dans les LAL, l'utilisation de plusieurs sondes différentes [36].

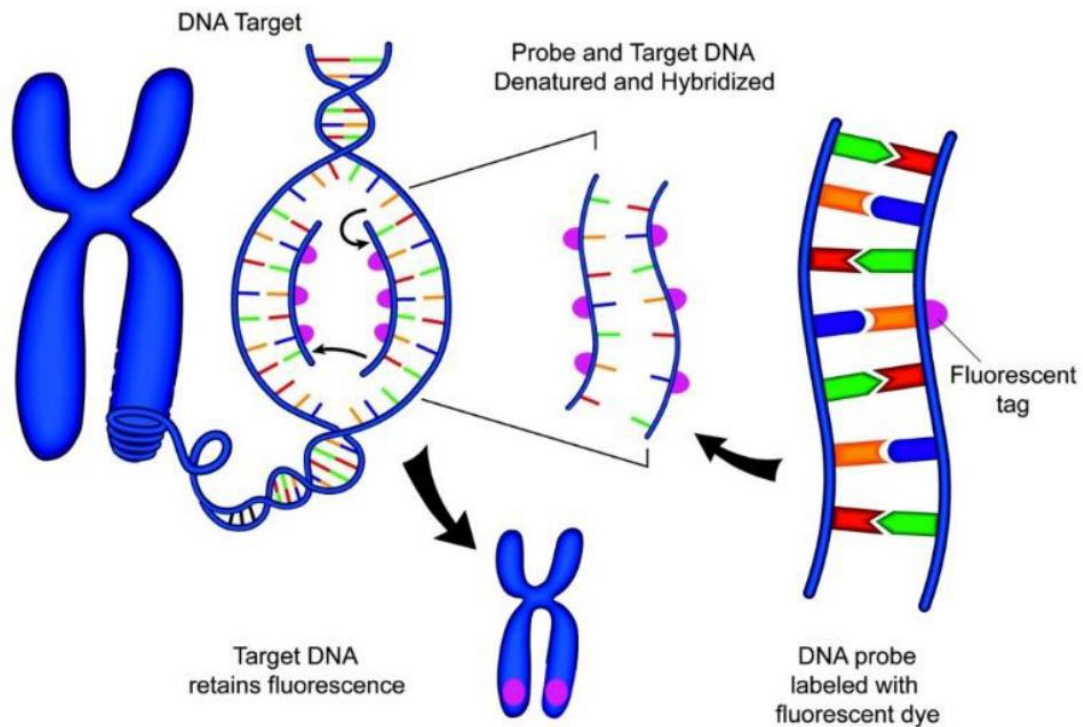


Figure 2 : Principes de l'hybridation fluorescente *in situ*

2.3.3 SNP-Array

La SNP-Array est une analyse pangénomique à haute résolution (de l'ordre du kilobase). Après une étape préalable de PCR, l'ADN va être fragmenté pour être hybridé sur une puce composée de millions d'oligonucléotides (Figure 3) permettant de mettre en évidence des « Single Nucleotide Polymorphisms » (SNPs) qui sont des variations d'une unique paire de bases réparties sur l'ensemble du génome (environ un SNP toutes les 100 à 1000 pb). Pour un locus donné, des oligonucléotides sont présents pour les 2 allèles du SNP permettant ainsi la détection de pertes d'hétérozygotie à nombre de copie neutre (CN-LOH). La SNP-Array permet également de détecter les déséquilibres du nombre de copies de l'ADN. Cependant, elle ne permet pas de mettre en évidence les anomalies équilibrées.

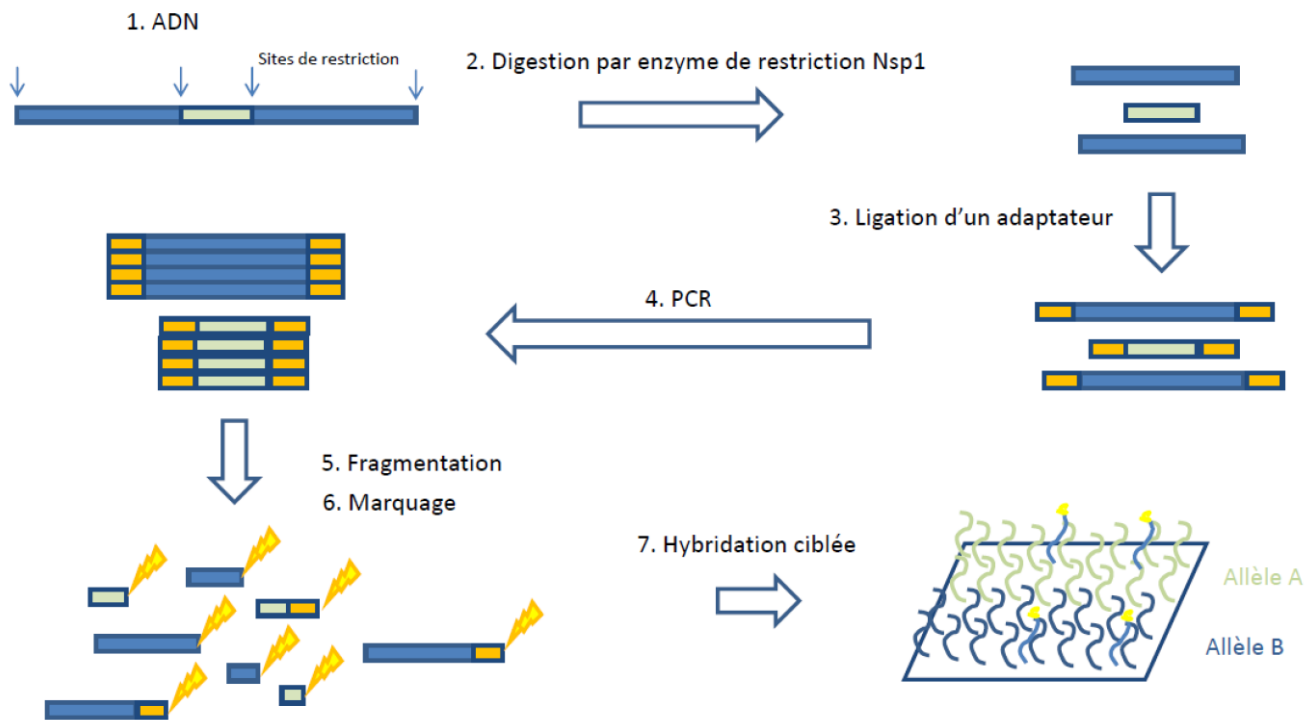


Figure 3 : Etapes de la SNP-Array (sur Cytoscan HD par Affimetrix)

2.3.4 RT-MLPA

La RT-MLPA nécessite une étape préalable de transcription inverse afin d'obtenir, à partir de l'ARN contenu dans les cellules, de l'ADN complémentaire (ADNc) qui pourra être amplifié par PCR. Après ces premières étapes, une nouvelle PCR ligation-dépendante va être réalisée. Des sondes d'oligonucléotides, choisies pour reconnaître des séquences au niveau de gènes potentiellement impliqués dans la formation de transcrits de fusion vont pouvoir venir s'hybrider sur l'ADNc (Figure 4A). Si deux sondes sont adjacentes, une ligase va venir établir une liaison entre elles (Figure 4B). Lors de l'étape finale de PCR, l'amplification n'aura lieu que s'il y a eu une ligation entre deux sondes (Figure 4C). La RT-MLPA permet ainsi, en multipliant les sondes utilisées, la détection d'un grand nombre de réarrangements équilibrés conduisant à un transcrit de fusion, mais là encore, il s'agit d'une analyse ciblée, qui ne permet pas la détection de réarrangements équilibrés impliquant d'autres mécanismes oncogénétiques.

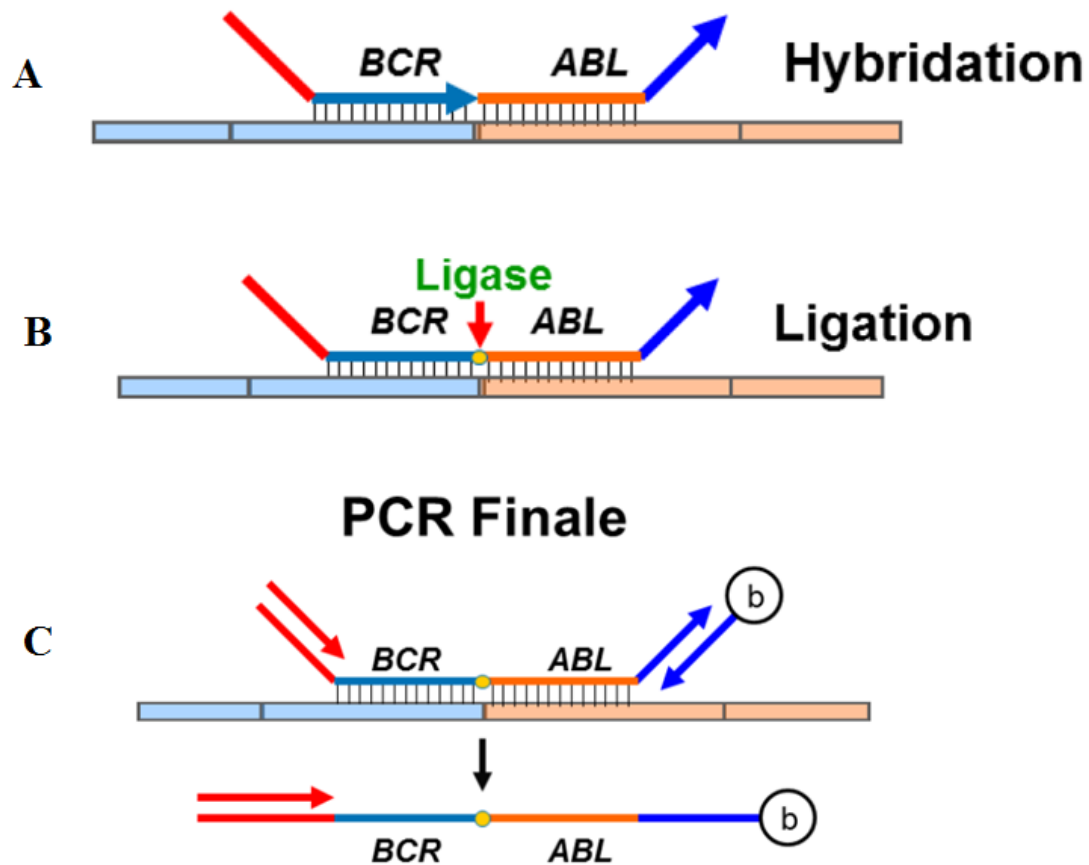


Figure 4 : Etapes de la RT-MLPA (exemple d'un transcrit *BCR-ABL1*)

La multiplication des techniques, mais aussi leur complexité et leur coût soulignent l'intérêt de développer une technique unique qui pourrait remplacer toutes celles actuellement utilisées en cytogénétique. Les technologies de séquençage à haut débit sont en passe d'offrir une telle solution, mais la détection d'anomalies structurales équilibrées reste coûteuse, avec une analyse des données très complexe, ce qui empêche leur utilisation en routine.

3 Cartographie optique du génome

3.1 Principes

La cartographie optique du génome (COG), initialement décrite par Schwartz *et al.* [37] dans les années 1990, est une technologie de reconstruction à haute résolution du génome à partir de molécules d'ADN marquées par une enzyme. Son principe est de cartographier les sites de liaison d'une enzyme sur l'ADN, ce qui permet de marquer spécifiquement chaque région du génome et de l'identifier de manière unique (à la manière d'un code barre). Cette technique permet aujourd'hui une analyse à haute résolution de l'ensemble du génome, permettant la

détection des anomalies de nombre aussi bien que de structure, des plus simples aux plus complexes, y compris les réarrangements équilibrés [38,39].

La dernière itération de cette technique, commercialisée par Bionano Genomics, se base sur l'utilisation d'ADN de très haut poids moléculaire (ADN THPM), c'est-à-dire d'au moins 150 kpb et pouvant aller jusqu'à 1 voire 2 Mpb (Figure 5A). Ces très longues molécules d'ADN sont ensuite marquées par fluorescence sur une séquence spécifique de 6 nucléotides (CTTAAG), reconnue par une enzyme, DLE-1 (Figure 5B). Cette séquence, qui se répète environ 15 fois par tranche de 100 kpb, va permettre un marquage de l'ensemble du génome, la succession de ces marquages fluorescents permettant d'identifier de manière unique la région chromosomique étudiée. Les molécules d'ADN marquées sont ensuite chargées sur une puce où un système nanofluidique va permettre la linéarisation de l'ADN (Figure 5C). La lecture des molécules devant un laser va permettre l'acquisition des images sur le scanner (Figure 5D).

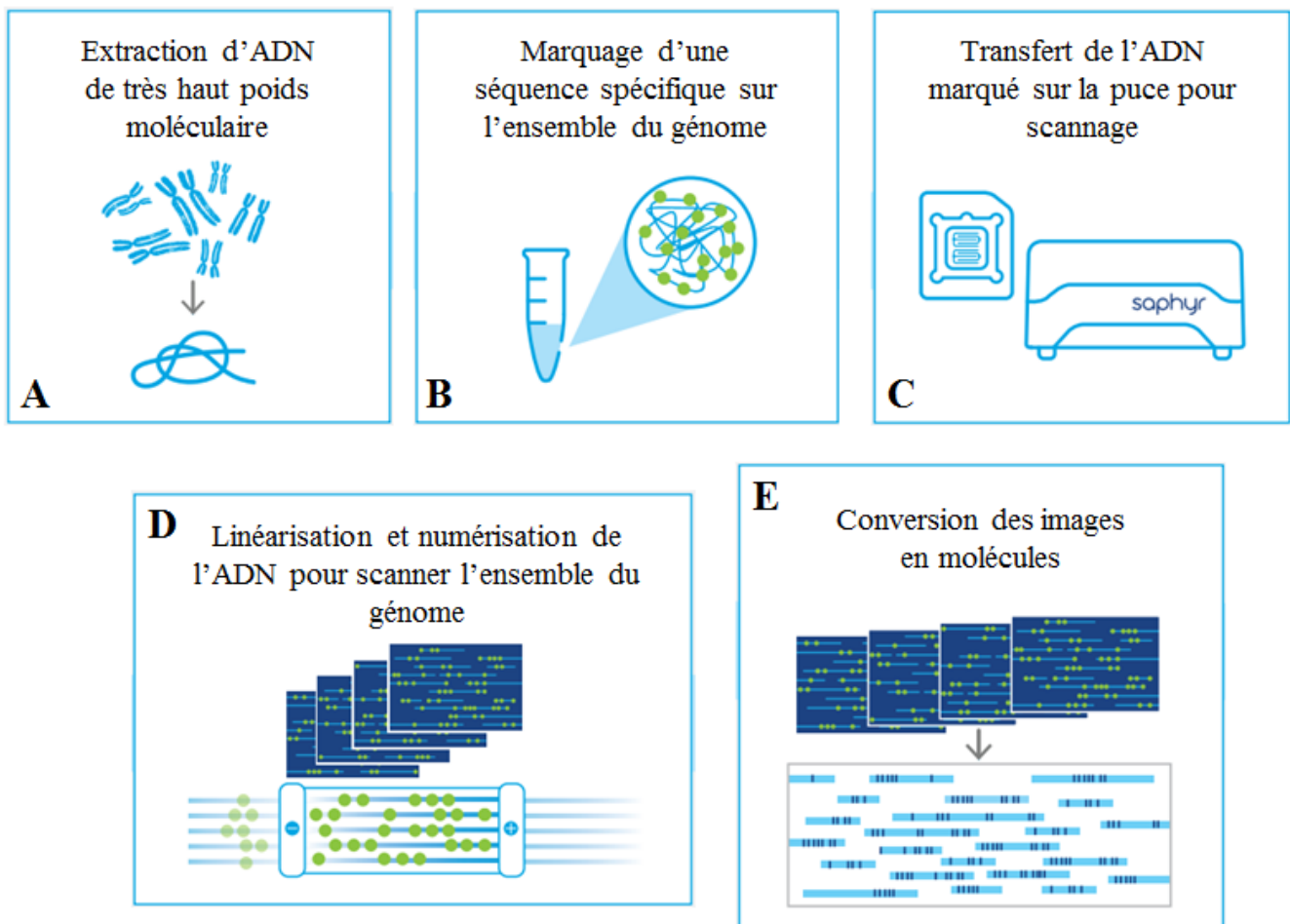


Figure 5 : Etapes de la cartographie optique du génome (adapté de Bionano)

Les images ainsi obtenues vont être numérisées pour être ensuite assemblées afin de reconstruire l'ensemble du génome (Figure 5E). Cette reconstruction peut se faire selon deux algorithmes. L'algorithme *de novo* permet un assemblage de proche en proche, par le chevauchement des molécules acquises (Figure 6). Cet assemblage est ensuite rapporté à un génome de référence. L'algorithme « Rare Variant Pipeline » (RVP) va quant à lui effectuer l'assemblage des molécules directement sur le génome de référence. C'est l'algorithme le plus adapté à la détection d'évènements rares. Enfin, l'analyse des données est faite simultanément par deux algorithmes complémentaires. L'algorithme CNV (Copy Number Variant) permet une analyse quantitative du génome, en se basant sur la densité de molécules obtenue au niveau de chaque région. L'algorithme SV (Structural Variant) permet quant à lui une analyse qualitative du génome et se base sur la séquence des marqueurs fluorescents et ses éventuelles variations par rapport au génome de référence. Il permet ainsi la détection des anomalies de structure de grande et de petite taille, qu'elles soient équilibrées ou déséquilibrées jusqu'à une résolution minimale théorique de 500 pb.

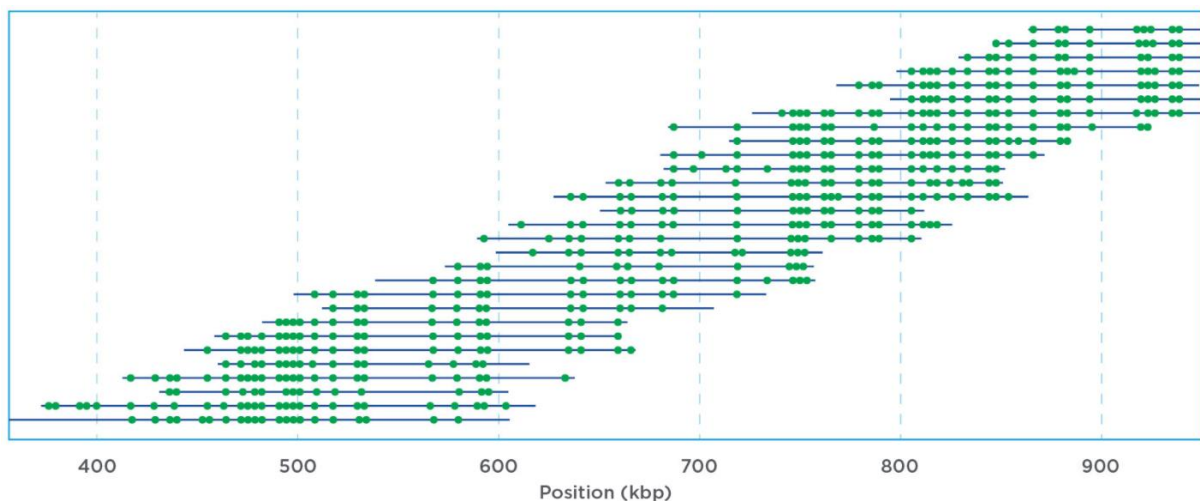


Figure 6 : Représentation la reconstruction du génome par chevauchement des molécules d'ADN marquées (d'après Bionano)

3.2 Apport en cytogénétique onco-hématologique

La COG a fait la preuve de son utilité dans de nombreux domaines de la génétique humaine. Initialement envisagée comme un outil de recherche, la COG a désormais vocation à prendre place parmi les outils disponibles dans la démarche diagnostique. Elle peut aussi être utilisée en complément des techniques de séquençage, permettant d'en guider l'assemblage [40,41]. La technologie de COG a montré des premiers résultats prometteurs pour la détection des variantes structurales et la caractérisation des pathologies constitutionnelles [42,43]. Plusieurs travaux

ont également montré la capacité de cette technologie à résoudre des anomalies complexes trouvées dans des tumeurs solides [44–46]. Enfin, la COG semble être une technique prometteuse pour la cytogénétique des hémopathies [47]. Une étude récente menée par Neveling *et al.* [48] au Pays-Bas sur 48 cas d’hémopathies malignes (leucémies aiguës myéloïdes, syndromes myélodysplasiques, leucémies myéloïdes chroniques et lymphomes B essentiellement) a permis une première évaluation de l’outil de COG dans le domaine de l’hématologie. Les cas complexes (au moins 5 anomalies) étaient distingués des cas simples (< 5 anomalies). Pour les cas simples, la COG permettait de rapporter l’intégralité des anomalies identifiées avec les techniques de référence (caryotype, FISH et SNP-Array). Les anomalies rapportées pour les cas complexes étaient elles aussi concordantes avec les techniques de références bien qu’incomplètes. La COG permettait également la caractérisation d’évènements complexes, type chromotripsis, et l’identification de nouvelles anomalies, incluant des translocations équilibrées pouvant aboutir à la description de nouveaux gènes de fusion (*RUNX1-AGBL4*, *BCR-EXOC2*). Une autre étude récente menée dans plusieurs centres aux Etats-Unis par Levy *et al.* [49] a permis de tester le potentiel de la COG sur 100 cas de leucémies aiguës myéloïdes. La COG permettait d’identifier l’ensemble des anomalies rapportées par les techniques standards (caryotype et FISH). Elle permettait également dans 11 % des cas la mise en évidence d’anomalies de structure non détectées par les techniques standards. Elle permettait de préciser la structure d’une anomalie dans 24 % des cas et a également permis dans 3 cas de mettre en évidence des translocations cryptiques comme *NSD1-NUP98*. Compte tenu de la complexité génétique des LAL, la COG pourrait avoir le potentiel de remplacer en une seule analyse l’ensemble des différentes techniques actuellement utilisées pour la caractérisation cytogénétique de ces hémopathies.

OBJECTIFS

L'objectif de notre étude était de comparer les performances techniques de la cartographie optique du génome avec les techniques utilisées en routine pour la caractérisation génétique des LAL. Nous avons sélectionné 10 échantillons de LAL-B ou T préalablement typés génétiquement par combinaison de caryotype, FISH, SNP-Array et RT-MLPA, afin de créer un panel représentatif des principales anomalies d'impact pronostique et de leur diversité structurelle. Pour la COG, nous avons utilisé le système Saphyr de Bionano et testé si les algorithmes et le logiciel développés par le fabricant pour traiter et filtrer les données étaient suffisamment fiables et faciles d'utilisation pour isoler les anomalies d'intérêt clinique des autres signaux et garantir un résultat fiable qui puisse être compatible avec une utilisation en routine hospitalière. Les variants détectés ont été triés, analysés et interprétés par deux opérateurs aveugles et indépendants, l'un ayant une formation en cytogénétique des hémopathies mais sans expérience en bioinformatique ni en cartographie optique et l'autre formé à l'utilisation de l'outil de COG mais sans connaissance en génétique des hémopathies, pour obtenir une estimation de l'impact de la profession de l'opérateur.

