HAL
open science

L'intégration du séquençage de nouvelle génération dans le diagnostic médical : application aux leucémies aiguës myéloïdes et syndromes myélodysplasiques

Virginie Jacques

To cite this version:

Virginie Jacques. L'intégration du séquençage de nouvelle génération dans le diagnostic médical : application aux leucémies aiguës myéloïdes et syndromes myélodysplasiques. Biotechnologies. 2013. dumas-01147522

HAL Id: dumas-01147522

https://dumas.ccsd.cnrs.fr/dumas-01147522v1

Submitted on 30 Apr 2015

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.

CONSERVATOIRE NATIONAL DES ARTS ET METIERS

TOULOUSE

MEMOIRE

Présenté en vue d'obtenir le

DIPLOME D'INGENIEUR CNAM

SPECIALITE : Sciences et Techniques du Vivant

OPTION : Génie Biologique

Par

JACQUES Virginie

L'intégration du séquençage de nouvelle génération dans le diagnostic médical : application aux leucémies aiguës myéloïdes et syndromes myélodysplasiques

Soutenu le : 09/12/2013

JURY :
Dr Virginie Bourlier (INSERM U1048, UPS)
Dr Thomas Cadoudal (INSERM U1048)
Pr Isabelle Castan-Laurell (INSERM U1048, UPS)
Pr Claude Knauf (INSERM U1048, UPS)
Pr Philippe Pochart (CNAM)

Remerciements

Je tiens à remercier pour ce mémoire :

Le Pr Eric Delabesse pour m’avoir accordé sa confiance et m’avoir transmis un savoir précieux sur l’application de la biologie moléculaire, mais surtout pour m’avoir soutenue dans la mise en œuvre de mes projets d’évolution professionnelle.

Le Dr Naïs Prade, pour m’avoir suivie, guidée et soutenue lors de mes recherches et ce, tout au long de l’élaboration de ce mémoire, pour m’avoir accordé du temps et également pour m'avoir transmis la méthodologie de travail de l'ingénieur que je cherche à devenir.

Sophie Valière ainsi que toute l’équipe de la plateforme INRA - GenoToul de Toulouse pour avoir passé de nombreuses heures à me former sur l’utilisation du MiSeq, le séquenceur nouvelle génération d’Illumina.

Mme Marty, Mme Cardonne et Mme Gilles, cadres du laboratoire, pour m’avoir accordé le temps nécessaire à la mise en œuvre de mes recherches sur mon temps de travail et avoir autorisé cette avancée professionnelle au sein même du secteur.

Mes collègues Cynthia, Magali et Sabrina pour avoir facilité ma prise de poste et avoir soutenu mon entreprise personnelle.

Mr Knauf, pour m’avoir poussée dans cette entreprise, pour m’avoir donné de précieux conseils et surtout pour m’avoir convaincue d’aller jusqu’au bout de ce parcours semé d’obstacles !

Sans oublier mon conjoint pour m’avoir facilité la vie pendant ces quatre années d’étude et m’avoir soutenue dans les moments difficiles…

Sommaire

I TABLE DES ABREVIATIONS 6

II INTRODUCTION 8

III OBJECTIFS GENERAUX ET PROBLEME POSE 10

1 Objectif principal 10

2 Planning 10

3 Gènes cibles 11

4 Accréditation des analyses biologiques 12

IV RECHERCHES BIBLIOGRAPHIQUES 13

1 Les Leucemies Aigües Myeloïdes et les Syndromes Myelodysplasiques 13

1.1 L'Hématopoïèse 13
1.1.1 Les cellules hématopoïétiques 13
1.1.2 Les cytokines 14
1.1.2.1 Les facteurs activateurs 14
1.1.2.2 Les facteurs répresseurs 15
1.1.3 Les facteurs de transcription 15

1.2 Les Leucémies Aigües Myéloïdes 15
1.2.1 Leur définition et épidémiologie 15
1.2.2 Les mécanismes oncogénétiques 16
1.2.3 La présentation clinique 17
1.2.4 Le diagnostic biologique 18
1.2.4.1 La numération et formule sanguine 18
1.2.4.2 Le myélogramme 18
1.2.4.3 L'immunophénotypage 18
1.2.4.4 Le caryotype 19
1.2.4.5 La Fluorescence in Situ par Hybridation 21
1.2.4.6 La biologie moléculaire 22
1.2.5 La classification 23
1.2.6 Le traitement 24
1.2.6.1 La phase d'induction 24
1.2.6.2 La phase de consolidation 24

1.3 Les Syndromes Myélodysplasiques 25
1.3.1 Leur définition et épidémiologie 25
1.3.2 La présentation clinique 25
1.3.3 Le diagnostic biologique 25
1.3.3.1 La numération et formule sanguine 25
1.3.3.2 Le myélogramme 25
1.3.3.3 Le caryotype 26
1.3.3.4 La biologie moléculaire 26
1.3.4 La classification 26
1.3.5 Le traitement 27

2 Les techniques moleculaires de recherche de mutations 27

2.1 Le séquençage Sanger 27

2.2 Le criblage de mutations par High Resolution Melting Curve PCR 28

2.3 Le séquençage de 2ème génération 29
2.3.1 La technologie 454 30
2.3.2 La technologie SOLiD 31
2.3.3 La technologie Ion Torrent / Ion Proton 32
2.3.4 La technologie Illumina 33

2.4 Le séquençage de 3ème génération 34
2.4.1 La technologie Oxford Nanopore 34
2.4.2 La technologie Pacific Biosciences 34

2.5 Les critères de qualité 35

3 Les 24 genes d'intéret majeur retenus pour notre etude mutationnelle 35

3.1 La diversité des gènes mutés **36**
3.1.1 Dans les Leucémies Aigües Myéloïdes 36
3.1.2 Dans les Syndromes Myélodysplasiques 37

3.2 Les mutations des facteurs de transcription **38**
3.2.1 La mutation « classique » de CEBPα 38
3.2.2 La mutation de WT1 38
3.2.3 La mutation d'ASXL1 38
3.2.4 La mutation de DNMT3A 38
3.2.5 La mutation de RUNX1 39
3.2.6 La mutation de TET2 39
3.2.7 La mutation de TP53 39
3.2.8 La mutation d'ETV6 40
3.2.9 La mutation de PHF6 40

3.3 Les mutations des récepteurs de cytokines et de leurs voies de signalisation **40**
3.3.1 La mutation « classique » de FLT3 40
3.3.2 La mutation de KIT 40
3.3.3 La mutation de KRAS et NRAS 41
3.3.4 La mutation de JAK2 41

3.4 Les autres mutations **41**
3.4.1 La mutation « classique » de NPM1 41
3.4.2 La mutation d'EZH2 41
3.4.3 Les mutations d'IDH1 et IDH2 42
3.4.4 La mutation des gènes du splicéosome (SRSF2, U2AF1 et SF3B1) 42
3.4.5 La mutation de SETBP1 43
3.4.6 La mutation du CBL 43

V OBJECTIF SPECIFIQUE DU TRAVAIL **44**

VI MATERIEL ET METHODES **45**

1 Glossaire technique **45**

2 Materiel **45**

2.1 La nature du matériel **45**

2.2 La qualité du matériel **46**

3 Methodes **46**

3.1 Les techniques d'enrichissement **46**
3.1.1 La capture 47
3.1.2 Le kit TruSeq Custom Amplicon 47
3.1.2.1 Le principe du kit 47
3.1.2.2 Le choix des gènes et des exons 48
3.1.2.3 Le Design Studio d'Illumina 48
3.1.3 La PCR multiplex 50
3.1.3.1 Principe de la technique 50
3.1.3.2 Le choix de la cohorte de patients 50
3.1.3.3 La recherche des mutations par les techniques « classiques » 51
3.1.3.4 Le design des amorces de NGS 55
3.1.3.5 La qualification des amorces de NGS 56
3.1.3.6 La génération des amplicons lors de la PCR 1 58
3.1.3.7 La quantification des produits de la PCR 1 58
3.1.3.8 La purification des produits de la PCR 1 59

3.2 Le NGS à proprement parlé **59**
3.2.1 L'intégration de l'index et des adaptateurs lors de la PCR 2 60
3.2.2 La purification et le dosage des produits de la PCR 2 60
3.2.3 La qualification des produits de PCR 2 sur puce 61
3.2.4 L'obtention de la librairie de séquençage 62
3.2.5 La PCR quantitative de contrôle de la librairie 62
3.2.6 Le séquençage sur le MiSeq 63
3.2.6.1 Le choix des paramètres du séquençage 63
3.2.6.2 Le logiciel Experiment Manager d'Illumina 64
3.2.6.3 La préparation des réactifs 65
3.2.6.4 Le chargement des réactifs et le lancement du run de séquençage 66

3.3 L'analyse des données de NGS **66**
3.3.1 L'étape préliminaire dans le MiSeq 66
3.3.2 Le logiciel de validation technique du run Sequencing Analysis Viewer 67

3.3.3 Le logiciel d'analyse MiSeq Reporter 68
3.3.4 La visualisation des données générées 69

VII RESULTATS 70

1 La validation technique 70

2 L'Analyse générale des séquences des témoins polyclonaux 70

2.1 La couverture et la profondeur 70

2.2 L'influence de la température et du nombre de cycles 71

2.3 L'influence de la réalisation de mix communs 72

3 L'Analyse détaillée des séquences des patients 72

3.1 La sélection des données 72

3.2 L'analyse des données 73

VIII DISCUSSION 74

1 L'Analyse financière 74

2 La Mise en place d'un « workflow » 74

3 L'enjeu de l'accréditation 77

4 Les Perspectives 77

IX CONCLUSIONS 78

1 L'application du NGS à l'activité de routine 78

2 L'évaluation possible de la maladie résiduelle 79

X ANNEXES 81

XI REFERENCES BIBLIOGRAPHIQUES 93

XIII LISTE DES FIGURES 97

XIV LISTE DES TABLEAUX 99

I TABLE DES ABREVIATIONS

ADN..... Acides Désoxyribonucléiques
ADNc ..... ADN complémentaire
AML1 ..... Acute Myeloid Leukemia 1 (=RUNX1 = CBFα)
ARN..... Acides Ribo-Nucléiques
ASXL1 ..... Additional Sex Combs Like 1
BCR/Abl..... Breakpoint Cluster Region/Abelson
C..... Cytosine
CALGB ..... Cancer and Leukemia Group B
CBFβ/MYH11 ..... Core Binding Factor β - Myosin Heavy chain 11
CBL ..... Casitas B-lineage lymphoma
CD ..... Cluster of Différenciation
CEBPa..... CCAAT/enhancer-binding protein alpha
CEBPβ..... CCAAT Enhancer Binding Protein Beta
CFU-B ..... Colony Forming Unit-Blast
CFU-E ..... CFU-Erythrocyte
CFU-Eo ..... CFU-Eosinophil
CFU-G ..... CFU-Granulocyte
CFU-GEMM..... CFU-Granulocyte Erythrocyte Macrophage Megakaryocyte
CFU-GM ..... CFU-Granulocyte Macrophage
CFU-L..... CFU-Lymphocyte
CFU-M..... CFU-Macrophage
CFU-MK ..... CFU-Megakaryocyte
CGH ..... Comparative Genomic Hybridization
CLP ..... Common Lymphocyte Progenitors
COG ..... Children's Oncology Group Study
CRCT ..... Centre de Recherches en Cancérologie de Toulouse
CSF ..... Colony Stimulating Factor
CSH..... Cellules Souches Hématopoïétique
CSV ..... Comma Separated Value
Ct..... Cycle threshold
DNMT3a ..... DNA (cytosine-5)-methyl-transferase 3 alpha
DNTP..... Désoxyribo Nucléotide Tri Phosphate
DDNTP ..... Di-Désoxyribo Nucléotide Tri Phosphate
E2A..... Immunoglobulin Enhancer binding factors E12/E47 (=TCF3)
EBSs ..... ETS Binding Sites
EPO..... Erythropoietine
ERG..... Erythroblastosis virus E26 oncogene homolog
ESG..... European Study Group
ETP ..... Early T-cell precursor
ETV6 ..... ETS Variant 6 (=TEL)
EZH2 ..... Histone-lysine N-methyltransferase
FAB ..... French American British
FC..... Flow cell
FISH ..... Fluorescence in Situ Hybridization
FLT3/ITD ..... FMS-like tyrosine kinase / Internal tandem duplication
G..... Guanine
GATA2..... GATA binding protein 2
G-CSF ..... Granulocyte CSF
GMAALL..... German Multicenter study group for Adult Acute Lymphoblastic Leukemia
GM-CSF ..... Granulocyte-Macrophage CSF
GRAALL ..... Group for Research in Adult Acute Lymphoblastic Leukemia
GTPase ..... Guanosine triphosphate hydrolase
HRM ..... High Resolution Melting
IDH1 et 2 ..... Isocitrate déshydrogénase 1 et 2
IG..... Immunoglobulin
IL ..... Interleukine
INCA..... Institut National du Cancer
INSERM ..... Institut National de la Santé et de la Recherche Médicale
JAK 2..... Janus Kinase 2
KIT..... Tyrosine-protein kinase Kit
KRAS..... v-Ki-ras2 Kirsten rat sarcoma viral
LAL..... Leucémie Aigüe Lymphoïde

LAM ..... Leucémie Aigüe Myéloïde
LLC..... Leucémie lymphoïde Chronique
LMC..... Leucémie Myéloïde Chronique
LMMC..... Leucémie Myélo-Monocytaire Chronique
M ..... Molaire
M-CSF ..... Macrophage CSF
MLL ..... Mixed Lineage Leukemia
MRD ..... Minimal Residual Disease
NGS..... New Generation Sequencing
NPM1 ..... Nucleophosmin 1
NRAS ..... Neuroblastoma RAS viral
OMS ..... Organisation Mondiale de la Santé
Pb..... Paire de Bases
PBL..... Peripheric Blood Lymphocytes
PBX1 ..... Pre-B cell leukemia transcription factor 1
PCR..... Polymerase Chain Reaction
PHF6 ..... (PHD)-like finger protein 6
PhiX..... Phage X (*Illumina*)
PML-RARα ..... Promyelocytic leukemia-Retinoic acid receptor alpha
PTEN..... Phosphatase and tensin homolog
RQ-PCR ..... Real Quantification PCR
RT-PCR..... Reverse Transcription PCR
RUNX1 ..... Runt-related Transcription factor 1 (=AML1)
SCF ..... Stem Cell Factor
SF3B1 ..... Splicing factor 3B subunit 1
SMD ..... Syndrome Myélo-Dysplasique
SMP..... Syndrome Myélo-Prolifératif
SMRT ..... Single Molecule Real Time
SNP ..... Single Nucleotide Polymorphism
SRSF2..... Serine/arginine-Rich Splicing Factor 2
STAT ..... Signal Transducer and Activator of Transcription
TCF3 ..... Transcription Factor 3 (=E2A)
TCR..... T Cell Receptor
TEL..... Translocated ETS Leukemia (= ETV6)
TET2..... Tet methylcytosine dioxygenase 2
TGF b ..... Transforming Growth Factor β
TNF a ..... Tumor Necrosis Factor α
TP53..... Tumor protein p53
TPO..... Thrombopoïétine
TP53..... Tumor Protein 53
U2AF1 ..... U2 small nuclear RNA Auxiliary Factor 1
UTR..... Untranslated region
WHO..... World Health Organisation
WT1..... Wilms tumor 1
ZMW..... Zero-Mode Waveguide

II INTRODUCTION

J'ai réalisé ce mémoire au sein de l'équipe de Biologie moléculaire appliquée aux Hémopathies malignes (secteur de Génétique des Hémopathies, laboratoire d'Hématologie, hôpital Purpan, Centre Hospitalier Universitaire de Toulouse).

La liste des bio marqueurs analysés en génétique somatique est amenée à s'accroître de manière significative à court et moyen terme en raison du nombre croissant de thérapies ciblées en cours de développement pour des groupes de patients définis en fonction des caractéristiques moléculaires de leur tumeur et du nombre croissant de marqueurs identifiés comme ayant un intérêt pronostique et permettant d'orienter la prise en charge des patients.

Cette augmentation régulière du nombre de tests à réaliser pour chaque patient conduit les plateformes à adopter des technologies permettant de multiplier les analyses, en particulier les nouvelles technologies de séquençage (NGS, acronyme anglais pour Next Generation Sequencing).

L'objectif principal de ce projet est de réaliser une analyse globale des mutations génétiques dans les Leucémies Aiguës Myéloïdes (LAM) et Syndromes Myélodysplasiques (SMD) et d'améliorer ainsi l'identification des patients ayant un fort risque d'évolutivité clinique ou présentant des mutations ciblées par des médicaments spécifiques.

Les LAM sont causées par un blocage et la prolifération d'une cellule souche transformée d'origine myéloïde (précurseurs des polynucléaires, des monocytes, des globules rouges et des plaquettes) ; cette prolifération conduit à une diminution des lignées hématopoïétiques produites normalement par la moelle.

Les SMD sont vus comme un état pré-leucémique, caractérisés par un maintien partiel de la différenciation et une forte mort intra-médullaire des précurseurs médullaires de la lignée myéloïde.

Notre objectif principal est critique pour une prise en charge correcte des patients car ces cancers sont associés à un mauvais pronostic vital malgré des traitements intensifs.

Les deux critères pronostics actuels admis pour la prise en charge clinique des LAM sont la présentation cytologique et les anomalies chromosomiques des cellules

leucémiques (caryotype), deux méthodes ayant démontré leur importance depuis plus de 50 ans.

Le développement récent de thérapies ciblées et la caractérisation moléculaire de nouvelles mutations récurrentes vont conduire à réévaluer progressivement la prise en charge de ces patients.

Ainsi, une récente étude américaine de Patel *et al [1]* incluant trois cent quatre-vingt-dix-huit patients de moins de 60 ans s'est intéressée aux mutations de dix-huit gènes qui, couplées à l'étude des anomalies cytogénétiques traditionnelles permettent d'affiner la stratification afin de mieux guider les décisions thérapeutiques.

De manière similaire, une étude portant sur deux-cent quatre-vingt-huit SMD a recherché l'impact de vingt-deux gènes dans la prise en charge de ces patients *[2]*.

III OBJECTIFS GENERAUX ET PROBLEME POSE

1 OBJECTIF PRINCIPAL

L'objectif de ce travail est de réorganiser le laboratoire de biologie moléculaire pour faire face aux nouvelles techniques de séquençage, qui modifie en profondeur le flux des analyses jusqu'ici réalisées (un échantillon avec une ou deux mutations à rechercher à un échantillon avec plusieurs dizaines de gènes à analyser dans le même délai), habituellement inférieur à 1 mois (durée de la période d'induction, basée sur une chimiothérapie associant 3 jours d'anthracyclines à 7 jours de cytarabine).

Ces analyses vont permettre de préciser la maladie permettant ainsi d'améliorer le pronostic et la prise en charge thérapeutique des patients.

Le NGS se développe et conduit à un très fort retentissement dans l'organisation de ces laboratoires.

Cette méthode contribue à diminuer le coût de séquençage par base (cependant le nombre de gènes étudiés étant fortement augmenté, le coût final risque d'augmenter notablement) et permet également d'augmenter la sensibilité de détection des mutations, permettant l'utilisation de cette technique sur des prélèvements non médullaires.

2 PLANNING

La première phase d'implantation du NGS pour le diagnostic des hémopathies malignes sera une évaluation de la faisabilité technique au sein du laboratoire pendant une période de six mois environ.

Durant cette période, nous évaluerons les différents aspects techniques :

- Détermination des régions d'intérêt (phases codantes, introns particuliers pour la duplication de *FLT3*, enhancer pour *GATA2*) et approche méthodologique pour enrichir ces régions (PCR, capture). Pour le choix des cibles, nous nous baserons sur les travaux récents cités précédemment cherchant à démontrer l'importance médicale de mutations retrouvées dans les LAM ainsi que dans les SMD. *[1] [2]*
- Les premiers tests d'évaluation seront réalisés rétrospectivement sur des échantillons pour lesquels le statut mutationnel de certains gènes est déjà

connu. Ces tests nous permettront de savoir si nous sommes capables de détecter les principaux types de mutations comme des substitutions, des insertions/délétions de différentes tailles ou des duplications. Nous pourrons également évaluer le pourcentage de faux-négatifs ou faux-positifs obtenus.

- Un autre aspect du développement de cette technique sera l'évaluation de la sensibilité de détection des anomalies. Ceci peut être évalué en réalisant une série de dilutions des échantillons d'ADN mutés dans de l'ADN non muté. Ce point nous permettra de savoir si nous nous pourrons détecter des mutations dans des sous-clones et si le NGS peut être utilisé pour l'évaluation rapide de la réponse de la maladie du patient à son traitement (maladie résiduelle).

La seconde phase de la mise en place du NGS sera une analyse prospective de nouveaux échantillons de diagnostic, réalisée en parallèle avec nos techniques classiques de criblage comme la PCR HRM (High Résolution Melting), le génotypage ou le séquençage classique de type Sanger.

Enfin après validation de la concordance de la technique de NGS avec nos techniques classiques, ces dernières seront abandonnées.

3 GENES CIBLES

Un profil moléculaire par NGS sera donc réalisé sur vingt-quatre gènes ayant un impact diagnostic, pronostic ou de thérapie ciblée démontré **(Tableau I)** : *ASXL1, CBL, CEBPA, DNMT3A, ETV6, EZH2, FLT3, IDH1, IDH2, JAK2, KIT, KRAS, NPM1, NRAS, PHF6, PTPN11, RUNX1, SETBP1, SF3B1, SRSF2, TET2, TP53, U2AF1* et *WT1*.

Après validation de ce premier panel, dans une seconde étape, nous rajouterons à la première version du panel de nouveaux gènes retrouvés mutés de façon récurrente dans les LAL (Leucémies Aiguës Lymphoïdes) comme *NOTCH1*, *FBXW7, PTEN, NRAS, KRAS* dans les LAL-T et *IKZF1* et *CRLF2* dans les LAL-B.

Enfin la dernière version du panel rajoutera les anomalies génétiques retrouvées de manière récurrente dans les Leucémies Lymphoïdes Chroniques (LLC).

Malgré des profils légèrement différents de mutations entre ces maladies, nous souhaitons développer un nombre restreint de panels car de nombreux gènes sont mutés de manière récurrente dans ces maladies (*TP53, NOTCH1*), les gènes de panel classiquement non mutés dans les autres maladies étudiés serviront de contrôle

Tableau I : Comparaison de la fréquence relative des mutations dans les LAM et MDS (Aajit Rampal et Ross L. Levine)

Frequency ↑	AML	MDS
	FLT3-ITD	*TET2*
	NPM1	*SF3B1*
	DNMT3A	*ASXL1*
	NRAS	*U2AF1*
	CEBPA	*SRSF2*
	TET2	*DNMT3A*
	WT1	*RUNX1*
	IDH2	*EZH2*
	IDH1	*JAK2*
	KIT	*NRAS*
	RUNX1	*TP53*
	MLL-PTD	*ETV6*
	ASXL1	*CBL*
	PHF6	*NPM1*
	KRAS	*IDH1*

négatif et enfin l'augmentation du nombre de patients permettra un flux régulier de patients permettant de conserver un débit suffisant compatible avec la prise en charge clinique.

Par la suite, chaque version évoluera en fonction des nouvelles données moléculaires.

4 ACCREDITATION DES ANALYSES BIOLOGIQUES

Notre projet doit également répondre à une politique d'accréditation des laboratoires (NF ISO 15189), répondant à des exigences en terme de qualité (protocoles bien définis ; contrôles de qualité internes et externes ; validation des méthodes).

Ainsi, je présenterai dans un premier temps la partie bibliographique basée sur les facteurs régulant l'hématopoïèse et les dysfonctionnements entrainant le développement des pathologies tels que les LAM et SMD. Puis je présenterai les techniques de NGS qui nous permettraient de mettre en évidence ces mutations afin de remplacer les techniques classiques de PCR et de HRM qui, couplées au séquençage Sanger conduisent actuellement à la détection de ces anomalies et enfin les gènes d'intérêt qui serviront de base à notre étude.

IV RECHERCHES BIBLIOGRAPHIQUES

1 LES LEUCEMIES AIGÜES MYELOÏDES ET LES SYNDROMES MYELODYSPLASIQUES

1.1 L'Hématopoïèse

1.1.1 Les cellules hématopoïétiques

L'hématopoïèse est le processus de formation des cellules sanguines tout au long de la vie d'un individu ; elle assure le renouvellement continu et régulé des différentes cellules sanguines.

Son siège varie durant l'embryogenèse : lors de la vie intra-utérine, l'hématopoïèse s'effectue au niveau du tissu conjonctif embryonnaire puis au niveau hépatique et splénique. L'hématopoïèse devient médullaire au début du développement des ébauches osseuses, pour se poursuivre exclusivement, après la naissance, dans la moelle osseuse.

Les cellules souches hématopoïétiques (CSH) sont capables d'auto-renouvellement et de différenciation en fonction du type de stimulation qu'elles reçoivent de de l'environnement (facteurs extrinsèques) comme les cytokines dont une des plus connues est l'érythropoïétine (EPO). Ces facteurs extrinsèques vont se fixer sur des récepteurs cibles (le récepteur de l'érythropoïétine dans notre exemple) qui vont conduire à l'activation de facteurs intrinsèques à la cellule hématopoïétique comme les facteurs de transcription (cascade JAK/STAT dans notre exemple).