Nous rapportons ici les résultats de notre étude qui visait à répondre aux objectifs suivants : (i) définir la corrélation entre la COG et les techniques standard ainsi qu'entre les deux opérateurs pour la détection des anomalies génétiques des LAL (ii) analyser les discordances éventuelles, (iii) identifier de potentielles anomalies génétiques non détectées par les techniques standards et identifiées uniquement avec la COG et (iv) évaluer la capacité de la COG à détecter les CN-LOH qui, bien qu'en théorie non détectables par cette technologie, peuvent être retrouvées à l'aide de méthodes indirectes comme l'ont récemment rapporté Neveling *et al.* [48].

MATERIEL ET METHODES

1 Sélection des échantillons

Dix échantillons de patients atteints de LAL-B ou T, traités au CHU Amiens-Picardie, ont été sélectionnés. Il s'agissait de patients pour lesquels un caryotype (complété si nécessaire par une ou plusieurs FISH), une SNP-Array et une recherche de transcrits de fusion par RT-MLPA avaient été effectués au moment du diagnostic et pour lesquels un culot de cellules DMSO (Diméthylsulfoxyde) ayant une quantité suffisante de cellules était disponible. Parmi les échantillons qui remplissaient ces conditions, ceux présentant des anomalies majeures d'intérêt et de structure diverses étaient sélectionnés, notamment une t(10;14) *TLX1-TRA*, une t(9;22) *BCR-ABL1* (transcrits e1a2 et e14a2), une t(10;11) *MLLT10-PICALM*, une t(11;19) *KMT2A-MLLT1*, un profil hyperdiploïde typique, une amplification intrachromosomique du chromosome 21 (iAMP21), une fusion *SET-NUP214*, et diverses délétions impliquant notamment *CDKN2A/CDKN2B*, *IKZF1*, *PAX5*, *PTEN*, *ERG* et *FBXW7*.

2 Caryotype

Les échantillons de moelle osseuse étaient cultivés pendant 24 heures à 37°C dans un milieu composé de RPMI (milieu du Roswell Park Memorial Institute) additionné de 18,75 % de sérum de veau fœtal, de pénicilline et de streptomycine. Après culture et blocage des mitoses en métaphase par addition de colchicine, un choc hypotonique était effectué avec une solution de KCl, suivi d'une préfixation et de 3 séries de fixation avec une solution composée de 3 volumes de méthanol pour 1 volume d'acide acétique. Après avoir étalé les culots de cytogénétique obtenus sur des lames, une dénaturation thermique de l'ADN en milieu salin (bandes R) était effectuée, puis les lames étaient déshydratées dans du méthanol avant d'être colorées au Giemsa. En l'absence d'anomalies clonales, au moins 20 métaphases étaient analysées pour chaque patient. Le caryotype était décrit selon la nomenclature ISCN 2016 (International System for human Cytogenomic Nomenclature).

3 FISH

La FISH interphasique était réalisée sur des étalements frais ou congelés obtenus après culture cellulaire. Les lames étaient déshydratées dans des bains successifs d'éthanol. Après dépôt des sondes (Metasystems, Cytocell) sur la zone à étudier, les lames étaient placées dans un automate ThermoBrite (Leica Biosystems) pour codénaturation et hybridation des sondes, selon les instructions du fournisseur. Après hybridation, les lames étaient lavées dans une solution saline

de citrate de sodium (SSC) avant d'être contre-colorées avec du DAPI (4'-6-diaminido-2-phénylindole). Les lames de FISH ainsi obtenues étaient lues au microscope à fluorescence par deux observateurs différents. Au moins 200 noyaux étaient analysés pour chaque mélange de sondes afin d'établir une formule selon la nomenclature ISCN 2016.

4 SNP-Array

L'analyse par SNP-Array était réalisée sur une puce Cytoscan HD (Affymetrix). Après isolement des cellules sur Ficoll, extraction de l'ADN sur colonne Qiagen et test, les échantillons d'ADN obtenus étaient amplifiés avec le kit Affymetrix selon les recommandations du fabricant. Les données étaient analysées sur le logiciel Chromosome Assay Suite Analysis (ChAS) v3.0. Les anomalies étaient retenues lorsque leur fraction allélique était $> 10\%$ et leur taille supérieure à 1 Mpb ou > 20 kpb et impliquant l'un des 186 gènes définis comme potentiellement pertinents dans la leucémogénèse (Annexe 1). Les CN-LOH n'étaient pas inclus dans l'étude en aveugle et ont fait l'objet d'une analyse dédiée.

5 RT-MLPA

L'ARN était extrait à partir d'isollements cellulaires séparés par Ficoll et conservés dans des tampons Trizol ou Maxwell. L'extraction était réalisée manuellement (Trizol) ou de manière automatisée (Maxwell). Après dosage, la quantité d'ARN était normalisée pour obtenir une solution d'ARN de 10 μ l et 100 ng/ μ l. Une étape de rétro-transcription de l'ARN en ADNc était effectuée. La qualité de l'ADN ainsi obtenu était évaluée par PCR quantitative en temps réel en utilisant *ABLI* comme gène de référence. La qualité de l'ADN était jugée insuffisante si le nombre de copies d'*ABLI* était inférieur à 10 000 copies. Les différentes étapes de la LD-RTPCR (Ligation Dependent RTPCR) étaient réalisées sur un thermocycleur à couvercle chauffant (Mastercycler, Eppendorf). L'ADNc était hybridé avec les sondes avant d'ajouter le mix permettant la ligation. Puis l'étape de PCR finale était réalisée. Le produit de PCR était analysé sur un gel d'agarose à 4 % et révélé sous lumière UV avant d'être séquençé par pyroséquençage. Le pyrogramme ainsi obtenu était analysé pour détecter la présence éventuelle de transcrits de fusion.

6 Cartographie optique du génome

Les cartes optiques ont été réalisées sur la plateforme GENTYANE INRAE de Clermont-Ferrand, en utilisant des culots cellulaires conservés dans du DMSO.

6.1 Extraction de l'ADN et collection des données

Pour chaque échantillon, un minimum de 1,5 million de cellules était utilisé pour extraire l'ADN de très haut poids moléculaire (THPM) conformément aux instructions du fabricant (Bionano Prep SP Cells in Freezing Medium, Bionano Prep SP Frozen Cell Pellet DNA Isolation Protocol, Bionano Genomics). Après numération, les leucocytes étaient mis en culot (2200 g pendant 2mn) et traités avec du tampon de lyse LBB et de la protéinase K pour libérer l'ADN génomique (ADNg). Après inactivation de la protéinase K par traitement au PMSF (fluorure de phénylméthylsulfonyle), l'ADNg était lié à un nano-disque, lavé et élué dans un tampon approprié. L'ADN THPM était ensuite laissé à homogénéiser à température ambiante pendant une nuit.

Le lendemain, les molécules d'ADN étaient marquées à l'aide du kit de marquage ADN DLS (Direct Label and Stain, Bionano Genomics). Sept cent cinquante nanogrammes d'ADNg étaient marqués en présence de l'enzyme DLE-1 par des fluorophores verts. Après nettoyage de l'excès de fluorophores et digestion rapide de l'enzyme DLE-1 restante par la protéinase K, l'ADN était contre-coloré pendant une nuit avant d'être quantifié et visualisé sur le scanner Saphyr (Bionano Genomics).

Un volume de 8,5 µl de solution d'ADNg marqué avec une concentration comprise entre 4 et 12 ng/µl était chargé sur la puce Saphyr et scanné sur l'instrument Saphyr (Bionano Genomics). Les puces Saphyr étaient exploitées à leur capacité maximale pour atteindre un rendement minimum de 1300 Gpb et une couverture du génome de 300x. Une puce comprend 3 cellules de flux et 2 puces par série peuvent être chargées simultanément sur l'instrument. La migration durait 2 jours par série.

6.2 Analyse par le Rare Variant Pipeline et sélection des variants de structure

L'analyse par le Rare Variant Pipeline (RVP) et la sélection des variants étaient effectués via le logiciel Bionano Access (v1.5.2). Le RVP traite les données selon deux algorithmes distincts : un pour les variantes structurelles (algorithme SV) et un autre pour les variations du nombre de copie (algorithme CNV). Le seuil de détection était fixé à 20 kpb pour les insertions, les délétions et les duplications détectées par l'algorithme SV, et à 275 kpb pour l'algorithme CNV. Les événements détectés par l'analyse RVP étaient ensuite filtrés par rapport à une base de données intégrée d'échantillons non pathologiques et seuls les variants absents de cette base de données étaient pris en compte pour l'analyse.

6.3 Comparaison des données

Les anomalies étaient préfiltrées selon des paramètres de filtrage personnalisés établis par l'étude de Neveling *et al.*[48], sauf pour les translocations qui étaient toutes étudiées (scores de confiance : Insertion : 0, Délétion : 0, Inversion : 0,01, Duplication : -1, Translocation : toutes, nombre de copies : 0,99). Les données étaient ensuite analysées en aveugle sur le logiciel Bionano Access et/ou sur des fichiers .csv téléchargés, par deux opérateurs indépendants, d'une part une ingénieure formée en COG sans connaissance en cytogénétique onco-hématologique (opérateur 2) et d'autre part, moi-même, interne de biologie médicale spécialisé en cytogénétique sans formation en COG (opérateur 1).

Les opérateurs avaient pour instruction de signaler toute anomalie répondant aux critères suivants : taille supérieure à 1 Mpb ou > 20 kpb et impliquant l'un des 186 gènes définis comme potentiellement pertinents dans la leucémogénèse (Annexe 1).

Pour étudier le chevauchement des coordonnées des anomalies détectées par SNP-Array et COG, trois comparaisons étaient effectuées. Les coordonnées de toutes les anomalies détectées avec l'algorithme SV étaient comparées à leur équivalent en SNP-Array. La même comparaison était effectuée pour les anomalies détectées avec l'algorithme CNV. Parmi ces anomalies, pour celles détectées à la fois par les algorithmes SV et CNV, le chevauchement des coordonnées obtenues par chacun de ces deux algorithmes était également étudié. Les coordonnées chromosomiques des différentes anomalies étaient collectées via le logiciel Bionano Access et le calcul du chevauchement était effectué sur la plateforme Cartagenia (Agilent Technologies).

Pour la mise en évidence des CN-LOH, nous avons utilisé le pipeline bioinformatique d'assemblage du génome *de novo* qui classe chaque SV en tant qu'hétérozygote ou homozygote. Tous les filtres utilisés précédemment ont été supprimés afin de révéler tous les variants identifiables par COG. Le nombre d'anomalies considérées comme homozygotes par le pipeline était rapporté au nombre total d'anomalies dans la région concernée. Ces ratios étaient calculés pour chaque région comportant une CN-LOH puis comparés aux ratios obtenus dans ces mêmes régions sur les patients indemnes de CN-LOH.

7 Confirmation des anomalies additionnelles

Les anomalies supplémentaires identifiées par COG et considérées comme ayant un intérêt clinique potentiel étaient confirmées par FISH, CGH, réanalyse des données brutes de SNP-Array ou PCR long-range selon la nature de l'anomalie.

RESULTATS

1 Caractéristiques des patients

L'âge moyen au diagnostic était de 14 ans (2-37 ans), dont 7 enfants (2-8 ans, moyenne 5,1 ans) et 3 adultes (32-37 ans, moyenne 34,7 ans). Il s'agissait de 9 hommes et d'une femme. Les échantillons de moelle osseuse étaient prélevés lors du diagnostic, entre mai 2016 et septembre 2019. Pour 6 patients le diagnostic de LAL-B était retenu (4 B-II et 2 B-III, selon la classification EGIL). Pour les 4 autres patients, le diagnostic de LAL-T était retenu (1 T-I et 3 T-III). Les caractéristiques des patients sont indiquées dans le Tableau 4Tableau 1.

Tableau 4 : Caractéristiques des patients

Patient	Age au diagnostic	Sexe	Classification EGIL
1	6	M	T-III
2	8	F	T-III
3	4	M	B-II
4	5	M	B-II
5	35	M	T-I
6	37	M	T-III
7	5	M	B-II
8	6	M	B-III
9	32	M	B-II
10	2	M	B-II

2 Données qualité de la cartographie optique du génome

La technique d'extraction de l'ADN THPM donnait des molécules avec un N50 moyen (≥ 150 kpb) de 286 kpb (258-334), un taux moyen de cartographie de 83,31 % (71,6-90,7), une densité moyenne de marqueurs de 14,25 marqueurs/100 kpb (12,54-15,28), un taux d'absence de

marqueurs moyen (Negative Label Variance : NLV) de 12,85 % (10,3-23,5) et une couverture moyenne de 339x (289,04-379,65). Les échantillons 6 et 8, présentant une NLV > 20 %, avaient un nombre très élevé de variants avec l'algorithme CNV (214 et 201 respectivement). Une majorité de ces signaux, interprétés comme du bruit de fond, ont été éliminés au cours de l'analyse en rejetant les délétions détectées par l'algorithme CNV si elles n'étaient pas confirmées par l'algorithme SV et avaient un nombre de copies < 1,5.

3 Identification des anomalies d'intérêt clinique par cartographie optique du génome

Au total, 80 anomalies ont été trouvées au moins par SNP-Array, pour l'ensemble des patients. Deux de ces anomalies étaient des CN-LOH et ont donc été analysées à part car théoriquement non détectables par COG. Sur les 78 anomalies restantes, 70 répondaient aux critères décrits ci-dessus pour être incluses dans la comparaison en aveugle. Dix autres anomalies non détectées par SNP-Array ont été détectées par caryotype conventionnel, FISH et/ou RT-MLPA. Ainsi, un total de 80 anomalies répondant aux critères de sélection de la SNP-Array et/ou trouvées avec les autres techniques standards, ont été retenues pour la comparaison en aveugle. Parmi ces 80 anomalies, 73 (91,25 %) ont été trouvées par l'opérateur 1 et 72 (90 %) par l'opérateur 2 (Figure 7). Toutes les translocations (8/8), 43/45 (95,6 %) délétions, 10/12 (83,3 %) gains ont été trouvés par les deux opérateurs, tandis que 12/15 (80 %) et 11/15 (73,3 %) gains de chromosomes entiers ont été trouvés par l'opérateur 1 et 2 respectivement (Tableau 5).

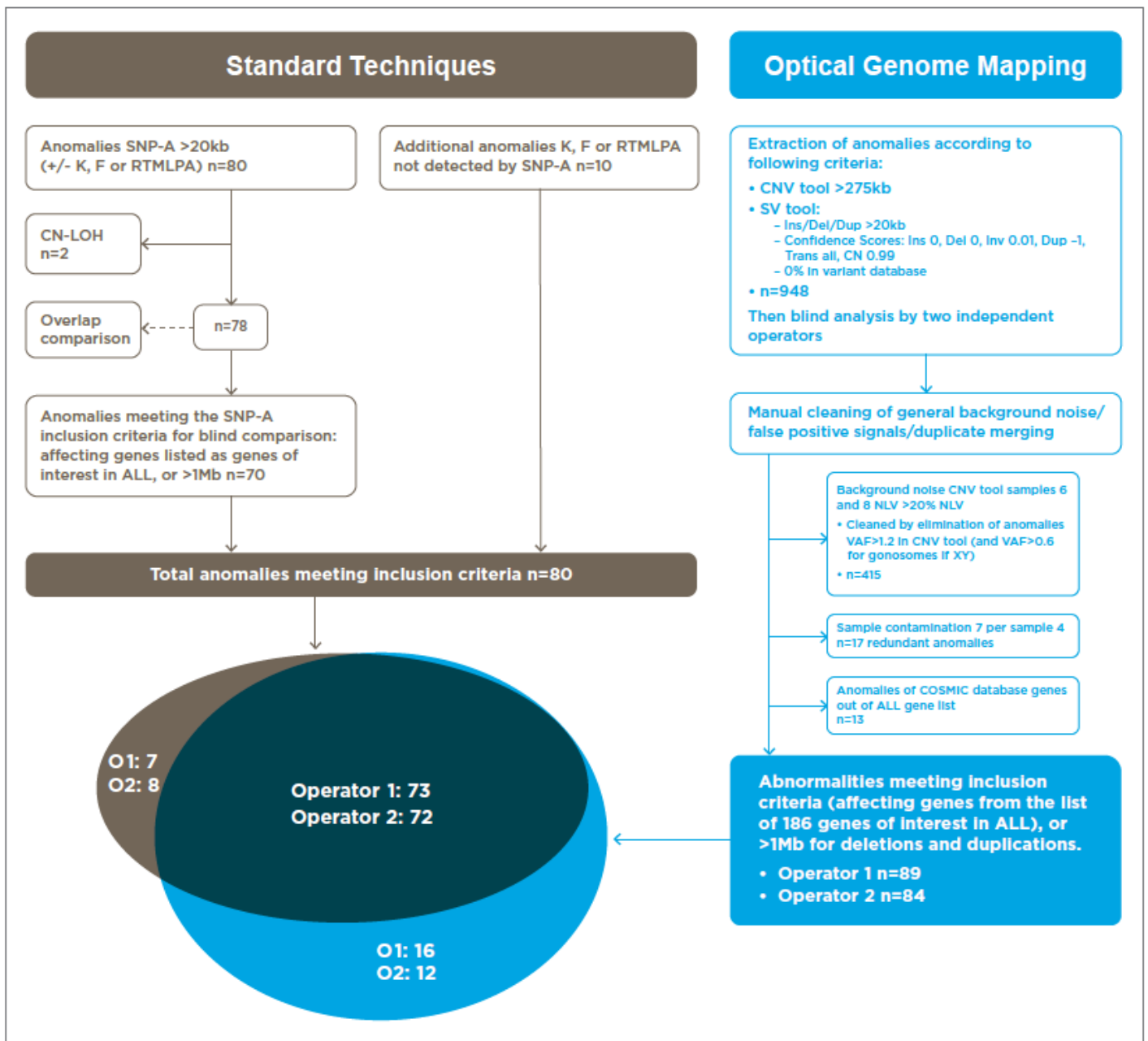


Figure 7 : Diagramme de flux des anomalies détectées par les techniques standards et par cartographie optique du génome

Tableau 5 : Anomalies détectées par les techniques standards et correctement identifiées par cartographie optique du génome

1	<p>Deletion 9p21.3 involving <i>CDKN2A</i>, <i>CDKN2B</i>, <i>MLLT3</i>, <i>MTAP</i></p> <p>Translocation t(10;14) <i>TLX1-TRD</i></p> <p>Deletion 13q14.2 involving <i>RBI</i></p> <p>Deletion 17p13.3-p13.1 involving <i>TP53</i></p> <p>Gain 20p13-p11.21 involving <i>SLX4IP</i></p> <p>Gain 20q11.21-q13.33</p>	<p>Deletion 2q34-q35 involving <i>IKZF2</i></p> <p>Deletion 5q13.1-q13.2 without genes of interest</p> <p>Gain 6q23.3 involving <i>MYB</i>, <i>AHII</i></p> <p>Deletion 9p21.3 involving <i>CDKN2A</i>, <i>CDKN2B</i>, <i>MTAP</i></p> <p>Deletion 9q34.11-q34.13 leading to <i>SET-NUP214</i> fusion</p> <p>Deletion 10q24.31-q24.32 involving <i>TLX1</i></p> <p>Deletion 10q25.1 involving <i>ADD3</i></p> <p>Deletion 11p13 involving <i>WT1</i></p> <p>Deletion 11q22.1-q22.2 without genes of interest</p> <p>Deletion 16p13.3 involving <i>CREBBP</i></p> <p>Deletion 16q11.2-q22.1 without genes of interest</p> <p>Gain 19p13.2-p11 involving <i>EPOR</i>, <i>LYLI</i></p>
2	<p>Deletion 1p33 leading to <i>STIL-TAL</i> fusion</p> <p>Deletion 4q31.3 involving <i>FBXW7</i></p> <p>Deletion 9p21.3 involving <i>CDKN2A</i>, <i>CDKN2B</i>, <i>MTAP</i></p> <p>Deletion 10q23.31 involving <i>PTEN</i></p> <p>Deletion 21q22.2 involving <i>ERG</i></p>	<p>Deletion 11q22.1-q22.2 without genes of interest</p> <p>Deletion 16p13.3 involving <i>CREBBP</i></p> <p>Deletion 16q11.2-q22.1 without genes of interest</p> <p>Gain 19p13.2-p11 involving <i>EPOR</i>, <i>LYLI</i></p>
3	<p>Deletion 1q42.2-q42.3 without genes of interest</p> <p>Deletion 1q44 without genes of interest</p> <p>Whole chromosome gain 3</p> <p>Whole chromosome gain 6</p> <p>Whole chromosome gain 10</p> <p>Whole chromosome gain 14</p> <p>Whole chromosome gain 17</p> <p>Whole chromosome gain 18</p> <p>Whole chromosome gain 21</p> <p>Whole chromosome gain X</p>	<p>Whole chromosome gain 5</p> <p>Deletion 6q13-q22.33 involving <i>CASP8AP2</i></p> <p>Deletion 7p12.2 involving <i>IKZF1</i></p> <p>Translocation t(7;21) <i>SERPINE1-DONSON</i></p> <p>Whole chromosome gain 10</p> <p>Translocation t(11;19) <i>KMT2A-MLLT1</i></p> <p>Gain 20q11.21-q13.33 without genes of interest</p> <p>Gain 21q11.2-q22.3 involving <i>ERG</i>, <i>RUNX1</i></p>
4	<p>Deletion 5q33.3 involving <i>EBF1</i></p> <p>Translocation t(7;9)</p> <p>Deletion 7p12.2 involving <i>IKZF1</i></p> <p>Deletion 9p24.3-p13.1 involving <i>PAX5</i>, <i>CDKN2A</i>, <i>CDKN2B</i>, <i>MLLT3</i>, <i>MTAP</i>, <i>JAK2</i></p> <p>Gain 9q32-q34.3 involving <i>ABL1</i>, <i>NUP214</i>, <i>NOTCH1</i></p> <p>Translocation t(9;22) <i>BCR-ABL1</i>, e1a2 transcript</p> <p>Deletion 11p12 involving <i>RAG2</i>, <i>RAG1</i></p> <p>Deletion 13q14.11 involving <i>SERP2</i></p> <p>Deletion 20p12.2 involving <i>SLX4IP</i></p> <p>Deletion 20q11.22-q13.33 without genes of interest</p>	<p>Deletion 13q14.2 (0,24Mb) involving <i>RBI</i></p> <p>Deletion 13q14.2 (0,09Mb) involving <i>RBI</i></p> <p>Deletion 13q14.2-q14.3 involving <i>DLEU1</i>, <i>DLEU2</i></p> <p>Deletion 21q11.2-q21.1 without genes of interest</p> <p>iAMP21 : Gain 21q21.1-q22.3 involving <i>ERG</i>, <i>RUNX1</i></p> <p>Deletion 21q22.3 without genes of interest</p> <p>Whole chromosome gain X</p>
5	<p>Gain 6q23.3 involving <i>MYB</i>, <i>AHII</i></p> <p>Translocation t(10;11) <i>PICALM-MLLT10</i></p> <p>Whole chromosome gain 13</p> <p>Deletion 16p13.3 involving <i>CREBBP</i></p> <p>Deletion 16p13.3-p13.13 without genes of interest</p> <p>Gain 17q25.2-q25.3 involving <i>SEPT9</i></p>	<p>Deletion 3q26.32 involving <i>TBL1XR1</i></p> <p>Deletion 7p22.3-p11.2 involving <i>IKZF1</i>, <i>CARD11</i></p> <p>Deletion 9p24.3-p21.3 involving <i>MLLT3</i>, <i>JAK2</i></p> <p>Deletion 9p21.3 involving <i>CDKN2A</i>, <i>CDKN2B</i>, <i>MTAP</i></p> <p>Deletion 9p21.3-p13.1 involving <i>PAX5</i>, <i>CDKN2B</i></p> <p>Translocation t(9;22) <i>BCR-ABL1</i>, e14a2 transcript</p>
		<p>Deletion 7p12.2 involving <i>IKZF1</i></p> <p>Highly complexity rearrangement involving chromosomes 7, 9, 13 and 16</p> <p>Deletion Xp21.1 involving <i>DMD</i></p>

3.1 Discordances entre les techniques

Huit anomalies étaient manquées par COG (Tableau 6). Quatre d'entre elles étaient liées à des problèmes de filtrage. Trois étaient des gains importants avec une faible fréquence allélique, clairement visibles dans les données brutes, mais rejetés par les filtres logiciels de l'algorithme CNV, totalement, ou partiellement pour une trisomie 3 ce qui causait la discordance entre les deux opérateurs (Annexes 2 et 3). La quatrième était une délétion du gène *VPREB1*, éliminée par erreur par les opérateurs parce qu'elle était incluse ("cachée") dans des zones de délétion récurrentes du locus *IGL*. Les 4 autres anomalies manquées étaient des anomalies réellement non détectées. Deux d'entre elles étaient des variations du nombre de copies dans la région pseudo-autosomale (PAR) de Xp22.33, une zone non couverte par la technologie de COG. La première était une microdélétion causant une fusion *P2RY8-CRLF2*. Les fusions *P2RY8-CRLF2* sont d'une importance clinique majeure dans les LAL-B et leur détection reste un défi. L'association de la RT-MLPA et de la SNP-Array permet de les détecter dans la majorité des cas, mais les faux négatifs restent fréquents. Un autre gain sous-clonal (trisomie 21) était trouvé dans 7/22 métaphases au caryotype mais pas par SNP-Array ni par COG, y compris dans les données brutes. Nous avons donc confirmé la présence de ce gain par FISH interphasique et l'avons quantifié à seulement 7 % des noyaux étudiés. On peut supposer que le clone trisomique avait un indice de prolifération *in vitro* élevé. La limite de détection de la SNP-Array et de la COG étant fixée à une fraction allélique de 10 %, ces technologies ne pouvaient pas mettre en évidence ce clone. La dernière anomalie était un sous-clone quasi tétraploïde (6/19 métaphases avec 80~95 chromosomes), trouvé uniquement au caryotype, qui pourrait être strictement tétraploïde et par définition indétectable par des techniques quantitatives.