On distingue quatre compartiments dans le système hématopoïétique en fonction de la maturation des cellules **(Figure 1)** :

- Le compartiment des cellules souches multipotentes dont les cellules sont quiescentes ; ces cellules possèdent un potentiel multipotent, de différenciation vers toutes les lignées hématopoïétiques;
- Le compartiment des progéniteurs hématopoïétiques dont les cellules sont engagées dans une ou plusieurs voies de différenciation et qui ont une capacité d'auto-renouvèlement réduite ; elles deviennent pluripotentes ;
- Le compartiment des précurseurs hématopoïétiques, dont les cellules sont morphologiquement reconnaissables et engagées dans une voie de différenciation unique; elles sont capables d'un nombre limité de mitoses ;

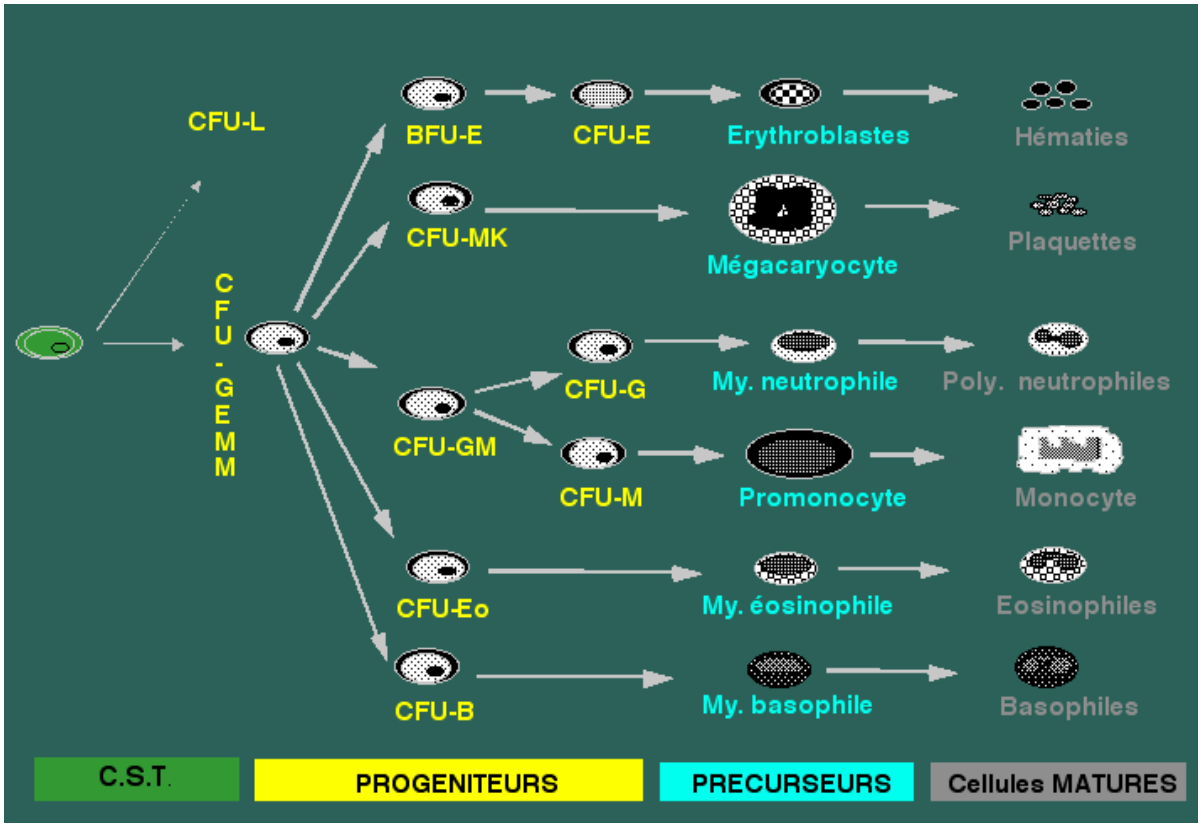


Figure 1 : Différenciation des cellules souches hématopoïétiques (Christian Binet, Faculté de Médecine de Tours)

BFU-E .....Burst Forming Unit-Erythrocyte

CFU-B .....Colony Forming Unit-Blast

CFU-E .....CFU-Erythrocyte

CFU-Eo .....CFU-Eosinophil

CFU-G.....CFU-Granulocyte

CFU-GEMM.....CFU-Granulocyte Erythrocyte Macrophage Megakaryocyte

CFU-GM.....CFU-Granulocyte Macrophage

CFU-L.....CFU-Lymphocyte

CFU-M.....CFU-Macrophage

CFU-MK .....CFU-Megakaryocyte

CST.....Cellules souche totipotentes

- Le compartiment des cellules matures fonctionnelles possédant le caractère spécifique d'une lignée donnée et pouvant passer dans le sang.

1.1.2 Les cytokines

Les cytokines, facteurs de croissance hématopoïétiques, sont des molécules cruciales pour la survie, la différenciation, la multiplication et la maturation des cellules hématopoïétiques.

Ce sont des glycoprotéines agissant comme des « hormones hématopoïétiques ». Ces dernières reconnaissent leurs cellules cibles par l'intermédiaire de récepteurs membranaires spécifiques.

Ces facteurs sont synthétisés par un grand nombre de cellules, à l'exception de l'EPO (production spécifiquement rénale) : stroma médullaire telles que les cellules endothéliales ou les fibroblastes, et cellules hématopoïétiques telles que monocytes, macrophages ou lymphocytes.

1.1.2.1 Les facteurs activateurs

Chez l'Homme, on distingue schématiquement trois types de facteurs de croissance selon leur lieu d'action au cours de l'hématopoïèse.

Les facteurs de promotion interviennent à des stades précoces de l'hématopoïèse et permettent l'engagement des cellules souches dans le cycle cellulaire en sensibilisant les CSH à l'action des autres facteurs de croissance. Ce sont principalement l'interleukine 6 (IL-6), le SCF (Stem Cell Factor) et le ligand du récepteur FLT3 (FMS-Like Tyrosine kinase 3).

Les facteurs multipotents permettent la survie et la différenciation des cellules souches les plus immatures lorsque celles-ci ont été sensibilisées par les facteurs de promotion. Ce sont principalement l'interleukine 3 (IL-3) et le GM-CSF (lignée granuleuse et macrophagique) intervenant sur les cellules souches et les progéniteurs précoces. **(Figure 2)**

Les facteurs restreints quant à eux, agissent sur les cellules souches engagées et favorisent la multiplication cellulaire et la maturation des précurseurs, cet effet étant limité à une lignée cellulaire **(Figure 3)**. Ce sont principalement le G-CSF (lignée granuleuse neutrophile), le M-CSF (lignée monocytaire), l'interleukine 5 (IL-5, lignée granuleuse éosinophile), l'interleukine 4 (IL-4, lignée granuleuse basophile), l'IL-6

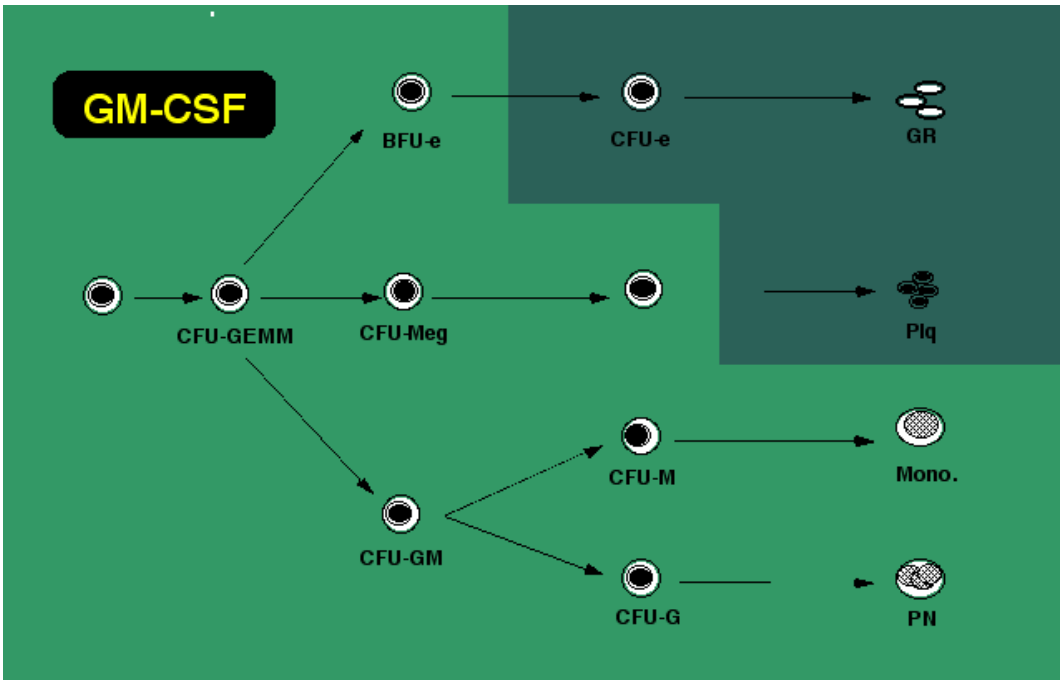


Figure 2 : Action des facteurs multipotents sur les CSH (Christian Binet, Faculté de Médecine de Tours)

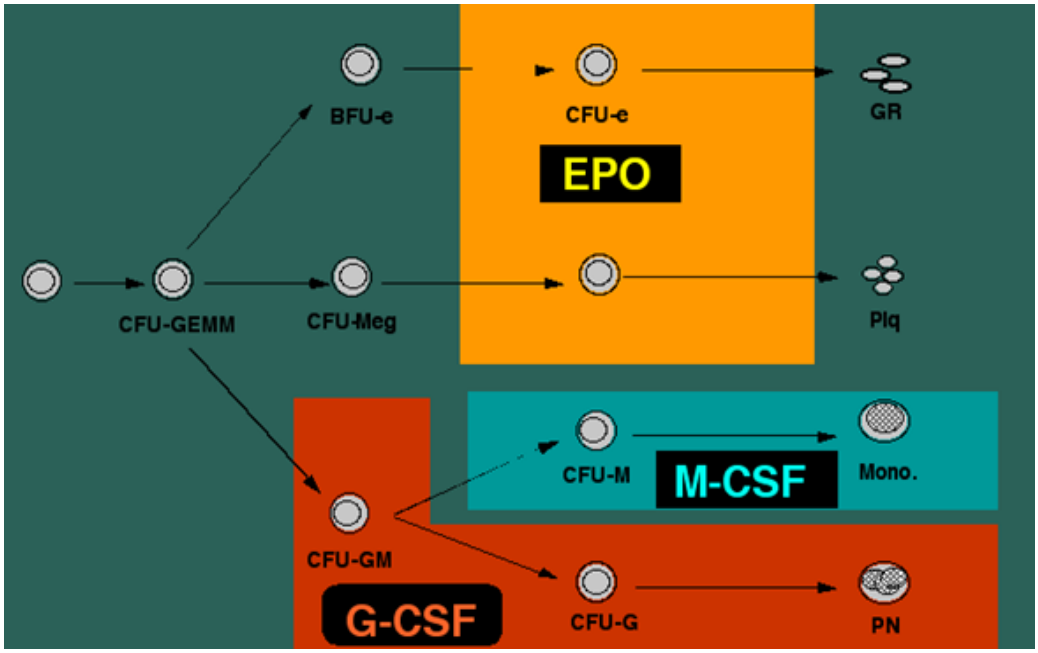


Figure 3 : Action des facteurs restreints sur les CSH (Christian Binet, Faculté de Médecine de Tours)

G-CSF.....Granulocyte CSF

M-CSF.....Macrophage CSF

PN.....Polynucléaire

Plq.....Plaquettes

GR.....Globule Rouge

(lignée mégacaryocytaire), l'EPO (lignée érythroïde) et la thrombopoïétine (TPO, mégacaryocytaire).

1.1.2.2 Les facteurs répresseurs

D'autres facteurs ont un effet inhibiteur sur l'hématopoïèse.

Ainsi, le Transforming Growth factor β (TGFβ) est un facteur de croissance inhibant la croissance des progéniteurs précoces et le Tumor Necrosis Factor α (TNFα) est produit par les monocytes et les lymphocytes T, il agit de manière différente suivant les facteurs de croissance associés.

Les interférons sont des cytokines essentielles à la communication entre les cellules immunitaires, produits par de nombreuses cellules dont les lymphocytes, notamment lorsqu'elles sont attaquées par un virus. Ils ont des effets antimitotiques et activent les macrophages et la phagocytose.

1.1.3 Les facteurs de transcription

La mise en place des programmes au sein des cellules hématopoïétiques, fait intervenir principalement les facteurs de transcription qui vont contrôler le programme génique de chaque lignée.

1.2 Les Leucémies Aigües Myéloïdes

1.2.1 Leur définition et épidémiologie

Les leucémies aiguës myéloïdes correspondent à des proliférations malignes aboutissant à l'accumulation dans la moelle, voire dans le sang, de précurseurs des cellules sanguines, appelés blastes, de nature myéloïde associée à un blocage à un stade précoce de leur différenciation. **(Figure 4)**

L'incidence globale des LAM est d'environ 3,5 nouveaux cas pour 100 000 habitants et par an. Cette incidence augmente progressivement avec l'âge, surtout après 50 ans.

La prévalence des LAM est en progression constante en raison de l'allongement de l'espérance de vie *[3]*. Plusieurs facteurs sociaux et/ou environnementaux semblent influencer le processus de leucémogenèse tels que :

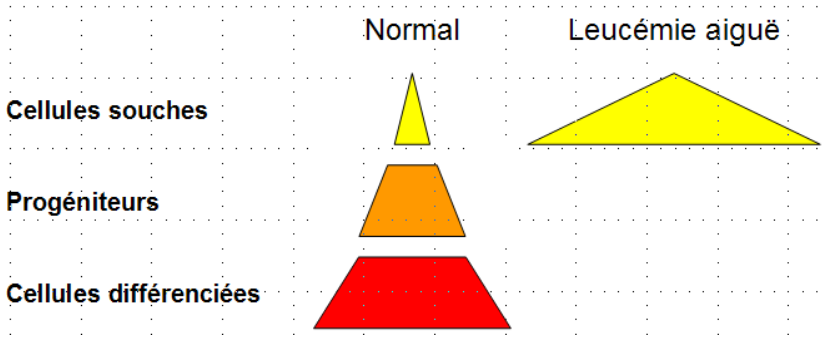


Figure 4 : Dérégulation de la différenciation cellulaire (Eric Delabesse).

- Les facteurs génétiques comme la trisomie 21, l'anémie de Fanconi ou le syndrome de Bloom ;
- L'exposition chronique à des toxiques comme le benzène ou les radiations ;
- La chimiothérapie anticancéreuse après exposition aux agents alkylants et aux inhibiteurs de l'activité topo-isomérase II ;
- Les hémopathies malignes comme les Syndromes Myéloprolifératifs (SMP) ou myélodysplasiques pouvant évoluer en LAM.

Les facteurs pronostics les plus importants des LAM sont l'âge *[4]* et l'existence d'anomalies chromosomiques.

(Voir paragraphe 1.2.2)

1.2.2 Les mécanismes oncogénétiques

L'étude des mécanismes moléculaires aboutissant à la transformation leucémique a permis d'identifier un processus en plusieurs étapes nécessitant la coopération d'au moins deux événements oncogéniques, un touchant la transcription (conduisant à un blocage de différenciation) et un la signalisation intracellulaire (conduisant à une prolifération accrue).

Il existe des mutations qui remanient des facteurs de transcription et induisent un blocage de la différenciation cellulaire. C'est le cas de la plupart des translocations chromosomiques récurrentes observées dans les LAM comme la translocation t(8;21)(q22;q22) conduisant à la mutation du facteur de transcription RUNX1 (Runt-related Transcription factor 1) dans la fusion RUNX1-RUNX1T1 (aussi connu comme AML1-ETO), l'inversion inv16(p13;q22) ou la translocation t(16;16)(p13;q22) conduisant à la mutation du facteur de transcription CBFβ dans la fusion CBFB-MYH11 (Core Binding Factor β - Myosin Heavy chain 11), la translocation t(15;17)(q22;q21) conduisant à la mutation du facteur de transcription RARA dans la fusion PML-RARA (Promyelocytic leukemia-Retinoic acid receptor alpha).

Ces mutations de facteur de transcription peuvent également se faire sous forme de mutations ponctuelles comme celles affectant *CEBPα* (CCAAT enhancer-binding protein alpha) *[5]*, *RUNX1* ou *GATA2* (GATA binding protein 2) *[6]*.

Des mutations de molécules importantes dans la signalisation intracellulaire sont retrouvées de manière récurrente dans les LAM touchant des récepteurs de

tyrosine kinase ou des enzymes clés de la signalisation intracellulaire. C'est le cas des récepteurs à activité tyrosine kinase membranaires comme *FLT3* (FMS-like tyrosine kinase 3), muté et constitutivement actif dans environ 25% des LAM, ou *KIT* (Tyrosine-protein kinase Kit) et des GTPases comme *NRAS* (Neuroblastoma RAS viral) et *KRAS* (Kirsten rat sarcoma viral).

L'amélioration des connaissances de ces maladies en particulier le séquençage du génome complet a permis de détailler les mutations des LAM, en identifiant des voies supplémentaires d'oncogenèse comme la méthylation de l'ADN (fusions MLL, mutations des gènes *IDH1, IDH2, DNMT3A* ou *TET2*).

Dans une étude récente comparant des LAM de type M1 (monoblastique, pour laquelle l'événement initiateur de la leucémie n'est pas connu) à des LAM de type M3 (promyélocytaire, pour laquelle l'événement initiateur est la fusion PML-RARA), parmi les cent-huit génomes étudiés, vingt-trois gènes comportaient des mutations récurrentes avec une moyenne de 2,2 mutations par génome dans les M1 (0 à 10) et 1,7 mutation par génome dans les M3 (1 à 4, y compris PML-RARA) *[7]*. Ce dernier résultat suggère que le nombre d'événements nécessaires à la transformation est limité, pouvant donc être évalué par un criblage limité.

1.2.3 La présentation clinique

La présentation clinique est fréquemment non spécifique et caractérisée surtout par l'insuffisance médullaire du fait de l'envahissement de la moelle par les blastes, avec des signes caractéristiques :

- D'insuffisance médullaire
  - Syndrome anémique (pâleur, tachycardie) ;
  - Syndrome infectieux (fièvre isolée ou associée à un point d'appel clinique) ;
  - Syndrome hémorragique (hémorragies des gencives, des muqueuses ou des tissus sous-cutanés).
- De manifestations tumorales
  - Hypertrophie des organes hématopoïétiques (hépatomégalie, splénomégalie ou adénopathie) ;
  - Lésions cutanées ;

- Hypertrophie gingivale (particulièrement dans les LAM 4 et 5) ;
- Atteinte osseuse peu fréquente.

1.2.4 Le diagnostic biologique

Le diagnostic biologique est habituellement simple *[8]*.

1.2.4.1 La numération et formule sanguine

Elle est toujours anormale et représente l'examen d'orientation majeur du diagnostic à partir duquel il est possible de noter :

- Une anémie presque constante, parfois sévère;
- Une thrombopénie très fréquente, parfois < 10G/L ;
- Une leucocytose variable allant de la leucopénie à l'hyperleucocytose (> 100G/L) ;
- Une neutropénie fréquente ;
- Des blastes circulant pouvant représenter l'ensemble des leucocytes mais étant parfois absents ou rares. **(Figure 5)**

1.2.4.2 Le myélogramme

Cet examen de la moelle osseuse par aspiration médullaire est l'examen clé du diagnostic, permettant d'identifier la blastose qui par définition est supérieure à 20%. Il est indispensable même s'il existe des blastes circulants et va permettre d'affirmer le diagnostic et de typer la leucémie *[8]*.

Les blastes contiennent généralement quelques granulations et parfois un ou plusieurs bâtonnets rouges (azurophiles) appelés corps d'Auer. **(Figure 6)**

1.2.4.3 L'immunophénotypage

Cet examen consiste en la recherche de marqueurs protéiques spécifiques exprimés soit à la surface des cellules leucémiques, soit à l'intérieure.

Ces marqueurs sont des protéines de la lignée myéloïde (CD13, CD33, CD117), des marqueurs d'immaturité (CD34) ou des marqueurs spécifiques de lignée (CD61 par exemple spécifique de la lignée mégacaryocytaire).

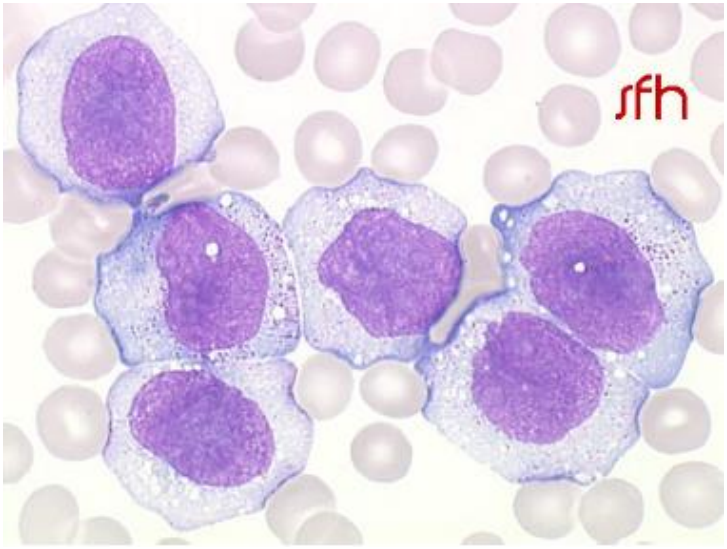


Figure 5 : Blastes typique d'une LAM5 visibles sur le frottis sanguin secondaire (Société Française d'Hématologie)

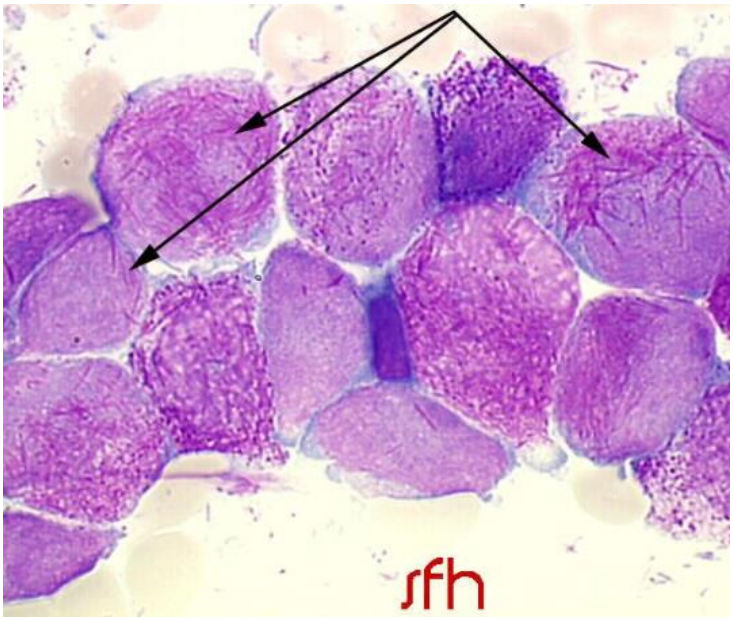


Figure 6 :

(Société Française d'Hématologie)

1.2.4.4 Le caryotype

Cet examen recherche les anomalies chromosomiques acquises des cellules leucémiques.

Du fait de l'ancienneté de cette technique (plus de 50 ans), les anomalies récurrentes sont bien connues et reliées à une entité homogène de LAM, permettant d'apprécier rapidement le pronostic de la maladie et donc la prise en charge thérapeutique du patient.

Les anomalies sont majoritairement acquises et non aléatoires ; elles sont fréquemment récurrentes (retrouvées dans plusieurs cas de LAM) et peuvent être spécifiques d'un type d'hémopathie.

On distingue les anomalies primaires (présentes dans toutes les cellules du clone leucémique) des anomalies secondaires (qui leur sont surajoutées et présentes dans un ou plusieurs sous-clones).

Une étude du caryotype portant sur cinq-mille huit-cent-soixante-seize patients âgés de 16 à 59 ans a permis d'identifier des pronostics précis pour des patients traités par des chimiothérapies conventionnels *[9]*, conduisant à une classification pronostique en trois groupes, un de pronostic favorable, un de pronostic défavorable et enfin un groupe de pronostic intermédiaire comprenant les cas non inclus dans les deux premiers ainsi que les caryotypes normaux : **(Tableau II)**

- Le groupe de pronostic favorable

  Les anomalies t(15;17)(q22;q21), t(8;21)(q22;q22) et inv(16)(p13q22)/t(16;16)(p13;q22) sont associés à un bon pronostic.

  - Le transcrit PML-RARA

    La LAM de type promyélocytaire (M3) est caractérisée par la translocation chromosomique t(15;17)(q22;q21) remaniant les gènes *PML* et *RARA* générant la protéine de fusion PML-RARA.

    Ce remaniement survient dans une cellule hématopoïétique déjà engagée dans le processus de maturation (promyélocytes).

Tableau II : Classifications cytogénétiques pronostiques anglaise et française des LAM (De Witte et Haferlach)

Groupe pronostique	Classification anglaise (MRC AML10)	Survie à 10 ans	Rechute à 10 ans	Classification GOELAMS	Survie à 5 ans	DFS à 5 ans
Bon pronostic	t(8;21) inv(16) t(15;17)	66 %	35 %	t(8;21) inv(16) t(15;17)	62,5 %	58 %
Pronostic intermédiaire	caryotype normal autres anomalies, y compris les anomalies en 11q23	41 %	51 %	caryotype normal autres anomalies, y compris t(9;11)	50 %	48 %
Mauvais pronostic	anomalies 3q -7 del5q/-5 t(9;22) t(6;9) caryotype complexe (>5 anomalies)	14 %	76 %	anomalies 3q del7q /-7 del5q /-5 t(9;22) t(6;9) caryotype complexe (≥ 3 anomalies) anomalies en 11q23 sauf t(9;11)	19 %	11,6 %

PML-RARA agit comme un dominant négatif sur la transcription des gènes contrôlés par RARA, notamment via le recrutement aberrant de complexes corépresseurs nucléaires (NCoR) *[10]*.

L'ATRA (Acide Tout Trans Rétinoïque) agit directement sur l'oncoprotéine via le gène RARα, ce qui induit le relargage du NCoR, puis la dégradation de l'oncoprotéine et à terme la levée du blocage de maturation.

Ce sous-groupe représente le premier exemple de traitement ciblé dans les LAM ; en effet, l'introduction de l'ATRA permet, en association avec la chimiothérapie, d'obtenir virtuellement 100% de rémission complète et une survie sans rechute d'environ 80 à 90% à 5 ans après consolidation par chimiothérapie exclusive et traitement d'entretien.

- Le groupe CBF

Le groupe CBF est défini par la présence dans les blastes des réarrangements chromosomiques t(8;21)(q22;q22), inv16(p13;q22) ou t(16;16)(p13;q22), présentes dans 10 à 15% des LAM du sujet jeune et dans moins de 5% des LAM du sujet âgé.

Elles sont communément regroupées sous le terme de CBF (Core Binding Factor Leukemia) car ces réarrangements affectent les deux sous unités du facteur de transcription CBF, interférant ainsi sur le processus de différenciation hématopoïétique.

Le complexe CBF est un facteur de transcription hétéro-dimérique composé d'une sous unité CBFα liant l'ADN au niveau des régions régulatrices de type « enhancer » et d'une sous unité unique CBFβ ne liant pas l'ADN. Chez l'homme, il existe trois gènes qui codent pour la sous unité *α* (CBFA1, CBFA2, CBFA3) et un qui code pour la sous unité *β* (CBFB) *[11]*.

➢ Le groupe de pronostic défavorable

D'autres anomalies comme l'inv(3)(q21q26), la t(3;3)(q21;q26), l'add(5q)/del(5q), , l'add(7q)/del(7q), les anomalies de la bande chromosomique 11q23 (à l'exception des t(9;11)(p21-22;q23) et le t(11;19)(q23;p13), les t(9;22)(q34;q11), les monosomies des chromosomes 5, 7 ou 17, sont associés à un mauvais pronostic.

Les patients ayant quatre anomalies non reliées ou plus ont également un mauvais pronostic (caryotype complexe).

Ce groupe représente 20 à 25% des sujets jeunes et 25 à 30% des sujets de plus de 60 ans.

Le pronostic est catastrophique quel que soit l'âge et les modalités thérapeutiques puisque même l'allogreffe n'a pas démontré un intérêt majeur dans ce sous-groupe ; en effet, le taux de rémission complète est significativement plus faible et la durée de rémission complète plus courte.

- Le groupe de pronostic intermédiaire

La majorité des caryotypes de LAM restent cependant sans anomalies cytogénétiques identifiées (caryotype dit normal) et ont un pronostic intermédiaire entre les bons et mauvais pronostics *[12]*.

Ce groupe représente 65% des LAM et reste très hétérogène ; il est composé à 40 - 45% par des patients présentant un caryotype normal et à 10 - 15% par des sujets présentant des anomalies chromosomiques qui n'ont pas de valeur pronostique péjorative ou favorable.