3.2 Discordances entre les opérateurs

La première discordance entre les opérateurs concernait un gain du chromosome 3 pour lequel l'extraction des données ne montrait qu'un gain partiel 3q, qui était interprété comme tel par l'opérateur 2. Le gain chromosomique complet était cependant visible dans les données brutes de la COG et a donc été interprété comme une trisomie 3 complète par l'opérateur 1 (Annexe 2). L'autre discordance entre les opérateurs concernait 4 anomalies supplémentaires, dont 3 délétions et 1 gain, qui étaient signalées par l'opérateur 1. Ces anomalies répondaient aux critères de taille (> 1 Mpb), n'impliquaient aucun des gènes d'intérêt listés en Annexe 1 et n'étaient observées qu'avec l'algorithme CNV. Cependant, ces 4 anomalies n'ont pas été retenues par l'opérateur 2, notamment en raison de leur faible fréquence allélique et ont été interprétées comme du bruit de fond.

Tableau 6 : Anomalies détectées par les techniques standards et manquées en cartographie optique du génome

Sample	Anomaly	Failure explanation
3	Whole chromosome gain 3 Whole chromosome gain 4	Detected in raw data : filtering issue Detected in raw data : filtering issue
4	Deletion 22q11.22-q11.23 involving <i>VPREB1</i> Gain 22q11.23-q13.33 involving <i>IL2RB</i>	Available in extraction list : operators failure Detected in raw data : filtering issue
5	Near tetraploid sub-clone 80~95<4n>	Non detected : probable true tetraploid clone
8	Gain Xp22.33 involving <i>SHOX</i> Deletion Xp22.33 leading to <i>P2RY8-CRLF2</i> fusion	Non detected : Pseudo Autosomal Region Non detected : Pseudo Autosomal Region
10	Sub-clonal whole chromosome gain 21	Non detected : below the limit of detection

4 Identification d'anomalies chromosomiques non détectées par les techniques standards

La COG a révélé 12 anomalies répondant aux critères de sélection qui n'avaient pas été décrites ou trouvées par les techniques standards (Tableau 7). Il s'agissait de 3 délétions, 1 gain, 1 duplication et 7 réarrangements inter-chromosomiques (dont 2 étaient des événements très complexes). Parmi les nouveaux réarrangements interchromosomiques identifiés en COG, certains conduisaient à des fusions *LMO2-TRA*, *TRB-MYC* et *IGH-CEBPB*, et un autre impliquait *DNMT3B*. Il est intéressant de noter que d'autres anomalies identifiées impliquaient des gènes d'intérêt tels que *ZMYM2*, *LEF1*, *CBFA2T3*, *TLX1* et *SETD2* et pourraient être pertinentes en termes de leucémogénèse.

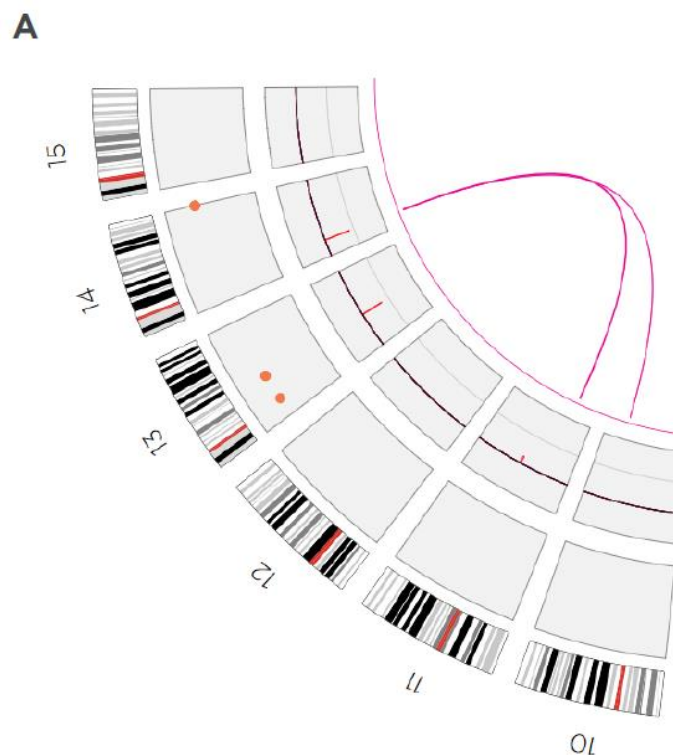
Tableau 7 : Anomalies décrites par cartographie optique et non détectées par les techniques standards

Sample	Anomaly	Size (kpb)	Anomaly confirmation
1	Interchromosomal rearrangement involving chromosomes 10, 11 and 14 leading to <i>TRA-LMO2</i> fusion	NA	<i>TRAD & LMO2</i> FISH
	Deletion 11q13.3-q13.4 without genes of interest	1 762,807	False positive signal
	Deletion 13q12.11 near <i>ZMYM2</i>	79,292	SNP-Array raw data
2	Deletion 4q25 involving <i>LEF1</i>	25,712	SNP-Array raw data
	Interchromosomal rearrangement involving chromosomes 7 and 8, leading to <i>TRB-MYC</i> fusion	NA	<i>TRB & MYC</i> FISH
3	Deletion 2q12.3-q13 without genes of interest	1 282,087	False positive signal
	Gain 16q24.3 involving <i>CBFA2T3</i>	186,705	SNP-Array raw data
4	Gain 6q27 without genes of interest	1 280,768	False positive signal
5	Duplication 10q24.31 near <i>TLX1</i>	53,614	SNP-Array raw data
	High complexity rearrangement involving chromosomes 9, 12, 16, 17 and 19	NA	Not verified
7	Deletion 2p25.3 without genes of interest	2 038,497	False positive signal
	High complexity rearrangement involving chromosomes 7, 10, 20 and 21	NA	Not verified
	Interchromosomal rearrangement involving chromosomes 20 and 21, and <i>DNMT3B</i>	NA	Under investigation by long range PCR
8	Interchromosomal rearrangement involving chromosomes 4 and 5, leading to <i>RELL1-TSLP</i> fusion	NA	Not verified
9	Interchromosomal rearrangement involving chromosomes 14 and 20, leading to <i>IGH-CEBPB</i> fusion	NA	<i>IGH & CEBPB</i> FISH
10	Deletion 3p21.31 involving <i>SETD2</i>	198,852	Not confirmed by CGH

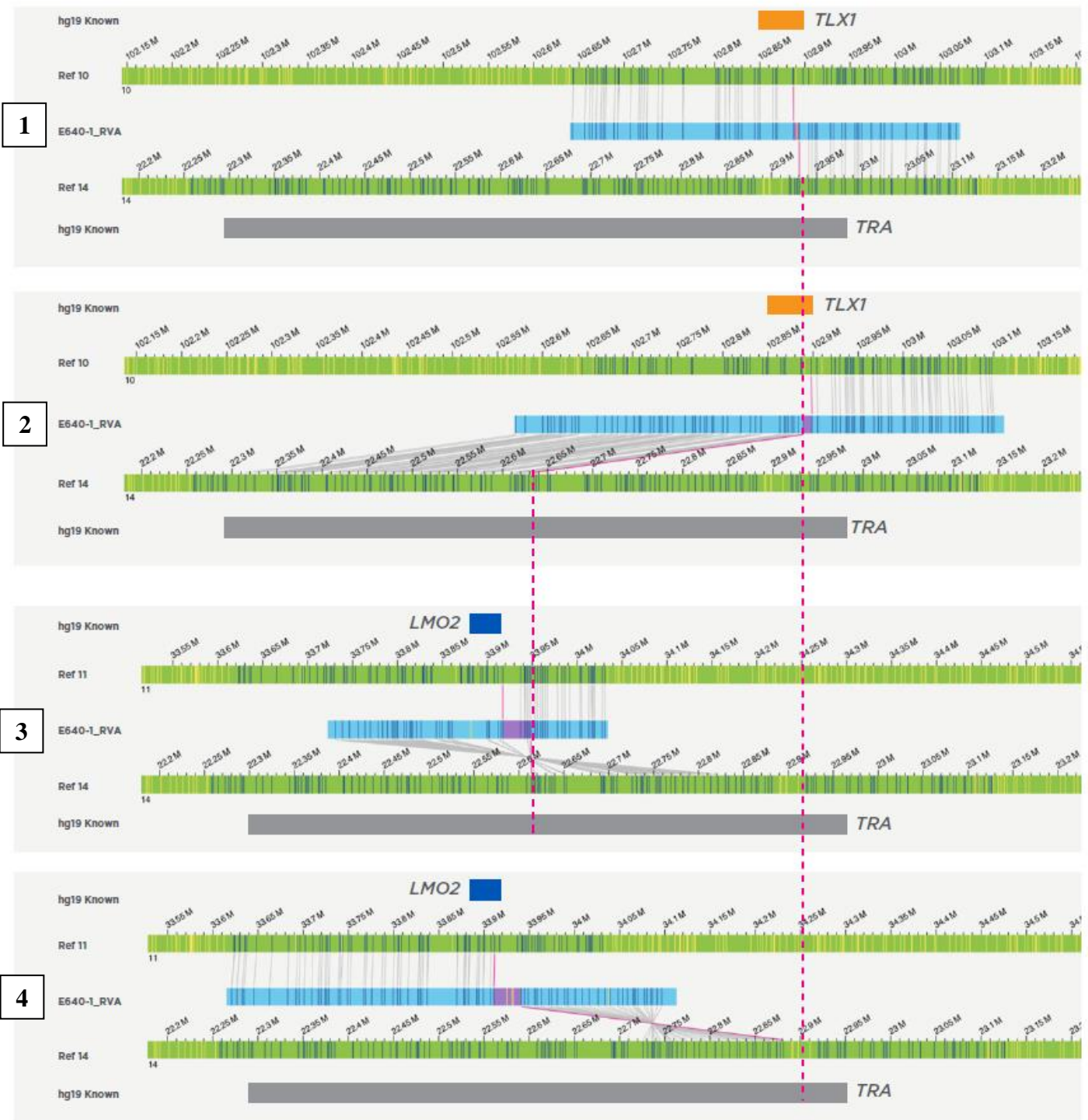
Le réarrangement chromosomique impliquant les gènes *TRA* et *LMO2* était en fait plus complexe et étroitement lié à la translocation *TLX1-TRA* déjà décrite par le caryotype conventionnel. Ces deux réarrangements étaient en fait un événement complexe à 4 points de cassure impliquant les chromosomes 10, 11, 14 respectivement aux loci *TLX1*, *LMO2* et *TRA*, résultant en une insertion cryptique d'une partie du locus *TRA* dans le gène *LMO2* (Figure 8). L'analyse de ce réarrangement par la technologie de COG permettait également de comprendre pourquoi seule la translocation de *TLX1* était visible au caryotype.

Concernant la translocation t(20;21) nouvellement découverte, le point de cassure étant situé proche de *DNMT3B* sur le chromosome 20, nous avons jugé intéressant d'étudier cette anomalie. Une PCR long range avec des sondes de part et d'autre des points de cassure chromosomiques doit être effectuée pour mettre en évidence un potentiel dérivé de translocation.

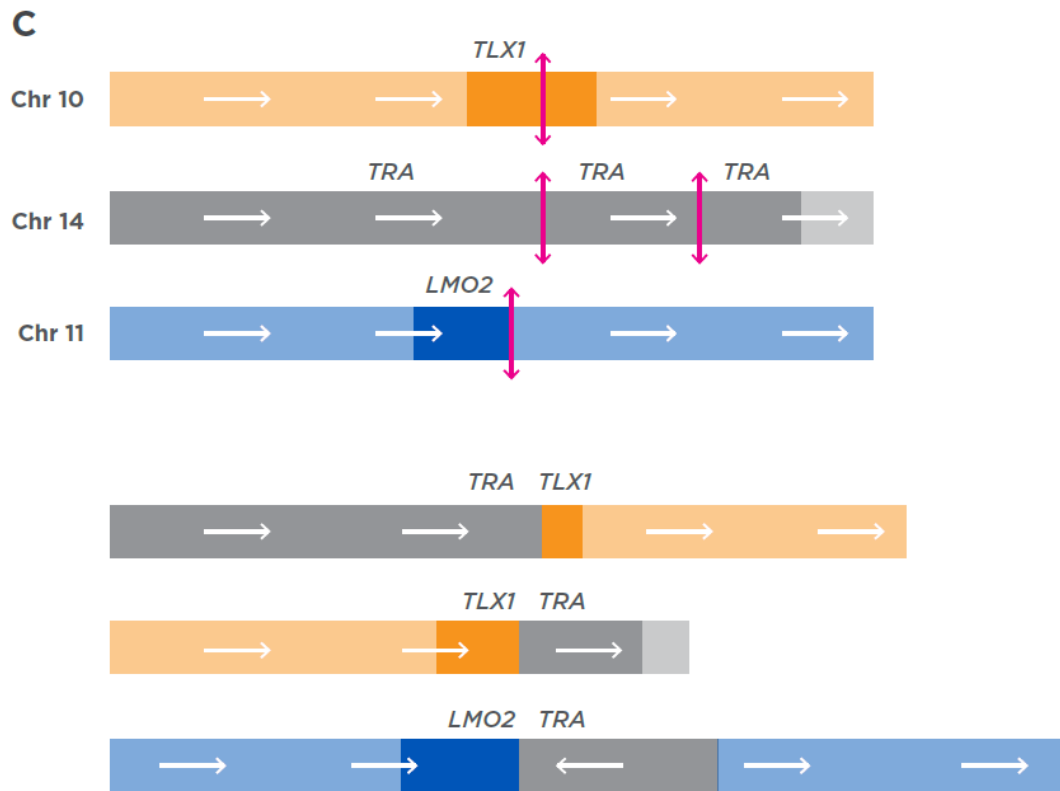
Figure 8 : Caractérisation par cartographie optique du génome d'un réarrangement complexe des chromosomes 10, 11 et 14 et impliquant les gènes *TLX1*, *LMO2* et *TRA*



A. Circos plot de la COG représentant le réarrangement entre les chromosomes 10, 11 et 14 (lignes roses).

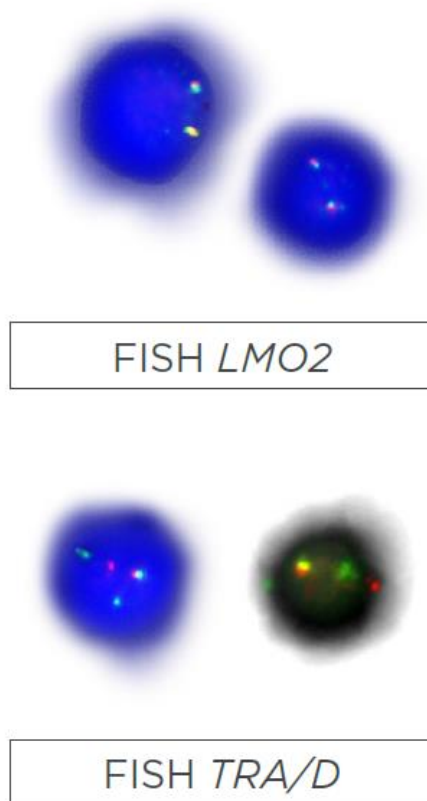
B

B. Cartes de consensus représentant les points de cassure au niveau des chromosomes 10, 11 et 14 (lignes pointillées roses) et les gènes impliqués (logiciel Bionano Access). Les cartes vertes représentent les cartes de référence des chromosomes 10, 11 et 14. Les cartes bleues sont le résultat de la COG. Elles représentent les anomalies retrouvées pour le patient et les différents dérivés du réarrangement. Les cadres 1 et 2 représentent les deux dérivés de la translocation $t(10;14)$ *TLX1-TRA*. Les cadres 3 et 4 représentent les deux dérivés du réarrangement 11-14 *LMO2-TRA*. Ces cartes montrent qu'il existe deux points de cassure différents dans *TRA*, chacun étant alignés avec l'un des dérivés *TLX1* et *LMO2*.



C. Schématisation des points de cassure dans les régions chromosomiques impliquées (en haut). Le double point de cassure du locus *TRA* est à l'origine d'une translocation *TLX1-TRA* et d'un réarrangement *LMO2-TRA* par insertion et inversion (en bas).

D



D. FISH interphasique avec des sondes de séparation *LMO2* (en haut) et *TRA* (en bas). La sonde de FISH *LMO2* ne montre pas de séparation (2 signaux de fusion). La sonde de FISH *TRA* montre un signal de fusion (chromosome 14 non réarrangé), un signal de séparation (1 signal vert et 1 signal rouge séparés) et un extra signal vert. Le signal de séparation correspond aux dérivés de la translocation *TLX1-TRA*. L'extra signal vert de la sonde de FISH *TRA* est compatible avec l'insertion d'une partie de *TRA* dans *LMO2*.

5 Corrélation entre les coordonnées chromosomiques des anomalies décrites en COG et en SNP-Array

Au total, 41 anomalies étaient décrites à la fois par SNP-Array et par l'algorithme SV de la COG. Pour ces anomalies, le chevauchement médian des coordonnées chromosomiques était de 94 % (de 18 à 99 %, SEM 3 %), avec 8/41 anomalies < 75 %. Cinquante-trois anomalies étaient décrites à la fois par SNP-Array et par l'algorithme CNV de la COG. Leur chevauchement médian était de 97 % (de 24 à 99 %, SEM 2 %), avec 4/53 anomalies < 75 %. Parmi les anomalies pour lesquelles une comparaison avec la SNP-Array était possible, 21 étaient trouvées parallèlement par les deux algorithmes d'analyse (SV et CNV) de la COG. Le chevauchement médian entre ces anomalies était de 95 % (de 18 à 100 %, SEM 6 %), avec 4/21 anomalies < 75 %.

6 Détection des pertes d'hétérozygotie en cartographie optique du génome

Dans la première région de CN-LOH [9q21.11q34.3(71,013,799-141,025,328)], en moyenne 231 (210-246) SV par échantillon étaient détectés. Pour l'échantillon 1, qui portait cette région de CN-LOH, 97,1 % (204/210) de SV homozygotes étaient observés, comparé à une moyenne de 67,7 % (148 à 167 variants homozygotes pour un total entre 210 et 246) dans les autres échantillons. Dans la deuxième région de CN-LOH [9p24.3p21.1(192,128-33,190,239)], en moyenne 85 (80-89) SV par échantillon étaient détectés. Pour l'échantillon 2, qui portait ce CN-LOH, 94,05 % (79/84) de SV homozygotes étaient observés par rapport à une moyenne de 56,6 % (38 à 54 variants homozygotes pour un total entre 80 et 89) dans les autres échantillons. Il convient de noter que l'échantillon 10 a été exclu de cette deuxième comparaison de données car il présentait une délétion 9p24.3p13.1 avec un nombre de copies de 1,05.

DISCUSSION

Dans cette étude, nous avons cherché à comparer les performances de la COG par rapport aux examens standards, dans le domaine de la cytogénétique onco-hématologique. Nous avons choisi d'étudier les LAL car elles comportent un nombre élevé d'anomalies génétiques d'une grande diversité structurelle. Il existe un intérêt médical, organisationnel et économique considérable à simplifier le workflow des analyses génétiques des LAL. Bien que le nombre de patients soit limité, les échantillons sélectionnés présentaient un grand nombre d'anomalies structurelles d'intérêt, représentatives de celles que l'on trouve habituellement dans les cas de LAL-B et T. Ces anomalies étaient de type et de taille variés (délétions, translocations équilibrées ou complexes, duplications, insertions, inversions, aneuploïdies), permettant d'explorer le spectre des variants détectables et les limites de la technique.

1 Efficacité de la COG pour la détection des anomalies chromosomiques des LAL

Tout d'abord, nous avons constaté une excellente corrélation entre les techniques standards et la COG en tant qu'examen unique pour la détection des anomalies chromosomiques liées aux LAL ; en effet, une grande majorité (90,5 %) des anomalies identifiées comme pertinentes et décrites par les techniques standards a été identifiée à l'aide de la COG. Cependant, 8 anomalies ont été manquées par l'analyse des données de COG. Un examen approfondi des données brutes et des filtres a permis d'expliquer chacune de ces défaillances. La moitié de ces défaillances étaient détectées sur les données brutes mais n'étaient pas retenues en raison des filtres utilisés, ce qui souligne la nécessité d'améliorer l'outil de filtrage afin d'éviter un examen visuel et manuel du graphique CNV alors que la COG est capable de détecter ces anomalies de ploïdie sous-clonales. Cependant, 4 étaient réellement manquées, soit en relation avec une faible fréquence allélique, inférieure à 10 %, qui est la limite de sensibilité de la technique en CNV, soit en raison d'une faible couverture de certaines régions peu accessibles à la COG, en particulier les régions génomiques contenant une forte variabilité interindividuelle ou des séquences très répétées telles que les centromères, les télomères et les PAR. Cette limitation rend impossible la détection des anomalies récurrentes survenant dans ces régions, comme les réarrangements *CRLF2*, bien décrits dans les LAL et dont l'impact pronostique est encore débattu [10].

2 Accessibilité de la technologie de COG

Les discordances observées entre les deux opérateurs étaient rares, faciles à expliquer et d'une importance mineure dans l'interprétation globale des dossiers. La première discordance entre les opérateurs, concernant la trisomie 3, était davantage liée à la technique qu'à son interprétation. Enfin, des anomalies supplémentaires signalées uniquement par l'opérateur 1 et considérées comme du bruit de fond par l'opérateur 2 n'ont pas non plus été trouvées dans les données brutes de SNP-Array. Ces anomalies n'étaient décrites qu'avec l'algorithme CNV et avaient une fréquence allélique relativement faible. Dans ces 4 cas, aucun gène d'intérêt n'était impliqué, ce qui limite l'impact clinique de ces anomalies. Cependant, ces signaux, considérés finalement comme faux positifs, devraient conduire à relativiser l'interprétation de l'algorithme CNV et peut-être à limiter son utilisation aux seules anomalies de très grande taille. La corrélation globale entre les deux observateurs était cependant excellente, malgré des parcours professionnels totalement différents, le premier étant interne de biologie médicale, spécialisé en hématologie, le second ingénieur spécialiste en COG mais sans compétence en hématologie maligne. Les technologies émergentes en matière d'analyse génétique nécessitent souvent la mobilisation d'opérateurs ayant une formation et des compétences avancées en bio-informatique que les biologistes médicaux ne possèdent pas toujours. Nos données montrent clairement que la COG est accessible à tout biologiste formé à la biologie moléculaire et à la cytogénétique.

3 Identification par COG d'anomalies non détectées par les techniques standards

Douze anomalies étaient identifiées par la COG mais pas par les autres techniques ; 7 d'entre elles ont pu être confirmées. Dans la majorité des cas, les gains et les délétions étaient confirmés par une nouvelle analyse des données brutes de SNP-Array et les translocations l'étaient par FISH. La COG a donc la capacité de détecter des anomalies passées inaperçues avec les techniques standards, y compris celles à haute résolution. Parmi ces anomalies, certaines étaient bien identifiées comme récurrentes dans les LAL, mais passées inaperçues avec les techniques standards. Nous avons ainsi identifié un réarrangement conduisant à la fusion de l'oncogène *MYC* avec *TRB*, qui pourrait ainsi être hyperexprimé, dans un cas de LAL-T (Annexe 4). Dans un autre cas avec un remaniement *TLX1-TRA*, la haute résolution de la COG a permis de mettre en évidence un réarrangement plus complexe impliquant un troisième gène, *LMO2* qui pourrait être surexprimé secondairement à l'insertion du *TRA*. Les LAL-T avec une dérégulation de l'expression de *LMO2* et *TLX1* sont classiquement classées dans des sous-groupes génétiques

bien distincts et ne sont pas associées [50,51]. Il s'agit à notre connaissance du premier cas faisant état d'un remaniement concomitant de ces deux gènes dans une LAL-T.

4 Description de nouvelles anomalies potentiellement impliquées dans la leucémogénèse

Dans notre série, nous avons pu identifier certaines anomalies impliquant des gènes qui peuvent avoir un rôle direct dans le processus leucémique. En effet, dans le cas d'une LAL-B, nous avons identifié le gène *DNMT3B*, régulateur épigénétique important pour la maintenance tumorale [52], comme un partenaire potentiel dans un remaniement entre les chromosomes 20 et 21 qui n'était pas détecté par les autres techniques. Cette anomalie est en cours de vérification et pourrait mener à une surexpression de *DNMT3B*. Par ailleurs, en examinant de plus près les variants rejetés identifiés par COG, mais non sélectionnés parce qu'ils n'étaient pas inclus dans notre liste de 186 gènes d'intérêt, nous avons trouvé 16 gènes référencés dans la base de données COSMIC (Catalogue Of Somatic Mutations In Cancer)[53] et qui pourraient avoir un rôle potentiel dans la leucémogénèse des cellules B/T : *CELA3B*, *CDH4*, *OTP*, *KDM2B*, *PSG11*, *PSG2*, *PSG5*, *PSG4*, *SOX4*, *P2RY2*, *P2RY6*, *SOX2*, *NRG2*, *ABRA1*, *CDK13*, *NARS2* (Annexe 5). En particulier, l'une de ces anomalies impliquait une délétion du chromosome 6, au niveau des gènes des histones, centrée notamment sur *HIST1H2BE* et *HIST1H4D*. Bien qu'ils ne fassent pas partie de notre panel de gènes d'intérêt, ce type de délétion a également été décrit dans les LAL [54]. Tous ces exemples montrent que les nouvelles technologies à haute résolution telles que la COG pourraient identifier de nouvelles voies moléculaires impliquées dans la leucémogénèse et donc ouvrir de nouveaux champs de recherche dans ce domaine.

5 Capacité de la COG pour la détection des CN-LOH

La COG n'est pas conçue pour la détection de CN-LOH car il ne s'agit pas d'une technologie basée sur la séquence et elle n'est pas capable de différencier deux allèles issus d'un même gène. Cependant, nous avons réussi à reproduire la méthode utilisée par Neveling *et al.* [48] pour détecter les CN-LOH avec l'algorithme d'assemblage du génome *de novo*. Cette méthode nous a permis de mettre en évidence une majorité de variants identifiés comme "homozygotes" dans les régions et les échantillons concernés. Nous avons pu prouver que ces résultats n'étaient pas reproduits en l'absence de CN-LOH en répétant cette analyse sur les échantillons qui en étaient dépourvus. Ces résultats suggèrent qu'il serait possible de détecter des anomalies type CN-LOH avec la technologie de COG, bien qu'ils doivent être reproduits sur une plus grande cohorte. Une automatisation de cette analyse par le développement d'un algorithme de recherche dédié serait alors nécessaire.

6 Interférences pré-analytiques et analytiques de la COG

Nous avons trouvé des signes évidents de contamination de l'échantillon 7 par l'échantillon 4 en COG : il y avait 17 variants de tailles et de coordonnées identiques (N=17) détectables dans les deux échantillons, toujours avec un nombre de molécules détectées beaucoup plus faible dans l'échantillon 7. Cette contamination a été confirmée par les données obtenues avec les techniques de référence. Les deux échantillons avaient été traités sur la même cellule de flux. Cet incident montre l'absolue nécessité d'introduire des moyens systématiques et/ou automatisés de détection de ce type d'événement afin d'envisager l'utilisation de la COG dans des procédures de diagnostic médical. Cela pourrait se faire via une comparaison entre les échantillons, par série ou par cellule de flux, ou l'utilisation d'un système d'identification de l'ADN spécifique au patient de type codes-barres génétiques, comme cela se fait habituellement dans le séquençage à haut débit par exemple. Il est à noter qu'il n'a pas été tenu compte de ces 17 variants surnuméraires dans l'échantillon 7 pour la comparaison en aveugle.

Pour deux échantillons, le taux de marqueurs manquants (NLV) a dépassé la valeur supérieure recommandée par le fournisseur. Cela s'est accompagné d'une densité moyenne de marqueurs plus faible, ce qui a entraîné un bruit de fond important avec l'algorithme CNV, sous la forme de nombreuses fausses petites délétions réparties dans tout le génome. L'algorithme SV ne semblait pas impacté par ce type de signaux. Ces signaux faux positifs ont été facilement éliminés par les opérateurs, sans incidence sur la détection des anomalies d'intérêt, mais cela pourrait théoriquement provoquer l'absence de détection de sous-clones existants. Il faudrait analyser une série d'échantillons frais pour estimer si les cas de NLV élevées sont moins fréquents sur ce type de matériel que sur les culots congelés en DMSO.

CONCLUSION

Pour conclure, nous estimons que la COG est une alternative innovante, rapide et rentable aux différentes techniques cytogénétiques actuellement utilisées au diagnostic pour la caractérisation génétique des LAL. Nous avons toutefois identifié des obstacles qui doivent encore être levés, tels que les filtres de l'algorithme CNV, qui génèrent parfois des faux négatifs pour la détection de grandes anomalies sous-clonales ou des signaux de type bruit de fond, ainsi que l'absence d'aides à la détection des contaminations inter-échantillons. Par ailleurs, la COG pourrait aider à identifier des événements génétiques qui ne sont pas observés avec les techniques actuelles, mais qui sont potentiellement directement impliqués dans le processus leucémique et donc potentiellement pertinents en matière de stratification pronostique, ainsi qu'en tant que cibles thérapeutiques potentielles. Avant son utilisation comme outil diagnostic de routine pour les hémopathies malignes, des séries multicentriques plus importantes sont nécessaires pour confirmer et compléter ces données, en particulier pour identifier les cas dans lesquels des outils complémentaires seraient encore nécessaires tels que les gains de chromosomes entiers sous-clonaux et les anomalies impliquant les télomères, les centromères et les PAR.

Cette étude pousse également à envisager une évaluation de ce nouvel outil pour la caractérisation d'autres hémopathies malignes. Nous avons démontré que la COG pouvait rendre accessible la compréhension des mécanismes cytogénétiques complexes intervenant dans les LAL, qui font partie des hémopathies malignes les plus difficiles à caractériser en routine clinique. Dans le domaine des syndromes lymphoprolifératifs, certaines entités font également preuve d'une complexité parfois difficile à caractériser. C'est le cas des lymphomes B de haut grade, tels que les lymphomes B diffus à grandes cellules, dont la caractérisation génétique reste encore aujourd'hui parfois complexe. La COG pourrait ainsi représenter un outil prometteur pour la caractérisation cytogénétique de cette catégorie de pathologies.

REFERENCES BIBLIOGRAPHIQUES

1. Le Guyader-Peyrou S, Monnereau A. Estimations nationales de l'incidence et de la mortalité par cancer en France métropolitaine entre 1990 et 2018. Volume 2 – Hémopathies malignes. Étude à partir des registres des cancers du réseau Francim.
2. Simone JV. History of the treatment of childhood ALL: a paradigm for cancer cure. *Best Pract Res Clin Haematol.* 2006;19(2):353-9.
3. Freireich EJ. The history of leukemia therapy--a personal journey. *Clin Lymphoma Myeloma Leuk.* déc 2012;12(6):386-92.
4. Hunger SP, Mullighan CG. Acute Lymphoblastic Leukemia in Children. Longo DL, éditeur. *N Engl J Med.* 15 oct 2015;373(16):1541-52.
5. Terwilliger T, Abdul-Hay M. Acute lymphoblastic leukemia: a comprehensive review and 2017 update. *Blood Cancer J.* 30 2017;7(6):e577.
6. Meyer C, Burmeister T, Gröger D, Tsaour G, Fechina L, Renneville A, et al. The MLL recombinome of acute leukemias in 2017. *Leukemia.* févr 2018;32(2):273-84.
7. Smith M, Arthur D, Camitta B, Carroll AJ, Crist W, Gaynon P, et al. Uniform approach to risk classification and treatment assignment for children with acute lymphoblastic leukemia. *J Clin Oncol.* janv 1996;14(1):18-24.
8. Vora A, Goulden N, Mitchell C, Hancock J, Hough R, Rowntree C, et al. Augmented post-remission therapy for a minimal residual disease-defined high-risk subgroup of children and young people with clinical standard-risk and intermediate-risk acute lymphoblastic leukaemia (UKALL 2003): a randomised controlled trial. *Lancet Oncol.* juill 2014;15(8):809-18.
9. Roberts KG. Genetics and prognosis of ALL in children vs adults. *Hematology.* 30 nov 2018;2018(1):137-45.
10. Moorman AV. New and emerging prognostic and predictive genetic biomarkers in B-cell precursor acute lymphoblastic leukemia. *Haematologica.* avr 2016;101(4):407-16.
11. Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, et al. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. *Blood.* 19 mai 2016;127(20):2391-405.
12. Jaso J, Thomas DA, Cunningham K, Jorgensen JL, Kantarjian HM, Medeiros LJ, et al. Prognostic Significance of Immunophenotypic and Karyotypic Features of Philadelphia positive B-Lymphoblastic Leukemia in the Era of Tyrosine Kinase Inhibitors. *Cancer.* 1 sept 2011;117(17):4009-17.

13. Moorman AV, Enshaei A, Schwab C, Wade R, Chilton L, Elliott A, et al. A novel integrated cytogenetic and genomic classification refines risk stratification in pediatric acute lymphoblastic leukemia. *Blood*. 28 août 2014;124(9):1434-44.
14. Russell LJ, Enshaei A, Jones L, Erhorn A, Masic D, Bentley H, et al. *IGH@* Translocations Are Prevalent in Teenagers and Young Adults With Acute Lymphoblastic Leukemia and Are Associated With a Poor Outcome. *JCO*. 10 mai 2014;32(14):1453-62.
15. Fischer U, Forster M, Rinaldi A, Risch T, Sungalee S, Warnatz H-J, et al. Genomics and drug profiling of fatal TCF3-HLF-positive acute lymphoblastic leukemia identifies recurrent mutation patterns and therapeutic options. *Nat Genet*. sept 2015;47(9):1020-9.
16. Heerema NA, Carroll AJ, Devidas M, Loh ML, Borowitz MJ, Gastier-Foster JM, et al. Intrachromosomal Amplification of Chromosome 21 Is Associated With Inferior Outcomes in Children With Acute Lymphoblastic Leukemia Treated in Contemporary Standard-Risk Children's Oncology Group Studies: A Report From the Children's Oncology Group. *JOURNAL OF CLINICAL ONCOLOGY*. :6.
17. Dastugue N, Suciú S, Plat G, Speleman F, Cavé H, Girard S, et al. Hyperdiploidy with 58-66 chromosomes in childhood B-acute lymphoblastic leukemia is highly curable: 58951 CLG-EORTC results. *Blood*. 28 mars 2013;121(13):2415-23.
18. Nachman JB, Heerema NA, Sather H, Camitta B, Forestier E, Harrison CJ, et al. Outcome of treatment in children with hypodiploid acute lymphoblastic leukemia. 2007;110(4):4.
19. Mullighan CG. The genomic landscape of acute lymphoblastic leukemia in children and young adults. *Hematology*. 5 déc 2014;2014(1):174-80.
20. Mullighan CG, Su X, Zhang J, Radtke I, Phillips LAA, Miller CB, et al. Deletion of IKZF1 and Prognosis in Acute Lymphoblastic Leukemia. *N Engl J Med*. 29 janv 2009;360(5):470-80.
21. Clappier E, Auclerc M-F, Rapon J, Bakkus M, Caye A, Khemiri A, et al. An intragenic ERG deletion is a marker of an oncogenic subtype of B-cell precursor acute lymphoblastic leukemia with a favorable outcome despite frequent IKZF1 deletions. *Leukemia*. janv 2014;28(1):70-7.
22. Enshaei A, Schwab CJ, Konn ZJ, Mitchell CD, Kinsey SE, Wade R, et al. Long-term follow-up of ETV6-RUNX1 ALL reveals that NCI risk, rather than secondary genetic abnormalities, is the key risk factor. *Leukemia*. nov 2013;27(11):2256-9.
23. Moorman AV, Harrison CJ, Buck GAN, Richards SM, Secker-Walker LM, Martineau M, et al. Karyotype is an independent prognostic factor in adult acute lymphoblastic leukemia (ALL): analysis of cytogenetic data from patients treated on the Medical Research Council (MRC) UKALLXII/Eastern Cooperative Oncology Group (ECOG) 2993 trial. *Blood*. 15 avr 2007;109(8):3189-97.

24. Van Vlierberghe P, Ferrando A. The molecular basis of T cell acute lymphoblastic leukemia. *J Clin Invest*. 1 oct 2012;122(10):3398-406.
25. Graux C, Cools J, Michaux L, Vandenberghe P, Hagemeijer A. Cytogenetics and molecular genetics of T-cell acute lymphoblastic leukemia: from thymocyte to lymphoblast. *Leukemia*. sept 2006;20(9):1496-510.
26. Van Vlierberghe P, van Grotel M, Beverloo HB, Lee C, Helgason T, Buijs-Gladdines J, et al. The cryptic chromosomal deletion del(11)(p12p13) as a new activation mechanism of LMO2 in pediatric T-cell acute lymphoblastic leukemia. *Blood*. 15 nov 2006;108(10):3520-9.
27. Clappier E, Cuccuini W, Kalota A, Crinquette A, Cayuela J-M, Dik WA, et al. The C-MYB locus is involved in chromosomal translocation and genomic duplications in human T-cell acute leukemia (T-ALL), the translocation defining a new T-ALL subtype in very young children. *Blood*. 15 août 2007;110(4):1251-61.
28. La Starza R, Borga C, Barba G, Pierini V, Schwab C, Matteucci C, et al. Genetic profile of T-cell acute lymphoblastic leukemias with MYC translocations. *Blood*. 4 déc 2014;124(24):3577-82.
29. Clappier E, Collette S, Gardel N, Girard S, Suarez L, Brunie G, et al. NOTCH1 and FBXW7 mutations have a favorable impact on early response to treatment, but not on outcome, in children with T-cell acute lymphoblastic leukemia (T-ALL) treated on EORTC trials 58881 and 58951. *Leukemia*. déc 2010;24(12):2023-31.
30. Cauwelier B, Cave H, Gervais C, Lessard M, Barin C, Perot C, et al. Clinical, cytogenetic and molecular characteristics of 14 T-ALL patients carrying the TCRb-HOXA rearrangement: a study of the Groupe Francophone de Cytogéne'tique He'matologique. :8.
31. Hagemeijer A, Graux C. *ABL1* rearrangements in T-Cell acute lymphoblastic leukemia. *Genes Chromosom Cancer*. 2010;NA-NA.
32. Van Vlierberghe P, van Grotel M, Stubbs A, Cools J, Nagata K, Fornerod M, et al. The recurrent SET-NUP214 fusion as a new HOXA activation mechanism in pediatric T-cell acute lymphoblastic leukemia. 2008;111(9):13.
33. Ben Abdelali R, Roggy A, Leguay T, Cieslak A, Renneville A, Touzart A, et al. SET-NUP214 is a recurrent gd lineage-specific fusion transcript associated with corticosteroid/chemotherapy resistance in adult T-ALL. 2014;123(12):4.
34. Moreau P, San Miguel J, Sonneveld P, Mateos MV, Zamagni E, Avet-Loiseau H, et al. Multiple myeloma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up†. *Annals of Oncology*. 1 juill 2017;28(suppl_4):iv52-61.

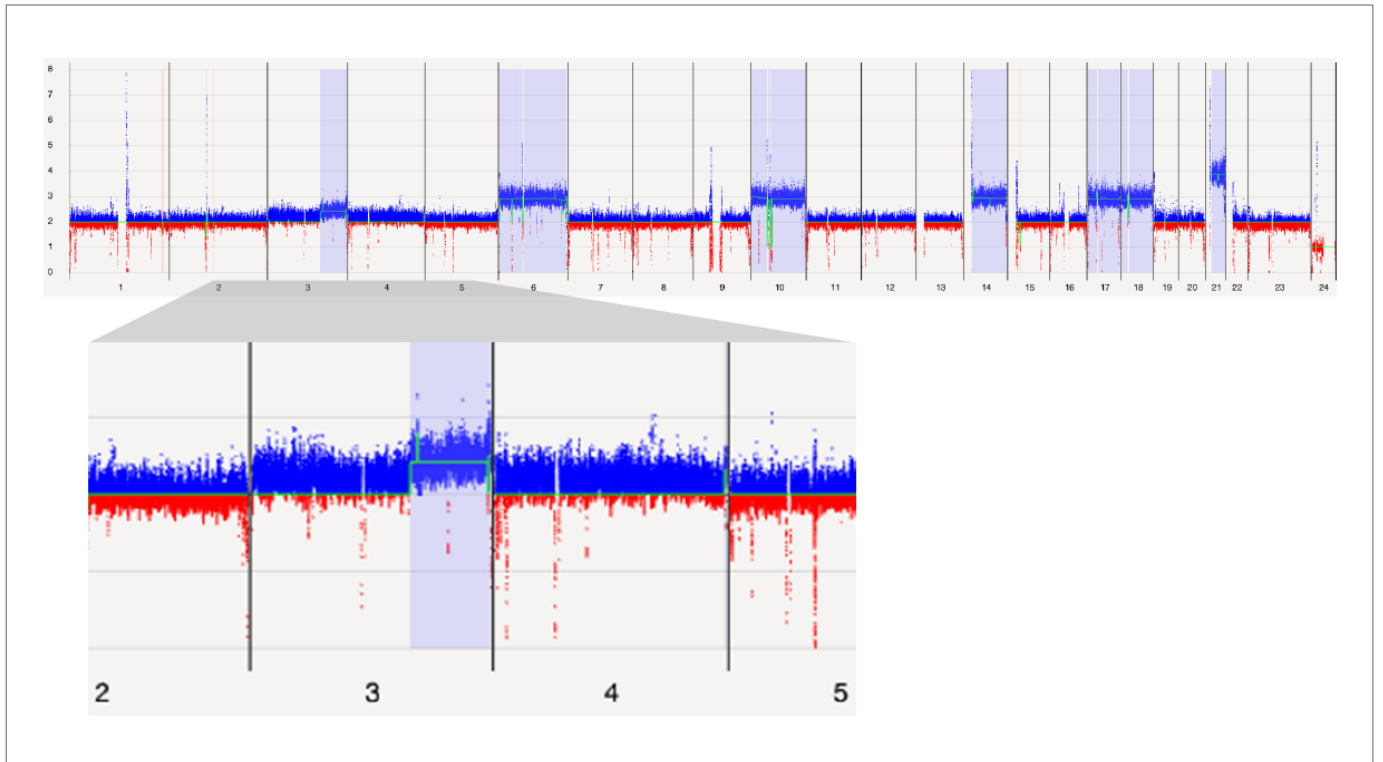
35. Rack KA, van den Berg E, Haferlach C, Beverloo HB, Costa D, Espinet B, et al. European recommendations and quality assurance for cytogenomic analysis of haematological neoplasms. *Leukemia*. août 2019;33(8):1851-67.
36. Baranger L, Cuccuini W, Lefebvre C, Luquet I, Perot C, Radford I, et al. Cytogenetics in the management of children and adult acute lymphoblastic leukemia (ALL): an update by the Groupe francophone de cytogénétique hématologique (GFCH). *Annales de biologie clinique*. sept 2016;74(5):547-60.
37. Schwartz D, Li X, Hernandez L, Ramnarain S, Huff E, Wang Y. Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping. *Science*. 1 oct 1993;262(5130):110-4.
38. Lam ET, Hastie A, Lin C, Ehrlich D, Das SK, Austin MD, et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat Biotechnol* [Internet]. août 2012 [cité 1 sept 2020];30(8). Disponible sur: <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3817024/>
39. Mak ACY, Lai YYY, Lam ET, Kwok T-P, Leung AKY, Poon A, et al. Genome-Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays. *Genetics*. janv 2016;202(1):351-62.
40. Pendleton M, Sebra R, Pang AWC, Ummat A, Franzen O, Rausch T, et al. Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat Methods*. août 2015;12(8):780-6.
41. Chaisson MJP, Sanders AD, Zhao X, Malhotra A, Porubsky D, Rausch T, et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun*. déc 2019;10(1):1784.
42. Mantere T, Neveling K, Pebrel-Richard C, Benoist M, van der Zande G, Kater-Baats E, et al. Next generation cytogenetics: genome-imaging enables comprehensive structural variant detection for 100 constitutional chromosomal aberrations in 85 samples [Internet]. *Genetics*; 2020 juill [cité 28 oct 2020]. Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/2020.07.15.205245>
43. Barseghyan H, Tang W, Wang RT, Almalvez M, Segura E, Bramble MS, et al. Next-generation mapping: a novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis. *Genome Med*. déc 2017;9(1):90.
44. Dixon JR, Xu J, Dileep V, Zhan Y, Song F, Le VT, et al. Integrative detection and analysis of structural variation in cancer genomes. *Nat Genet*. oct 2018;50(10):1388-98.
45. Luebeck J, Coruh C, Dehkordi SR, Lange JT, Turner KM, Deshpande V, et al. AmpliconReconstructor integrates NGS and optical mapping to resolve the complex structures of focal amplifications. *Nat Commun*. déc 2020;11(1):4374.