Des anomalies moléculaires permettent de mieux préciser le pronostic de ce sous-groupe ont été récemment décrites et contribuent déjà à adapter les traitements de consolidation comme les mutations de *FLT3*, *NPM1* (Nucléophosmine 1) et *CEBPα*.

(Voir paragraphe 1.2.4.6)

Les taux de rémissions complètes sont de l'ordre de 85% et, en fonction des séries et des traitements, la survie sans maladie à 3 ans est de 40 à 60% *[13]*.

Ainsi une stratification des patients en fonction de la cytogénétique permet de distinguer trois groupes de pronostics différents avec influence sur la survie sans rechute mais également la survie globale *[14]*.

1.2.4.5 La Fluorescence in Situ par Hybridation

A côté du caryotype classique cherchant à obtenir des métaphases de chromosomes de cellules en mitoses, l'analyse de type FISH (Fluorescent In Situ

Hybridization) permet de confirmer les remaniements de gènes de fusion dont la présence est parfois difficile à affirmer en cytogénétique conventionnelle.

1.2.4.6 La biologie moléculaire

- Les gènes de fusion

En complément de la cytogénétique conventionnelle et des techniques FISH, les techniques de biologie moléculaire peuvent caractériser la présence de gènes de fusions identifiés au caryotype permettant également d'obtenir un marqueur de maladie résiduelle au cours du traitement du patient.

- Les mutations ponctuelles

C'est le groupe des LAM à caryotype normal (moitié des LAM environ) qui a le plus bénéficié des avancées moléculaires récentes ajoutant un impact pronostique notable de ces LAM.

Les trois principales anomalies moléculaires actuellement recherchées sont les mutations des gènes *FLT3, NPM1 et CEBPα*. La recherche de ces anomalies permet actuellement de réserver les indications d'allogreffe des LAM à caryotype normal aux formes associant deux facteurs péjoratifs sur les trois (*NPM1-, FLT3+, CEBPα-*).

- Les mutations de *FLT3*

Le gène *FLT3* code pour un récepteur à activité tyrosine kinase localisé sur le13q12 du FLT3-ligand. Deux types de mutations ont été mis en évidence.

Une duplication interne du domaine juxta-membranaire (*FLT3-ITD*) *[15]* est retrouvée dans 24% des LAM dont un tiers ayant un caryotype normal *[16]* aboutissant à une activation constitutive de l'activité tyrosine kinase.

A côté de cette duplication, une substitution d'un acide aspartique par un résidu tyrosine le plus souvent en position 835 au niveau du domaine catalytique tyrosine kinase (*FLT3-TKD*) a été observée avec une incidence de 7% des LAM dont 14% avec un caryotype normal *[17]*.

Les mutations de *FLT3* sont associée à une valeur pronostique péjorative, en particulier celle de type ITD *[17]*.

- Les mutations de *NPM1*

Les mutations de l'exon 12 du gène *NPM1* consistent en une insertion de quatre bases qui conduit à un décalage du cadre de lecture et à la perte du signal de localisation nucléaire *[18]*. La protéine mutante a ainsi une localisation cytoplasmique aberrante ;

Ces mutations sont présentes dans environ 50% des LAM à caryotype normal et sont associées à un bon pronostic, indépendant de la survie globale et de la survie sans événement *[18]* en particulier les patients n'exprimant pas la mutation *FLT3*-ITD *[19]* et *[20].*

- Les mutations de *CEBPα*

Les mutations du gène *CEBPα* localisé en 19q13 sont retrouvées dans 9 % des LAM et associées à une meilleure survie sans rechute ou survie globale *[21]*.

Les mutations sont localisées soit du côté N, soit du côté C-terminal et peuvent être bi-alléliques.

Plusieurs études ont montré récemment que seule la présence de deux mutations de CEBPα était associée à un pronostic favorable, contrairement à la présence d'une seule mutation *[22] et [23]*.

- Les autres mutations

Le développement du NGS appliqué aux LAM a permis d'identifier de nouvelles mutations comme celles d'*IDH1* (Isocitrate déshydrogénase 1) *[24]*, *IDH2 [25]* et *DNMT3A [26]* pour les plus fréquentes. Ces mutations sont fréquemment associées, comme celles de *DNMT3A,* aux mutations de *FLT3-ITD* et *NPM1 [26]*.

1.2.5 La classification

Les classifications sont des tentatives de regroupements des cas de LAM afin de définir des groupes homogènes de maladie.

Une première tentative a été réalisée en 1976 sur des bases cytologiques et cytochimiques, la classification FAB (proposée par sept cytologistes franco-américano-britanniques) *[27]*. **(Tableau III)**

La caractérisation d'anomalies cytogénétiques récurrentes pronostiques a conduit cette classification a évolué vers une prise en compte de ces anomalies dans la classification récente de l'OMS (Organisation Mondiale de la Santé) *[8]*. **(Tableau IV)**

Au-delà des gènes de fusion identifiés par la cytogénétique, les mutations ponctuelles n'ont pas encore pris une place prépondérante dans les classifications du fait de leur manque de recul clinique.

1.2.6 Le traitement

Le traitement par chimiothérapie intensive d'un sujet atteint de LAM comporte deux phases : une phase d'induction de la rémission, suivie d'une phase de consolidation cherchant à maintenir une rémission prolongée.

1.2.6.1 La phase d'induction

Le traitement d'induction a pour objectif d'induire une rémission complète du malade, c'est-à-dire la disparition des blastes médullaires avec normalisation de l'hémogramme et éventuellement du syndrome tumoral initial.

Ce traitement repose toujours sur l'association d'une anthracycline pendant 3 jours et de la cytosine arabinoside (Ara C) pendant 7 jours.

1.2.6.2 La phase de consolidation

Après obtention de la rémission complète après un mois environ, le traitement de consolidation repose habituellement sur la chimiothérapie intensive (utilisation d'Ara C à forte dose pendant trois ou quatre cures) suivie ou non d'une greffe allogénique ou autologue en fonction du risque de rechute.

L'allogreffe de CSH reste le meilleur traitement pour la prévention des rechutes mais reste limitée par la disponibilité d'un donneur HLA compatible ainsi que par l'âge du sujet.

Tableau III : Classification FAB des LAM (Bennett JM et al.)

Type FAB	Nom	Proportion
LAM 0	LAM indifférenciée	3 à 5 %
LAM 1	LAM peu différenciée	15 à 20 %
LAM 2	LAM avec différenciation	25 à 30 %
LAM 3	LA promyélocytaire	5 à 10 %
LAM 4	LAM myélo-monocytaire	20 à 30 %
LAM 5	LA monoblastique 5a/5b = sans/avec différenciation	10 %
LAM 6	LA érythroblastique	3 à 5 %
LAM 7	LA mégacaryoblastique	5 %

Tableau IV : Révision 2008 de la classification OMS des leucemies aigues et neoplasies associees (Vey N et al.)

LAM avec anomalies cytogénétiques récurrentes	LAM avec t(8;21)(q22;q22) ; *RUNX1–RUNX1T1* LAM avec inv(16)(p13.1q22) ou t(16;16)(p13.1q22) ; *CBFB–MYH11* LAP avec t(15;17)(q22;q12) ; *PML–RARA* LAM avec t(9;11)(p22;q23) ; *MLLT3–MLL* LAM avec t(6;9)(p23;q34) ; *DEK–NUP214* LAM avec inv(3)(q21q26.2) ou t(3;3)(q21;q26.2) ; *RPN1–EVI1* LAM (mégacaryocytaire) avec t(1;22)(p13;q13) ; *RBM15–MKL1* Entités à venir : LAM avec mutation de *NPM1*, LAM avec mutation de *CEBPA*
LAM avec critères de myélodysplasie	
LAM et MDS secondaires à des thérapeutiques	
LAM n'entrant pas dans les catégories précédentes	LAM avec différenciation minimale (M0-FAB) LAM sans maturation LAM avec maturation LAM avec différenciation myélo-monocytaire LAM monocytaire et monoblastique LAM érythroblastique (LAM érythrocytaire pure et érythroleucémie) LAM mégacaryoblastique LAM à différenciation basophile LAM avec myélofibrose
Sarcome granulocytaire	
Myéloproliférations associées à la trisomie 21	Myélopoïèse anormale transitoire Leucémie myéloïde associée à la trisomie 21
Leucémie à cellules dendritiques plasmacytoïdes	

1.3 Les Syndromes Myélodysplasiques

1.3.1 Leur définition et épidémiologie

Les SMD sont des hémopathies myéloïdes malignes du sujet âgé formant un groupe de pathologies extrêmement hétérogènes aux niveaux clinique et biologique *[8]*.

Ils sont caractérisés par la différenciation et la maturation anormale de cellules souches hématopoïétiques, entraînant une ou plusieurs cytopénies.

Les SMD ont une incidence de 3 à 5 cas pour 100 000 qui augmente avec l'âge (20 cas pour 100 000 chez les sujets de plus de 70 ans).

1.3.2 La présentation clinique

Dans 90 % des cas, les signes révélateurs reflètent l'anémie. On peut également observer des troubles hémorragiques liés à une thrombopénie profonde, ou de troubles infectieux liés à une neutropénie *[8]*.

1.3.3 Le diagnostic biologique

1.3.3.1 La numération et formule sanguine

Elle va révéler une ou des cytopénies ; l'anémie est macrocytaire ou normocytaire, normochrome.

La neutropénie et la thrombopénie peuvent aussi apparaître isolément.

L'examen du frottis peut révéler des signes de dysplasie comme la dégranulation de la lignée granuleuse *[8].*

1.3.3.2 Le myélogramme

Le myélogramme affirme le diagnostic et objective des anomalies morphologiques caractéristiques *[8].*

La moelle est de cellularité normale ou augmentée, contrastant avec les cytopénies périphériques, reflétant l'hématopoïèse inefficace.

Les anomalies morphologiques ont une valeur diagnostique variable ; la dysérythropoïèse a une faible valeur diagnostique, alors que la dysgranulopoïèse et la dysmégacaryopoïèse ont une valeur diagnostique plus importante. **(Figure 7)**

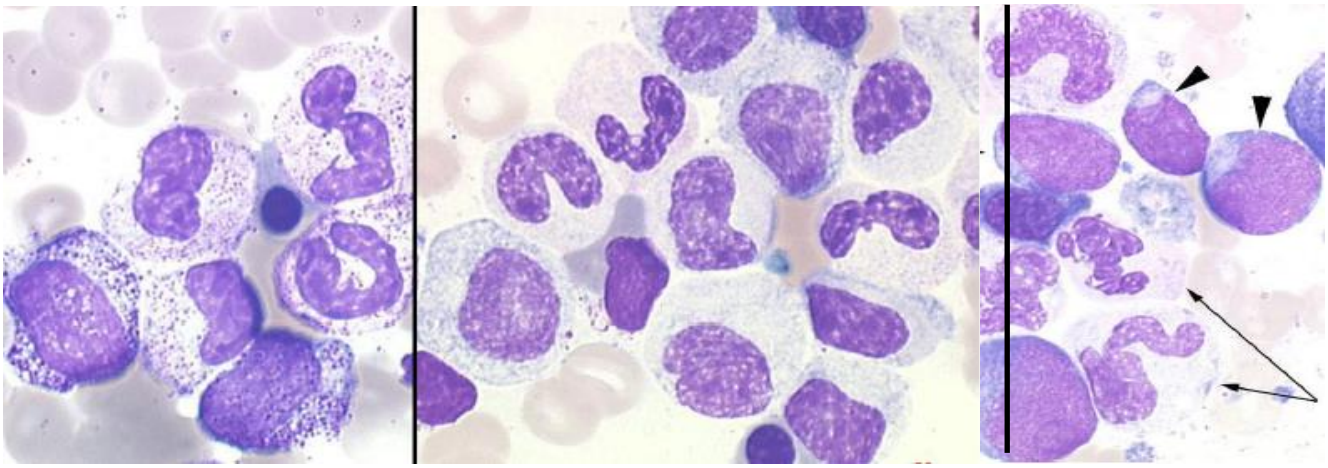


Figure 7 : Anomalies des cellules typiques de SMD visibles sur le frottis médullaire (SFH)

1.3.3.3 Le caryotype

La cytogénétique tient une place prépondérante dans l'évaluation des SMD au niveau du diagnostic et du pronostic.

Des anomalies cytogénétiques clonales sont observées dans environ la moitié des cas. Ainsi, la délétion du bras long du chromosome 5 ou del(5q) isolée est associée à la présence de mégacaryocytes hypo-lobés voire non lobés, une anémie réfractaire macrocytaire, un taux de plaquettes normal ou augmenté, et une évolution clinique favorable constituant une entité à part dans la classification OMS.

Les caryotypes complexes (au moins trois anomalies) sont généralement associés à une évolution clinique défavorable.

1.3.3.4 La biologie moléculaire

La biologie moléculaire n'est encore pas utilisée en routine pour aider à la prise en charge des patients au diagnostic de SMD. Cependant, des données récentes de séquençage à haut débit a permis d'identifier des mutations spécifiques des SMD comme les mutations des constituants du splicéosome *[28]* qui permettent là également d'avoir une indication de pronostic mais également d'affirmer le diagnostic de SMD.

Ainsi, l'impact de l'identification de mutations de vingt-deux gènes a été démontré dans deux-cent quatre-vingt-huit SMD à bas risque d'évolution selon l'IPSS (International Prognosis Scoring System).

(Voir paragraphe 1.3.4)

Les mutations d'*EZH2*, *RUNX1*, *TP53* et *ASXL1* étaient associées avec un pronostic péjoratif significatif *[2]*.

1.3.4 La classification

La classification OMS *[8]* représente un consensus sur les SMD et complète la classification FAB. **(Tableau V)**

Les critères pris en compte par cette classification sont :

- Le nombre de cytopénies périphériques ;
- Le nombre de lignées dysplasiques ;

Tableau V : Classification OMS – 2008 des SMD (Vardiman et al.)

Maladie	Caractéristiques sanguines	Caractéristiques médullaires
Cytopénie réfractaire avec dyplasie unilignée (CRDU) : Anémie réfractaire (AR) Neutropénie réfractaire (NR) Thrombopénie réfractaire (TR)	Cytopénie unique ou bicytopénie[1] Absence ou rares blastes (<1 %)[2]	Dysplasie unilignée : ≥ 10 % des cellules d'une lignée myéloïde < 5 % de blastes < 15 % de sidéroblastes en couronne
Anémie réfractaire avec sidéroblastes en couronne (ARS)	Anémie Absence de blastes	≥ 15 % de sidéroblastes en couronne Dysplasie de la lignée érythroblastique uniquement < 5 % de blastes
Cytopénie réfractaire avec dysplasie multilignée	Cytopénie(s) Absence ou rares blastes (< 1 %)[2] Pas de corps d'Auer Monocytes < 1 G/L	Dysplasie ≥ 10 % sur ≥ 2 lignées (neutrophile et/ou érythroblastique et/ou mégacaryocytaire) < 5 % de blastes Absence de corps d'Auer +/- 15 % de sidéroblastes en couronne
Anémie réfractaire avec excès de blastes de type 1 (AREB-1)	Cytopénie(s) < 5 % de blastes[2] Absence de corps d'Auer Monocytes < 1 G/L	Dysplasie uni- ou multilignée 5 à 9 % de blastes[2] Absence de corps d'Auer
Anémie réfractaire avec excès de blastes de type 2 (AREB-2)	Cytopénie(s) 5 à 19 % de blastes +/- présence de corps d'Auer[3] Monocytes < 1 G/L	Dysplasie uni- ou multilignée 10 à 19 % de blastes +/- présence corps d'Auer[3]
Syndrome myélodysplasique inclassable (SMD-I)	Cytopénies ≤ 1 % de blastes[2]	Dysplasie marquée mais < 10 % Présence d'une anomalie cytogénétique considérée comme présomptive du diagnostic d'un SMD < 5 % blastes
Syndrome myélodysplasique associé à la délétion 5q isolée (Syndrome 5q-)	Anémie Taux de plaquettes normal ou augmenté Absence ou rares blastes (≤ 1 %)	Nombre de mégacaryocytes normal à augmenté avec noyau hypolobé < 5 % de blastes Anomalie cytogénétique isolée : del(5q) Absence de corps d'Auer

[1] Une bicytopénie peut occasionnellement être observée. Les cas sans cytopénie doivent être classés en SMD-I.

[2] Si le pourcentage de myéloblastes est inférieur à 5 %, mais que 2 à 4 % de myéloblastes sont retrouvés dans le sang, la pathologie est alors classée en AREB-1. Les cas de CRDU et CRDM avec 1 % de myéloblastes dans le sang doivent être classés comme SMD-I

[3] Les cas avec corps d'Auer et < 5 % de myéloblastes dans le sang et < 10 % de myéloblastes dans la moelle doivent être classés en AREB-2.

- Le pourcentage de blastes circulants et médullaires ;
- La présence de sidéroblastes en couronne ;
- La présence éventuelle de corps d'Auer ;
- La présence d'une anomalie cytogénétique particulière.

Le caryotype est considéré comme un facteur pronostic indépendant dans les SMD.

Une autre classification dite IPSS est utilisée en pratique clinique pour la prise en charge thérapeutique des patients *[29]*.

Les anomalies chromosomiques y sont réparties en trois groupes dits, de bon pronostic, de pronostic intermédiaire ou de mauvais pronostic. Toutefois, en raison de la grande hétérogénéité des anomalies cytogénétiques, de la combinaison de ces anomalies entre elles et du nombre limité de patients à caryotypes anormaux de la cohorte ayant permis d'établir l'IPSS, seul l'impact pronostique des anomalies chromosomiques les plus fréquentes (-5/5q-, -7/7q-, +8, 20q-, -Y, et caryotypes complexes) a été établi de manière rigoureuse. **(Tableau VI)**

1.3.5 Le traitement

Le traitement est délicat du fait de l'âge avancé des patients atteints de SMD.

Il est d'abord symptomatique, visant à pallier aux cytopénies, et repose sur l'administration d'EPO ou de culots érythrocytaires. L'utilisation d'antibiothérapies à large spectre est réservée en cas d'infection suite à une neutropénie.

Un traitement chélateur du fer permet de lutter contre la surcharge en fer chez les patients régulièrement transfusés.

Dans les SMD de faible grade, lorsque les cytopénies restent modérées et asymptomatiques, l'abstention thérapeutique est de mise.

2 LES TECHNIQUES MOLECULAIRES DE RECHERCHE DE MUTATIONS

2.1 Le séquençage Sanger

Le séquençage de l'ADN consiste à déterminer l'ordre d'enchaînement des nucléotides pour un fragment d'ADN donné.

Tableau VI : Score IPSS (Germing et al.)

Score	0	0,5	1	1,5	2
Pourcentage de blastes médullaires	0-4	5-10	-	11-20	21-29
Nombre de cytopénies [1]	0-1	2-3	-	-	-
Groupe cytogénétique de risque	Bas	Intermédiaire	Haut	-	-
Groupe de risque	**Score**				
Bas risque	0				
Risque intermédiaire 1	0,5-1				
Risque intermédiaire 2	1,5-2				
Haut risque	≥ 2,5				
[1] Plaquettes < 100 G/L ; Hémoglobine < 10 g/dL ; Polynucléaires Neutrophiles < 1,8 G/L					

En 1977, Maxam et Gilbert publient une technique basée sur le marquage radioactif de fragments et leur coupure sélective par dégradation chimique, or en parallèle *[30]*, Sanger effectue une tout autre approche de synthèse enzymatique qui consiste à initier la polymérisation de l'ADN à l'aide d'un petit oligonucléotide ou amorce complémentaire à une partie du fragment d'ADN à séquencer *[31]*.

L'élongation de l'amorce est réalisée par une ADN polymérase dépourvue d'activité exo nucléase en présence d'un mélange des quatre désoxynucléotides (dATP / dCTP / dGTP / dTTP) et une faible concentration de quatre di-désoxynucléotides (ddATP / ddCTP / ddGTP / ddTTP), chacun associé à un marqueur fluorescent différent permettant le mélange des quatre ddNTP (initialement, le marquage était radioactif, la réaction comportant un seul ddNTP marqué par réaction). Une fois incorporés dans le nouveau brin synthétisé, ces di-désoxynucléotides empêchent la poursuite de l'élongation ; il en résulte un mélange de fragments d'ADN de taille croissante **(Figure 8)**. Ces fragments sont ensuite séparés par électrophorèse capillaire sur un gel de polyacrylamide, ce qui permet ainsi de lire l'enchainement des bases dans la séquence.

Cette technique dite de Sanger est devenu le standard permettant d'atteindre une sensibilité de l'ordre de 20% ; son coût est de l'ordre de 4€ par échantillon.

2.2 Le criblage de mutations par High Resolution Melting Curve PCR

Cette technique permet la détection de séquences variantes hétérozygotes. Elle repose sur l'amplification d'une région de 100 à 200pb par PCR suivie d'une courbe de dénaturation en temps réel en présence d'un fluorochrome intercalant spécifique de l'ADN double brin après formation d'hétéro-duplexes par dénaturation prolongée et refroidissement rapide. **(Figures 9 et 9bis)**

La courbe de fusion permet de différencier les amplicons strictement complémentaires des amplicons non appariés en totalité (hétérozygotes comportant une variation de séquence, mutation ou polymorphisme) se dénaturant plus rapidement.

En cas de possibles variations, le contenu du puits de PCR est prélevé et envoyé directement pour séquençage sans ré-amplification du produit.

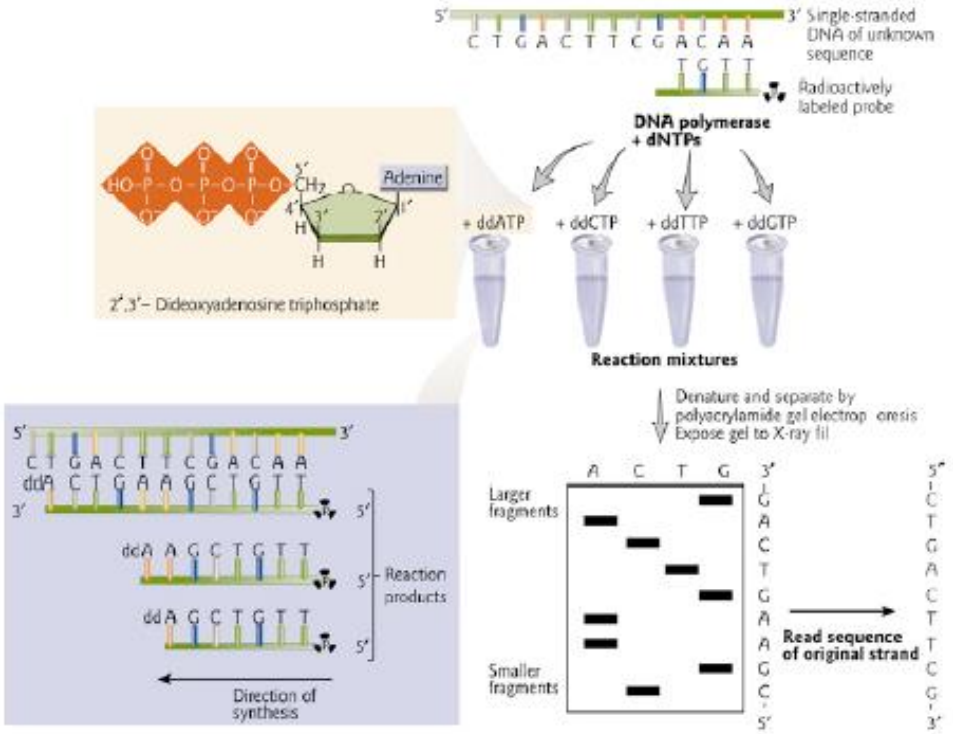
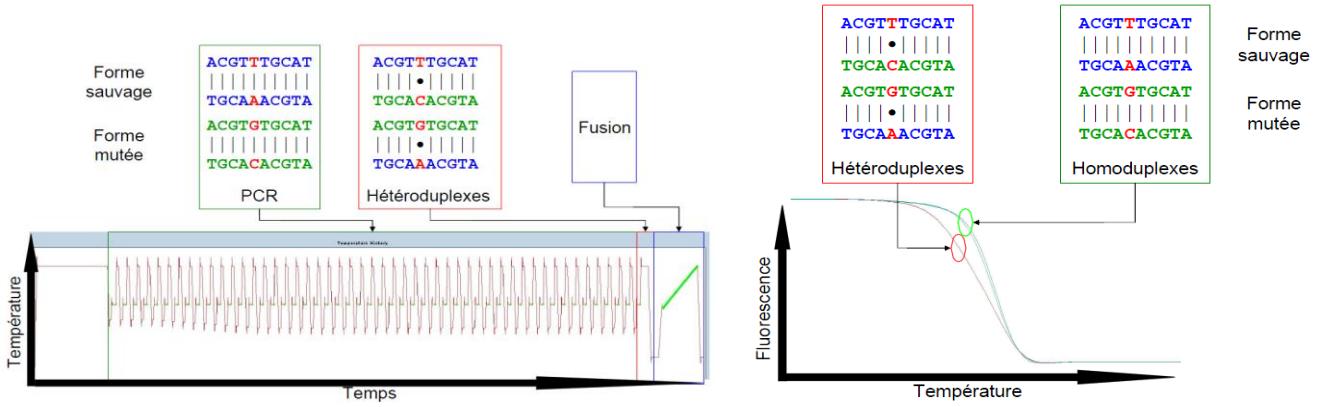


Figure 8 : Principe du séquençage enzymatique Sanger (Allison, Blackwell Publishing, 2007)



1. PCR et amplification de l'ADN d'intérêt
2. Dénaturation et ajout du marqueur fluorescent qui s'intercale entre les 2 brins d'ADN
3. Dénaturation et perte de la fluorescence des brins d'ADN simple.

Figure 9 : Principe du HRM – Dénaturation et courbe de fusion.

Figure 9bis: Principe du HRM – Mutation.

Cette technique permet de rechercher des mutations pour des coûts très faibles avec une sensibilité de l'ordre de 10%.

Cependant, si la technique est fiable pour des mutations bien définies dont la localisation sur un codon reste stable, la fiabilité est difficile à affirmer pour des gènes dont les mutations sont éparpillées comme pour *TP53*. De plus, cette technique s'avère difficile à mettre en œuvre pour le criblage d'un grand nombre d'exons (supérieur à dix).

2.3 Le séquençage de 2ème génération

Depuis 2005, l'émergence de nouvelles technologies de séquençage a permis de séquencer avec des débits qui évoluent de façon phénoménale *[32]*. **(Figure 10)**

Cette augmentation de débit est associée avec une diminution drastique des coûts en approchant de l'objectif du génome à 1000 $ *[33]* même si l'inclusion des coûts des réactifs et de la main d'œuvre, l'amortissement des instruments, le stockage et le management des données conduit un cout global bien supérieur *[34]*. **(Figure 11)**

Ainsi, ces nouvelles techniques permettent de combiner :

- Re-séquençage de génomes entiers (Whole Genome Seq) ou de régions spécifiques (comme la totalité des exons constituant 44Mb chez l'homme, ou Exome sequencing) afin de détecter des changements par rapport à une séquence de référence ;
- Génotypage : identification de polymorphismes SNP (Single Nucleotide Polymorphism) ;
- Analyse du niveau d'expression et identification d'ARN de fusion (RNA-seq) ;
- Identification des régions immuno-précipités (ChIP-Seq) ;
- Séquençage *de novo* pour la caractérisation de nouveaux génomes ;
- Re-séquençage de régions ciblées préalablement capturées tel que le séquençage d'exomes ;
- Méta génomique afin de caractériser l'ensemble des génomes dans un mélange de populations et de communautés.