46. Chan EKF, Cameron DL, Petersen DC, Lyons RJ, Baldi BF, Papenfuss AT, et al. Optical mapping reveals a higher level of genomic architecture of chained fusions in cancer. *Genome Res.* mai 2018;28(5):726-38.
47. Xu J, Song F, Schleicher E, Pool C, Bann D, Hennessy M, et al. An Integrated Framework for Genome Analysis Reveals Numerous Previously Unrecognizable Structural Variants in Leukemia Patients' Samples [Internet]. *Genomics*; 2019 févr [cité 25 mai 2020]. Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/563270>
48. Neveling K, Mantere T, Vermeulen S, Oorsprong M, van Beek R, Kater-Baats E, et al. Next generation cytogenetics: comprehensive assessment of 48 leukemia genomes by genome imaging [Internet]. *Genomics*; 2020 févr [cité 25 mai 2020]. Disponible sur: <http://biorxiv.org/lookup/doi/10.1101/2020.02.06.935742>
49. Levy B, Baughn LB, Chartrand S, LaBarge B, Claxton D, Lennon A, et al. A National Multicenter Evaluation of the Clinical Utility of Optical Genome Mapping for Assessment of Genomic Aberrations in Acute Myeloid Leukemia. *medRxiv.* 1 janv 2020;2020.11.07.20227728.
50. Meijerink JPP. Genetic rearrangements in relation to immunophenotype and outcome in T-cell acute lymphoblastic leukaemia. *Best Practice.* 2010;12.
51. Van Vlierberghe P, Pieters R, Beverloo HB, Meijerink JPP. Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. *British Journal of Haematology.* oct 2008;143(2):153-68.
52. Poole CJ, Zheng W, Lodh A, Yevtodiyenko A, Liefwalker D, Li H, et al. DNMT3B overexpression contributes to aberrant DNA methylation and MYC-driven tumor maintenance in T-ALL and Burkitt's lymphoma. *Oncotarget.* 29 sept 2017;8(44):76898-920.
53. Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research.* 8 janv 2019;47(D1):D941-7.
54. Loudin MG, Wang J, Leung H-CE, Gurusiddappa S, Meyer J, Condos G, et al. Genomic profiling in Down syndrome acute lymphoblastic leukemia identifies histone gene deletions associated with altered methylation profiles. *Leukemia.* oct 2011;25(10):1555-63.

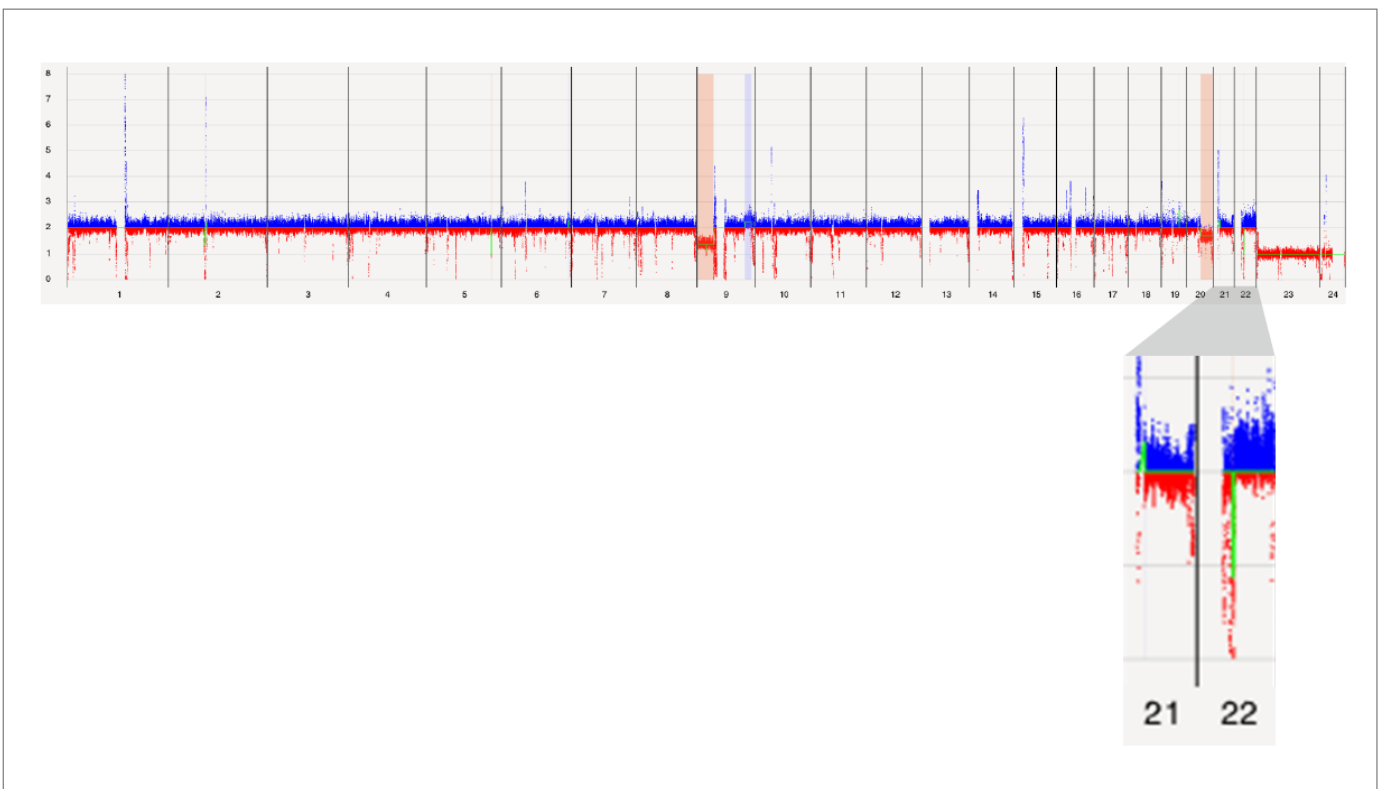
ANNEXE 1 : Liste des 186 gènes d'intérêts utilisés pour la sélection des anomalies

Gène	Locus	Gène	Locus	Gène	Locus	Gene	Locus
<i>SFPQ</i>	1p34	<i>DEK</i>	6p22	<i>LMO1</i>	11p15	<i>CREBBP</i>	16p13
<i>STIL</i>	1p33	<i>CASP8AP2</i>	6q15	<i>NUP98</i>	11p15	<i>MYH11</i>	16p13
<i>TAL1</i>	1p33	<i>MYB</i>	6q23	<i>WT1</i>	11p13	<i>GLIS2</i>	16p13
<i>EPS15</i>	1p32	<i>AHI1</i>	6q23	<i>RAG1</i>	11p13	<i>NDE1</i>	16p13
<i>RBM15</i>	1p13	<i>AFDN</i>	6q27	<i>LMO2</i>	11p13	<i>CBFB</i>	16q22
<i>TPM3</i>	1q21	<i>CEP43</i>	6q27	<i>CD44</i>	11p13	<i>CBFA2T3</i>	16q24
<i>ARNT</i>	1q21	<i>HOXA13</i>	7p15	<i>SLC1A2</i>	11p13	<i>TP53</i>	17p13
<i>PDE4DIP</i>	1q21	<i>HOXA9</i>	7p15	<i>RAG2</i>	11p12	<i>NF1</i>	17q11
<i>MLLT11</i>	1q21	<i>IKZF1</i>	7p12	<i>ARHGEF17</i>	11q13	<i>SUZ12</i>	17q11
<i>PBX1</i>	1q23	<i>CARD11</i>	7p22	<i>C2CD3</i>	11q13	<i>MLLT6</i>	17q12
<i>RCSD1</i>	1q24	<i>PIK3CG</i>	7q22	<i>PICALM</i>	11q14	<i>IKZF3</i>	17q21
<i>ABL2</i>	1q25	<i>TRIM24</i>	7q33-4	<i>MAML2</i>	11q21	<i>RARA</i>	17q21
<i>IGK</i>	2p11	<i>TRB</i>	7q34	<i>DDX10</i>	11q22	<i>STAT5B</i>	17q21
<i>AFF3</i>	2q11	<i>ZC3HAV1</i>	7q34	<i>ARHGEF12</i>	11q23	<i>SPAG9</i>	17q21
<i>RANBP2</i>	2q12	<i>EZH2</i>	7q36	<i>KMT2A</i>	11q23	<i>HLF</i>	17q22
<i>ZEB2</i>	2q22	<i>MNX1</i>	7q36	<i>ZBTB16</i>	11q23	<i>SEPTIN9</i>	17q25
<i>IKZF2</i>	2q34	<i>PCM1</i>	8p22	<i>CBL</i>	11q23	<i>PTPN2</i>	18p11
<i>INPP5D</i>	2q37	<i>PTK2B</i>	8p21	<i>DCPS</i>	11q24	<i>TCF3</i>	19p13
<i>SETD2</i>	3p21	<i>FGFR1</i>	8p11	<i>ETV6</i>	12p13	<i>EPOR</i>	19p13
<i>FHIT</i>	3p14	<i>TOX</i>	8q12	<i>ATF7IP</i>	12p13	<i>TYK2</i>	19p13
<i>PTPRG</i>	3p14	<i>PAG1</i>	8q21	<i>KDM5A</i>	12p13	<i>LYL1</i>	19p13
<i>FOXP1</i>	3p13	<i>RUNX1T1</i>	8q21	<i>BCAT1</i>	12p12	<i>MLLT1</i>	19p13
<i>BTLA</i>	3q13	<i>RAD21</i>	8q24	<i>RASSF3</i>	12q14	<i>ELL</i>	19p13
<i>CD200</i>	3q13	<i>MYC</i>	8q24	<i>BTG1</i>	12q21	<i>SLX4IP</i>	20p12
<i>CIP2A</i>	3q13	<i>JAK2</i>	9p24	<i>SH2B3</i>	12q23	<i>MAPRE1</i>	20q11
<i>MLF1</i>	3q25	<i>CDKN2A</i>	9p21	<i>PAN3</i>	13q12	<i>TOP1</i>	20q12
<i>TBL1XR1</i>	3q26	<i>CDKN2B</i>	9p21	<i>ZMYM2</i>	13q12	<i>ERG</i>	21q22
<i>PDGFRA</i>	4q12	<i>MLLT3</i>	9p21	<i>ELF1</i>	13q14	<i>RUNX1</i>	21q22
<i>CHIC2</i>	4q12	<i>MTAP</i>	9p21	<i>RB1</i>	13q14	<i>VPREB1</i>	22q11
<i>FIP1L1</i>	4q12	<i>PAX5</i>	9p13	<i>DLEU1</i>	13q14	<i>BCR</i>	22q11
<i>AFF1</i>	4q22	<i>CNTRL</i>	9q33	<i>SERP2</i>	13q14	<i>IGL</i>	22q11
<i>RAP1GDS1</i>	4q23	<i>ABL1</i>	9q34	<i>DGKH</i>	13q14	<i>SEPTIN5</i>	22q11
<i>LEF1</i>	4q25	<i>NUP214</i>	9q34	<i>DLEU2</i>	13q14	<i>IL2RB</i>	22q12
<i>NR3C2</i>	4q31	<i>NOTCH1</i>	9q34	<i>TRD</i>	14q11	<i>MNI</i>	22q12
<i>FBXW7</i>	4q31	<i>SET</i>	9q34	<i>TRA</i>	14q11	<i>MRTFA</i>	22q13
<i>MEF2C</i>	5q14	<i>MLLT10</i>	10p12	<i>STRN3</i>	14q12	<i>SHOX</i>	Xp22
<i>SSBP2</i>	5q14	<i>CCDC6</i>	10q21	<i>BCL11B</i>	14q32	<i>CRLF2</i>	Xp22
<i>TSLP</i>	5q22	<i>TET1</i>	10q21	<i>IGH</i>	14q32	<i>CSF2RA</i>	Xp22
<i>SNX2</i>	5q23	<i>ZMIZ1</i>	10q22	<i>EML1</i>	14q32	<i>IL3RA</i>	Xp22
<i>NR3C1</i>	5q31	<i>PTEN</i>	10q23	<i>ATP10A</i>	15q11	<i>P2RY8</i>	Xp22
<i>AFF4</i>	5q31	<i>PIK3API</i>	10q24	<i>SPRED1</i>	15q14	<i>DMD</i>	Xp21
<i>ARHGAP26</i>	5q31	<i>TLX1</i>	10q24	<i>MGA</i>	15q15	<i>KDM6A</i>	Xp11
<i>CSF1R</i>	5q32	<i>ADD3</i>	10q25	<i>KNL1</i>	15q15	<i>GATA1</i>	Xp11
<i>PDGFRB</i>	5q32			<i>PML</i>	15q24	<i>ATRX</i>	Xq21
<i>EBF1</i>	5q33			<i>NTRK3</i>	15q25	<i>SEPTIN6</i>	Xq24
<i>TLX3</i>	5q35					<i>BCORL1</i>	Xq25
<i>NPM1</i>	5q35					<i>SH2D1A</i>	Xq25
<i>NSD1</i>	5q35					<i>STAG2</i>	Xq25
						<i>PHF6</i>	Xq26
						<i>ARHGEF6</i>	Xq26

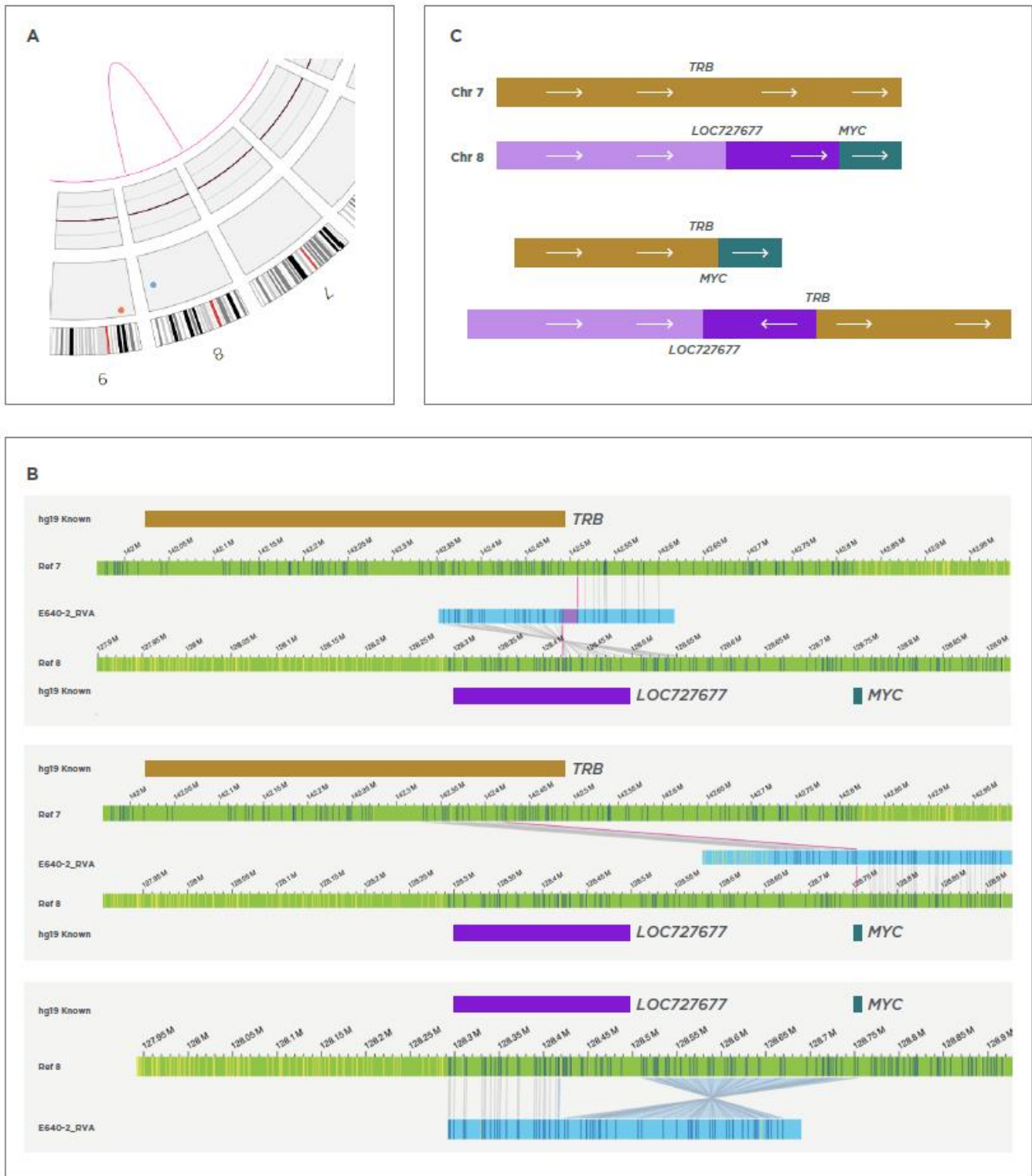
ANNEXE 2 : Analyse de l'échantillon 3 en cartographie optique par l'algorithme CNV



ANNEXE 3 : Analyse de l'échantillon 4 en cartographie optique par l'algorithme CNV



ANNEXE 4 : Caractérisation par cartographie optique du génome d'un réarrangement des chromosomes 7 et 8 impliquant les gènes *MYC* et *TRB*



- A.** Circos plot de la COG représentant le réarrangement entre les chromosomes 7 et 8 (ligne rose).
- B.** Cartes de consensus représentant les points de cassure au niveau des chromosomes 7 et 8 et les gènes impliqués (logiciel Bionano Access). Les cartes vertes représentent les cartes de référence des chromosomes 10,11 et 14. Les cartes bleues sont le résultat de

la COG. Elles représentent les anomalies retrouvées pour le patient et les différents dérivés du réarrangement. Les deux premiers cadres représentent les deux dérivés de du réarrangement 7-8 *MYC-TRB*. Le troisième cadre représente l'inversion au niveau du chromosome 8.

- C. Schématisation des chromosomes 7 et 8 non remaniés au niveau des régions chromosomiques impliquées (en haut). Ce remaniement est à l'origine d'un réarrangement *MYC-TRB* avec inversion d'une partie du chromosome 8 (en bas).

ANNEXE 5 : Anomalies détectées en cartographie optique, ne répondant pas aux critères d'inclusion mais impliquant des gènes de la base de données COSMIC

Echantillon	Anomalie	Gène(s) COSMIC impliqué(s)	Taille (kpb)
1	Délétion 1p36.12	<i>CELA3B</i>	23,178
2	Délétion 20q13.33	<i>CDH4</i>	293,101
3	Gain 5q14.1	<i>OTP</i>	32,364
	Délétion 6p22.2	<i>HIST1H2BE & HIST1H4D</i>	100,381
4	Délétion 12q24.31	<i>KDM2B</i>	32,286
	Gain 19q31.31	<i>PSG11, PSG2, PSG5, PSG4</i>	277,171
5	Gain 6q22.3	<i>SOX4</i>	93,214
	Gain 11q13.4	<i>P2RY2 et P2RY6</i>	172,411
6	Délétion 3q26.33	<i>SOX2</i>	421,090
9	Délétion 5q31.2	<i>NRG2</i>	25,380
	Délétion 6q24.1	<i>ABRA1</i>	21,298
	Délétion 7p14.1	<i>CDK13</i>	167,562
10	Délétion 11q14.1	<i>NARS2</i>	31,322

ANNEXE 6 : Manuscrit de l'article de thèse, en préparation pour soumission

Optical genome mapping, a promising alternative to gold standard cytogenetic approaches in a series of acute lymphoblastic leukemias

Valentin Lestringant, Nicolas Duployez, Dominique Penther, Isabelle Luquet, Nathalie Gardel, Anne Lutun, Claude Preudhomme, Michaela West, Joyce Lee, Catherine Devoldere, Jean-Pierre Marolleau, Loïc Garçon, Guillaume Jedraszak and Yann Ferret.

Abstract

Acute lymphoblastic leukemias (ALL) are characterized by a large number of genetic abnormalities of prognostic and therapeutic interest. Their detection requires the use of several conventional and molecular cytogenetic techniques in order to establish the most exhaustive genetic profile: conventional karyotype, FISH, SNP-Array and RT-MLPA for the detection of fusion transcripts. Optical genome mapping technique is based on analysis of ultra-high molecular weight that provides a high-resolution genome-wide analysis highlighting structural anomalies, including balanced translocations, and copy number anomalies. We compared optical mapping to standard techniques in 10 selected B or T-ALL patients carrying at least one major genetic abnormality. Abnormalities on SNP-Arrays were retained when larger than 1Mbp or > 20 kbp and involving one of the 186 genes defined as potentially relevant in leukemogenesis. A total of 80 abnormalities meeting these criteria in SNP-array and/or found by other standard techniques were identified (45 deletions, 15 whole chromosome gains, 12 duplications or partial gains, and 8 translocations). Eight discrepancies between optical mapping and standard techniques were identified, 2 of them were copy number variations in Xp22.33 (pseudoautosomal region), a non-covered area with optical mapping technology. Three other were low allelic frequency large gains, which were clearly visible in the raw data but rejected by software's CNV filters. Another sub-clonal gain (trisomy 21), found by karyotype and FISH (quantified at 7% of all nuclei) but not by SNP-array, was missed, even in raw data. The two remaining discrepancies were a deletion of the *VPREB1* gene, erroneously eliminated by the operators because included in the recurrent deletion zones of the *IGL* locus, and a tetraploid subclone, found only on karyotype, and by definition undetectable by quantitative techniques. Interestingly, new anomalies well described in ALL but not detected by standard techniques, were revealed using optical mapping, including a complex rearrangement of chromosomes 11 and involving *LMO2* and *TRA*, a rearrangement with inversion of chromosomes 7 and 8, involving *TRB* and *MYC* and a rearrangement of chromosomes 14 and 20, involving *IGH* and *CEBPB*. Thus, optical mapping represents an innovative, time and cost effective alternative to the different cytogenetic techniques currently performed routinely for ALL genetic characterization. It enables identification of complex cytogenetic events, including those currently inaccessible to standard techniques such as chromotripsis or certain complex translocations.

Introduction

Acute lymphoblastic leukemias (ALL) are neoplasms resulting from an early blockage of cells committed to B or T-lymphoid lineage. Although rare (1.7 cases/year/100,000 inhabitants) in the general population, they are the most common cancer in children, with an incidence approximately 3 times higher than in adults.¹ ALL are pathologies of great genetic diversity²⁻⁴ which routinely require the use of several conventional and molecular cytogenetic techniques to establish the most exhaustive genetic profile required for accurate prognosis and therapeutic stratification.^{5,6} This genetic profiling generally consists in combination of conventional karyotyping, Fluorescent *in situ* Hybridization (FISH), Comparative Genomic Hybridization (CGH) or Single Nucleotide Polymorphism (SNP) arrays and a gene expression profiling for example by Reverse Transcription Multiplex Ligation-dependent Probe Amplification (RT-MLPA) or transcriptome. All these techniques are complementary. Karyotyping is a whole-genomic analysis that allows the detection of balanced and unbalanced structural anomalies, but its resolution is limited (of the order of 10 to 20 Mbp) and its success depends on the cell culture phase, as well as on the quality of the metaphases obtained. Especially in ALL, the low proliferative index of blasts is the cause of frequent karyotype failure. FISH allows the detection of many different anomalies with a resolution of around 200 kbp depending on the probes used, but it is a targeted and limited in throughput analysis, which often requires, particularly in ALL, the use of several different probes. The CGH and the SNP-arrays are also genome-wide analyses with a high resolution; SNP allows the detection of copy neutral loss of heterozygosity (CN-LOH). However these two techniques do not allow the detection of balanced anomalies. Finally, RT-MLPA allows the detection of a large number of balanced rearrangements leading to a fusion transcript, but again this is a targeted analysis, which does not allow the detection of balanced

rearrangements involving other oncogenetic mechanisms. The multiplication of techniques, but also their complexity and cost underline the interest in developing a single test that could replace all those currently used in cytogenetics. High-speed sequencing technologies are on the way to offering such a solution, but the detection of balanced or large structural anomalies is still too costly, with data analysis remaining very complex, preventing their use in routine.

Optical genome mapping (OM) is a technology for high-resolution reconstruction of the genome from single enzyme-labelled DNA molecules. Its aim is to map binding sites of an enzyme on DNA, resulting in the specific marking of each region of the genome ("barcode") and allowing it to be uniquely identified. This technique, initially described by Schwartz *et al.*⁷ in the 1990s, allows today high-resolution genome-wide analysis, enabling the detection of both number and structural anomalies, from the simplest to the most complex, including balanced rearrangements.^{8,9} Used as a complement to sequencing techniques, it enables to guide the assembly.^{10,11} OM technology has shown promising initial results for the detection of structural variants and characterization of constitutional pathologies.^{12,13} This technology is also highlighted by its ability to resolve complex anomalies found in solid tumors.¹⁴⁻¹⁶ Finally, OM seems to be a promising technique for cytogenetics in hemopathies.^{17,18} For example, in AML, OM has recently proven its equivalence with standard cytogenetic techniques in a series of 100 cases.¹⁹ In this study, we aimed to compare the technical performances of OM with standard procedures in a series of 10 previously genetically typed B or T-ALL by combination of karyotype, FISH, SNP-array and RT-MLPA, in order to create a representative panel of major prognostic impact anomalies and their large structural diversity. For OM, we used a Saphyr® System-Bionano and tested whether the algorithms and software developed by the manufacturer to process and filter the data were reliable and user-friendly enough to isolate anomalies of clinical interest from other signals and guarantee a safe result that may be compatible with routine hospital use. The detected variants were sorted, analyzed and interpreted by two blind and independent operators, one with training in hemopathy cytogenetics but without experience in neither bioinformatics nor optical mapping and the other trained in the use of the OM tool but without knowledge in the genetics of hemopathies to get an estimate of the impact of the operator's profession. We report here the results of our study that aimed to respond to the following objectives: (i) defining the correlation between OM and standard techniques as well as between the two operators in terms of ALL genetic anomalies (ii) deciphering the discrepancies that may have occurred, (iii) identifying potential genetic anomalies not noticed using standard tools but identified only using OM and (iv) evaluating the capacity of OM to detect CN-LOH which, although in theory not identified by this technology, may be seen using indirect methods as recently reported by Neveling¹⁸ *et al.*

Methods

Sample selection

Ten samples of patients with B or T-ALL, treated at the CHU Amiens-Picardie were selected. These were patients for whom a karyotype (supplemented if necessary by one or more FISH), a SNP-array and a search for fusion transcripts by RT-MLPA had been performed at diagnosis. Bone marrow samples were collected at diagnosis between May 2016 and September 2019. Among the samples that met these conditions, those with major and diverse abnormalities of interest were selected, including t(10;14) *TLX1-TRA*, t(9;22) *BCR-ABL1* (e1a2 and e14a2 transcripts), t(10;11) *MLLT10-PICALM*, t(11;19) *KMT2A-MLLT1*, typical hyperdiploid profile, intrachromosomal amplification of *RUNX1* (iAMP21), *SET-NUP214* fusion, various deletions involving *CDKN2A/CDKN2B*, *IKZF1*, *PAX5*, *PTEN*, *ERG* and *FBXW7*.

DNA extraction, labelling and data collection for optical mapping

The optical maps were performed on the GENTYANE INRAE Clermont Platform, using cell pellets preserved in DMSO. For each sample, a minimum of 1.5 million cells was used to purify ultra-high molecular weight (UHMW) DNA according to manufacturer instructions (Bionano Prep SP Cells in Freezing Medium, Bionano Prep SP Frozen Cell Pellet DNA Isolation Protocol Bionano Genomics, San Diego USA). Briefly, after counting, white blood cells were pelleted (2200g for 2mn) and treated with LBB lysis buffer and proteinase K to release genomic DNA (gDNA). After inactivation of proteinase K by PMSF treatment, genomic DNA was bound to a nanobind disk, washed and eluted in an appropriate buffer. UHMW DNA was left to homogenize at room temperature overnight. The next day, DNA molecules were labeled using the DLS (Direct Label and Stain) DNA Labeling Kit (Bionano genomics, San Diego USA). Seven hundred and fifty nanograms of gDNA were labelled in presence of Direct Label Enzyme (DLE-1) and DL-green fluorophores. Thus, fluorescent labels are attached by DLE-1 enzyme to a specific 6 bp sequence motif (CTTAAG), occurring about 15 times per 100 kbp of UHMW DNA. After clean-up of the excess of DL-Green fluorophores and rapid digestion of the remaining DLE-1 enzyme by proteinase K, DNA backbone was counterstained overnight before quantitation and visualization on a Saphyr instrument. A volume of 8.5µl of labelled gDNA solution of concentration between 4 and 12ng/µl was loaded on Saphyr chip and scanned on the Saphyr instrument (Bionano genomics, San Diego USA) where NanoChannels allow only a

single linearized DNA molecule to travel through by electrophoretic migration. Saphyr chip were ran at maximum capacity to reach a minimum yield of 1300 Gbp and a 300x genome coverage. A chip includes 3 flow cells and 2 chips can be loaded per instrument run. Migration lasts 2 days per run.

Rare variant analysis and structural variant calling

Rare Variant Pipeline (RVP) and variant calling were performed via Bionano Access software (v1.5.2). RVP processes data according to two distinct algorithms: one for structural variants (SV tool) and another for CNVs (CNV tool). The cutoff was set at 20kbp for insertions, deletions and duplications detected through the map alignments (SV tool), and 275 kbp for the CNV tool. Events detected by the Rare Variants Analysis Pipeline were subsequently filtered against an embedded normal samples database and only variants absent of that database were considered for analysis.

Data comparison

Anomalies were pre-filtered according to the custom filtering parameters established by Neveling *et al.* in a recent work¹⁸, except for translocations which were all studied (confidence scores: Insertion: 0, Deletion: 0, Inversion: 0.01, Duplication: -1, Translocation: all, Copy Number: 0.99). The data were then blindly analyzed on software and/or on downloaded .csv files by two independent operators, one medical biology student specializing in cytogenetics with no training in OM (operator 1) and the other OM engineer without knowledge in onco-hematological cytogenetics (operator 2). Operators were instructed to report any anomalies that met the following criteria: larger than 1Mbp or > 20 kbp and involving one of the 186 genes defined as potentially relevant in leukemogenesis (Supplemental table A). Concerning the coordinates overlap of the anomalies detected in both SNP-array and OM, 3 comparisons were made. The coordinates of all anomalies detected with SV tool were compared to their SNP-array equivalent. The same comparison was performed for the anomalies detected with CNV tool. Among these anomalies, for those detected by both the SV and CNV tools, the overlap of the coordinates obtained by these two algorithms was also compared. The chromosomal coordinates of the different anomalies were collected and the calculation of the overlap was performed on the Cartagenia platform (Agilent Technologies). For highlighting of CN-LOH, we used the bioinformatics pipeline Genome *De novo* assembly which classifies each SV as heterozygous or homozygous. All previously used filters were removed in order to reveal all the variants identifiable by OM. The number of anomalies considered homozygous by the pipeline was related to the total number of anomalies in the concerned region. These ratios were calculated in each region affected by a CN-LOH and then compared to the ratios obtained in the same regions on the other patients.

RQ-PCR, Karyotype, FISH, SNP Array and RT-MLPA are detailed in Supplemental Methods

Results

Patients' characteristics

The mean age at diagnosis was 14 years (range 2-37), including 7 children (2-8 years, mean 5.1 years) and 3 adults (32-37 years, mean 34.7 years). They were 9 men and 1 woman. Six patients were diagnosed with B-ALL (4 B-II and 2 B-III, according to the EGIL classification) and four with T-ALL (1 T-I and 3 T-III). Patients' characteristics are shown on Table 1.

OM molecules quality data

The UHMW DNA extraction technique provided molecules with an average N50 (≥ 150 kbp) of 286 kbp (258-334), an average map rate of 83.31 % (71.6-90.7), an average label density of 14.25 labels/100kbp (12.54-15.28), an average negative label variance (NLV) of 12.85 % (10.3-23.5) and an average coverage of 339X (289.04-379.65). Samples 6 and 8, showing an NLV >20%, had a very high number of calls with the CNV tool (214 and 201 respectively). A majority of these signals, interpreted as background noise, were eliminated during the analysis by rejecting the deletions detected by the CNV tool if they were neither confirmed by the SV tool nor characterized by a copy number <1.5.

Table 1 : Patients' characteristics

Patient	Age at diagnosis	Sex	EGIL Classification
1	6	M	T-III
2	8	F	T-III
3	4	M	B-II
4	5	M	B-II
5	35	M	T-I
6	37	M	T-III
7	5	M	B-II
8	6	M	B-III
9	32	M	B-II
10	2	M	B-II

OM identifies the majority of abnormalities of clinical interest.

A total of 80 abnormalities were found at least by SNP-array, for all patients. Two of these abnormalities were CN-LOH and were therefore considered separately since in theory not detectable using OM. Of the 78 remaining abnormalities, 70 met the criteria retained for inclusion. Ten additional abnormalities not detected by SNP-Array were detected by conventional karyotype, FISH or RT-MLPA. Thus a total of 80 anomalies meeting the selection criteria in SNP-array and/or found with the other standard techniques, were retained for the blind comparison. Among these 80 anomalies, 73 (91.25%) were found by operator 1 and 72 (90%) by operator 2 (Supplemental figure A). All (8/8) translocations, 43/45 (95.6%) of deletions, 10/12 (83.3%) of gains were found by both operators whereas 12/15 (80%) and 11/15 (73.3%) of whole chromosome gains were found by operator 1 and 2 respectively (Table 2).

Discrepancies between techniques

Eight anomalies were missed using OM (Table 3). Four of them were filtering issues. Three were low allelic frequency large gains clearly visible in raw data but rejected by software's CNV tool filters, totally, or partially for a trisomy 3 that caused the discrepancy between the two observers. The fourth was a deletion of *VPREB1* gene, erroneously eliminated by operators because included ("hidden") in recurrent deletion zones of *IGL* locus. The 4 other missed abnormalities were truly undetected anomalies. Two of them were copy number variations in the pseudo autosomal region (PAR) in Xp22.33, an area not covered with OM technology. One was a microdeletion causing a *P2RY8-CRLF2* fusion. Another sub-clonal gain (trisomy 21) was found in 7/22 metaphases by karyotype but neither by SNP-array nor OM even in raw data. We therefore confirmed the presence of this gain by interphasic FISH and quantified it in only 7% of the cores studied. It can be assumed that the trisomic clone had a high *in vitro* proliferation index. The detection limit of SNP-A and OM being an allelic fraction of 10%, these technologies could not highlight this clone. The last discrepancy was a near-tetraploid subclone (6/19 metaphases with 80~95 chromosomes), found only on karyotype, that might be strictly tetraploid and by definition undetectable by quantitative techniques.

Table 2 : Anomalies detected by standard techniques, correctly identified in optical mapping.

<p>1</p> <p>Deletion 9p21.3 involving <i>CDKN2A, CDKN2B, MLLT3, MTAP</i> Translocation t(10;14) <i>TLX1-TRD</i> Deletion 13q14.2 involving <i>RBI</i> Deletion 17p13.3-p13.1 involving <i>TP53</i> Gain 20p13-p11.21 involving <i>SLX4IP</i> Gain 20q11.21-q13.33</p>	<p>Deletion 2q34-q35 involving <i>IKZF2</i> Deletion 5q13.1-q13.2 without genes of interest Gain 6q23.3 involving <i>MYB, AHII</i> Deletion 9p21.3 involving <i>CDKN2A, CDKN2B, MTAP</i> Deletion 9q34.11-q34.13 leading to <i>SET-NUP214</i> fusion Deletion 10q24.31-q24.32 involving <i>TLXI</i></p>
<p>2</p> <p>Deletion 1p33 leading to <i>STIL-TAL</i> fusion Deletion 4q31.3 involving <i>FBXW7</i> Deletion 9p21.3 involving <i>CDKN2A, CDKN2B, MTAP</i> Deletion 10q23.31 involving <i>PTEN</i> Deletion 21q22.2 involving <i>ERG</i></p>	<p>6</p> <p>Deletion 10q25.1 involving <i>ADD3</i> Deletion 11p13 involving <i>WT1</i> Deletion 11q22.1-q22.2 without genes of interest Deletion 16p13.3 involving <i>CREBBP</i> Deletion 16q11.2-q22.1 without genes of interest Gain 19p13.2-p11 involving <i>EPOR, LYLI</i></p>
<p>3</p> <p>Deletion 1q42.2-q42.3 without genes of interest Deletion 1q44 without genes of interest Whole chromosome gain 3 Whole chromosome gain 6 Whole chromosome gain 10 Whole chromosome gain 14 Whole chromosome gain 17 Whole chromosome gain 18 Whole chromosome gain 21 Whole chromosome gain X</p>	<p>7</p> <p>Whole chromosome gain 5 Deletion 6q13-q22.33 involving <i>CASP8&AP2</i> Deletion 7p12.2 involving <i>IKZF1</i> Translocation t(7;21) <i>SERPINE1-DONSON</i> Whole chromosome gain 10 Translocation t(11;19) <i>KMT2A-MLLT1</i> Gain 20q11.21-q13.33 without genes of interest Gain 21q11.2-q22.3 involving <i>ERG, RUNXI</i></p>
<p>4</p> <p>Deletion 5q33.3 involving <i>EBF1</i> Translocation t(7;9) Deletion 7p12.2 involving <i>IKZF1</i> Deletion 9p24.3-p13.1 involving <i>PAX5, CDKN2A, CDKN2B, MLLT3, MTAP, JAK2</i> Gain 9q32-q34.3 involving <i>ABL1, NUP214, NOTCH1</i> Translocation t(9;22) <i>BCR-ABL1</i>, e1a2 transcript Deletion 11p12 involving <i>RAG2, RAG1</i> Deletion 13q14.11 involving <i>SERP2</i> Deletion 20p12.2 involving <i>SLX4IP</i> Deletion 20q11.22-q13.33 without genes of interest</p>	<p>8</p> <p>Deletion 13q14.2 (0,24Mb) involving <i>RBI</i> Deletion 13q14.2 (0,09Mb) involving <i>RBI</i> Deletion 13q14.2-q14.3 involving <i>DLEU1, DLEU2</i> Deletion 21q11.2-q21.1 without genes of interest iAMP21 : Gain 21q21.1-q22.3 involving <i>ERG, RUNXI</i> Deletion 21q22.3 without genes of interest Whole chromosome gain X</p>
<p>5</p> <p>Gain 6q23.3 involving <i>MYB, AHII</i> Translocation t(10;11) <i>PICALM-MLLT10</i> Whole chromosome gain 13 Deletion 16p13.3 involving <i>CREBBP</i> Deletion 16p13.3-p13.13 without genes of interest Gain 17q25.2-q25.3 involving <i>SEPT9</i></p>	<p>9</p> <p>Deletion 3q26.32 involving <i>TBLIXR1</i> Deletion 7p22.3-p11.2 involving <i>IKZF1, CARD11</i> Deletion 9p24.3-p21.3 involving <i>MLLT3, JAK2</i> Deletion 9p21.3 involving <i>CDKN2A, CDKN2B, MTAP</i> Deletion 9p21.3-p13.1 involving <i>PAX5, CDKN2B</i> Translocation t(9;22) <i>BCR-ABL1</i>, e14a2 transcript</p>
<p>5</p> <p>Gain 6q23.3 involving <i>MYB, AHII</i> Translocation t(10;11) <i>PICALM-MLLT10</i> Whole chromosome gain 13 Deletion 16p13.3 involving <i>CREBBP</i> Deletion 16p13.3-p13.13 without genes of interest Gain 17q25.2-q25.3 involving <i>SEPT9</i></p>	<p>10</p> <p>Deletion 7p12.2 involving <i>IKZF1</i> Highly complexity rearrangement involving chromosomes 7, 9, 13 and 16 Deletion Xp21.1 involving <i>DMD</i></p>

Table 3 : Anomalies detected by standard techniques, missed in optical mapping analysis.

Sample	Anomaly	Failure explanation
3	Whole chromosome gain 3	Detected in raw data : filtering issue
	Whole chromosome gain 4	Detected in raw data : filtering issue
4	Deletion 22q11.22-q11.23 involving <i>VPREB1</i>	Available in extraction list : operators failure
	Gain 22q11.23-q13.33 involving <i>IL2RB</i>	Detected in raw data : filtering issue
5	Near tetraploid sub-clone 80~95<4n>	Non detected : probable true tetraploid clone
8	Gain Xp22.33 involving <i>SHOX</i>	Non detected : Pseudo Autosomal Region
	Deletion Xp22.33 leading to <i>P2RY8-CRLF2</i> fusion	Non detected : Pseudo Autosomal Region
10	Sub-clonal whole chromosome gain 21	Non detected : below the limit of detection

Discrepancies between operators

The first discordance between operators concerned the chromosome 3 gain for which the data extraction showed only a partial 3q gain and interpreted as such by operator 2. The complete chromosome gain was however visible in the raw OM data and therefore interpreted as a complete trisomy 3 by operator 1 (supplemental figure B). The other discrepancy between operators concerned 4 additional anomalies, including 3 deletions and 1 gain, which were reported by operator 1. These anomalies met the size criteria (>1 Mbp) without genes of interest and were only seen with CNV tool. However, these 4 anomalies were not retained by operator 2, in particular because of their low allelic frequency and interpreted as noise.

Genetic anomalies unseen using standard techniques

OM revealed 12 anomalies meeting the selection criteria that were not found by any other technique (table 4). These included 3 deletions, 1 gain, 1 duplication and 7 inter-chromosomal rearrangements (of which 2 were highly complex events). New inter-chromosomal rearrangements identified in OM include some that lead to *LMO2-TCRA*, *TCRB-MYC* and *IGH-CEBPB* fusions and another involving *DNMT3B*. Interestingly, other anomalies identified involving genes of interest such as *ZMYM2*, *LEF1*, *CBFA2T3*, *TLX1* and *SETD2* that could be relevant in terms of leukemogenesis.

The chromosomal rearrangement involving *TRA* and *LMO2* genes was actually more complex and intertwined with the *TLX1-TRA* translocation already described by the karyotype. These two rearrangements were in fact a complex 4-breakpoint event involving chromosomes 10, 11, 14 respectively at *TLX1*, *LMO2* and *TRA* loci, resulting in a cryptic insertion of part of *TRA* locus into *LMO2* gene (figure 1). The analysis of this rearrangement using OM technology thus allowed us to understand why only *TLX1* translocation was visible at the karyotype.

Concerning the newly found t(20;21) translocation, as the breakpoint is located close to *DNMT3B* on chromosome 20, we considered interesting to investigate this anomaly. We therefore performed a PCR with probes on either side of the breakpoints.

Loss of heterozygosity detection by optical mapping

Two CN-LOH detected using SNP were identified and analyzed separately since in theory invisible using OM. In the first CN-LOH region [9q21.11q34.3(71,013,799-141,025,328)], a mean of 231 (210-246) SVs per sample was found. For sample 1, which carried this CN-LOH, 97.1% (204/210) homozygous SVs were found, compared to a mean of 67.7% (158/231, homozygous anomalies from 148 to 167, total anomalies from 210 to 246) in other samples. In the second CN-LOH region [9p24.3p21.1(192,128-33,190,239)] a mean of 85 (80-89) SVs per sample was found. For sample 2, which carried this CN-LOH, 94.05% (79/84) of homozygous SVs were found compared to a mean of 56.6% (48/85, homozygous anomalies from 38 to 54, total anomalies from 80 to 89) in other samples. It should be noted that sample 10 was excluded from this second data comparison because it showed a 9p24.3p13.1 deletion with a copy number of 1.05 copies.

Table 4 : Newly found anomalies in optical mapping.