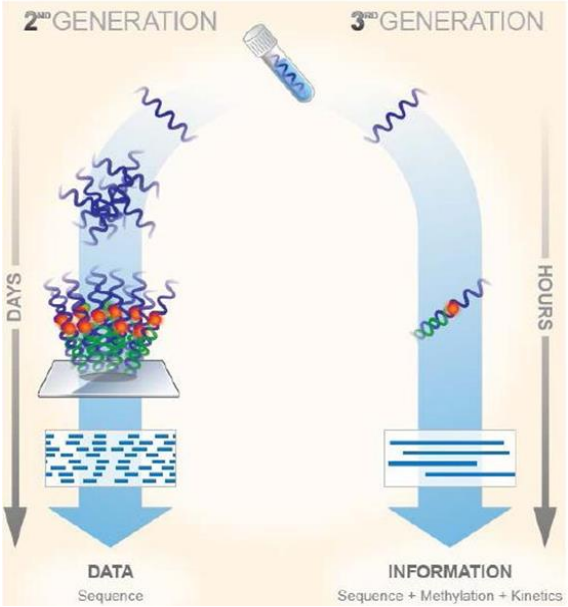
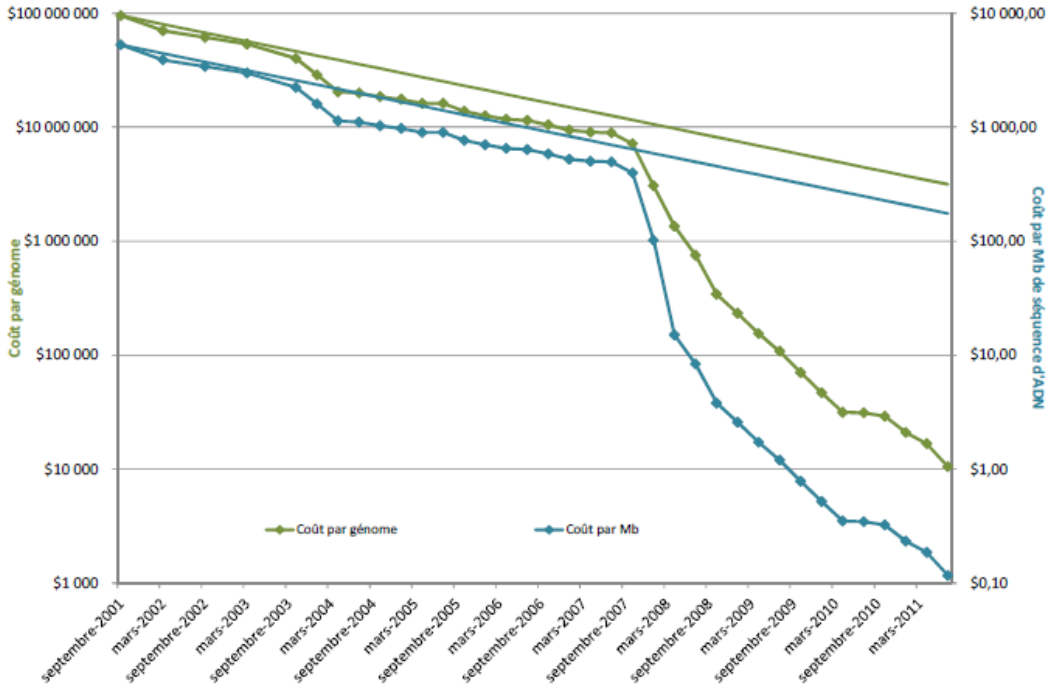


Figure 10 : Amélioration des techniques de séquençage de nouvelle génération (*Biorigami*).



En vert : le cout du séquençage d'un génome de la taille du génome humain ;
En bleu : le cout du séquençage d'1 million de bases.

Les profondeurs utilisées sont les suivantes : 6x pour du séquençage Sanger, 10x pour le 454, 10x pour Illumina et le SOLiD.
Les droites représentent l'évolution théorique selon la loi de Moore pour comparaison.
Le décrochage à partir du 1er trimestre 2008 coïncide avec la période à laquelle les centres de séquençage se sont dotés de séquenceurs de 2nde génération.
Ces couts englobent la main-d'œuvre, les réactifs et consommables, les instruments et leur amortissement sur 3 ans, un service informatique minimum, la construction des librairies et la soumission des données à des bases de données publiques. Ils ne prennent pas en compte le développement technologique, le contrôle-qualité, le matériel et le développement des outils informatiques ni l'analyse complète des données.

Figure 11 : Evolutions du prix du séquençage (www.genome.gov)

Ces nouveaux outils permettront de faire à la fois la détection de nouvelles mutations pressenties au diagnostic et à la rechute, mais également aider au suivi moléculaire spécifique de ces anomalies.

Les quatre technologies de séquençage qui dominent actuellement le marché *[35]* se démarquent les unes des autres par les chimies qui les constituent **(Tableau VII)**. Elles se décomposent cependant toutes en quatre grandes étapes principales :

- La préparation des librairies qui contient une étape d'amplification par PCR ;
- Les cycles de réactions de séquençage ;
- La prise d'image après chacun de ces cycles pour déterminer le nucléotide correspondant ;
- L'analyse des données.

Ces nouvelles générations de machines ont pour avantage leur capacité à analyser de grands génomes à haute résolution grâce à la « parallélisation » des réactions.

2.3.1 La technologie 454

Cette technique de pyro-séquençage commercialisée par *Roche* permet d'obtenir jusqu'à 900 Mb de données en une dizaine d'heures, soit quinze mille fois plus qu'avec la première génération de séquenceurs.

Elle a été utilisée récemment dans une étude des altérations moléculaires de gènes (mutations et amplifications) *[36]*.

La spécificité de cette technologie repose sur une PCR en émulsion pour l'amplification des fragments à séquencer : la PCR a lieu dans une microgoutte renfermant une microbille d'agarose en phase aqueuse, séparée des autres billes (plusieurs millions) par de l'huile. On obtient ainsi des copies d'un seul fragment d'ADN par bille.

Chacune des billes est ensuite déposée dans un des 1,6 millions de puits d'un support solide appelé *PTP* (PicoTiter Plate). Des réactions de pyro-séquençage ont alors lieu à l'échelle du pico litre dans chaque puits : un flux de nucléotides (chaque nucléotide l'un après l'autre) traverse la PTP et lorsque l'un d'entre eux est incorporé

Tableau VII : Caractéristiques des technologies de séquençage (Glenn 2011)

	1ère génération	2e génération													
Plateforme	3730	454			Illumina					SOLiD			Ion Torrent		
Entreprise actuelle	Life Technologies	Roche			Illumina					Life Technologies			Life Technologies		
Entreprise d'origine	Applied Biosystems	454			Solexa					Applied Biosystems			Ion Torrent		
Commercialisé depuis	2002	2005			2006					2007			2010		
Renommée	Représentatif de la 1ère génération	1er de la nouvelle génération, longs *reads*			1er séquenceur à courts *reads*, leader actuel					2e séquenceur à courts *reads*, faible taux d'erreurs			1er séquenceur sans système de détection optique, 1er à moins de 100.000$		
Méthode de séquençage	Sanger	Synthèse (pyroséquençage)			Synthèse					Ligation			Synthèse (détection de H^+)		
Méthode d'amplification	PCR	PCR en émulsion			PCR en ponts					PCR en émulsion			PCR en émulsion		
Instrument	3730xl	GS Junior Titanium	FLX Titanium	FLX+	MiSeq	GA IIx	HiSeq 1000	HiSeq 2000	HiSeq 2000 v3	SOLiD 4	SOLiD 5500	SOLiD 5500xl	Ion PGM, puce 314	Ion PGM, puce 316	Ion PGM, puce 318
Coût à l'achat ($ US)	376 000	108 000	500 000	500 000	125 000	250 000	560 000	690 000		475 000	349 000	595 000	49 500		
Temps de *run*[1]	2h	10h	10h	18-20h	26h	14 jours	8 jours	8 jours	10 jours	12 jours	8 jours	8 jours	2h		
Millions de *reads/run*	0,000096	0,10	1	1	3,4	320	500	1000	≤ 3000	> 840	> 700	> 1410	0,10	1	4-8
Nb bases/*read*	650	400	400	700	150+150	150+150	100+100	100+100	100+100	50+35	75+35	75+35	100	>100	>100
Rendement Mb/*run*	0,06	50	500	900	1 020	96 000	100 000	200 000	≤ 600 000	71 400	77 000	155 100	>10	>100	>1000
Coût des réactifs/*run*[2]	96	1 100	6 200	6 200	750	11 524	10 220	20 120	23 470	8 128	6 101	10 503	500	750	925
Coût des réactifs/Mb	1500	22	12,4	7	0,74	0,12	0,1	0,1	<0,04	<0,11	<0,08	<0,07	<50	7,5	0,93
Erreurs principales	Substitutions	Insertions, délétions			Substitutions					Biais A-T			Insertions, délétions		
Taux d'erreur (%)[4]	0,1-1	1			>0,1					>0,06	>0,01	>0,01	~1		
Applications principales		a*, b, c*, d, g, h*			a*, b, c*, d, e, f, g, h					c*, e, f, h			a, b, c, d, h		

1 : pour la longueur de lecture maximale
2 : de la préparation de la librairie au séquençage. Pour le séquençage capillaire, uniquement le séquençage
3 : instruments et réactifs ne sont désormais plus commercialisés
4 : pourcentage d'erreur par base sur les reads de longueur maximale (voir ligne nombre de bases/read)

nd: non disponible

Tous les coûts sont donnés en $ US

a : séquençage *de novo* de génomes microbiens
b : caractérisation du transcriptome
c : reséquençage ciblé
d : réquençage de novo de génomes animaux et végétaux
e : reséquençage
f : détection de mutation
g : métagénomique
h : autres applications (ChIP-Seq, MeDIP-Seq, RNA-Seq, ...)
* : indexage nécessaire pour une utilisation optimale de cette application

par la polymérase, un PPi (Pyrophosphate) est libéré. Deux enzymes, contenues dans un autre type de billes beaucoup plus petites, permettent alors une cascade enzymatique : en présence du PPi libéré, une sulphurylase convertit de l'APS (Adénosine Phospho-sulphate) en ATP (Adénosine Triphosphate). L'ATP permet ensuite la conversion de luciférine en oxy-luciférine par une luciférase. Cette dernière réaction produit de la lumière et l'intensité de la réponse lumineuse est directement proportionnelle à la quantité de nucléotides incorporés. **(Figure 12)**

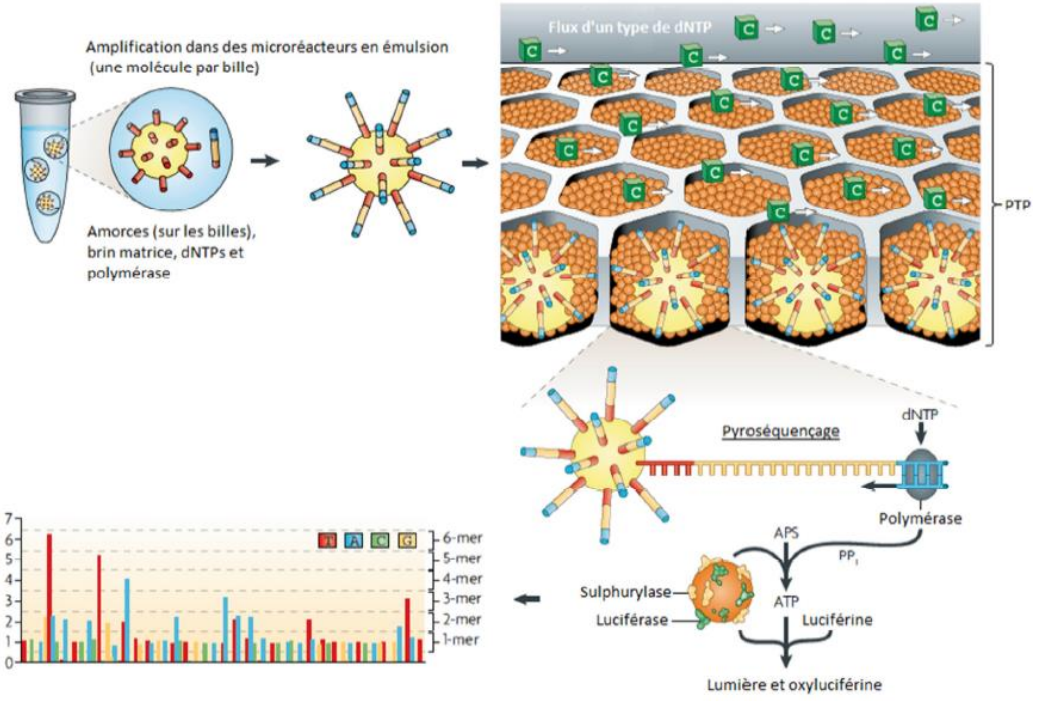
Cette technique ne nécessite pas de clonage, et permet une lecture directe de la séquence obtenue après le séquençage ; une seule machine génère, avec plus de 99% d'exactitude, autant de données que cent séquenceurs capillaires à haut débit *[37] [38].*

2.3.2 La technologie SOLiD

La technologie *SOLiD* (Sequencing by Oligonucleotide Ligation and Detection) est commercialisée par *Life Technologies.* Elle repose également sur une PCR en émulsion sur billes ; le séquençage ne s'effectue pas par synthèse comme sur les plateformes précédentes mais par ligation.

Une amorce de séquençage universelle se fixe sur l'adaptateur puis des oligonucléotides dégénérés de huit bases, marqués par fluorescence, sont ajoutés **(Figure 13)**. Dès que l'un d'entre eux correspond à la séquence adjacente de l'amorce, la ligase le fixe et de la fluorescence est émise, permettant d'identifier l'oligonucléotide fixé et d'interroger ainsi ses deux premières bases. Un clivage chimique retire les bases 6 à 8 ainsi que le fluorophore puis les oligonucléotides sont alors à nouveau ajoutés : on identifie les bases 6 et 7 de notre fragment à séquencer, puis dans un troisième temps les bases 11 et 12, et ainsi de suite jusqu'aux bases 31 et 32.

Un deuxième cycle de ligation est alors entamé avec une amorce universelle se fixant en position n-1 et on identifie ainsi les bases en position 0 et 1, 5 et 6, et ainsi de suite jusqu'aux bases 30 et 31 ; trois nouveaux cycles de ligation sont effectués grâce à des amorces se fixant en n-2, n-3 puis n-4. Le nombre de cycles de ligation, détection et clivage détermine ainsi la longueur de lecture. Chaque base est lue deux fois avec cette technologie, ce qui explique sa grande précision et qui la rend adaptée au re-séquençage ou à l'analyse de polymorphismes.



PTP PicoTiter Plate
PPi pyrophosphate inorganique
APS adenosine phosphosulphate
ATP Adenosine triphosphate

Figure 12: Principe de la technologie du séquenceur 454 (Metzker - 2008).

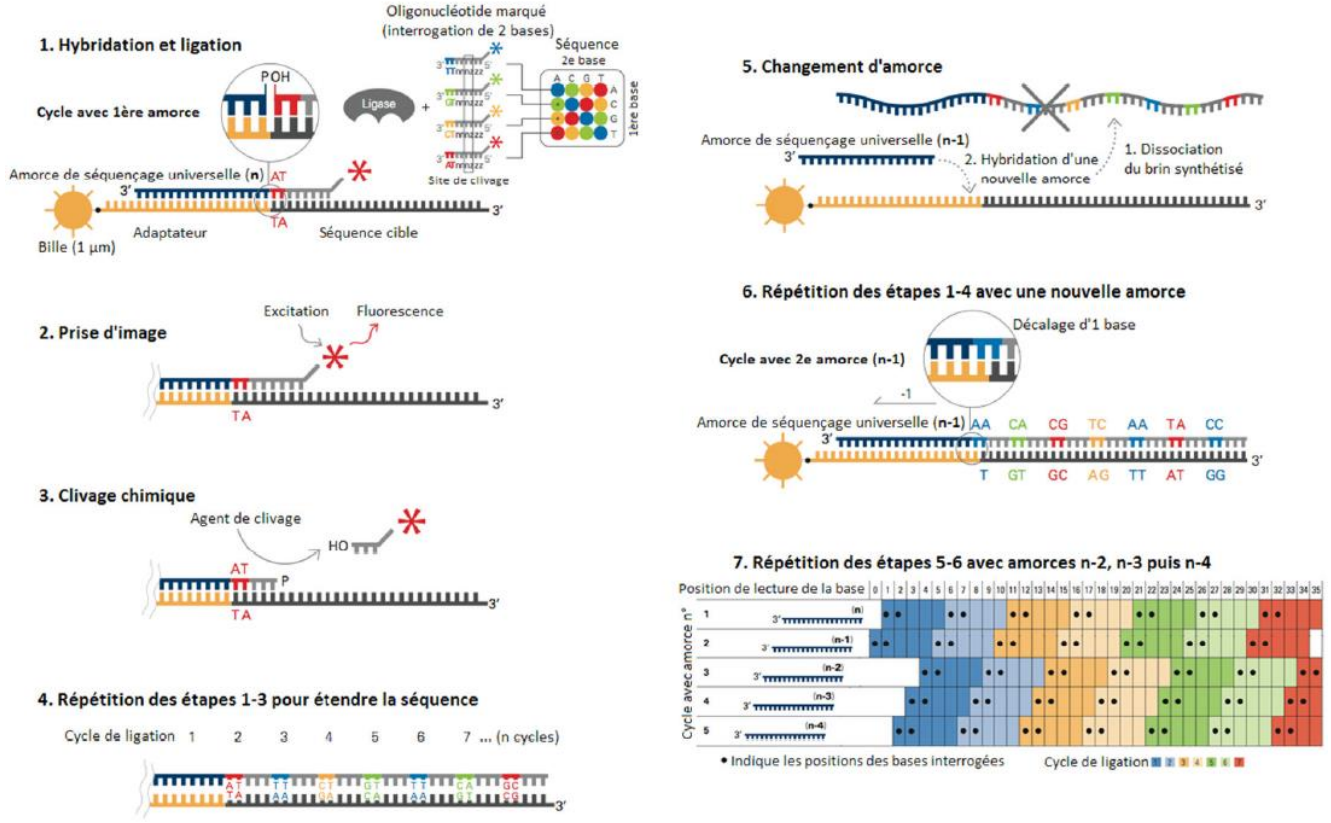


Figure 13 : Principe de la technologie *SOLID* (Elaine R Mardis)

Néanmoins, la complexité de fonctionnement et l'impossibilité de lire des régions longues ou répétées de cette technologie en constituent des inconvénients majeurs qui ont empêché un développement de cet équipement.

2.3.3 La technologie Ion Torrent / Ion Proton

En 2010, *Life Technologies* lance un séquenceur le *PGM-Ion Torrent* ; sa technologie repose sur la libération naturelle d'un ion H^+ après incorporation d'un nucléotide par une polymérase ; ce phénomène entraine une modification du pH pouvant être détectée par une puce de silicium semi-conductrice composée de plusieurs millions de transistors.

Le capteur mesure le pH dans chacune des cavités, ce qui indique la présence d'une ou de plusieurs bonnes bases qui ont été intégrées dans l'ADN en cours d'analyse ; chaque cavité permet de séquencer un brin d'ADN d'environ 200pb. **(Figure 14)**

Cette technologie est qualifiée de « PostLight » car aucun intermédiaire de lumière n'est utilisé ; il s'agit d'une modification chimique qui entraine la création du signal. L'absence de marquage fluorescent et de système de détection optique ou encore l'utilisation de micro-puces standards explique le faible cout de cette machine.

Cependant cet équipement présente un taux d'erreur légèrement supérieur à celui de la technologie *Illumina*.

(Voir paragraphe 2.3.4)

La dernière version de ces séquenceurs permet désormais de lire 1Gb de données avec une grande précision en seulement 2h et la longueur de lecture atteint les 400pb.

En 2012, *Life Technologies* a commercialisé une version plus évoluée, l'*Ion Proton*, qui permettrait le séquençage d'un génome entier d'un individu en quelques heures avec une capacité de séquençage mille fois supérieure au premier format de puce PGM Ion torrent soit 10 Gb. Compte tenu de l'apparente similarité entre les deux plateformes, l'évolution réside essentiellement dans l'architecture de la puce avec une augmentation du nombre de puits.

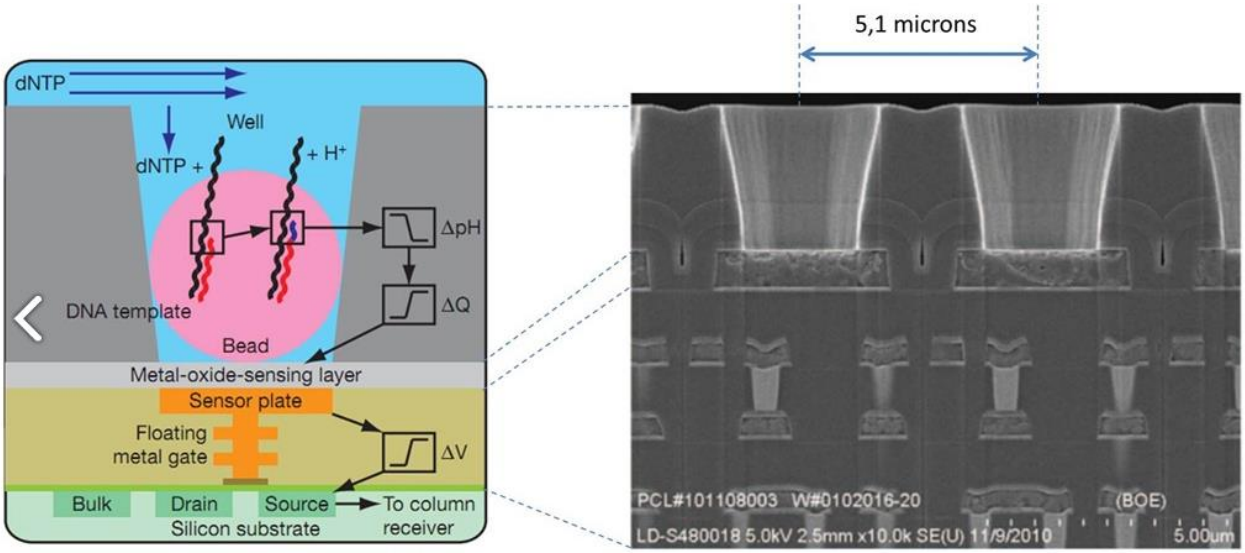


Figure 14 : Principe de la technologie *Ion Torrent* (Olivier Ezratty

2.3.4 La technologie Illumina

Il s'agit d'une technologie de séquençage sur support solide, basée sur une variante de la technique de Sanger (Cyclic Reversible Termination). La spécificité de cette technologie repose sur une amplification en pont (« *bridge PCR »)* des fragments à séquencer qui a lieu sur une surface de verre appelée *flowcell*, similaire a une lame de microscope, divisée en huit lignes. **(Figure 15)**

Les fragments de la librairie à séquencer possèdent des adaptateurs à leurs extrémités ; ces derniers vont leur permettre de se fixer de façon aléatoire sur la *flowcell*, par hybridation sur les amorces qui en couvrent la surface. Un nouveau brin est alors synthétisé par une polymérase : il est fixe de façon covalente à la *flowcell*. Le brin d'origine est alors éliminé par dénaturation et l'extrémité libre du brin restant s'hybride a une amorce adjacente pour former un pont. La polymérase synthétise à nouveau le brin complémentaire pour former un pont d'ADN double brin puis les deux copies sont libérées par dénaturation.

Le cycle d'amplification en pont recommence pour former à terme un « spot » d'ADN clonal en une zone appelée *cluster*. Les brins anti-sens sont ensuite clivés au cours de la linéarisation. L'extrémité 3' libre des fragments d'ADN est bloquée et l'amorce de séquençage s'y hybride. **(Figure 16)**

Le séquençage s'effectue sur des centaines de millions de *clusters* simultanément, grâce à une chimie de terminateurs réversibles marqués par fluorescence. Lorsque le nucléotide complémentaire est incorporé, la fluorescence est émise. Le fluorophore est ensuite clivé, permettant l'ajout d'un nouveau nucléotide **(Figure 17)**. A chaque cycle d'incorporation, une base peut être déterminée.

Illumina est maintenant le leader sur le marché du séquençage à très haut débit, notamment avec le lancement de son appareil *HiSeq2000*, qui peut produire jusqu'a 600 Gb en 10 jours, soit sept-cent fois plus que le séquenceur 454 de Roche, grâce à une méthode optique qui permet de travailler sur les deux surfaces de la *flowcell* et sur deux *flowcell* en parallèle *[37] [38]*.

Illumina a développé un équipement plus modeste, plus adapté à une pratique clinique, le *MiSeq* **(Figure 18)**, avec un flux de travail simplifié. Cette plateforme permet de réaliser des analyses rapides et économiques et les données récupérées peuvent aller jusqu'à 16Gb.