Sample	Anomaly	Size (kbp)	Anomaly confirmation
1	Interchromosomal rearrangement involving chromosomes 10, 11 and 14 leading to <i>TRA-LMO2</i> fusion	NA	<i>TRAD & LMO2</i> FISH <i>LMO2</i> expression
	Deletion 11q13.3-q13.4 without genes of interest	1 762,807	False positive signal
	Deletion 13q12.11 near <i>ZMYM2</i>	79,292	SNP-Array raw data
2	Deletion 4q25 involving <i>LEF1</i>	25,712	SNP-Array raw data
	Interchromosomal rearrangement involving chromosomes 7 and 8, leading to <i>TRB-MYC</i> fusion	NA	<i>TRB & MYC</i> FISH
3	Deletion 2q12.3-q13 without genes of interest	1 282,087	False positive signal
	Gain 16q24.3 involving <i>CBFA2T3</i>	186,705	SNP-Array raw data
4	Gain 6q27 without genes of interest	1 280,768	False positive signal
5	Duplication 10q24.31 near <i>TLX1</i>	53,614	SNP-Array raw data <i>TLX1</i> expression
	High complexity rearrangement involving chromosomes 9, 12, 16, 17 and 19	NA	Not verified
7	Deletion 2p25.3 without genes of interest	2 038,497	False positive signal
	High complexity rearrangement involving chromosomes 7, 10, 20 and 21	NA	Not verified
	Interchromosomal rearrangement involving chromosomes 20 and 21, and <i>DNMT3B</i>	NA	Long range PCR <i>DNMT3B</i> expression
8	Interchromosomal rearrangement involving chromosomes 4 and 5, leading to <i>RELL1-TSLP</i> fusion	NA	Not verified
9	Interchromosomal rearrangement involving chromosomes 14 and 20, leading to <i>IGH-CEBPB</i> fusion	NA	<i>IGH & CEBPB</i> FISH
10	Deletion 3p21.31 involving <i>SETD2</i>	198,852	Not confirmed by CGH

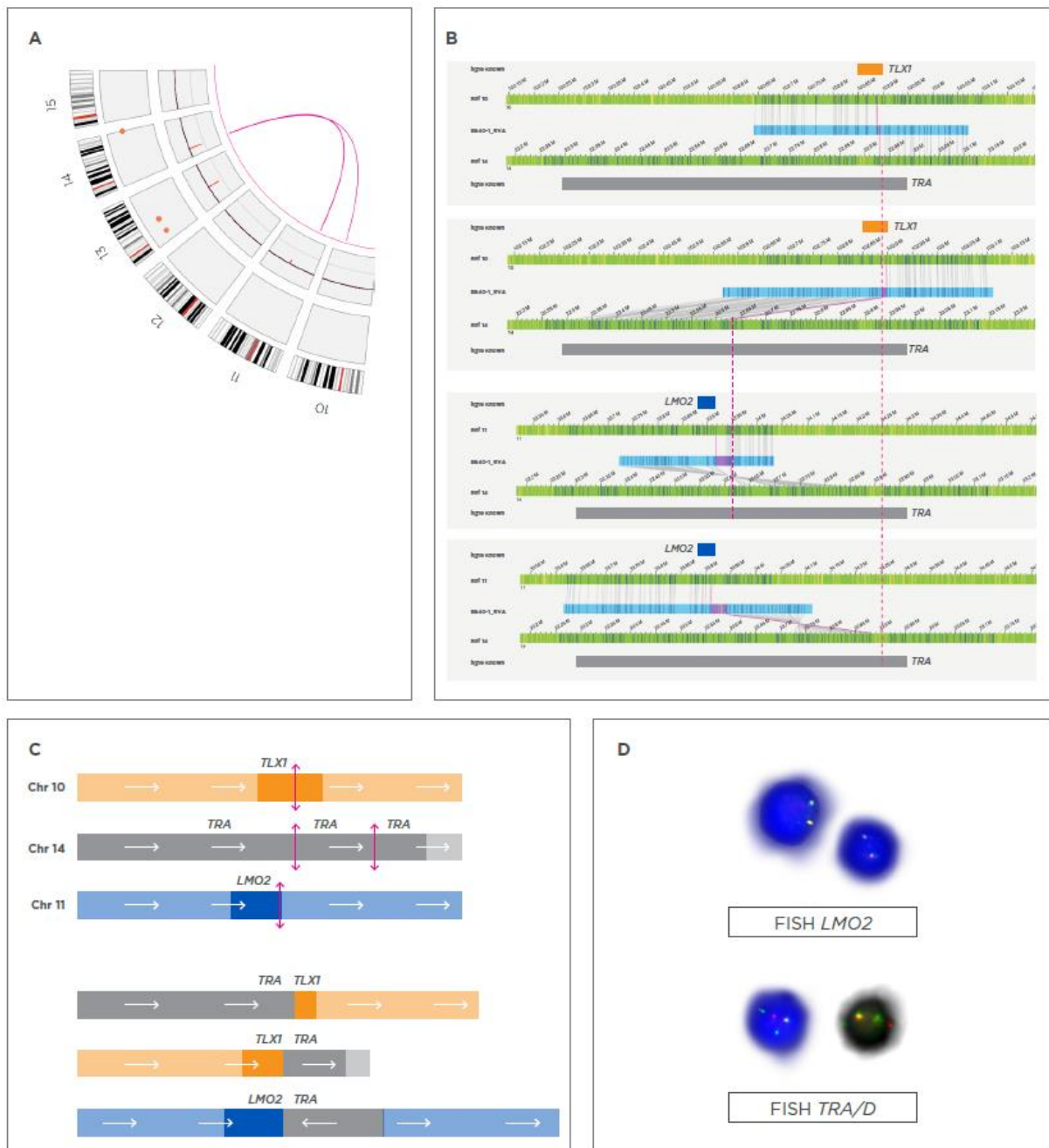


Figure 1: Optical genome mapping characterization of a complex rearrangement of chromosomes 10, 11 and 14 involving *TLX1*, *LMO2* and *TRA*.

- OGM circos plot representing the rearrangement between chromosomes 10, 11 and 14 (pink lines).
- OGM consensus maps representing the breakpoints of chromosomes 10, 11 and 14 (pink dotted line) and the genes involved. The green maps represent the reference maps of chromosomes 10, 11 and 14. The blue maps are the consensus maps obtained by OGM. They represent the abnormalities found in patients and the different derivatives of the rearrangement. These maps show that there are two breakpoints in *TRA*, each aligned with a *TLX1* and a *LMO2* derivative.
- Schematization of breakpoints at chromosome regions involved. The double breakpoint in *TRA* causes a *TLX1-TRA* translocation and a *LMO2-TRA* rearrangement by insertion.
- Interphasic FISH with separation probes *LMO2* (top) and *TRA* (bottom). The *LMO2* FISH probe shows no separation (2 fusion signals). The *TRA* FISH probe shows 1 fusion signal (chromosome 14 not rearranged), one separation (1 green and 1 red signal) and an extra green signal. The separation signal corresponds to the *TLX1-TRA* translocation. The extra green signal of the *TRA* probe is compatible with an insertion of part of *TRA* into *LMO2*.

Discussion

In this study, we aimed to compare the performance of OM compared to the standard workflow in the field of onco-hematological cytogenetics. We chose ALL because they show a high number and diversity of genetic abnormalities. Thus, it would be a tremendous medical, organizational and economic interest in simplifying ALL genetic analysis workflow. Although a limited number of patients, the selected samples had a large number of structural abnormalities of interest, representative of those typically found in B- and T-ALL. These anomalies were various in type and length (deletions, balanced or complex translocations, duplications, insertions, inversions, aneuploidies), allowing to explore the spectrum of detectable variants and blind spots. Our results allow us to respond to our initial objectives:

Is OM efficient to detect ALL-related genomic anomalies?

We found an excellent correlation between standard techniques and OM as a single tool in detecting ALL-related genetic anomalies; indeed, a large majority (90.5%) of the anomalies identified as relevant and found by the standard techniques were identified using OM. However, eight abnormalities were missed in OM analysis. A close examination of raw data and filters brought explanation for each of these failures. Half of them were detected on raw data but not retained because of the filters used, underlining the requirement for an improvement of the filtering tool to avoid visual and manual review of copy number graph when OM is able to detect this sub-clonal ploidy anomalies. However, 4 were truly missed, either because on a low VAF, under 10%, which represents the limit of sensitivity, either because of a low coverage of some regions poorly accessible to OM, particularly genomic regions containing high inter-individual variability or highly repeated sequences such as centromeres, telomeres, and PARs. This limitation makes it impossible to detect recurrent anomalies occurring in these regions, such as *CRLF2* rearrangements, well described in ALL but which detection still remains a challenge. RT-MLPA and SNP-array combined make it possible to detect them in a majority of cases but false negatives remain frequent and their prognostic impact is still in debate.²⁰

Is OM a technology compatible with a routine laboratory or does it require specific skills?

Discrepancies between the two operators were rare, easily explained and of minor importance in the overall interpretation of the cases. The first discordance between operators, concerning trisomy 3, was more related to the technique than to its interpretation. Finally, additional anomalies reported only by operator 1 and considered as noise by operator 2 were also not found in SNP-array raw data. These anomalies were found only with the CNV tool and had a relatively low allelic frequency. In these 4 cases, no gene of interest was involved limiting the clinical impact of these anomalies. However, these signals, considered finally as false positive, should lead to relativize the interpretation of the CNV tool and perhaps to limit its use to very large abnormalities only. The overall correlation between the two observers was however excellent, despite totally different professional background, the first one being hematologist, the second specialist in OM but with no skill in hematological malignancies. Emerging technologies in genetic analysis often require the mobilization of operators with advanced training and skills in bioinformatics that medical biologists do not always possess. Our data clearly show that OM is accessible to any hematologist trained in molecular biology and cytogenetics.

Can OM identify so far unrecognized anomalies that may be involved in leukemogenesis?

Twelve anomalies were identified by OM but not by any other techniques; 9 among them were confirmed. In the majority of cases, gains and deletions were confirmed by reanalysis of the SNP-array raw data and translocations by FISH. OM thus has the ability to detect anomalies that fell under the scope of standard techniques, including high resolution ones. In our series, we were able to identify involvement of genes that may have a role in the leukemic process. Indeed, we identified *DNMT3B*, an epigenetic regulator important for tumor maintenance,²¹ as a potential partner in one case of ALL-B carrying a t(20;21) undetected with other techniques. We also identified a translocation leading to the *MYC* oncogene overexpression in one case of ALL-T (supplemental figure C). In another case with a *TLX1-TRA*, the high OM resolution allowed to decipher a more complex rearrangement involving 3 genes, including *LMO2* that was overexpressed secondary to *TRA* insertion (figure 1). T-ALL with dysregulated *LMO2* and *TLX1* expression being classically classified in distinct, non-overlapping genetic subgroups, this is to our knowledge the first case reporting involvement of both genes in the ALL-T leukemogenesis.^{22,23} At last, having a closer look to rejected variants identified with OM, but not selected because not included in our list of 186 genes of interest, we found 16 genes referenced in the COSMIC database that could have a potential role in B/T cells leukemogenesis: *CELA3B*, *CDH4*, *OTP*, *KDM2B*, *PSG11*, *PSG2*, *PSG5*, *PSG4*, *SOX4*, *P2RY2*, *P2RY6*, *SOX2*, *NRG2*, *ABRACL*, *CDK13*, *NARS2* (Supplemental table B). In particular, one of these abnormalities involved a chromosome 6 deletion in the histone genes, centered particularly on *HIST1H2BE* and *HIST1H4D*. Although not part of our panel

of genes of interest, this type of deletion has also been described in ALL²⁴. All these examples highlight that new high-resolution technologies such as OM could identify new molecular pathways involved in leukemogenesis and therefore open a new field in that research area.

OM ability to identify CN-LOH

OM is not designed for detection of CN-LOH since not a sequence-based technology and unable to differentiate two alleles of the same gene. However, we succeeded in replicating the method used by Neveling *et al*¹⁸ to detect CN-LOH with the Genome *de novo* assembly algorithm. This method allowed us to highlight a majority of variants labelled "homozygous" in the zones and samples concerned. We were able to prove that these results were not reproduced in the absence of CN-LOH by repeating this analysis on the samples that were free. These results suggest that it would be possible to detect CN-LOH anomalies with OM technology, although these results need to be replicated on a larger cohort.

OM pre analytic and analytic interferences

We found clear signs of contamination of sample 7 by sample 4 in OM: there was 17 variants with identical breakpoints and sizes (N=17) detectable in both samples, always with a much lower single molecule count in sample 7. This contamination was confirmed by reference data. The two samples had been processed on the same flowcell. This incident shows the absolute necessity of introducing systematic and/or automated means of detection of this type of event in order to envisage its use in diagnostic procedures. This could involve inter-sample comparison by series or flowcell, or the use of a patient-specific DNA identification system such as genetic barcodes as is usually done in high throughput sequencing for example. It should be noted that we considered these 17 supernumerary variants in sample 7 outside the blind comparison. For two samples, the negative label variance exceeded the upper value recommended by the supplier. This was coupled with a lower average label density, leading to a significant background noise in CNV tool, in the form of numerous false small deletions distributed throughout the genome. These false positive signals were easily eliminated by the operators without impacting the detection of any anomaly of interest, but it might theoretically cause real sub-clones to be missing. A series of fresh samples would have to be analyzed to estimate whether cases of high NLVs are less frequent on this type of material than on frozen DMSO pellets.

In summary, we report OM as an innovative, time and cost-effective alternative to the different cytogenetic techniques currently performed at diagnosis for ALL genetic characterization. We identified however stumbling blocks that still need to be addressed such as filters for the CNV tool, sometimes causing false negatives for the detection of large sub-clonal anomalies or background noise signals, as well as the lack of solution in case of inter-sample contamination. Of note, OM could help to identify genetic events not seen with any current techniques but potentially directly involved in the leukemic process and potentially relevant in terms of prognosis and therapeutic stratification. Before its use as a routine tool in malignant hematology, larger multicentric series are required to confirm and complete these data, particularly in identifying cases for which complementary tools would still be needed such as whole chromosomes sub-clonal gains and anomalies involving telomeres, centromeres and PARs.

Author contributions

Conception and design: YF, GJ and LG. Administrative support: LG, JPM and CP. Clinical data management: AL, CD and JPM. Karyotype & FISH analysis: YF, DP and IL. SNP-array & RT-MLPA analysis: ND and NG. OM data acquisition: MW. OM data analysis: VL and JL. Data comparison and analysis: VL and YF; and all authors read, commented on, and approved the manuscript.

Acknowledgments

We are grateful to Prof. Sandrine Castelain and CHU Amiens-Picardie for their financial support.

References

1. Howlader N, Noone A, Krapcho M, et al. SEER Cancer Statistics Review, 1975-2017, National Cancer Institute. Bethesda, MD. https://seer.cancer.gov/csr/1975_2017/, based on November 2019 SEER data submission, posted to the SEER web site, April 2020.
2. Iacobucci I, Mullighan CG. Genetic Basis of Acute Lymphoblastic Leukemia. *JCO*. 2017;35(9):975–983.
3. Liu Y-F, Wang B-Y, Zhang W-N, et al. Genomic Profiling of Adult and Pediatric B-cell Acute Lymphoblastic Leukemia. *EBioMedicine*. 2016;8:173–183.
4. Van Vlierberghe P, Ferrando A. The molecular basis of T cell acute lymphoblastic leukemia. *J Clin Invest*. 2012;122(10):3398–3406.
5. Hoelzer D, Bassan R, Dombret H, et al. Acute lymphoblastic leukaemia in adult patients: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology*. 2016;27:v69–v82.
6. Rack KA, van den Berg E, Haferlach C, et al. European recommendations and quality assurance for cytogenomic analysis of haematological neoplasms. *Leukemia*. 2019;33(8):1851–1867.
7. Schwartz D, Li X, Hernandez L, et al. Ordered restriction maps of *Saccharomyces cerevisiae* chromosomes constructed by optical mapping. *Science*. 1993;262(5130):110–114.
8. Lam ET, Hastie A, Lin C, et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. *Nat Biotechnol*. 2012;30(8):.
9. Mak ACY, Lai YYY, Lam ET, et al. Genome-Wide Structural Variation Detection by Genome Mapping on Nanochannel Arrays. *Genetics*. 2016;202(1):351–362.
10. Pendleton M, Sebra R, Pang AWC, et al. Assembly and diploid architecture of an individual human genome via single-molecule technologies. *Nat Methods*. 2015;12(8):780–786.
11. Chaisson MJP, Sanders AD, Zhao X, et al. Multi-platform discovery of haplotype-resolved structural variation in human genomes. *Nat Commun*. 2019;10(1):1784.
12. Mantere T, Neveling K, Pebrel-Richard C, et al. Next generation cytogenetics: genome-imaging enables comprehensive structural variant detection for 100 constitutional chromosomal aberrations in 85 samples. *Genetics*; 2020.
13. Barseghyan H, Tang W, Wang RT, et al. Next-generation mapping: a novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis. *Genome Med*. 2017;9(1):90.
14. Dixon JR, Xu J, Dileep V, et al. Integrative detection and analysis of structural variation in cancer genomes. *Nat Genet*. 2018;50(10):1388–1398.
15. Luebeck J, Coruh C, Dehkordi SR, et al. AmpliconReconstructor integrates NGS and optical mapping to resolve the complex structures of focal amplifications. *Nat Commun*. 2020;11(1):4374.
16. Chan EKF, Cameron DL, Petersen DC, et al. Optical mapping reveals a higher level of genomic architecture of chained fusions in cancer. *Genome Res*. 2018;28(5):726–738.
17. Xu J, Song F, Schleicher E, et al. An Integrated Framework for Genome Analysis Reveals Numerous Previously Unrecognizable Structural Variants in Leukemia Patients' Samples. *Genomics*; 2019.
18. Neveling K, Mantere T, Vermeulen S, et al. Next generation cytogenetics: comprehensive assessment of 48 leukemia genomes by genome imaging. *Genomics*; 2020.
19. Levy B, Baughn LB, Chartrand S, et al. A National Multicenter Evaluation of the Clinical Utility of Optical Genome Mapping for Assessment of Genomic Aberrations in Acute Myeloid Leukemia. *medRxiv*. 2020;2020.11.07.20227728.
20. Moorman AV. New and emerging prognostic and predictive genetic biomarkers in B-cell precursor acute lymphoblastic leukemia. *Haematologica*. 2016;101(4):407–416.
21. Poole CJ, Zheng W, Lodh A, et al. DNMT3B overexpression contributes to aberrant DNA methylation and MYC-driven tumor maintenance in T-ALL and Burkitt's lymphoma. *Oncotarget*. 2017;8(44):76898–76920.
22. Meijerink JPP. Genetic rearrangements in relation to immunophenotype and outcome in T-cell acute lymphoblastic leukaemia. *Best Practice*. 2010;12.
23. Van Vlierberghe P, Pieters R, Beverloo HB, Meijerink JPP. Molecular-genetic insights in paediatric T-cell acute lymphoblastic leukaemia. *British Journal of Haematology*. 2008;143(2):153–168.
24. Loudin MG, Wang J, Leung H-CE, et al. Genomic profiling in Down syndrome acute lymphoblastic leukemia identifies histone gene deletions associated with altered methylation profiles. *Leukemia*. 2011;25(10):1555–1563.

25. Tate JG, Bamford S, Jubb HC, et al. COSMIC: the Catalogue Of Somatic Mutations In Cancer. *Nucleic Acids Research*. 2019;47(D1):D941–D947.

Supplemental Methods

Karyotype

Bone marrow samples were cultured for 24 hours at 37°C in a medium consisting of RPMI supplemented with 18.75% fetal calf serum, penicillin and streptomycin. After culture and metaphase mitosis blockade by addition of colchicine, hypotonic shock was performed with a KCl solution, followed by prefixation and 3 fixation series with a solution composed of 3 volumes of methanol to 1 volume of acetic acid. After spreading the obtained cytogenetic pellets on slides, a thermal denaturation of the DNA in saline medium (R bands) was performed, then the slides were dehydrated in methanol before being stained with Giemsa. In the absence of clonal abnormalities, at least 20 metaphases were analyzed for each patient. The karyotype was described according to the ISCN 2016 nomenclature.

FISH

Interphasic fluorescent in situ hybridization (FISH) was performed using fresh or frozen spreads obtained after cell culture. Slides were dehydrated in successive baths of ethanol. After deposition of the probes (METASYSTEMS, CYTOCELL) on the area to be studied, the slides were placed in a Thermobrite automaton for codenaturation and hybridization of the probes according to the supplier's instructions. After hybridization, the slides were washed in Saline-Sodium Citrate (SSC) solution before being counter-stained with DAPI (4'-6-diaminido-2-phenylindole). The resulting FISH slides were read under fluorescence microscopy by 2 different observers. At least 200 nuclei were analyzed for each mix of probes in order to establish a formula according to the ISCN 2016 nomenclature.

SNP Array

The SNP-array analysis was performed on a Cytoscan HD chip (Affymetrix). After cell isolation on Ficoll, DNA extraction on QIAGEN column and assay, the resulting DNA samples were amplified with the Affymetrix kit according to the manufacturer's recommendations. The data were analyzed on the Chromosome Assay Suite Analysis (ChAS) software v3.0. Abnormalities were retained when allele fraction was > 10% and larger than 1Mbp or > 20 kbp and involving one of the 186 genes defined as potentially relevant in leukemogenesis (Supplemental table 1). CN-LOHs were not included in the blinded study and were considered separately.

RT-MLPA

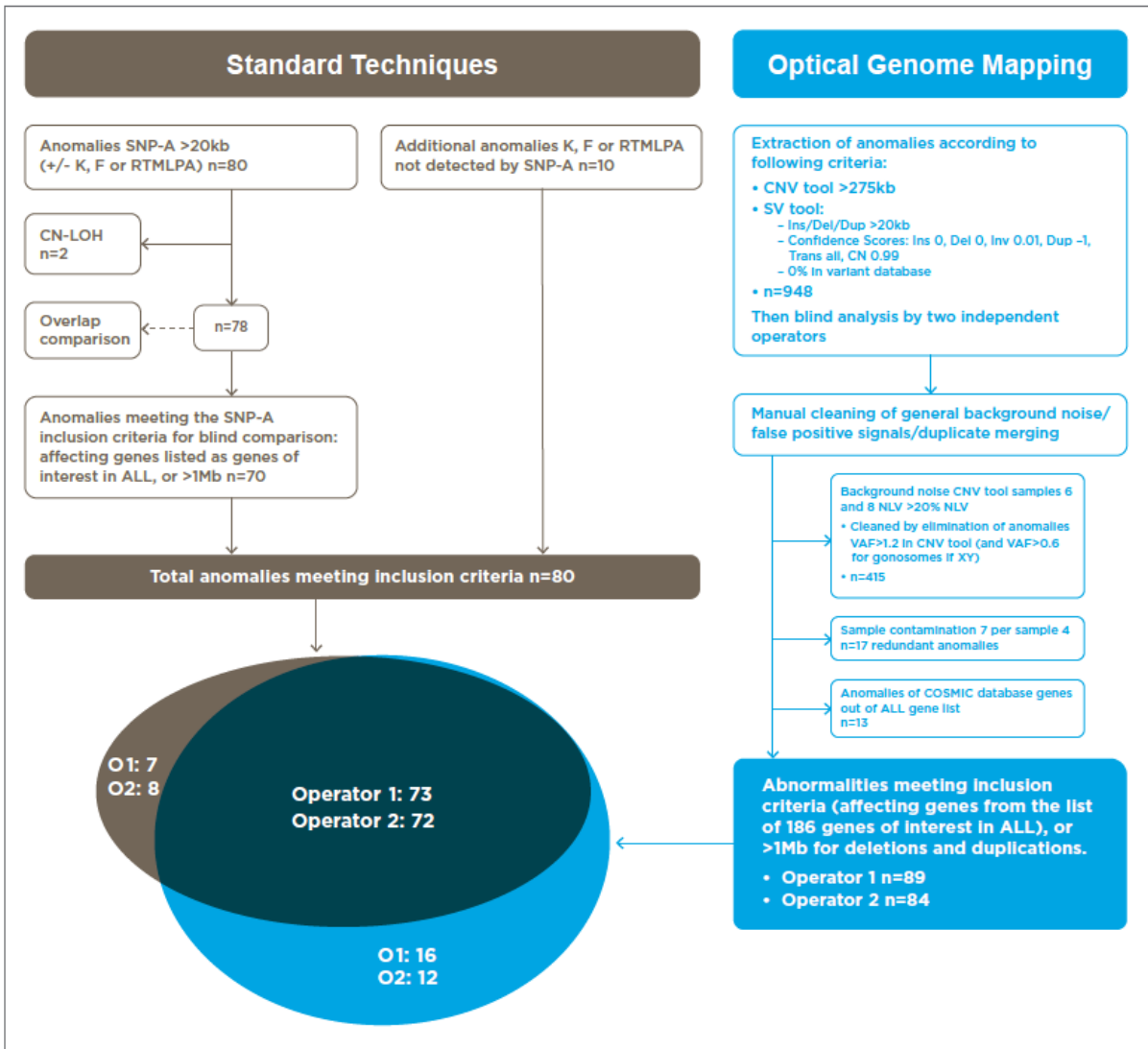
RNA was extracted from cell isolations separated by Ficoll and stored in Trizol or Maxwell buffers. Extraction was performed manually (Trizol) or automatically (Maxwell). After dosing, the amount of RNA was normalized to obtain an RNA solution of 10 µl and 100 ng/µl. A step of retro-transcription of the RNA into cDNA was performed. The quality of the DNA thus obtained was evaluated by real-time quantitative PCR using *ABL1* as the reference gene. The quality of the DNA was considered insufficient if the copy number of *ABL1* was less than 10,000 copies. The different steps of the LD-RTPCR (Ligation Dependent RTPCR) were performed on a thermal cycler with a heated lid (Mastecycler, Eppendorf). The cDNA was hybridized with the probes before adding the ligation mix. Then the final PCR step was performed. The PCR product was analyzed on 4% agarose gel and revealed under UV light before being sequenced by pyrosequencing. The resulting pyrogram was analyzed for the presence of fusion transcripts.