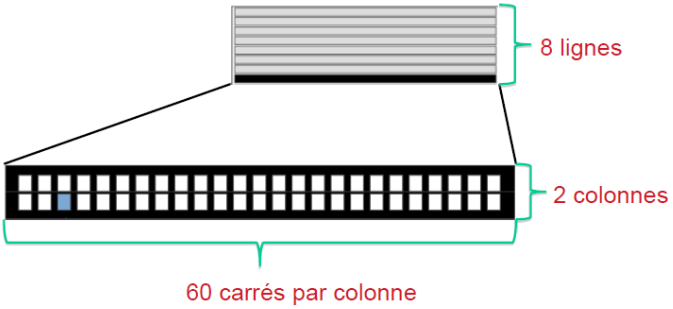


Figure 15 : Exemple de la structure d'une *flowcell* utilisée sur le *HiSeq* d'I*llumina* (Illumina)

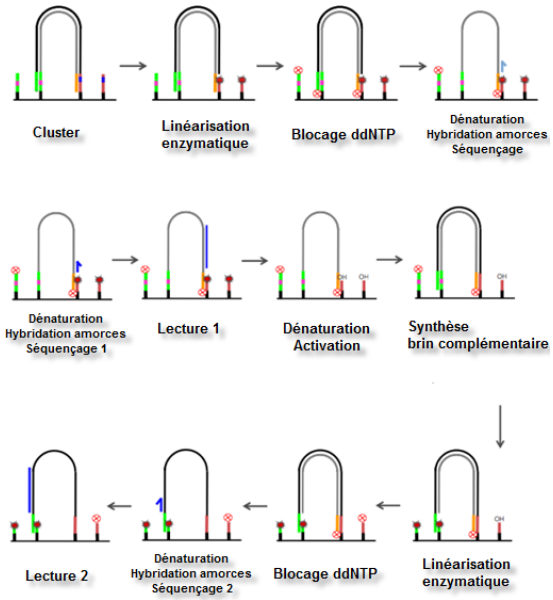


Figure 16 : Principe de la technologie du séquenceur *Illumina* (Illumina)

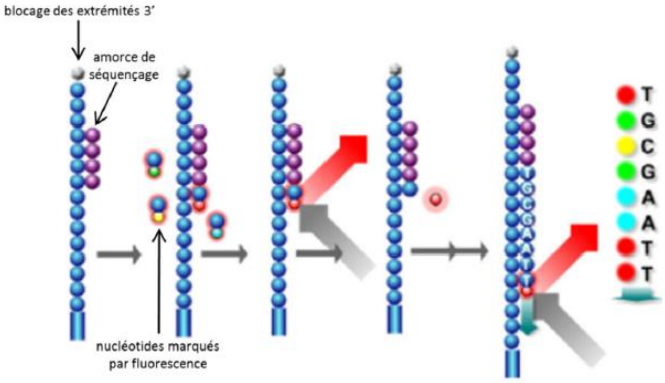


Figure 17 : Principe de l'émission de fluorescence lors du séquençage *Illumina* (Illumina)

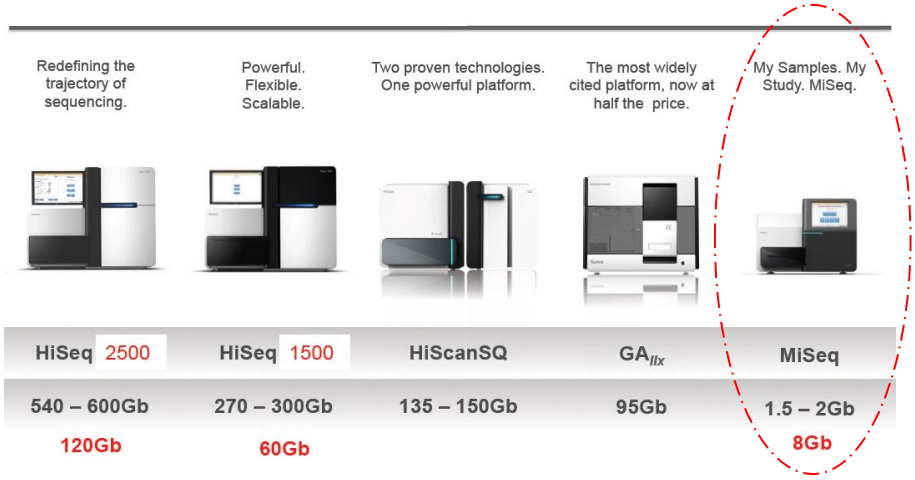


Figure 18 : Gamme de séquenceurs *Illumina* (Illumina)

2.4 Le séquençage de 3ème génération

La différence majeure entre la deuxième et la troisième génération de séquenceurs réside dans leur capacité à séquencer directement des molécules d'ADN de façon individuelle sans aucune amplification préalable, permettant d'accéder aux modifications des bases comme la méthylation directement avec une très longue processivité (capacité à séquencer plusieurs milliers de bases d'un même fragment).

Cette technologie n'est pas encore développée en pratique courante.

2.4.1 La technologie Oxford Nanopore

Cette technologie consiste à faire passer des brins d'ADN au travers de nanopores, protéines lipidiques déposées sur un substrat, et à lire les bases qui les traversent une par une.

L'ADN simple brin en traversant des pores de 2nm de diamètre modifie le courant électrique, variation du courant dépendante du type de nucléotide. C'est une technologie de séquençage non optique relativement rapide. **(Figure 19)**

L'avantage de cette technologie est qu'aucun traitement, ni marquage préalable ne sont nécessaires, de faibles quantités d'ADN suffisent et est capable d'identifier les modifications épi-génétiques, une 5-methylcytosine pouvant être différenciée d'une cytosine.

Les défis actuels de cette technologie restent la gestion d'un brin d'ADN qui « traverse » trop rapidement l'entrée ou à l'intérieur du nanopore.

2.4.2 La technologie Pacific Biosciences

Pacific Biosciences a développé un instrument capable de séquencer une molécule unique en temps réel par lecture directe sans amplifier le matériel de départ.

Le *PacBio RS* utilise une structure composée de cellules SMRT (Single Molecule Real Time), contenant chacune 75 000 nanostructures appelées détecteurs ZMW (Zero-Mode Waveguide) de 100nm de diamètre, c'est-à-dire plus petits que les longueurs d'onde utilisées (532 et 643 nm). La lumière ne peut donc pas s'y propager. Chacun de ces ZMW contient une polymérase qui y est immobilisée et qui incorpore des nucléotides liés à un fluorophore, libérant ainsi celui-ci a l'extérieur du ZMW. **(Figure 20)**

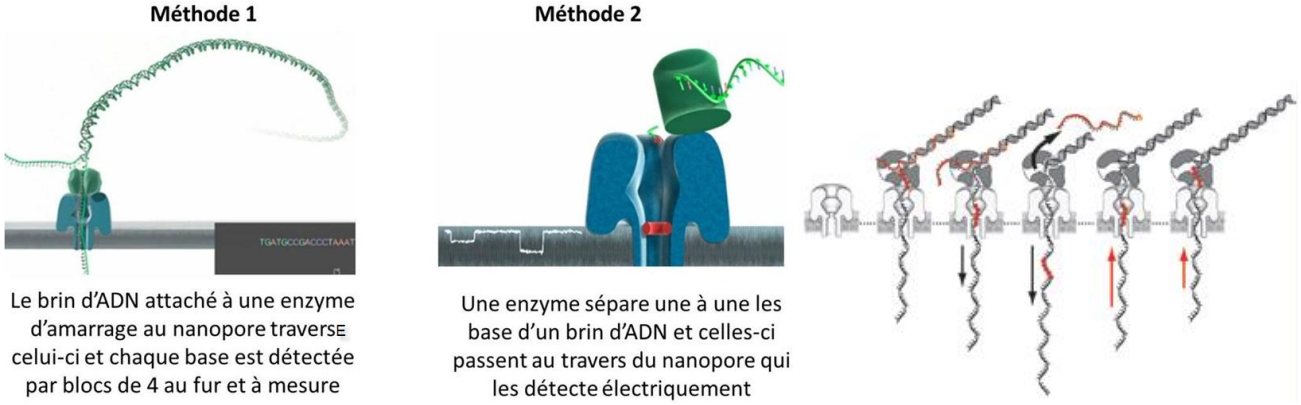


Figure 19 : Principe de la technologie des Nanopores (Olivier Ezratty et Cherf *et al*, *Nature Biotech*)

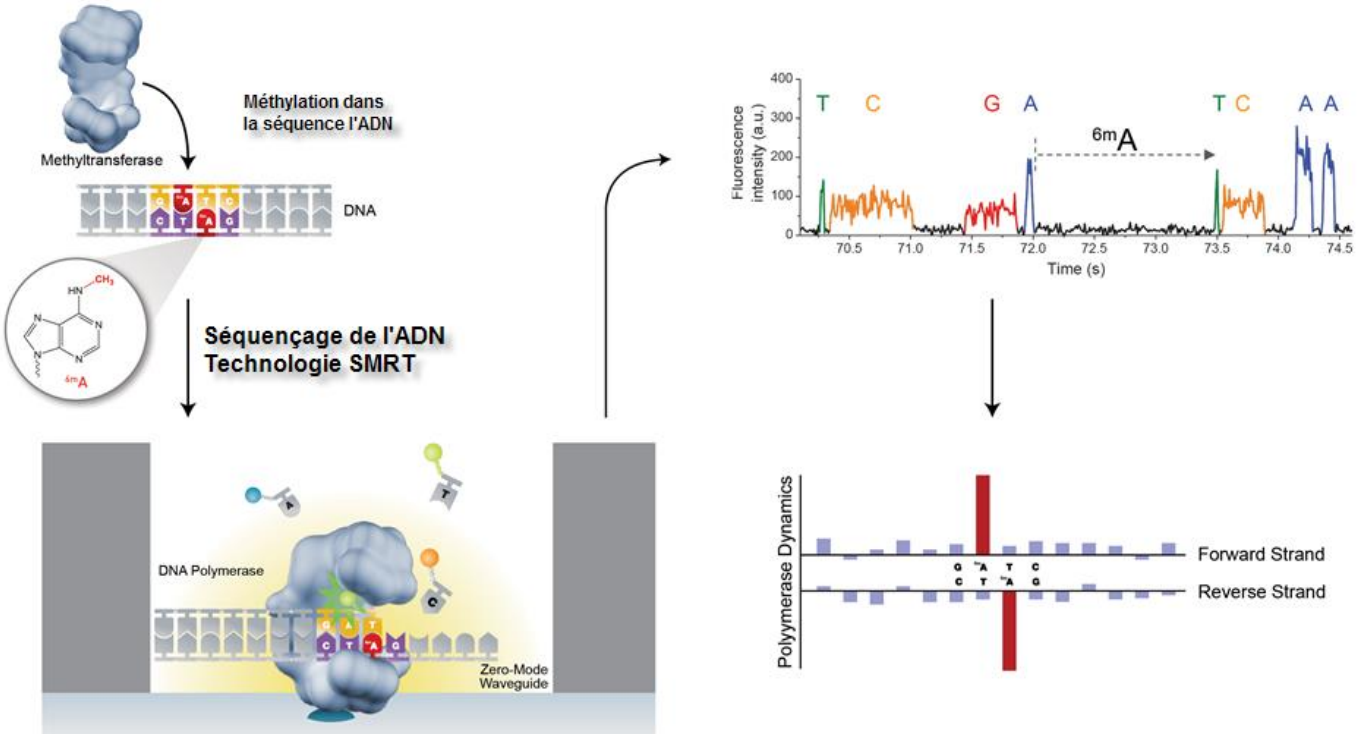


Figure 20 : Principe de la technologie *PacBio RS* (Pacific Biosciences)

L'intervalle de temps entre chaque pic de fluorescence, ainsi que sa durée pic sont propres à chaque nucléotide et permettent ainsi leur identification. Ces paramètres sont également différents pour des nucléotides méthylés et il est donc possible, sans traitement préalable, de différencier une cytosine d'une methyl-cytosine ou encore d'une hydroxymethyl-cytosine *[39]*.

La longueur de lecture moyenne serait de plus de 1000pb, la préparation de l'échantillon durerait environ 30 minutes et les données seraient produites en quelques minutes seulement, ce qui en ferait une technologie très prometteuse.

2.5 Les critères de qualité

Les appels de bases sont effectués directement en fonction des mesures d'intensité du signal au cours de chaque cycle, pouvant varier dans les régions de séquences répétitives et au sein des homopolymères.

Il est important d'associer à chaque base définie par ces séquenceurs un score de qualité dépendant de la qualité du signal détecté par rapport au bruit de fond.

Le « Q-score » ou score de qualité, est proportionnel de façon logarithmique à la probabilité d'erreur d'appel de base selon la formule :

$$Q = -10\mathrm{Log}_{10}(e)$$

Où « e », est la probabilité estimée que la base identifiée soit fausse.

Plus ce score est élevé, plus le risque d'erreur est faible ; ainsi, un score de 20 signifie un risque d'erreur de 1 base sur 100 et un score de 30 de 1 base sur 1000 **(Tableau VIII)**. Quand la qualité de séquençage approche d'une valeur de 30, les séquences de l'ordre de 400 bases seront théoriquement « parfaites », sans erreurs ni ambiguïté. La proportion de séquences ayant un score de Q30 est donc prise comme « référence » qualité pour le NGS.

3 LES 24 GENES D'INTERET MAJEUR RETENUS POUR NOTRE ETUDE MUTATIONNELLE

Nous avons choisi une première série de gènes pour notre projet en se basant sur des recherches bibliographiques listant les gènes retrouvées mutés dans les LAM et les SMD.

Tableau VIII : Q-score et précision (Illumina)

Phred Quality Score	Probability of Incorrect Base Call	Base Call Accuracy
10	1 in 10	90%
20	1 in 100	99%
30	1 in 1,000	99.9%
40	1 in 10,000	99.99%
50	1 in 100,000	99.999%

3.1 La diversité des gènes mutés

3.1.1 Dans les Leucémies Aigües Myéloïdes

Les réarrangements de la structure chromosomique sont des marqueurs pronostics et diagnostics, suggérant que ces anomalies génétiques acquises et récurrentes ont un rôle essentiel dans la pathogénicité *[40] [41]*. Cependant 50% des échantillons de LAM ont un caryotype normal ; ainsi la majorité des patients atteints de LAM présentent un profil de risque intermédiaire et pour cette raison, des études récentes se sont lancées dans le décryptage de nouveaux bio-marqueurs pouvant être utilisés pour établir une meilleure classification du groupe de risque intermédiaire *[42] [43].*

Les nouveaux algorithmes de classification incorporent *FLT3, NPM1, CEBPα* et *KIT [1]* [44]. Plus récemment, de nouveaux gènes se sont retrouvés mutés dans les LAM tels que *DNMT3A* [26]*, IDH1 et IDH2 [24], et TET2 [45]* et pourraient fournir des informations pronostiques pour des patients de risque intermédiaire présentant un caryotype normal.

Une étude du *Cancer Genome Atlas Research Network [46]* en 2013 a mis en évidence les principaux gènes mutés parmi une cohorte de deux-cents patients majoritairement de risque intermédiaire **(Figure 21)**. Cette étude a permis d'identifier des combinaisons ou exclusions de mutations dans les échantillons **(Figure 22)** et de regrouper les mutations pouvant avoir un rôle au niveau de certaines voies de signalisation. En effet, sur les deux-cents patients de la cohorte, cent quatre-vingt-dix-neuf (soit plus de 99%) présentaient au moins une mutation d'un des neuf facteurs définis comme ayant une fonction biologique ou un rôle confirmé dans la pathogénicité des LAM :

- Les facteurs de transcription (18%) ;
- Le gène *NPM1* (27%) ;
- Les gènes suppresseurs de tumeurs (16%) ;
- Les gènes en relation avec la méthylation de l'ADN (44%) ;
- Les gènes d'activation de la signalisation (59%) ;
- Les gènes de modification de la chromatine (30%) ;
- Les gènes codant pour les facteurs de transcription myéloïdes (22%) ;

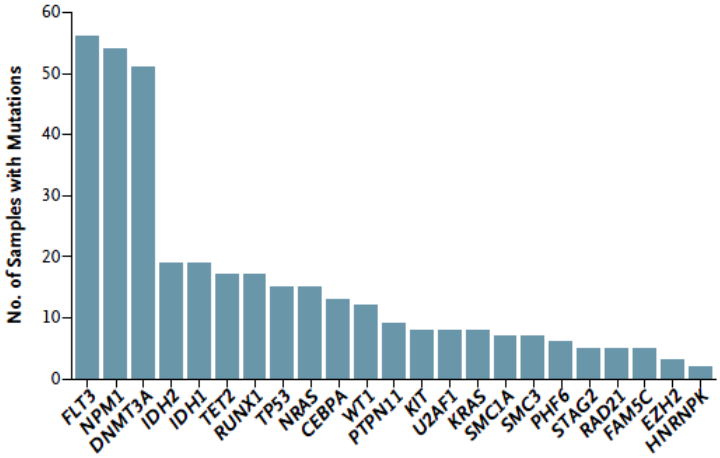


Figure 21 : Proportion des gènes mutés dans une cohorte de 200 patients (CGAR)

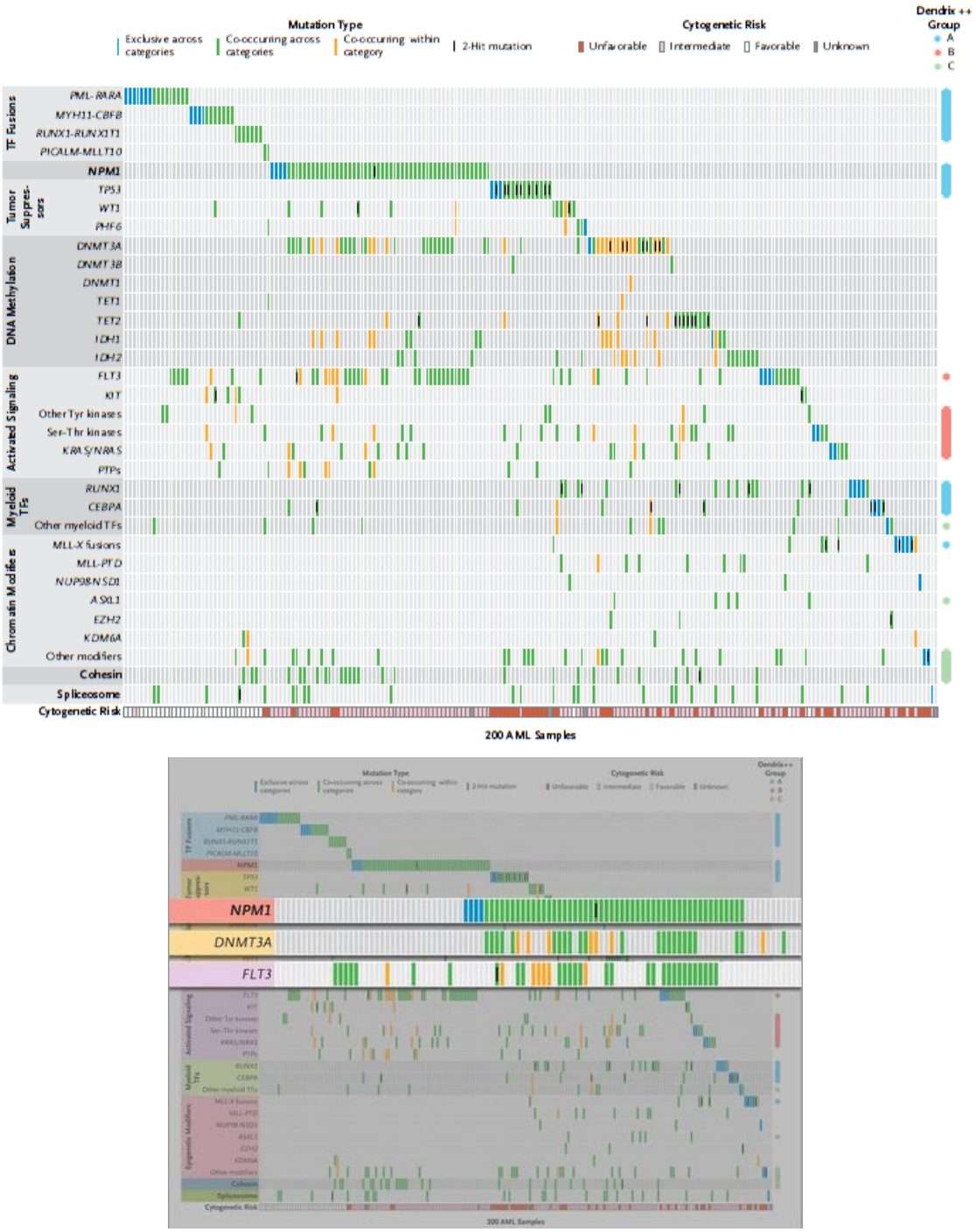


Figure 22 : Regroupement des types de mutations dans une cohorte de 200 patients et correspondances (CGAR)

- Les gènes du complexe de la cohésine (13%) ;
- Les gènes du complexe du splicéosome (14%). **(Figure 23)**

Ces différentes études montrent qu'il existe un nombre relativement faible de variations nucléotidiques somatiques récurrentes dans les cellules de LAM. Cela suggère que la LAM ne résulte pas d'une instabilité génomique, mais serait plutôt associée à un faible nombre de mutations oncogéniques conductrices, la grande majorité étant des mutations passagères, qui jouent un rôle passif dans le développement de la maladie. **(Figure 24)** [47]

3.1.2 Dans les Syndromes Myélodysplasiques

Des anomalies de la méthylation ont été mises en évidence dans les SMD, entraînant la répression de gènes suppresseurs de tumeurs, de gènes régulateurs du cycle cellulaire, de gènes impliqués dans la réparation de l'ADN et de gènes de l'apoptose *[2]*.

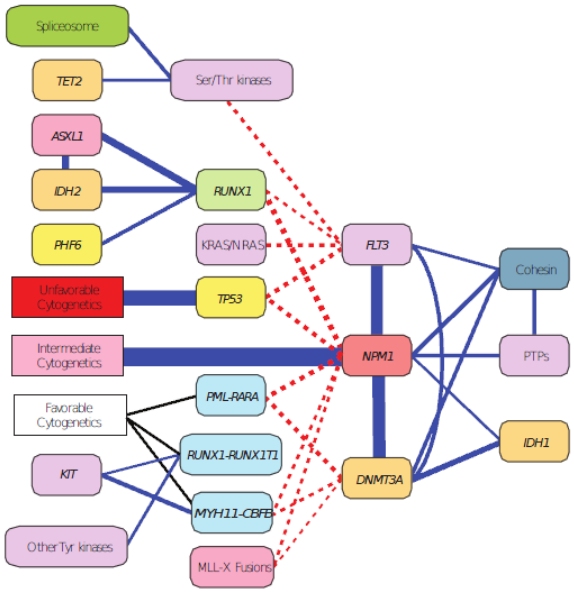
De nombreuses mutations ont été retrouvées dans les SMD, impliquant les gènes *RAS*, *TP53*, *RUNX1*, *ASXL1*, *CBL*, *DNMT3A*, *IDH1 et 2*, *TET2* et *EZH2*, recouvrant une partie des gènes mutés dans les LAM.

La présence de certaines mutations (*TP53, EZH2, ETV6, RUNX1, ASXL1*) était prédictive d'une survie globale inférieure *[48]*. **(Tableau IX)**

Dans des SMD de bas risque avec 5q-, la technologie du NGS a permis la détection de petits clones mutés de *TP53* avec une sensibilité pouvant descendre jusqu'à 1% *[49]*. Ces mutations de *TP53*, même à un taux faible, influencent le pronostic des patients de façon défavorable *[47].*

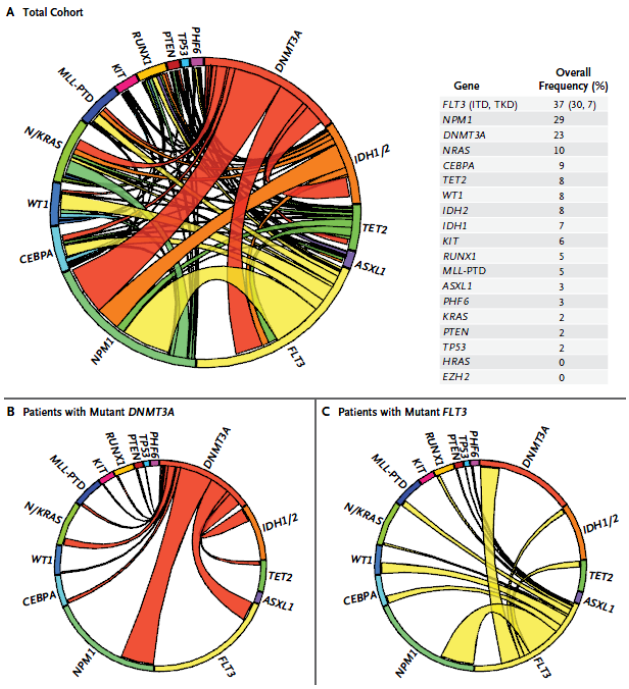
Mais ces mutations n'expliquent pas entièrement la pathogénèse des SMD car ces anomalies sont retrouvées dans d'autres pathologies au-delà des LAM.

Dans environ 20% des SMD, aucune anomalie génétique n'est retrouvée. L'analyse de vingt-neuf cas de SMD, par séquençage du génome entier a révélé de façon inattendue des mutations de la machinerie d'épissage de l'ARN, tels que des mutations des gènes *U2AF35, ZRSR2, SRSF2* et *SF3B1 [28].* Ces mutations sont retrouvées à une fréquence très importante chez les patients atteints de SMD et semblent corrélées à la classification OMS. Ainsi, les mutations de *SF3B1* sont associées aux anémies réfractaires avec sidéroblastes en couronne, alors que les



Traits bleus	connectent des nœuds qui témoignent d'évènements s'étant produits dans un nombre significatif d'échantillons
Traits rouges	connectent des nœuds s'excluant mutuellement
Traits noirs	signalent des mutations définies comme cytogénétiquement favorables.

Figure 23 : Interconnexion des mutations de gènes et groupes de risques cytogénétiques (CGAR)



Gene	Overall Frequency (%)
FLT3 (ITD, TKD)	37 (30, 7)
NPM1	29
DNMT3A	23
NRAS	10
CEBPA	9
TET2	8
WT1	8
IDH2	8
IDH1	7
KIT	6
RUNX1	5
MLL-PTD	5
ASXL1	3
PHF6	3
KRAS	2
PTEN	2
TP53	2
HRAS	0
EZH2	0

Figure 24 : Complexité des mutations retrouvées dans les LAM (Patel et al.)

Tableau IX : Impact pronostic des mutations retrouvées dans les SMD (Cluzeau – SFH 2011)

Gènes	Fréquence	Fonction probable	Valeur pronostique
TET2	20-25 %	Conversion 5MC en 5HMC	Favorable ?
EZH2	6 %	Méthylation H3K27	Défavorable
ASXL1	10 %	Cofacteur du récepteur RAR	Défavorable
DNMT3	8 %	Méthylation de l'ADN	Défavorable
IDH1/IDH2	5-10 %	Nécessaire au cycle de Krebs pour la formation de l' α-kétoglutarate	Défavorable
UTX	2-5 %	Méthylation H3K27	Non définie
Nras/Kras	10-15 %	Signal de transduction	Plutôt défavorable
TP53	5-10 %	Réparation ADN, apoptose, régulation du cycle cellulaire	Défavorable
RUNX1/AML1	10-15 %	Facteur de transcription	Défavorable
NPM1	5 %	Transport nucléaire, fonction p53, facteur de transcription	Non définie
JAK2	2-5 %	Tyrosine kinase, signal de transduction intracellulaire	Favorable ?
FLT3	2-5 %	Récepteur de facteur de croissance	Non définie
CEBPα	1-4 %	Facteur de transcription	Non définie
EVI1	2 %	Facteur de transcription	Non définie

mutations de *SRSF2* sont plutôt associées aux leucémies myélo-monocytaires chroniques.

3.2 Les mutations des facteurs de transcription

3.2.1 La mutation « classique » de CEBPα

(Voir paragraphe 1.2.4.6)

3.2.2 La mutation de WT1

Le gène *WT1* (Wilms' Tumor 1), localisé en 11p13, code pour une protéine régulatrice de la transcription.

Initialement découvert du fait de son altération dans la tumeur de Wilms, le gène *WT1* s'est avéré être aussi impliqué dans les hémopathies malignes, en particulier les LAM dans environ 10% des cas.

3.2.3 La mutation d'ASXL1

ASXL1 (Additional Sex Comb-like 1), localisé en 20q11, est fréquemment muté dans les hémopathies myéloïdes incluant les MDS, les LAM, la LMC, la LMMC ainsi que d'autres SMP ; plusieurs études ont aussi montré que ces mutations avaient un impact négatif au plan pronostique.

Les mutations d'ASXL1 sont retrouvées dans 17% des LAM avec risque cytogénétique intermédiaire *[50]*.