Supplemental Table A: Genes of interest

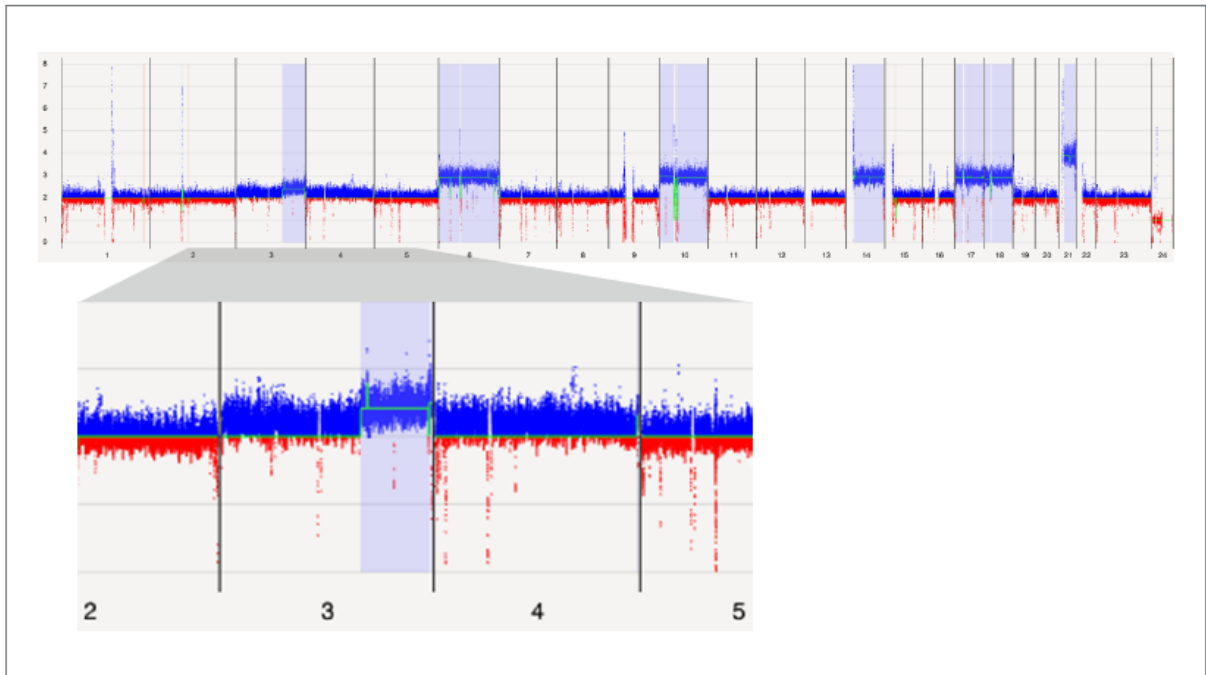
Gene	Locus	Gene	Locus	Gene	Locus	Gene	Locus
<i>SFPQ</i>	1p34	<i>DEK</i>	6p22	<i>LMO1</i>	11p15	<i>CREBBP</i>	16p13
<i>STIL</i>	1p33	<i>CASP8AP2</i>	6q15	<i>NUP98</i>	11p15	<i>MYH11</i>	16p13
<i>TAL1</i>	1p33	<i>MYB</i>	6q23	<i>WT1</i>	11p13	<i>GLIS2</i>	16p13
<i>EPS15</i>	1p32	<i>AH11</i>	6q23	<i>RAG1</i>	11p13	<i>NDE1</i>	16p13
<i>RBM15</i>	1p13	<i>AFDN</i>	6q27	<i>LMO2</i>	11p13	<i>CBFB</i>	16q22
<i>TPM3</i>	1q21	<i>CEP43</i>	6q27	<i>CD44</i>	11p13	<i>CBFA2T3</i>	16q24
<i>ARNT</i>	1q21	<i>HOXA13</i>	7p15	<i>SLC1A2</i>	11p13	<i>TP53</i>	17p13
<i>PDE4DIP</i>	1q21	<i>HOXA9</i>	7p15	<i>RAG2</i>	11p12	<i>NF1</i>	17q11
<i>MLLT11</i>	1q21	<i>IKZF1</i>	7p12	<i>ARHGEF17</i>	11q13	<i>SUZ12</i>	17q11
<i>PBX1</i>	1q23	<i>CARD11</i>	7p22	<i>C2CD3</i>	11q13	<i>MLLT6</i>	17q12
<i>RCSD1</i>	1q24	<i>PIK3CG</i>	7q22	<i>PICALM</i>	11q14	<i>IKZF3</i>	17q21
<i>ABL2</i>	1q25	<i>TRIM24</i>	7q33-4	<i>MAML2</i>	11q21	<i>RARA</i>	17q21
<i>IGK</i>	2p11	<i>TRB</i>	7q34	<i>DDX10</i>	11q22	<i>STAT5B</i>	17q21
<i>AFF3</i>	2q11	<i>ZC3HAV1</i>	7q34	<i>ARHGEF12</i>	11q23	<i>SPAG9</i>	17q21
<i>RANBP2</i>	2q12	<i>EZH2</i>	7q36	<i>KMT2A</i>	11q23	<i>HLF</i>	17q22
<i>ZEB2</i>	2q22	<i>MNX1</i>	7q36	<i>ZBTB16</i>	11q23	<i>SEPTIN9</i>	17q25
<i>IKZF2</i>	2q34	<i>PCM1</i>	8p22	<i>CBL</i>	11q23	<i>PTPN2</i>	18p11
<i>INPP5D</i>	2q37	<i>PTK2B</i>	8p21	<i>DCPS</i>	11q24	<i>TCF3</i>	19p13
<i>SETD2</i>	3p21	<i>FGFR1</i>	8p11	<i>ETV6</i>	12p13	<i>EPOR</i>	19p13
<i>FHIT</i>	3p14	<i>TOX</i>	8q12	<i>ATF7IP</i>	12p13	<i>TYK2</i>	19p13
<i>PTPRG</i>	3p14	<i>PAG1</i>	8q21	<i>KDM5A</i>	12p13	<i>LYL1</i>	19p13
<i>FOXP1</i>	3p13	<i>RUNX1T1</i>	8q21	<i>BCAT1</i>	12p12	<i>MLLT1</i>	19p13
<i>BTLA</i>	3q13	<i>RAD21</i>	8q24	<i>RASSF3</i>	12q14	<i>ELL</i>	19p13
<i>CD200</i>	3q13	<i>MYC</i>	8q24	<i>BTG1</i>	12q21	<i>SLX4IP</i>	20p12
<i>CIP2A</i>	3q13	<i>JAK2</i>	9p24	<i>SH2B3</i>	12q23	<i>MAPRE1</i>	20q11
<i>MLF1</i>	3q25	<i>CDKN2A</i>	9p21	<i>PAN3</i>	13q12	<i>TOP1</i>	20q12
<i>TBLXRI</i>	3q26	<i>CDKN2B</i>	9p21	<i>ZMYM2</i>	13q12	<i>ERG</i>	21q22
<i>PDGFRA</i>	4q12	<i>MLLT3</i>	9p21	<i>ELF1</i>	13q14	<i>RUNX1</i>	21q22
<i>CHIC2</i>	4q12	<i>MTAP</i>	9p21	<i>RBI</i>	13q14	<i>VPREB1</i>	22q11
<i>FIP1L1</i>	4q12	<i>PAX5</i>	9p13	<i>DLEU1</i>	13q14	<i>BCR</i>	22q11
<i>AFF1</i>	4q22	<i>CNTRL</i>	9q33	<i>SERP2</i>	13q14	<i>IGL</i>	22q11
<i>RAP1GDS1</i>	4q23	<i>ABL1</i>	9q34	<i>DGKH</i>	13q14	<i>SEPTIN5</i>	22q11
<i>LEF1</i>	4q25	<i>NUP214</i>	9q34	<i>DLEU2</i>	13q14	<i>IL2RB</i>	22q12
<i>NR3C2</i>	4q31	<i>NOTCH1</i>	9q34	<i>TRD</i>	14q11	<i>MNI</i>	22q12
<i>FBXW7</i>	4q31	<i>SET</i>	9q34	<i>TRA</i>	14q11	<i>MRTFA</i>	22q13
<i>MEF2C</i>	5q14	<i>MLLT10</i>	10p12	<i>STRN3</i>	14q12	<i>SHOX</i>	Xp22
<i>SSBP2</i>	5q14	<i>CCDC6</i>	10q21	<i>BCL11B</i>	14q32	<i>CRLF2</i>	Xp22
<i>TSLP</i>	5q22	<i>TET1</i>	10q21	<i>IGH</i>	14q32	<i>CSF2RA</i>	Xp22
<i>SNX2</i>	5q23	<i>ZMIZ1</i>	10q22	<i>EML1</i>	14q32	<i>IL3RA</i>	Xp22
<i>NR3C1</i>	5q31	<i>PTEN</i>	10q23	<i>ATP10A</i>	15q11	<i>P2RY8</i>	Xp22
<i>AFF4</i>	5q31	<i>PIK3API</i>	10q24	<i>SPRED1</i>	15q14	<i>DMD</i>	Xp21
<i>ARHGAP26</i>	5q31	<i>TLX1</i>	10q24	<i>MGA</i>	15q15	<i>KDM6A</i>	Xp11
<i>CSF1R</i>	5q32	<i>ADD3</i>	10q25	<i>KNL1</i>	15q15	<i>GATA1</i>	Xp11
<i>PDGFRB</i>	5q32			<i>PML</i>	15q24	<i>ATRX</i>	Xq21
<i>EBF1</i>	5q33			<i>NTRK3</i>	15q25	<i>SEPTIN6</i>	Xq24
<i>TLX3</i>	5q35					<i>BCORL1</i>	Xq25
<i>NPM1</i>	5q35					<i>SH2D1A</i>	Xq25
<i>NSD1</i>	5q35					<i>STAG2</i>	Xq25
						<i>PHF6</i>	Xq26
						<i>ARHGEF6</i>	Xq26

Supplemental Table B : Anomalies found in optical mapping, not meeting inclusion criteria but involving a gene from the COSMIC database.²⁵

Sample	Anomaly	COSMIC genes involved	Size (kbp)
1	Deletion 1p36.12	<i>CELA3B</i>	23,178
2	Deletion 20q13.33	<i>CDH4</i>	293,101
3	Gain 5q14.1	<i>OTP</i>	32,364
	Deletion 6p22.2	<i>HIST1H2BE & HIST1H4D</i>	100,381
4	Deletion 12q24.31	<i>KDM2B</i>	32,286
	Gain 19q31.31	<i>PSG11, PSG2, PSG5, PSG4</i>	277,171
5	Gain 6q22.3	<i>SOX4</i>	93,214
	Gain 11q13.4	<i>P2RY2 et P2RY6</i>	172,411
6	Deletion 3q26.33	<i>SOX2</i>	421,090
9	Deletion 5q31.2	<i>NRG2</i>	25,380
	Deletion 6q24.1	<i>ABRACL</i>	21,298
	Deletion 7p14.1	<i>CDK13</i>	167,562
10	Deletion 11q14.1	<i>NARS2</i>	31,322

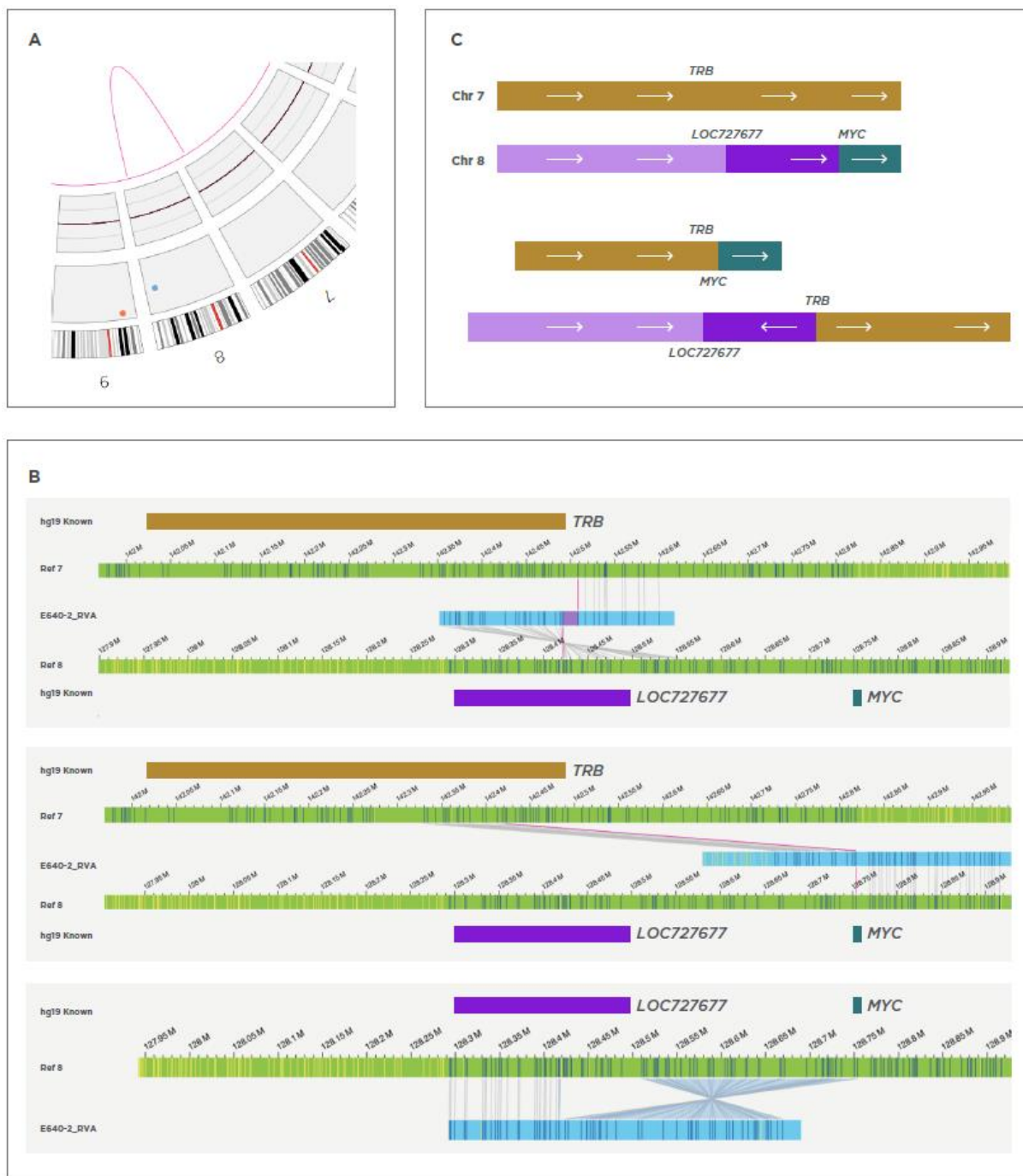


Supplemental Figure A: Flow chart of detected anomalies using standard techniques and optical genome mapping.



Supplemental Figure B: Optical genome mapping CNV tool analysis of sample 3.

OGM raw data show a whole gain of chromosomes 3 and 4 (zoom in). However, only the region 3q22.1q29 (highlighted in blue), which higher copy number gain, passed the filters for anomalies extraction.



Supplemental figure C: Optical genome mapping characterization of a rearrangement of chromosomes 7 and 8 involving *MYC* and *TRB*.

- A. OGM circos plot representing the rearrangement between chromosomes 7 and 8 (pink lines).
- B. OGM consensus maps representing the breakpoints of chromosomes 7 & 8 and the genes involved. The green maps represent the reference maps of chromosomes 7 & 8. The blue maps are the consensus maps obtained by OGM. They represent the abnormalities found in patients and the different derivatives of the rearrangement. The bottom maps on chromosome 8 represent an inversion.
- C. Schematization of breakpoints at chromosome regions involved. The double breakpoint in chromosome 8 causes a *MYC-TRB* rearrangement with inversion.

ANNEXE 7 : Serment d'Hippocrate

Au moment d'être admis à exercer la médecine, je promets et je jure d'être fidèle aux lois de l'honneur et de la probité.

Mon premier souci sera de rétablir, de préserver ou de promouvoir la santé dans tous ses éléments, physiques et mentaux, individuels et sociaux.

Je respecterai toutes les personnes, leur autonomie et leur volonté, sans aucune discrimination selon leur état ou leurs convictions. J'interviendrai pour les protéger si elles sont affaiblies, vulnérables ou menacées dans leur intégrité ou leur dignité. Même sous la contrainte, je ne ferai pas usage de mes connaissances contre les lois de l'humanité.

J'informerai les patients des décisions envisagées, de leurs raisons et de leurs conséquences. Je ne tromperai jamais leur confiance et n'exploiterai pas le pouvoir hérité des circonstances pour forcer les consciences.

Je donnerai mes soins à l'indigent et à quiconque me les demandera. Je ne me laisserai pas influencer par la soif du gain ou la recherche de la gloire.

Admis dans l'intimité des personnes, je tairai les secrets qui me seront confiés. Reçu à l'intérieur des maisons, je respecterai les secrets des foyers et ma conduite ne servira pas à corrompre les mœurs.

Je ferai tout pour soulager les souffrances. Je ne prolongerai pas abusivement les agonies. Je ne provoquerai jamais la mort délibérément.

Je préserverai l'indépendance nécessaire à l'accomplissement de ma mission. Je n'entreprendrai rien qui dépasse mes compétences. Je les entretiendrai et les perfectionnerai pour assurer au mieux les services qui me seront demandés.

J'apporterai mon aide à mes confrères ainsi qu'à leurs familles dans l'adversité. Que les hommes et mes confrères m'accordent leur estime si je suis fidèle à mes promesses ; que je sois déshonoré et méprisé si j'y manque.

LA CARTOGRAPHIE OPTIQUE DU GENOME, UNE ALTERNATIVE PROMETTEUSE AUX APPROCHES CYTOGENETIQUES STANDARDS DANS LES LEUCEMIES AIGUES LYMPHOBLASTIQUES : ETUDE COMPARATIVE EN AVEUGLE SUR 10 PATIENTS

Introduction : Les leucémies aiguës lymphoblastiques (LAL) sont caractérisées par une grande diversité d'anomalies génétiques. Leur détection en routine requiert l'emploi de nombreuses techniques complémentaires : caryotype, hybridation fluorescente *in situ* (FISH), Single nucleotide polymorphism Array (SNP-Array) et recherche de transcrits de fusion. La cartographie optique du génome (COG) est une technologie permettant l'étude de l'ensemble du génome à haute résolution ainsi que la détection des anomalies de nombre et de structure, y compris les translocations équilibrées. L'objectif était d'évaluer les performances de cette nouvelle technologie face aux techniques de standards.

Matériel et méthodes : Une COG Bionano a été réalisée pour 10 patients atteints de LAL-B ou T au diagnostic. Les données étaient analysées en aveugle par deux opérateurs indépendants puis comparées à celles obtenues par caryotype, FISH, SNP-Array et Reverse Transcription Multiplex Ligation-dependent Probe Amplification (RT-MLPA).

Résultats : Sur les 80 anomalies décrites par les techniques standards, 73 étaient retrouvées par les opérateurs en COG.

Discussion : Les 10 anomalies non retrouvées en COG étaient dues à des problèmes de filtre des algorithmes ou à un défaut de détection de la technologie. Douze nouvelles anomalies étaient décrites en COG, dont 7 ont pu être confirmées par une autre technique, certaines impliquant des gènes d'intérêt. La COG permettait également la résolution d'évènements chromosomiques complexes.

Conclusion : La COG semble être une alternative innovante, rapide et rentable aux techniques de cytogénétique standards pour la caractérisation génétique des LAL.

Mots-clefs : Leucémie aiguë lymphoblastique, cytogénétique, cartographie chromosomique, hémopathies malignes, hématologie, génétique, biologie

OPTICAL GENOME MAPPING, A PROMISING ALTERNATIVE TO GOLD STANDARD CYTOGENETIC APPROACHES IN ACUTE LYMPHOBLASTIC LEUKEMIAS: A BLIND COMPARISON ON 10 PATIENTS

Introduction: Acute lymphoblastic leukemia (ALL) is characterized by a large number of genetic abnormalities. Their routine detection requires the use of many complementary techniques: karyotyping, fluorescent *in situ* hybridization (FISH), Single nucleotide polymorphism Array (SNP-Array) and the search for fusion transcripts. Optical genome mapping (OGM) is a technology that allows a high-resolution genome-wide analysis as well as the detection of copy number anomalies and structure anomalies, including balanced translocations. The objective was to evaluate the performance of this new technology compared to standard techniques.

Material and methods: A Bionano OGM was performed for 10 patients with B or T-ALL at diagnosis. The data were blindly analyzed by two independent operators and compared to those obtained by karyotype, FISH, SNP-Array and by Reverse Transcription Multiplex Ligation-dependent Probe Amplification (RT-MLPA).

Results: Among the 80 abnormalities described by standard techniques, 73 were found by operators in OGM.

Discussion: The 10 anomalies not found in OGM were due to algorithms' filter issues or to a detection defect in the technology. Twelve anomalies were newly described in OGM, 7 of which could be confirmed by another technique and some involving genes of interest. OGM also allowed the resolution of complex chromosomal events.

Conclusion: OGM appears to be an innovative, rapid and cost-effective alternative to standard cytogenetic techniques for the genetic characterization of ALL.

Keywords: Acute lymphoblastic leukemia, cytogenetics, chromosome mapping, hematological malignancies, hematology, genetics, biology