La fréquence des mutations était associée à un âge avancé, des antécédents de SMD ainsi qu'avec un immunophénotype immature. Ces mutations étaient fréquemment associées à une mutation de *RUNX1* et, à l'opposé, moins fréquemment associées à une mutation de *NPM1,* de *FLT3* ou de *DNMT3A*.

Les patients avec une mutation d'*ASXL1* avaient une survie globale significativement plus courte ; l'analyse multi-variée identifiait les mutations d'*ASXL1* comme un facteur pronostic négatif indépendant pour la survie globale.

3.2.4 La mutation de DNMT3A

DNMT3A (DNA [cytosine-5]- Methyltransferase 3A), localisé en 2p23, est muté dans 18% des patients de moins de 60 ans *[51]*.

Ces patients avaient souvent un caryotype normal et présentaient des mutations associées de *NPM1, FLT3* et *IDH1.* Le mauvais pronostic associé à ces mutations a été retrouvé dans une autre étude *[26]*.

3.2.5 La mutation de RUNX1

Les mutations de *RUNX1* (Runt-related transcription factor 1), localisé en 21q22, ont été identifiées dans les SMD et les LAM. Elles sont associées à de sévères thrombopénies et à une proportion croissante de blastes *[2]*.

Les mutations somatiques de *RUNX1* sont prédictives d'une faible survie globale chez les patients atteints de MDS, indépendamment des facteurs de risque, tout comme les mutations de *TP53, EZH2, ETV6 et ASXL1*.

3.2.6 La mutation de TET2

La recherche d'anomalies moléculaires communes à différentes pathologies myéloïdes (SMD, SMP, LAM de novo et secondaires) a permis d'identifier une perte d'hétérozygotie de la région 4q24, mise en évidence par SNP et correspondant à une délétion de 325kb détectée par CGH (Hybridation Génomique Comparative).

TET2 (Tet methylcytosine dioxygenase 2), localisé en 4q24 est le seul gène codant localisé à cet endroit ; il appartient à une famille de trois protéines possédant deux domaines hautement conservés, dont la fonction est mal connue.

Le séquençage de patients atteints de diverses pathologies myéloïdes retrouve des mutations du gène *TET2* dans 15 % des cas *[45]*.

Une autre étude confirme la présence de mutations de *TET2* dans 21 % des SMD, en particulier dans le groupe à faible risque *[52]*.

La valeur pronostique de ces mutations reste à déterminer ; les premières études montrent des résultats variables en fonction de la pathologie initiale avec un impact pronostique favorable dans les SMD [53] et défavorables dans les LAM *[54]*.

3.2.7 La mutation de TP53

Des mutations de *TP53* (Tumor Protein p53), localisé en 17p13, ont été identifiées dans les MDS et les LAM. Elles sont associées à de sévères thrombopénies et à une proportion croissante de blastes *[2]*.

Ces mutations dans les SMD sont associées à un mauvais pronostic et les caryotypes complexes. Avec l'arrivée du NGS, des mutations de *TP53* ont été retrouvées dans des SMD de faible risque avec délétion 5q isolée (syndrome 5q-) *[49]*.

Dans les LAM à caryotype complexe, la mutation de *TP53* était associée à un plus mauvais pronostic, au sein de ce groupe de patients qui a déjà un très mauvais pronostic *[55]*.

Dès lors, la recherche de la mutation de *TP53* semble importante dans les syndromes 5q- où elle semble corrélée à un mauvais pronostic et pourrait permettre de décider d'un traitement renforcé et, d'autre part, dans les SMD de haut risque à caryotype complexe où la présence de la mutation semble aggraver encore le pronostic.

3.2.8 La mutation d'ETV6

Le gène *ETV6* (ETS variant 6), localisé en 12p13 code pour un facteur de transcription de la famille ETS. Une étude récente a détecté des mutations d'*ETV6* dans 2% des SMD *[2]*.

3.2.9 La mutation de PHF6

PHF6 (PHD Finger Protein 6), localisé en Xq26 code pour un facteur de transcription à deux domaines PHD en doigts de zinc caractéristiques d'un rôle dans la régulation transcriptionnelle. Les mutations de ce gène sont corrélées à une réduction de la survie globale *[1]*.

3.3 Les mutations des récepteurs de cytokines et de leurs voies de signalisation

3.3.1 La mutation « classique » de FLT3

(Voir paragraphe 1.2.4.6)

3.3.2 La mutation de KIT

Les mutations du gène sont très fréquemment regroupées sur l'exon 17 qui code pour le domaine kinase et l'exon 8, codant pour une région très conservée de la région extracellulaire du récepteur.

Plusieurs études ont montré l'impact pronostique défavorable sur la survie sans événement et la survie globale des mutations de l'exon 17 dans les LAM t(8;21), aussi

bien chez l'adulte que chez l'enfant *[56]*. Cependant ce résultat est controversé pour les mutations de l'exon 8 de *KIT* dans les LAM inv(16) *[57]*.

Il est important de détecter les mutations de *KIT* de façon systématique dans les LAM CBF pour la stratification thérapeutique, mais également de déterminer le type de mutation, les mutations de l'exon 8 étant sensibles aux inhibiteurs des tyrosines kinases comme l'Imatinib.

3.3.3 La mutation de KRAS et NRAS

Les gènes *KRAS* (Kirsten-Ras), localisé en 12p12, et *NRAS* (Neuroblastoma-Ras), localisé en 1p13, codent pour de petites protéines G monomériques à activité GTPase (Guanosine triphosphate hydrolase).

Les mutations de ces gènes sont fréquentes, retrouvées dans 18% des patients présentant une LAM *[58]* et 11% des SMD *[59]*.

Les mutations des gènes *RAS* ne sont pas des facteurs pronostics par eux-mêmes mais peuvent servir de marqueurs thérapeutiques pour le suivi de la maladie résiduelle. *[58]*

3.3.4 La mutation de JAK2

JAK2 (Janus Kinase 2) localisé en 9p24 code pour une tyrosine kinase essentielle de la cascade de signalisation.

JAK2 est principalement muté dans une entité très particulière des SMD, l'anémie réfractaire avec présence de sidéroblastes en couronne *[60]* La détection de ces mutations est importante du fait de l'existence d'inhibiteurs spécifiques de la fonction tyrosine kinase de JAK2 comme le *Ruxolitinib*.

3.4 Les autres mutations

3.4.1 La mutation « classique » de NPM1

(Voir paragraphe 1.2.4.6)

3.4.2 La mutation d'EZH2

EZH2 (Enhancer Of Zeste Homolog 2), localisée en 7q35-36, code pour une histone méthyl-transférase. Elle est rarement mutée dans les LAM mais plus

fréquemment dans les SMD (8%) où sa détection permettrait d'améliorer la prise en charge des SMD de bas risque *[2]*.

3.4.3 Les mutations d'IDH1 et IDH2

L'analyse du génome d'une LAM à caryotype normal a permis d'identifier de mutations du gène *IDH1* (Isocitrate Déshydrogénase 1), localisé en 2q33. Cette mutation a été dans quinze LAM sur cent-quatre-vingt-sept analysées *[2]*.

Des mutations d'*IDH2* (Isocitrate Déshydrogénase 2), localisé en 15q26, ont été également identifiées.

IDH1 et *IDH2* participent au métabolisme du citrate *[61]*. Ces mutations sont retrouvées dans 17% des LAM et sont associées à un caryotype normal et aux mutations de *NPM1 [62]*. Elles conduisent à une perte de fonction et surtout à l'acquisition nouvelle de la capacité à catalyser la réduction d'alpha-kétoglutarate en hydroxyglutarate *[25]*.

3.4.4 La mutation des gènes du splicéosome (SRSF2, U2AF1 et SF3B1)

Les mutations des composants du splicéosome sont fréquentes dans les SMD *[28] [63]*, en particulier *U2AF1* (U2 small nuclear RNA auxiliary factor 1), localisé en 21q22, *SRSF2* (Sérine/arginine-rich splicing factor 2), localisé en 17q25, et *SF3B1* (Splicing Factor 3b, Subunit 1) localisé en 2q33.

Ces mutations sont hautement spécifiques des pathologies comportant une composante myélodysplasique. Elles sont retrouvées dans :

- 85% des SMD avec sidéroblastes en couronne, principalement SF3B1;
- 44% des SMD sans sidéroblastes en couronne ;
- 55% des LMMC, principalement SRSF2 dans ce cas;
- 26% des LAM secondaires à une chimiothérapie ou à un SMD.

En revanche, ces mutations sont plus rares dans les LAM *de novo* (6,6 %) et les SMP (9,4 %).

Ces mutations sont mutuellement exclusives, suggérant un impact sur la machinerie de l'épissage et potentiellement la pathogénèse des SMD. A la différence des mutations de *SF3B1*, les mutations de *SRSF2* et *U2AF1* n'étaient pas liées à la

présence de sidéroblastes et les mutations de ces deux gènes étaient retrouvées quel que soit le type SMD, le profil de risque ou encore les anomalies cytogénétiques retrouvées.

Les mutations de *SRSF2* étaient associées aux mutations de *RUNX1* ou *IDH1* alors que celles d'*U2AF1* étaient associées aux mutations d'*ASXL1* et *DNMT3A [64].* Ces dernières sont associées à une survie globale plus faible et à une acutisation en LAM plus rapide et fréquente *[64]*.

3.4.5 La mutation de SETBP1

SETBP1 (SET Binding Protein 1) est localisé en 18q21. Ce gène est rarement muté dans les LAM et SMD *[65]*. Ses mutations sont fréquemment associées à celles de *SRSF2*.

3.4.6 La mutation du CBL

Le gène *CBL* (Casitas B-lineage Lymphoma), localisé en position 11q23, code pour une E3 ubiquitine ligase permettant ainsi la dégradation de protéines spécifiques via le protéasome. Ce gène est muté dans environ 10% des SMD *[66]* et plus rarement dans les LAM (<2%) *[67]*.

V OBJECTIF SPECIFIQUE DU TRAVAIL

Le NGS représente un saut évolutif dans la biologie moléculaire, après la PCR (Polymérase Chain Reaction) et la PCR quantitative.

Après la recherche de mutations sur un seul gène par échantillon grâce à la technique de séquençage Sanger, ces nouvelles technologies permettent d'analyser n gènes pour un même échantillon, voire n gènes pour n échantillons.

L'utilisation de ces nouvelles technologies dans le cadre de l'activité diagnostique impose, dans un contexte d'évolution technologique rapide, d'assurer une qualité optimale des examens. Elle va entraîner des réorganisations au sein des laboratoires et rendre indispensable l'acquisition de nouvelles compétences pour la préparation des échantillons et l'analyse des résultats ; en particulier, le recours à la bio-informatique sera indispensable pour l'analyse et la gestion des données issues du NGS.

Mais le recours à cette approche impose de continuer, en parallèle, à utiliser les techniques habituelles pour la recherche d'altérations moléculaires autres que les mutations.

Les points sur lesquels je dois donc travailler sont :

- Le choix de la technique d'enrichissement en termes d'organisation technique ;
- Le choix de la plateforme de NGS ;
- L'analyse bio-informatique et la validation des données ;
- Le traitement des données de séquençage ;
- Le rendu des résultats aux cliniciens.

Ceci afin de définir un « workflow » adapté à une activité de routine au sein d'un CHU et pouvant répondre aux normes de qualité en vigueur.

Afin de se familiariser avec le NGS et de travailler sur ces différents points, nous avons réalisé une première analyse de NGS pour un nombre restreint de gènes et pour quelques patients dont le statut mutationnel a été évalué par séquençage Sanger préalablement. Même si nous avons exploré les autres possibilités, pour ce premier test nous avons choisi d'utiliser la technique de préparation des librairies de séquençage par PCR.

VI MATERIEL ET METHODES

1 GLOSSAIRE TECHNIQUE

Nous avons décidé de rédiger un glossaire technique succinct afin de présenter les termes et anglicismes couramment utilisés pour le séquençage haut débit.

- Base calling : attribution des bases à chaque cycle en fonction du signal fluorescent détecté.
- Cluster : réaction de séquence ayant lieu sur la *flowcell* et donnant un spot lumineux.
- Design : dessin d'amorces.
- Enhancer : réactif formulé pour améliorer l'amplification par PCR des régions riches en GC.
- Flowcell : lame de verre sur laquelle a lieu la réaction de séquence.
- Index : séquence oligonucléotidique de 6pb spécifique de chaque échantillon, rajoutée aux adaptateurs Illumina et permettant le mélange (ou multiplexage) de plusieurs échantillons dans un même « run » de séquençage. La lecture de cette séquence permettra lors de l'analyse des données de séparer les séquences correspondant à chaque échantillon.
- Q-score : score de qualité.
- Paired-end : séquençage à double sens de lecture (sens et anti-sens) avec possibilité de recoupement des lectures.
- Read : séquence de lecture crée à l'issue du *base-calling*.
- Run : lancement d'un processus complet de séquençage haut débit (ou autre), de la formation des clusters sur la flowcell à la réaction de séquence à proprement parlé.
- Single-end : séquençage à simple sens de lecture.
- Workflow : mise en place d'une organisation la mieux adaptée au processus et à l'environnement.

2 MATERIEL

2.1 La nature du matériel

Nous travaillerons sur de la moelle ou du sang frais de patients, à partir desquels nous extrairons de l'ADN.

L'extraction se fait avec l'aide d'un kit, *Wizard Genomic DNA Purification kit* de chez *Promega* à partir des cellules mononuclées du sang ou de la moelle osseuse préalablement isolées par gradient de densité et conservées en culot sec congelé. Elle repose sur le principe que toute cellule nucléée renferme dans son noyau l'ADN du génome entier et dans son cytoplasme l'ARN. Les étapes générales de l'extraction sont au nombre de quatre :

- Lyse des cellules ;
- Protéolyse ;
- Dé-protéinisation ;
- Précipitation et purification de l'ADN génomique.

2.2 La qualité du matériel

Après extraction de l'ADN, nous réalisons un dosage et nous vérifions la pureté de l'ADN par mesure de l'absorbance à 260, 280 et 230nm sur le *NanoDrop*.

Il s'agit d'un spectrophotomètre qui ne nécessite pas l'utilisation de cuvette et où 1µL de l'échantillon est déposé directement à l'extrémité d'un câble de fibre optique ; lorsque celui-ci est mis en contact avec un second câble par fermeture de l'appareil, l'échantillon est maintenu par une tension de surface et traversé par un signal lumineux.

L'inconvénient majeur de ce type de mesure réside dans le fait que toute molécule parasite qui absorbe la lumière à 260nm est prise en compte dans la mesure de la concentration en ADN.

Afin d'évaluer une possible contamination en protéines (qui absorbent à 280nm) ou en solvants (qui absorbent à 230nm), le *NanoDrop* calcule le ratio de l'absorbance à 260 sur 280nm et à 260 sur 230nm qui doivent être compris entre 1,8 et 2.

3 METHODES

3.1 Les techniques d'enrichissement

L'enrichissement est une technique permettant d'obtenir le matériel nécessaire à la préparation de la librairie afin de réaliser le séquençage haut débit des régions d'intérêt.

Parmi les techniques d'enrichissement, il existe deux technologies : la capture et les techniques d'amplicons.Parmi ces dernières, il existe, entre autres, le kit *TruSeq Custom Amplicon* d'*Illumina* et la PCR classique que nous développerons.

3.1.1 La capture

La capture de séquence est une technologie basée surtout sur la propriété d'hybridation de brins d'ADN ayant des séquences complémentaires, sur support solide (capture sur lame) ou en suspension (capture en phase liquide). Elle permet un enrichissement spécifique d'une librairie génomique en fragments d'ADN issus de régions ciblées et pré-définies.

Cette technique est surtout utilisée pour des tailles de génome de l'ordre du mégabase, ce qui est nettement supérieur à la taille de l'ensemble des exons que nous souhaitons séquencer et correspondrait plutôt à une utilisation pour du séquençage d'exome ; nous n'utiliserons ni ne testerons donc cette technique d'enrichissement non adaptée à notre étude.

3.1.2 Le kit TruSeq Custom Amplicon

Nous avons pensé que le kit TruSeq Custom Amplicon proposé par Illumina serait adéquate pour débuter ; il s'agit d'une trousse commerciale permettant de générer une importante librairie d'ADN génomique par capture et amplification des régions d'intérêt, préconisé pour le multiplexage d'amplicons.

3.1.2.1 Le principe du kit

Le kit permet la création d'une librairie pouvant comprendre de seize à mille cinq-cents trente-six amplicons (quatre tailles d'amplicons possibles). La technique se décompose en trois étapes **(Figure 25)**:

- Hybridation de sondes de part et d'autre de chaque région à séquencer sur le génome. Les sondes sont composées d'une partie spécifique de la région d'intérêt et d'une queue flottante identique pour toutes les sondes.
- Synthèse d'un brin d'ADN complémentaire de la région d'intérêt entre les deux sondes et ligation.
- Ajout des index et des adaptateurs Illumina nécessaires au séquençage par réaction de PCR sur les queues flottantes des fragments générés à l'étape précédente.

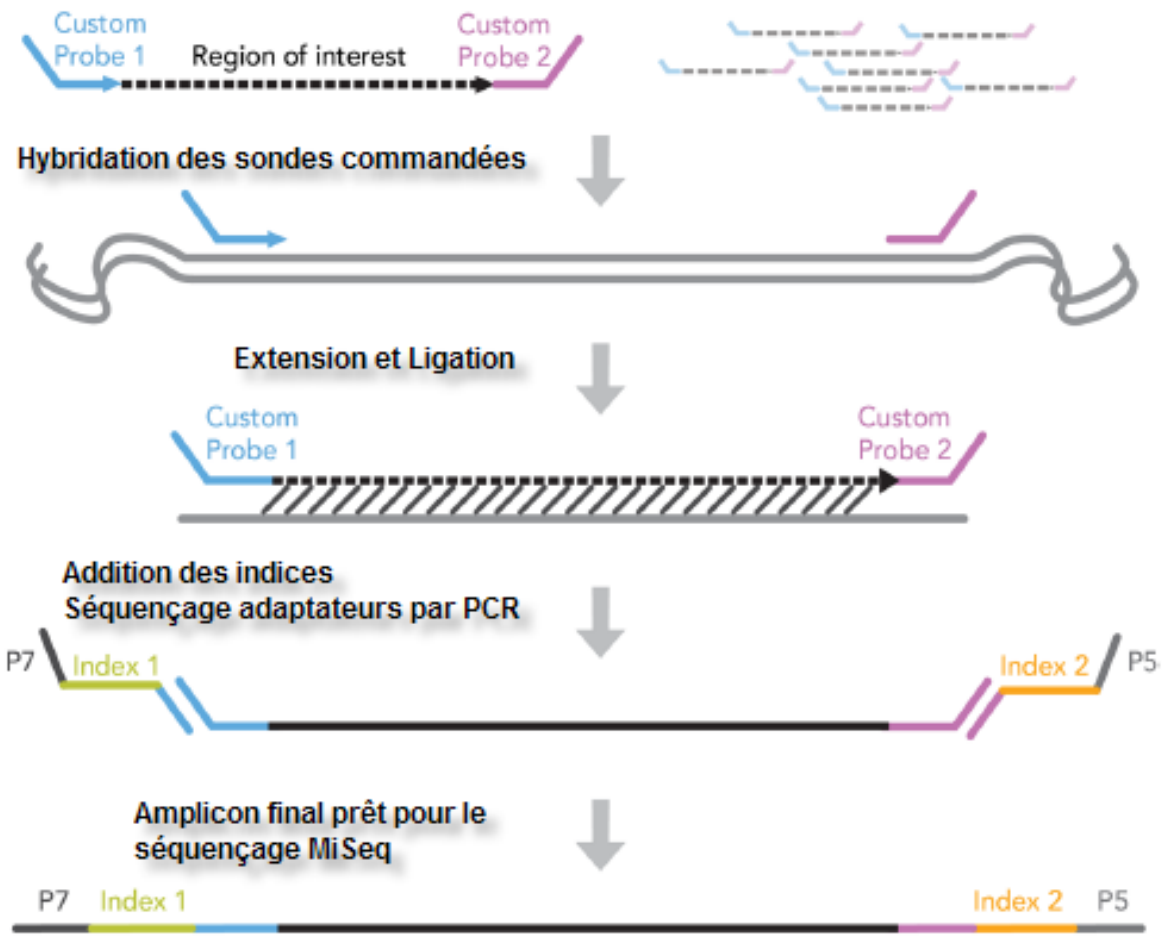


Figure 25 : Etapes du kit *TruSeq* (Illumina)

La société *Illumina* nous fournira donc un set de sondes permettant de générer les amplicons des régions d'intérêt que l'on veut séquencer. Le design de ce set de sondes à partir des coordonnées des régions d'intérêt se fait sur une application de leur site internet dans le *Design Studio*.

3.1.2.2 Le choix des gènes et des exons

Nous avons ainsi choisi le kit *TruSeq Custom Amplicon* afin de tester notre premier panel de gènes que nous appellerons NGS_V1 et qui est constitué de vingt-quatre gènes pour lesquels nous avons choisi de séquencer tous les exons de chacun des gènes excepté les parties non traduites 5' et 3'-UTR.

A l'aide du site *ENSEMBL* (http://www.ensembl.org/index.html), nous avons pu récupérer la séquence des gènes (introns et exons) et les coordonnées de chaque exon que nous avons saisies dans le *Design Studio* d'*Illumina*, permettant ainsi au logiciel de calculer le nombre d'amplicons à dessiner pour couvrir l'intégralité des gènes.

3.1.2.3 Le Design Studio d'Illumina

Les étapes du design sont les suivantes :

- Créer un compte sur le site *Illumina*.
- Se connecter au *Design Studio* (dans *Tools)*.
- Entrer le nom du nouveau projet à créer, NGS_V1 avec les informations concernant :
    - Le kit que l'on souhaite utiliser ; ici il s'agit du *TruSeqCustom Amplicon* ;
    - L'espèce ; humaine dans notre étude ;
    - La source de récupération des données du génome ; ici UCSC en version *hg19* ;
    - La taille des amplicons ; nous avons choisi des amplicons de 250pb afin que la couverture soit maximale sur un kit 2x250 (250pb séquencées en paired-end) ou 2x300pb comme cela est proposé pour les *flowcells* version 2 ou 3.

- Ajouter les régions d'intérêt ou les exons à l'aide de leurs coordonnées sur le génome de façon manuelle l'un après l'autre pour chacun des gènes ou en important un fichier « CSV » sur le format : « Name, Chromosome, StartCoordinate, StopCoordinate, TargetType, Padding,Labels ».
- Lancer le design des amplicons : nous aboutissons à un total de cinq cents amplicons.
- Vérifier le pourcentage de couverture : pour notre panel, nous avons obtenu une couverture à 96% des régions d'intérêt.

  Dans un second temps, les bio-informaticiens d'*Illumina* ont pris en charge les exons pour lesquels la couverture était faible et ont ainsi corrigé notre fichier pour obtenir une couverture à 100% de tous les exons.
- Lancer le calcul du coût pour le nombre de patients voulu : nous obtenons un devis de l'ordre de 16 000 euros pour un kit *TruSeq Custom Amplicon* de quatre-vingt-seize échantillons.

Les étapes réalisées pour cette première version du panel NGS V1 de vingt-quatre gènes nous a permis d'estimer l'importance du travail à fournir afin de définir les coordonnées de chacun des gènes mais également la collaboration nécessaire avec les équipes d'*Illumina* afin d'aboutir à la couverture maximale pour ne pas passer à côté de la moindre mutation lors de nos recherches.

Cela nous a également permis d'estimer le coût pour chaque patient et de le comparer avec une technique « manuelle » par PCR que nous avons testée en pratique ; cette technique de design assisté nous permet de nous affranchir des étapes de choix des amorces ainsi que du test de chacun des couples d'amorces, ce qui n'est pas négligeable en termes de temps et de coût de mise au point.

Nous avons décidé de refaire ces étapes pour une version 2 à quatre-vingt-un gènes communs aux LAL / LAM / SMD **(Tableau X)** afin d'avoir une idée du nombre d'amplicons de 425pb pour ce panel, et du coût de revient par patient ; ceci afin de savoir si un tel panel d'amplicons est possible à créer dans un même kit.

Après que le calcul ait été réalisé avec les données rentrées sur le site *Illumina* pour cette seconde version, nous arrivons à un total de mille-deux-cent-cinquante-sept amplicons avec une couverture de 88% (là encore nous aurons besoin de l'assistance

Tableau X : Evolution de la liste de gènes à une V2 comprenant 81 gènes

Liste des gènes V1 - 24 gènes

ASXL1	PHF6
CBL	PTEN
CEBPA	RUNX1
DNMT3A	SF3B1
ETV6	SRSF2
EZH2	TET2
FLT3	TP53
IDH1	U2AF1
IDH2	WT1
JAK2	
KIT	
KRAS	
MLL	
NPM1	
NRAS	

Liste des gènes V2 - 81 gènes

ADARB2	DNM2	ID3	MPL	RB1	TP53
ARID1A	DNMT3A	IDH1	MUC4	RELN	U2AF1
ARID5B	EBF1	IDH2	NF1	RUNX1	U2AF2
ASMTL	ECT2L	IKZF1	NOTCH1	SETBP1	VPREB
ASXL1	EED	IKZF2	NPM1	SETD2	WT1
BCL11B	EP300	IKZF3	NR3C1	SF1	ZRSR2
BCOR	ETV6	IL7R	NR3C2	SF3A1	
BLNK	EZH2	IRF8	NRAS	SF3B1	
BRAF	FBXW7	JAK1	NXPE2	SH2B3	
CADM1	FLT3	JAK2	PAG1	SRSF2	
CBL	GATA1	JAK3	PAX5	STAG2	
CDKN2A	GATA2	KIT	PHF6	SUZ12	
CEBPA	GATA3	KRAS	PRPF40B	TBL1XR1	
CEBPE	HFE	MAP3K1	PTEN	TCF3	
CREBBP	HRAS	MLL	PTPN11	TET2	

d'*Illumina* afin de maximiser la couverture), le coût serait alors de l'ordre de 20 000 euros.

Il reste encore à collaborer avec les bio-informaticiens d'*Illumina* pour que chaque région d'intérêt présente une couverture optimale mais il semble que les recherches de mutations avec ce panel ne soit pas utopique bien que le nombre d'amplicons soit colossal.

3.1.3 La PCR multiplex

Comme énoncé précédemment, dans un premier temps et afin de s'approprier la technique de NGS, nous avons choisi de réaliser un premier test de séquençage haut débit en préparant notre librairie d'amplicons par une technique de PCR. En effet, nous réalisons des PCR quotidiennement au laboratoire et le coût de mise au point est relativement faible. Nous ne choisirons que quelques gènes pour lesquels les mutations sont recherchées par des techniques classiques couplées au séquençage Sanger.

3.1.3.1 Principe de la technique

Le principe général est basé sur l'amplification par PCR des régions du génome à séquencer grâce à des amorces spécifiques comportant une queue flottante (PCR 1). Par la suite, les index et les adaptateurs *Illumina* **(Figure 26)** sont ajoutés lors d'une seconde PCR (PCR 2), avec un couple d'amorces reconnaissant les queues flottantes ajoutées à l'étape précédente.

La PCR 1 peut être réalisée en simplex avec un seul couple d'amorces par PCR ou en multiplex (plusieurs couples d'amorces dans la même PCR). Nous avons choisi ici une PCR multiplex afin que la réalisation ne soit pas trop lourde en termes de manipulation et dans l'optique de faire passer cette PCR en routine.

La première étape sera de dessiner des amorces couvrant nos régions d'intérêt.

Une fois les conditions et le nombre de PCR multiplex établis, nous réaliserons un gel de vérification ainsi qu'une purification de ces produits de PCR, avant de réaliser la PCR 2 pour l'ajout des adaptateurs *Illumina*.

3.1.3.2 Le choix de la cohorte de patients

Nous avons choisi des patients pour lesquels le diagnostic était bien défini et les techniques de recherche de mutations pour ces pathologies déjà réalisées en routine

Figure 26 : Composition des adaptateurs *Illumina* (Illumina)

SP	Amorce (SP+P5/P7 = 95pb)
P5	Adaptateur commun
P7	Adaptateur commun contenant l'index
Index	Fragment spécifique de chaque patient (6pb)
Insert	Fragment d'intérêt

pour certains exons, et ce afin de pouvoir comparer les résultats obtenus en routine et en NGS.

Nous avons ainsi choisi :

- Quatre patients atteints de LAM dont deux LAM 4 et deux LAM 1 ;
- Trois patients atteints de LLC.

En effet, actuellement en routine, nous ne recherchons aucune mutation pour les patients présentant un SMD.

En routine, pour les patients LLC, nous recherchons les mutations des exons 4 à 10 de *TP53* et pour les LAM, nous recherchons les mutations des gènes : *FLT3* (exons 13 et 20) / *NPM1* (exon 12) / *IDH1 et 2* (exon 4) / *KIT* (exon 17) / *CEBPα* / *DNMT3* (exon 23).

3.1.3.3 La recherche des mutations par les techniques « classiques »

Les séquences des amorces utilisées pour ces techniques se trouvent en annexe. **(Annexe 1)**

- Les mutations recherchées par PCR :
  - *FLT3* (ITD) / *NPM1*

    Le gène *FLT3* est muté dans 20 à 30% des LAM de l'adulte et les mutations de *NPM1* sont détectées dans 30 à 40% des LAM de l'adulte à caryotype normal.Ces mutations doivent être recherchées dans toute LAM à caryotype de risque intermédiaire au diagnostic et ont une valeur pronostique. Pour le clinicien, l'intérêt est une stratification thérapeutique adaptée au risque.

    L'analyse de l'exon 13 du gène *FLT3* se fait par PCR ; il s'agit pour ce gène de rechercher des duplications de bases pouvant être très longues or, comme ces mutations ne sont pas constantes, il est difficile de dessiner des amorces pour encadrer de petits fragments et « découper » l'exon comme c'est le cas en HRM.

    L'analyse de l'exon 12 de *NPM1* se fait également par PCR. Comme *NPM1* et *FLT3* sont les deux gènes dont les mutations sont recherchées en

systématique dans le diagnostic de LAM, nous réalisons en routine une PCR multiplex pour cette recherche.

La PCR génère des fragments de l'ordre de 450pb pour *FLT3* et 250pb pour *NPM1*, et se fait avec le kit *AmpliTaq Gold DNA Polymerase* de chez *Life Technologies* selon le protocole suivant :

	Concentration initiale	**Concentration finale**
TAMPON	10X	1X
$MgCl_2$	25mM	5mM
dNTP	10mM	0,75mM
Taq Gold	5U/µL	1,25U
Amorce F (FLT3 / NPM1)	10µM	0,3µM
Amorce R (FLT3 / NPM1)	10µM	0,3µM
ADN Patient	100ng/µL	10ng/µL
EAU qsp 25µL		

Le programme de PCR utilisé est le suivant :

1 cycle	8mn	95°C	Pré-incubation
35 cycles	30s	95°C	Amplification
	30s	54°C	
	30s	72°C	
1 cycle	10min	72°C	Cooling

- *CEBPα*

La recherche de mutations de *CEBP*α est réalisée dans toute LAM à caryotype de risque intermédiaire de sujets de moins de 65 ans.

Les mutations sont éparpillées sur le gène qui ne comporte qu'un seul exon donc deux produits PCR, PP1 et PP2, permettent de couvrir le gène entier : le produit PCR PP1 couvre la partie N-terminale de la protéine et le produit PCR PP2 couvre la partie C-terminale ; les mutations peuvent être mono-alléliques ou bi-alléliques (en général une sur PP1 et une sur PP2) *[5]*.

Le gène CEBPα est riche en GC et donc difficilement amplifiable ; nous utilisons le kit *AmpliTaq Gold 360 GC* de chez *Life Technologies* comprenant un *enhancer* pour faciliter l'amplification de régions riches en GC.

Nous utilisons le protocole de PCR suivant :

	Concentration initiale	Concentration finale
ENHANCER 360	5X	1X
AMPLITAQ GOLD 360 CG	2X	1X
Amorce F (PP1 / PP2)	10µM	0,3µM
Amorce R (PP1 / PP2)	10µM	0,3µM
ADN Patient	100ng/µL	10ng/µL
EAU qsp 25µL		

Le programme de PCR utilisé est le suivant :

1 cycle	10mn	95°C	Pré-incubation
35 cycles	30s	95°C	Amplification
	30s	56°C	
	1mn	72°C	
1 cycle	7min	72°C	Cooling

Les tailles attendues sont de l'ordre de 580pb et 720pb pour PP1 et PP2.

- DNMT3a

La recherche des mutations de l'exon 23 du gène *DNMT3* n'est pas effectuée en routine mais nous avons recherché ces mutations chez les patients ayant une LAM, en suivant le protocole utilisé pour *CEBPα* en utilisant un seul couple d'amorce couvrant la totalité de l'exon et générant un fragment de l'ordre de 230pb.

A la fin des PCR, les amplicons générés sont séquencés par la méthode Sanger.

➢ Les mutations recherchées par HRM

Les gènes pour lesquels les mutations sont recherchées par HRM sont *FLT3* exon 20 (TKD), *KIT* exon 17 et *IDH1/2* exon 4 dans les LAM, et *TP53* exons 4 à 10 dans les LLC.

Le protocole indiqué ci-dessous, est identique pour ces différents exons et nous utilisons le kit *LC480 HRM High Resolution Melting* de chez *Roche* comprenant le master mix et le $MgCl_2$, en ajoutant, pour chaque exon le couple d'amorces spécifique de la région à amplifier.

	Concentration initiale	Concentration finale
LC 480 HRM Master Mix	2X	1X
$MgCl_2$	25mM	3mM
Amorce F	10µM	0,2µM
Amorce R	10µM	0,2µM
ADN Patient	10ng/µL	2ng/µL
EAU qsp 10µL		

La PCR en temps réel est réalisée sur le *Light cycler 480* de chez *Roche*, selon le programme suivant :

1 cycle	10mn	95°C	Dénaturation
50 cycles	15s	95°C	Amplification
	15s	63°C	
	25s	72°C	
1 cycle	1mn	95°C	Fusion
	1mn	40°C	
	5s	70°C	
		70 à 95°C	Augmentation progressive de la température

Après l'analyse HRM, les patients retrouvés positifs sont séquencés en Sanger.

- L'analyse des séquences

La première étape consiste à vérifier les séquences à l'aide du logiciel *FINCH TV* en ouvrant la séquence à analyser « AB1 ».

Les pics de l'électrophorégramme doivent être bien distincts les uns des autres et doivent présenter une bonne intensité par rapport au bruit de fond ; il faut également vérifier la longueur de la séquence d'après la taille attendue des produits PCR.

L'analyse se poursuit sur le logiciel *Mutation surveyor* permettant d'aligner les séquences sur la séquence de référence et de détecter les mutations présentes.

Le logiciel présente son analyse sous forme de tableau avec des « contig » correspondant à l'assemblage des séquences des différents patients et leur alignement sur le génome afin de faire apparaître les variations présentes **(Figure 27)**. Dans le cas où une variation apparaît

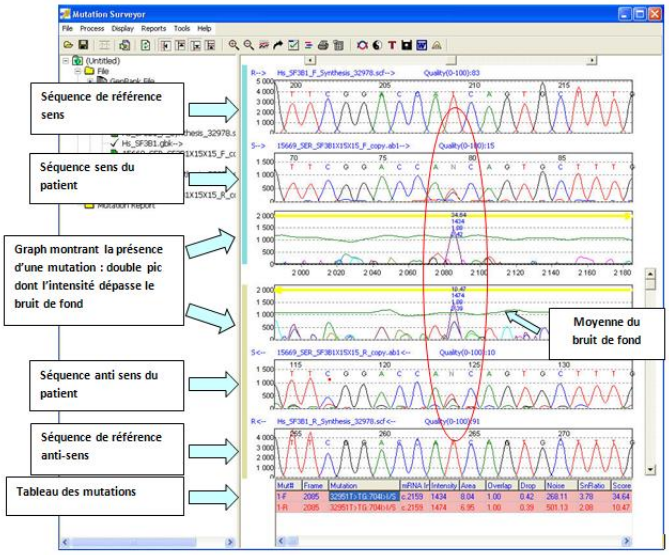


Figure 27 : Visualisation d'une mutation sur le logiciel d'analyse Mutation Surveyor

(insertion / délétion / substitution), il faut vérifier l'alignement de la séquence par rapport à la séquence de référence.

Pour chaque variation de séquence détectée dans *Mutation Surveyor*, il convient de vérifier si cette mutation a déjà été décrite et s'il ne s'agit pas d'un polymorphisme connu (ou SNP) correspondant à une variation de nucléotide existant à l'état physiologique et germinale, n'entrainant pas de conséquences pathologiques.

Il faut donc ouvrir l'outil d'analyse *BLAT* de *Ensembl Genome Browser* (http://www.ensembl.org/Multi/blastview), un algorithme d'alignement qui permet de coller une séquence comprenant la variation en format « FASTA ». Si aucun SNP n'est décrit à l'endroit où se trouve la variation on peut rechercher si cette mutation a déjà été décrite à l'aide d'une seconde base de donnée : *COSMIC* (http://www.sanger.ac.uk/perl/genetics/CGP/cosmic) qui répertorie les mutations somatiques connues.

Enfin, pour décrire précisément la mutation avec la nomenclature internationale, on utilise *Mutalyser* (https://mutalyzer.nl/) en entrant le nom de la séquence de référence du gène d'intérêt, puis en sélectionnant « Coding DNA » ; il s'agit ensuite de choisir le type de mutation dans le menu déroulant et de compléter les informations demandées.

- Les résultats

Nous avons regroupé les résultats de séquence pour les différents gènes testés dans un tableau en association avec les informations correspondant à chacun des patients de la cohorte. **(Tableau XI)**

3.1.3.4 Le design des amorces de NGS

La séquence d'ADN génomique du gène d'intérêt pour lequel nous souhaitons concevoir les amorces de PCR, peut être obtenue grâce au site de l'*UCSC* en fournissant le nom du gène d'intérêt ou les coordonnées chromosomiques de la région recherchée sur l'assemblage *hg19* du génome humain *[68]*.

Tableau XI : Résultats des recherches de mutations par les techniques « classiques » sur notre cohorte de patients tests

Nom	Sexe	DDN	Date Examen	Nature	Cible de PCR qualitative	Résultat PCR qualitative	Séquençage Sanger	Pathologie
MOUCH	M	21/12/58	26/05/05	MOELLE	FLT3_X13_ITD	Muté		LAM4
					FLT3_X20_TKD	Non muté		
					NPM1_X12	Muté	Type A	
					IDH1_X04	Muté	R132H	
					IDH2_X04	Non muté		
					KIT_X17	Non muté		
					CEBPA	Muté	Ins/del 590F/568R; 679_684het_dupCGCACC	
					DNMT3A_X23	Non muté		
DELCO	F	29/07/61	18/06/05	MOELLE	FLT3_X20_TKD	Muté	D835Y	LAM4
					FLT3_X13_ITD	Non muté		
					NPM1_X12	Muté	Type D	
					IDH1_X04	Muté	R132H	
					IDH2_X04	Non muté		
					KIT_X17	Non muté		
					CEBPA	Non muté		
					DNMT3A_X23	Muté	R882H	
DESMI	M	12/11/26	25/04/08	MOELLE	FLT3_X13_ITD	Muté	+42 pb, Normal>Tumoral	LAM1
					NPM1_X12	Muté	Type A	
					IDH1_X04	Non muté		
					IDH2_X04	Non muté		
					KIT_X17	Non muté		
GERJE	M	20/02/52	02/11/05	MOELLE	FLT3_X13_ITD	Muté		LAM1
					FLT3_X20_TKD	Non muté		
					NPM1_X12	Non muté		
					KIT_X17	Muté	D816V	
					DNMT3A_X23	Non muté		
LAGAN	M	07/04/42	07/03/11	SANG	TP53_X04	Non muté		LLC
					TP53_X05	Muté	K132N ; mutation hétérozygote	
					TP53_X06	Non muté		
					TP53_X07	Non muté		
					TP53_X08	Non muté		
					TP53_X09	Non muté		
					TP53_X10	Non muté		
LIOMI	M	29/10/39	06/10/08	SANG	TP53_X04	Non muté		LLC
					TP53_X05	Non muté		
					TP53_X06	Non muté		
					TP53_X07	Non muté		
					TP53_X08	Muté	R273C	
					TP53_X09	Non muté		
					TP53_X10	Non muté		
CARMI	M	12/03/44	04/11/09	SANG	TP53_X04	Muté	517_528delACGGTTTCCGTC	LLC
					TP53_X05	Non muté		
					TP53_X06	Non muté		
					TP53_X07	Non muté		
					TP53_X08	Non muté		
					TP53_X09	Non muté		
					TP53_X10	Non muté		

Il s'agit ensuite d'encadrer la région d'intérêt par les amorces à dessiner sachant que nous souhaitons obtenir des amplicons dont la taille serait de l'ordre de 250pb.

Nous nous sommes basés sur les amorces utilisées pour les techniques classiques de recherche de mutations et nous avons redessiné les amorces pour les amplicons de taille trop éloignée de la taille souhaitée. Une fois ce premier design réalisé, nous avons ajouté les queues flottantes permettant l'ajout des adaptateurs *Illumina* aux amplicons lors de la seconde PCR de préparation des librairies :

- Queue amorce sens : CTTTCCCTACACGACGCTCTTCCGATCT ;
- Queue amorce anti-sens : GGAGTTCAGACGTGTGCTCTTCCGATCT.

3.1.3.5 La qualification des amorces de NGS

Notre objectif étant de réaliser une PCR multiplex avec l'ensemble des amorces, nous devons, dans un premier temps, tester individuellement les amorces dessinées afin de savoir si ces dernières permettent la synthèse des amplicons désirés et dans quelles conditions.

Pour cela nous avons à notre disposition un kit *AmpliTaq Gold 360GC* dans lequel un *enhancer* est compris et spécialement formulé pour aider à l'amplification des régions riches en GC.

Dans un premier temps, nous allons tester chaque couple d'amorces individuellement avec ce kit, en présence ou non de l'*enhancer,* afin de déterminer les conditions optimales pour chaque couple d'amorces.

Nous avons donc réalisé une RQ-PCR (Real quantification PCR) sur de l'ADN de PBL (Peripheric blood Lymphocyte) avec la Taq Polymerase qui sera utilisée pour les PCR multiplex, un agent intercalant fluorescent *(le Resolight)*, en présence ou non de l'*enhancer*, selon le protocole suivant :

	Concentration initiale	**Concentration finale**
RESOLIGHT	20X	1X
+/ - ENHANCER 360 GC	5X	1X
AMPLITAQ GOLD 360 GC	2X	1X
Couple amorces F/R	10ng/µL	0,2µM
ADN PBL	100ng/µL	20ng/µL
EAU qsp 25µL		

Nous avons réalisé la PCR sur le *LC480* (appareil de PCR en temps réel) selon le programme suivant :

1 cycle	5mn	95°C	Pré-incubation
55 cycles	10s	95°C	Amplification
	10s	56°C	
	10s	72°C	
1 cycle	5s	95°C	Melting curve
	1mn	65°C	
		65 à 95°C	Augmentation graduelle de la température
1 cycle	30s	40°C	Cooling

L'agent fluorescent utilisé va s'intercaler dans l'ADN double brin nouvellement formé lors de la PCR ; la fluorescence alors émise sera proportionnelle à la quantité d'ADN présente dans le milieu.

La fluorescence émise va alors franchir une valeur seuil, à un certain cycle de PCR appelé Ct (Cycle threshold), or ce Ct est proportionnel à la quantité initiale d'ADN présente.

Pour chaque couple, nous avons comparé les résultats obtenus avec ou sans *enhancer* afin de déterminer la meilleure condition ; dans le cas où aucune des deux conditions ne donnait satisfaction (Ct > 40), nous avons redessiné et re-testé les amorces.

Au final, seul le couple d'amorces de FLT3 conserve un Ct > 40 mais étant donné que l'amplicon généré est de grande taille (450pb), il aurait été nécessaire d'augmenter le temps d'élongation ; nous avons décidé de conserver les amorces telles quelles d'autant que les mutations de *FLT3* sont éparses.

Nous avons récapitulé les résultats dans un tableau. **(Tableaux XII)**

Malheureusement, il n'a pas été possible de trouver une unique condition idéale pour l'ensemble des amorces donc nous réaliserons deux PCR multiplex (l'une en présence de l'enhancer et l'autre sans) qui seront mélangées après la PCR pour chacun des échantillons. La répartition des couples d'amorces dans chacune des deux multiplex est présentée dans le tableau XV.

Tableau XII : Valeurs de Ct obtenues pour le test des couples d'amorces en SyBrGreen avec et sans Enhancer

Tests amorces AVEC/SANS Enhancer

	PBL 100ng/µL	
AMORCES F/R	***AVEC***	***SANS***
FLT3_X20	Ct = 39,55	Ct = 35,10
NPM1_X12	Ct = 47,4	Ct = 36
FLT3_NGS*	/	Ct = 47,1
PP1_bR	Ct = 32,62	Ct = 34,85
PP1_Fa	Ct = 33,25	/
PP2_cR	Ct = 36,45	Ct = 45,50
PP2_bisFa*	Ct = 36,40	Ct = 43,6
PP2_bb	Ct = 32,64	/
IDH1_X04	Ct = 37	Ct = 32,96
IDH2_X04	Ct = 36,67	Ct = 37,30
DNMT3_X23	Ct = 32,10	Ct = 32,60
Kit_X17	/	Ct = 39,10
TP53_X04_bis*	Ct = 50	Ct = 38,60
TP53_X05*	Ct = 48,40	Ct = 36,50
TP53_X06	Ct = 37,10	Ct = 36,75
TP53_X07	Ct = 33,45	Ct = 35,20
TP53_X08_bis*	Ct = 50	Ct = 35
TP53_X09	Ct = 38,25	Ct = 38,03
TP53_X10	Ct = 37,59	Ct = 37,50

** Amorces re-dessinées et re-commandées*

ROUGE : Amorces fonctionnant avec Enhancer
JAUNE : Amorces fonctionnant sans Enhancer
ORANGE : Amorces fonctionnant avec et sans Enhancer

La liste des séquences des amorces utilisées pour la PCR multiplex du NGS est ajoutée en annexe. **(Annexe 2)**

3.1.3.6 La génération des amplicons lors de la PCR 1

Nous avons donc réalisé pour chaque échantillon deux PCR multiplex, pour lesquelles la composition des mix de PCR est présentée en annexe **(Annexe 3)**. Pour les sept patients choisis ***(Voir paragraphe 3.1.3.1 de la partie Matériel et Méthodes)***, nous avons réalisé une PCR en conditions que nous appellerons « standard » sur 30 cycles et avec une température d'hybridation de 56°C.

Par ailleurs, nous avons voulu savoir si les conditions de PCR pouvaient influer sur les résultats du séquençage ; pour cela, nous avons réalisé des PCR sur des PBL en variant les paramètres pour tester :

- L'influence de la température d'hybridation sur quatre PCR de 30 cycles à 54°C / 56°C / 58°C ou 60°C comme température d'hybridation;
- L'influence du nombre de cycles de PCR sur trois PCR à 56°C avec 27 / 30 et 33 cycles ;

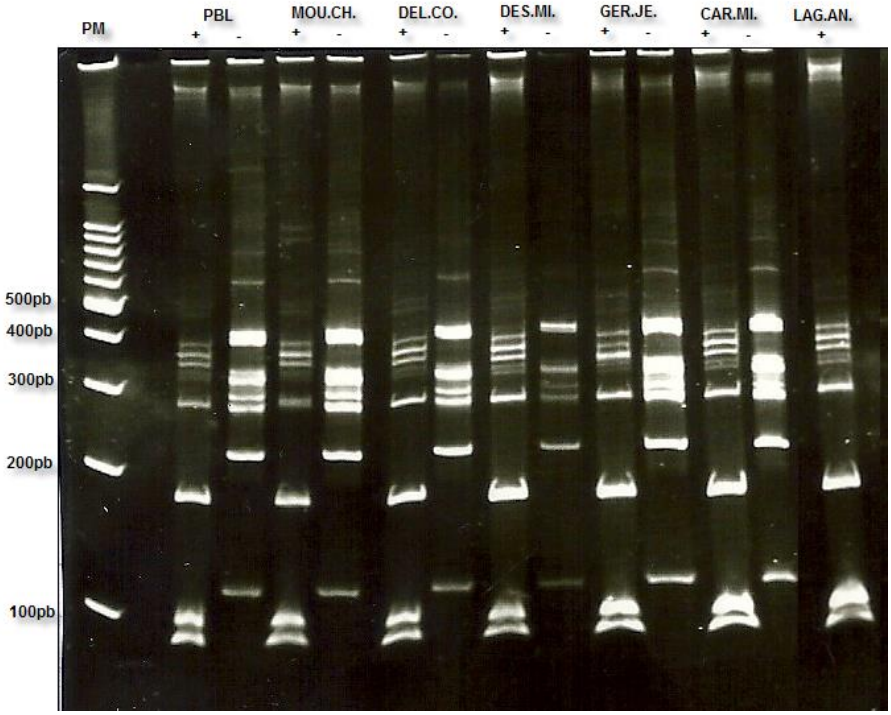
Enfin, nous avons choisi de tester un mix avec toutes les amorces (dix-neuf sets) en présence ou non d'*enhancer*, sur une PCR avec une température d'hybridation de 56°C en 30 cycles.

Le programme de PCR utilisé est le suivant :

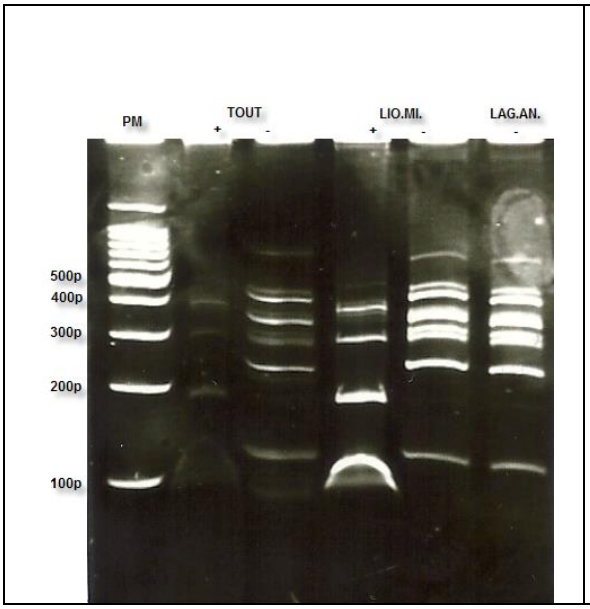
1 cycle	10mn	95°C
27/30/33 cycles	30s	95°C
	30s	54/56/58/60°C
	60s	72°C
1 cycle	7 mn	72°C

3.1.3.7 La quantification des produits de la PCR 1

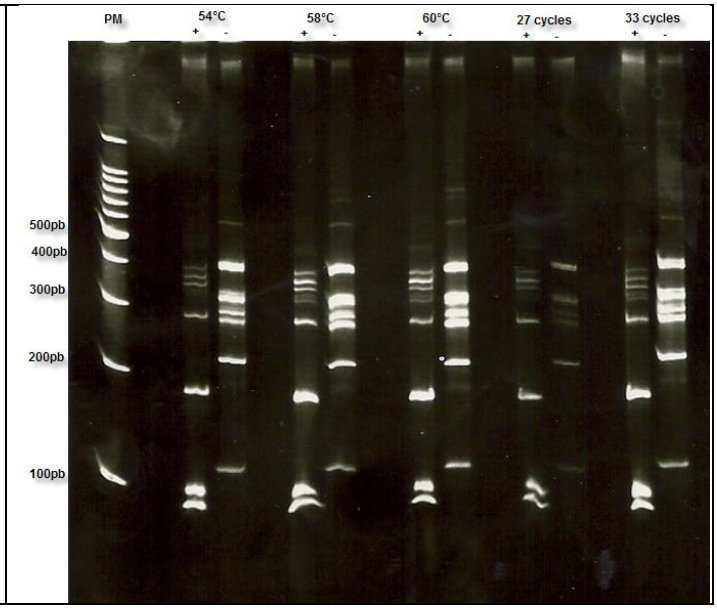
Nous avons ensuite réalisé un gel d'acrylamide afin de vérifier la présence de bandes pour chacun des vingt-huit tubes et d'estimer si leur taille correspondait à celles attendues **(Figure 28)**. L'analyse des gels révèle des fragments dont la taille était attendue mais également des bandes de taille très inférieure non spécifiques pouvant correspondre à des amorces croisant et amplifiant des fragments plus courts.



PCR 56°C et 30 cycles



PCR 56°C et 30 cycles (suite)



PCR diverses conditions

Figure 28 : Taille des fragments obtenus pour la PCR 1 de NGS sur gel d'acrylamide

Enfin, nous avons dosé chacun des échantillons post-PCR **(Tableau XIII)** afin d'avoir une idée de la quantité de matériel disponible et de mélanger les tubes de produits PCR avec et sans *enhancer* de chaque échantillon, en « quantité » égale ; au vu des dosages obtenus, nous avons mélangé volume à volume.

Au final, nous conservons quinze échantillons que nous devons purifier avant l'étape de PCR 2.

3.1.3.8 La purification des produits de la PCR 1

La purification est réalisée à l'aide de billes magnétiques du kit *AMPure XP* de chez *Bekman Coulter*. Ces billes ont la propriété de capter les fragments d'ADN à leur surface. L'utilisation d'un portoir magnétique permettra donc de purifier l'ADN, c'est-à-dire d'éliminer les tampons de la PCR.

Le protocole se décompose ainsi :

- Ajouter les billes magnétiques au produit de PCR et homogénéiser ;
- Déposer la plaque sur un portoir magnétique et éliminer le surnageant une fois le culot de billes formé contre la paroi du tube.
- Réaliser deux lavages en éthanol à 80% puis laisser sécher ;
- Re-suspendre le culot dans du tampon d'élution et homogénéiser ;
- Déposer à nouveau la plaque sur le portoir magnétique et transférer le maximum de surnageant dans une nouvelle plaque.

Un nouveau dosage sur *Nanodrop* est réalisé afin de pouvoir préparer les volumes à utiliser pour la seconde PCR d'intégration des index et des adaptateurs d'*Illumina*.

3.2 Le NGS à proprement parlé

Toutes les étapes post-PCR 1 sont effectuées sur la plateforme de séquençage GenoToul – INRA de Toulouse.

La première étape consiste à finaliser la préparation de la librairie par l'ajout des index et adaptateurs, puis les étapes se poursuivent jusqu'à la création des *clusters* et au lancement du *run* de séquençage sur le *MiSeq*.

Tableau XIII : Dosage des produits post PCR 1 pour chaque échantillon

Conditions PCR	Identification	Dosage (ng/µL)
56°C 30 cycles	PBL + Enh	382
	PBL - Enh	408
	MOU.CH. + Enh	397
	MOU.CH. - Enh	402
	DEL.CO. + Enh	385
	DEL.CO. - Enh	419
	DES.MI. + Enh	393
	DES.MI. - Enh	409
	GER.JE. + Enh	383
	GER.JE. - Enh	401
	CAR.MI. + Enh	382
	CAR.MI. - Enh	401
	LIO.MI. + Enh	394
	LIO.MI. - Enh	400
	LAG.AN. + Enh	391
	LAG.AN. - Enh	380
	TOUT + Enh	391
	TOUT - Enh	416
54°C 30 cycles	PBL + Enh	379
	PBL - Enh	402
58°C 30 cycles	PBL + Enh	384
	PBL - Enh	408
60°C 30 cycles	PBL + Enh	388
	PBL - Enh	424
56°C 27 cycles	PBL + Enh	375
	PBL - Enh	376
56°C 33 cycles	PBL + Enh	371
	PBL - Enh	381

3.2.1 L'intégration de l'index et des adaptateurs lors de la PCR 2

Comme nous allons mélanger les différents échantillons lors de la réaction de séquence, il est nécessaire de pouvoir différencier les séquences spécifiques de chaque échantillon pour interpréter les données. Pour cela nous rajoutons des index (séquence de 6 bases différente pour chaque échantillon) à la suite de chaque amplicon avant les adaptateurs.

L'ajout des adaptateurs P5 et P7 qui sont nécessaires à la fixation des amplicons sur la flowcell lors du séquençage, ainsi que de l'index se fait grâce à une seconde PCR (PCR 2) utilisant des amorces qui reconnaissent la queue flottante ajoutée lors de la PCR 1**(Figure 29)**. Cette PCR 2 utilise une *MTP Taq polymerase* de chez *Sigma Aldrich* qui comprend les réactifs suivants :

	Concentration initiale	**Concentration finale**
Tampon MTP	10X	1X
DNTP	10mM	0,75mM
P5	20µM	0,6µM
P7 spécifique	20µM	0,6µM
MTP DNA Taq Polymerase	5U/µL	1,25U
Produit PCR1 purifié	50 à 100ng/µL	5 à 10ng/µL
EAU qsp 25µL		

C'est la concentration définie lors du dosage qui permet de calculer le volume de produit de PCR 1 à ajouter au mix. **(Tableau XIV)**

La PCR est ensuite lancée sur le programme suivant :

1 cycle	60s	94°C	Pré-incubation
12 cycles	60s	94°C	Amplification
	60s	65°C	
	60s	72°C	
1 cycle	10mn	72°C	

3.2.2 La purification et le dosage des produits de la PCR 2

Les étapes sont identiques à celles réalisées à la suite de la PCR1. ***(Voir paragraphe 3.1.3.7 de la partie Matériel et Méthodes)***

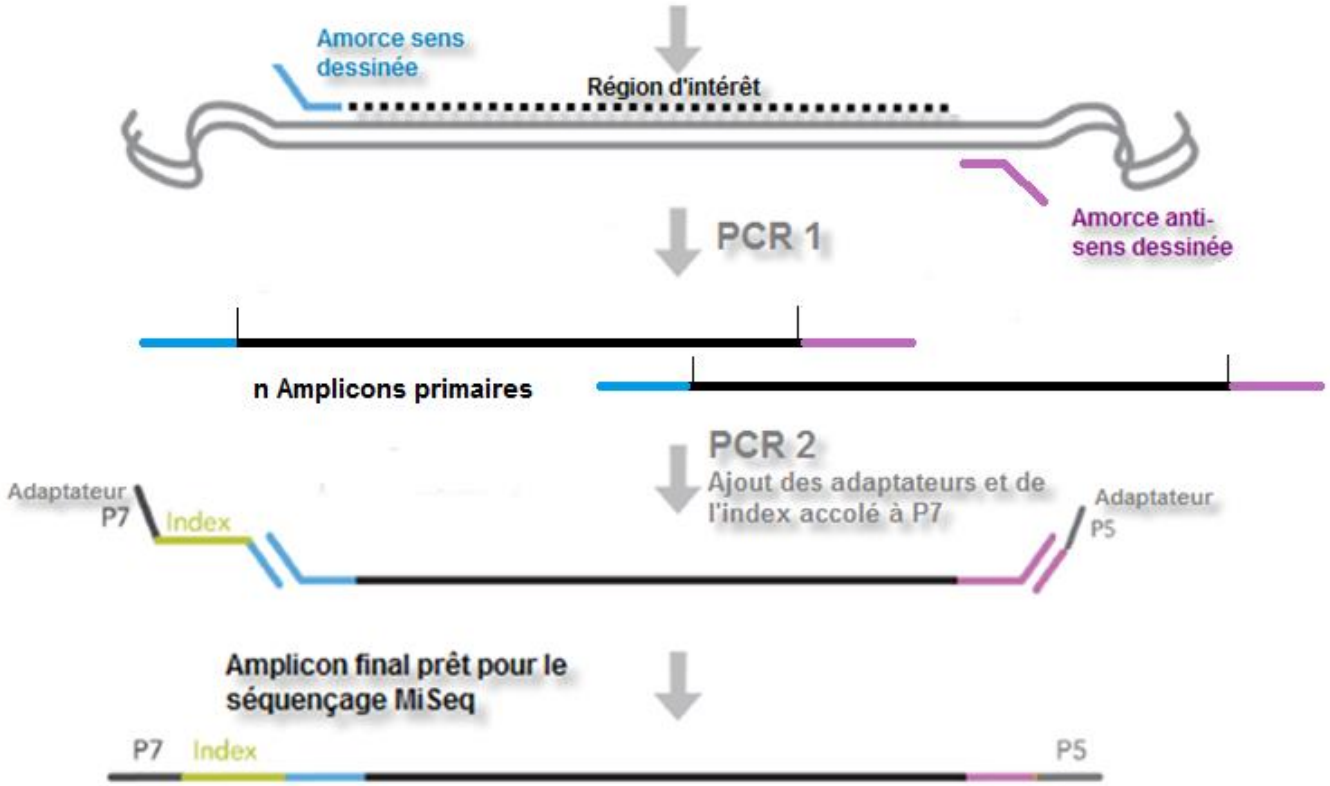


Figure 29 : Workflow de la préparation de la librairie en technique d'enrichissement par PCR mulitplex

Tableau XIV : Dosage des produits de PCR1 purifiée à ajouter dans le mix de PCR2 et concentration initiale de PCR2

Nom des échantillons	Dosage (ng/µL) de la PCR1	Volume d'ADN (µl) pour PCR2	Quantité initiale PCR2 (ng)	N° d'index associé à l'adaptateur P7
56_30_PBL	38,92	2	77,84	I01
56_30_MOU.CH.	34,02	2	68,04	I03
56_30_DEL.CO.	23,45	2	46,9	I04
56_30_DES.MI.	26,34	2	52,68	I05
56_30_GER.JE.	42,96	2	85,92	I06
56_30_CAR.MI.	45,88	2	91,76	I07
569_30_LIO.MI.	37,9	2	75,8	I08
56_30_LAG.AN.	49,24	2	98,48	I09
56_30_TOUT_ENH	18,76	2	37,52	I10
56_30_TOUT	26,21	2	52,42	I11
54_30_PBL	38,98	2	77,96	I12
58_30_PBL	37,95	2	75,9	I13
60_30_PBL	37,82	2	75,64	I14
56_27_PBL	23,63	2	47,26	I15
56_33_PBL	52,54	2	105,08	I16

3.2.3 La qualification des produits de PCR 2 sur puce

Cette étape ainsi que celle de la PCR quantitative correspondent au contrôle qualité de la librairie ainsi créée et à l'estimation de la qualité des échantillons.

Cette analyse se fait sur le *BioAnalyzer – Agilent* par lecture de la taille des fragments après préparation d'une puce *Agilent DNA 7500*.

La puce est composée de plusieurs puits, dont douze destinés aux échantillons, et un au marqueur de poids moléculaire qui sert d'étalon ; au niveau de ces puits, il existe des micro-réseaux. On dépose d'abord un gel filtré mélangé à un intercalant de l'ADN spécifique et fluorescent et ce dernier est réparti au niveau de chaque puits, et remplit les micro-réseaux. Un tampon est alors ajouté dans chaque puits, afin de placer ensuite les échantillons dans un milieu stable, à un pH donné. La puce est vortexée afin d'homogénéiser le tampon et les échantillons, puis placée dans l'appareil et la migration est lancée.

Une électrode est en contact avec chaque puits et au moment de l'analyse, cette électrode est soumise à un haut voltage, qui va permettre une séparation très rapide des brins d'ADN de l'échantillon en fonction de leur taille ; au moment de leur passage dans le gel, les molécules d'ADN vont se lier spécifiquement à l'intercalant.

Enfin, les molécules sont détectées par fluorescence : les molécules sont excitées par un rayonnement UV, et seuls les ADN liés à l'intercalant vont réémettre un rayonnement fluorescent. L'appareil mesure, en temps réel, l'intensité de la fluorescence émise en fonction du temps de rétention de l'échantillon au sein du micro-réseau. Ainsi, plus les ADN sont courts, plus ils migrent rapidement et donc plus leur temps de rétention est faible et plus l'intensité de fluorescence est importante, plus la quantité d'ADN est importante.

Les résultats obtenus **(Figure 30)** nous donnent une indication de la taille du fragment généré, de la concentration ainsi que de la molarité de l'échantillon ; ils montrent ici des profils similaires pour tous les échantillons avec quelques différences de pics qui peuvent être dues à la concentration de l'échantillon ainsi qu'à sa nature même. De plus, la majorité de nos amplicons générés ont une taille relativement proche de 250pb ce qui explique que nous obtenons des pics assez larges. Les profils diffèrent pour les deux échantillons pour lesquels nous avons mélangé la totalité de

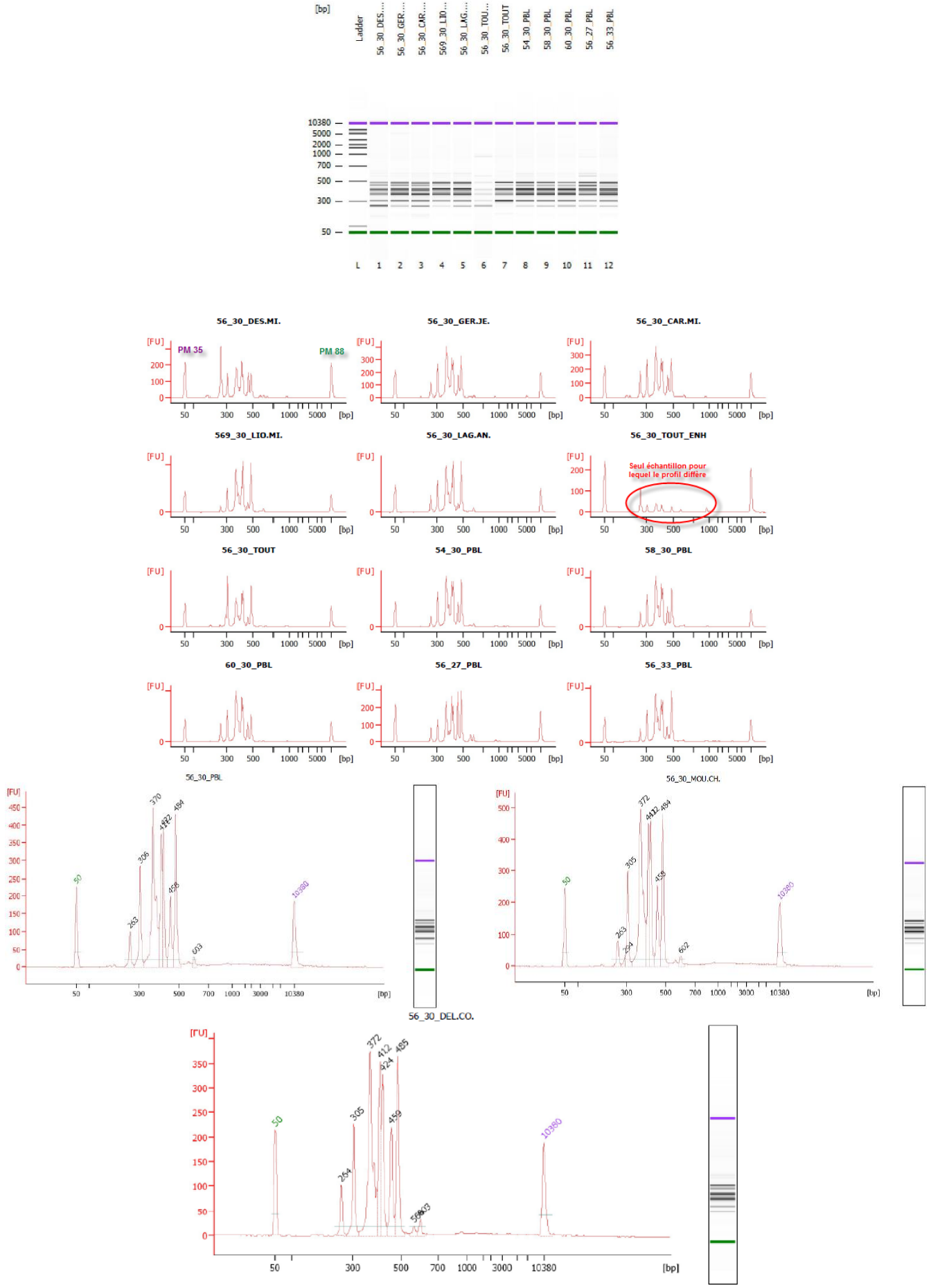


Figure 30: Résultats obtenus d'après la puce *Agilent* sur la PCR 2 de NGS.

nos amorces, avec une qualité nettement inférieure pour celui pour lequel le mix a été préparé avec l'*enhancer*.

Nous avons ainsi pu vérifier si les résultats concordaient avec nos résultats de gels et les tailles théoriques des amplicons générés **(Tableau XV)**, ce qui est le cas pour tous nos échantillons. Il faut bien sur considérer que les adaptateurs ajoutés lors de la PCR 2 ajoutent environ 120pb à la taille évaluée, ce qui est à rajouter aux 56pb des queues flottantes des amorces de NGS.

3.2.4 L'obtention de la librairie de séquençage

Afin d'obtenir la librairie de séquençage, il faut préparer un pool équimolaire de nos quinze échantillons post-PCR 2 à l'aide des concentrations issues du dosage au *Nanodrop* de la PCR 2. **(Tableau XVI)**

Le tube obtenu contient donc l'ensemble des amplicons à séquencer auxquels ont été ajouté les index et les adaptateurs *Illumina*, pour tous les échantillons de départ : il s'agit de la librairie de séquençage.

3.2.5 La PCR quantitative de contrôle de la librairie

Cette étape permet de « quantifier » la librairie et de connaitre la concentration des produits PCR ayant bien intégré les adaptateurs, seuls produits qui seront séquencés sur le *MiSeq*.

La quantification précise du nombre de molécules amplifiables dans une librairie est critique pour le bon déroulement d'une technique de NGS ; en effet, une surestimation de la concentration de cette librairie entrainerait une densité plus faible de *clusters* générés et une sous-estimation entrainerait la génération d'un trop grand nombre de *clusters* sur la *flowcell*, responsable d'une faible résolution.

Cette PCR quantitative en SybrGreen se fait à l'aide du *KAPA Library quantification kit* de *Cliniscience* et préconisé par *Illumina* pour un thermocycleur défini. Le kit contient une gamme de contrôles allant de 0,002pm à 20pm d'où la nécessité, lors de la préparation de la plaque de diluer notre librairie au centième, millième et dix-millième car la concentration de notre pool est de l'ordre du nano-molaire. Ce kit *KAPA* fournit également les amorces construites avec les adaptateurs *Illumina* selon les séquences suivantes :

- Primer P1 : 5'- AAT GAT ACG GCG ACC ACC GA - 3' ;

Tableau XV : Taille des amplicons attendue avec queue flottante et avec adaptateurs

COUPLES D'AMORCES	TAILLE AMPLICON THEORIQUE	TAILLE AMPLICON + QUEUE FLOTTANTE	TAILLE AMPLICON + QUEUE FLOTTANTE + ADAPTATEURS
FLT3_X20_NGS_F1 / FLT3_X20_NGS_R1	280	336	456
NPM1_X12_F1 / NPM1_X12_R1	243	299	419
FLT3_NGS_F / FLT3_NGS_R	468	524	644
PP1b_NGS_F / PP1_R	321	377	497
PP1_F / PP1a_NGS_R	301	357	477
PP2c_NGS_F / PP2_R	273	329	449
PP2_bis_F / PP2a_bis_R	238	294	414
PP2b_NGS_F / PP2b_NGS_R	347	403	523
IDH1_X4_F2 / IDH1_X4_R2	235	291	411
IDH2_F1 / IDH2_R2	353	409	529
DNMT3A_X23_F1 / DNMT3A_X23_R1	228	284	404
KIT _X17_NGS_F1 / KIT_X17_R3	326	382	502
TP53_X04_NGS_bis_F1 / TP53_X04_R22	251	307	427
TP53_X05_F12 / TP53_X05_NGS_R	284	340	460
TP53_X06_NGS_F1 / TP53_X06_R1	228	284	404
TP53_X07_NGS_F1 / TP53_X07_R1	237	293	413
TP53_X08_F1 / TP53_X08_NGS_bis_R1	273	329	449
TP53_X09_NGS_F1 / TP53_X09_R1	263	319	439
TP53_X10_NGS_F1 / TP53_X10_R1	273	329	449

Tableau XVI : Volume de PCR 2 à mélanger de façon équimolaire pour la PCR quantitative

Nom des échantillons	Dosage (ng/µL) de la PCR2	Quantité pour pool (ng)	Volume de librairie (µL) pour le pool
56_30_PBL	71,26	150	2,1
56_30_MOU.CH.	70,05	150	2,1
56_30_DEL.CO.	64,15	150	2,3
56_30_DES.MI.	44,35	150	3,4
56_30_GER.JE.	59,71	150	2,5
56_30_CAR.MI.	60,13	150	2,5
569_30_LIO.MI.	73,18	150	2,0
56_30_LAG.AN.	69,55	150	2,2
56_30_TOUT_ENH	7,567	150	19,8
56_30_TOUT	56,33	150	2,7
54_30_PBL	77,04	150	1,9
58_30_PBL	75,46	150	2,0
60_30_PBL	71,98	150	2,1
56_27_PBL	45,69	150	3,3
56_33_PBL	78,1	150	1,9

- Primer P2 : 5'- CAA GCA GAA GAC GGC ATA CGA - 3'.

Les résultats de la PCR quantitative obtenus nous ont permis de conclure que la concentration de notre librairie en amplicons séquençables était de 9,75nM.

3.2.6 Le séquençage sur le MiSeq

3.2.6.1 Le choix des paramètres du séquençage

Il y a un certain nombre de choix à réaliser sur les paramètres du séquençage qui vont influer sur la qualité et la profondeur de lecture des séquences obtenues :

- Le séquençage en *single-end* ou *paired-end* :

  Lors d'un séquençage en *paired-end*, il y a une lecture de chaque séquence à partir de chacune des extrémités de l'amplicon, c'est-à-dire que pour 500 cycles, il y aura 250 bases lues à partir de l'adaptateur P5 puis 250 bases lues à partir de l'adaptateur P7. Pour un amplicon de moins de 500 pb, il y aura donc un recouvrement des deux séquences ce qui augmente la qualité de lecture des bases.

  A l'inverse en « single-end », le séquençage ne se fait que par l'un des deux adaptateurs.

  Afin d'augmenter la qualité de nos séquences pour ce premier essai de séquençage haut débit nous avons choisi de réaliser un séquençage en *paired-end*.

  Avec un kit de 500 cycles, nos amplicons qui ont une taille proche de 250 pb bénéficieront d'un recouvrement quasi-total au cours des deux lectures.

- Le choix de la *flowcell* :

  Illumina propose plusieurs *flowcells* de capacités différentes allant de 1 million de « reads » ou séquences lues pour 500Mb données générées à 15 millions de séquences lues pour 7,5 Gb de données générées **(Tableau XVII).**

  En fonction de la capacité de la *flowcell* choisie (nombre de séquences lues), du nombre d'amplicons à séquencer et du nombre d'échantillons multiplexés, on obtiendra une profondeur de lecture théorique moyenne pour le run de séquençage.

Tableau XVII : Comparaison des différentes Flow Cells disponibles (Illumina)

Flow Cell	# of Reads	Kit Size (cycles)	Output (Max.)	2 x 75 Output (runtime)	2 x 150 Output (runtime)	2 x 250 Output (runtime)
Standard Flow Cell	15 M	50, 300, 500	7.5 Gb	2.25 Gb (~13 hrs)	4.5 Gb (~24 hrs)	7.5 Gb (~40 hrs)
Micro Flow Cell	4 M	300	1.2 Gb	600 Mb (~10 hrs)	1.2 Gb (~19 hrs)	--
Nano Flow Cell	1 M	300, 500	500 Mb	150 Mb (~9 hrs)	300 Mb (~17 hrs)	500 Mb (~28 hrs)

Ainsi, si l'on prend l'exemple de la *Standard Flowcell* (14 millions de séquences lues) et que l'on souhaite garantir une profondeur moyenne de 1000X avec un design de cinq-cents amplicons, nous pourrons séquencer simultanément vingt-huit échantillons :

14.000.000 / (500x1000) = 28

A l'inverse, si l'on souhaite séquencer vingt-quatre échantillons, la profondeur théorique moyenne sera de 1166X :

14.000.000 / 500 / 24 = 1166

Sur la plateforme du Genotoul, seule la *flowcell* standard était disponible. Dans nos conditions de dix-neuf amplicons pour quinze échantillons multiplexés sur la *flowcell* standard, la profondeur de lecture moyenne théorique attendue est d'environ 50 000X.

3.2.6.2 Le logiciel Experiment Manager d'Illumina

Ce logiciel va nous permettre de créer et d'importer sur le *MiSeq* nos données concernant le *run* de séquençage.

Il s'agit dans un premier temps de créer la « sample plate », récapitulant les différents échantillons avec les index utilisés pour les différencier, puis dans un second temps, la « sample sheet » associée ; pour cela dans le logiciel, nous devons rentrer :

- La technique utilisée, ici technique des amplicons ;
- Le nombre d'index de lecture, un seul en P7 dans ce cas (il y aura deux index dans le kit *TruSeq Custom Amplicon*) ;
- Le type de séquençage : *Paired-end*, comme c'est le cas ici, ou *Single-end* ;
- Le nombre de cycles défini par la version du kit, nous utilisons la version 2 en 2x250 cycles ;
- Le mode d'analyse automatique désiré pour l'analyse des données de séquençage et particulièrement l'algorithme utilisé pour la détection des variants ; nous choisirons le « somatic variant caller » pour commencer, qui est conseillé pour l'analyse des échantillons tumoraux.

L'étape suivante consiste à rentrer l'identification des échantillons avec les index utilisés, à partir de la « sample plate » créée précédemment.

Nous devrons aussi créer un fichier texte contenant les coordonnées, sur le génome de référence (version *hg19*), des différents amplicons à séquencer ; ce fichier sera utilisé lors de l'analyse pour l'alignement des séquences obtenues sur le génome.

Ces fichiers créés précédemment seront chargés dans la base de données du séquenceur.

3.2.6.3 La préparation des réactifs

Le kit utilisé pour notre *run* est le *MiSeq V2 Reagent kit* box1-2 qui contient :

- La cartouche de réactifs ;
- Le tampon d'hybridation HT1 (coffret n°1);
- La *flowcell* V2 dans une solution de conservation ;
- Le tampon d'incorporation PR2 (coffret n°2).

La première étape consiste à faire décongeler la cartouche de réactifs dans de l'eau et le tampon HT1 dans de la glace. Pendant ce temps, il faut réaliser trois rinçages successifs de la fluidique du *MiSeq* avec du tampon *tween*.

A partir de la librairie à 9,75nM, nous réalisons une série de dilutions de manière à ajuster la concentration de la librairie à 20pM, comme demandé par *Illumina*.

Il faut également réaliser une dilution d'un contrôle interne *Illumina*, le *PhiX* à 20pM. Ce dernier sera inclus à 10% dans notre pool final ; il s'agit d'un petit génome permettant d'avoir un contrôle qualité pour la génération des *clusters*, le séquençage et l'alignement, et un contrôle de calibration pour calculer la correction éventuelle à appliquer sur l'intensité lue pour le *base-calling*.

Illumina préconise ensuite de préparer de la soude à 0,2N dans du tampon de lavage afin de dénaturer la librairie et le phage séparément ; le tampon HT1 sera ajouté après dénaturation et permettra d'arrêter la réaction mais également de réaliser la dilution finale du pool et du *PhiX*.

Le pool et le *PhiX* à 20pM seront finalement mélangés dans des proportions 90% - 10% et 600µL d'une ultime dilution de ce mélange à 7pM dans du tampon HT1 sera déposée dans la cartouche de réactifs décongelée dans un puits dédié.

3.2.6.4 Le chargement des réactifs et le lancement du run de séquençage

Il reste alors à charger la cartouche avec le pool d'échantillons dans le séquenceur, le tampon PR2 ainsi que la *flowcell* qu'il est nécessaire de rincer et sécher avant utilisation.

La *flowcell* est déposée à l'intérieur du séquenceur au-dessus d'un prisme permettant de dévier le laser qui excitera les fluorophores à chaque cycle de séquençage. Les fragments sont hybridés sur la *flowcell* puis ils subissent l'amplification en pont, la linéarisation enzymatique, le blocage de leurs extrémités 3' et l'hybridation de la première amorce de séquençage.

Le « paired-end module » permet ensuite de retourner la molécule pour la séquencer par sa deuxième extrémité avec la deuxième amorce de séquençage.

Le séquençage dure environ 48h après lesquelles nous récupèrerons un dossier de données à analyser.

3.3 L'analyse des données de NGS

Il existe de très nombreux logiciels d'analyse disponibles gratuitement au téléchargement sur le site *Illumina,* ou payants. **(Figure 31)**

3.3.1 L'étape préliminaire dans le MiSeq

La surface de la *flowcell* utilisée sur le *MiSeq* est divisée en deux lignes composées de quatorze segments (carrés) chacune ; un segment est la surface correspondant à une prise d'image par la camera après chaque incorporation d'un nucléotide ; ainsi chaque segment donne lieu à quatre images par cycle de séquençage (une par nucléotide).

La première étape d'analyse réalisée sur la plateforme *Illumina*, consiste à identifier chaque cluster par ses coordonnées sur la *flowcell* et à lui attribuer les bases correspondantes à chaque cycle (c'est le *base-calling*) **(Figure 32)**. L'enchainement de ces bases fournit une séquence de lecture ; on transforme donc les fichiers images en

Analyse du signal base calling
Traitement des séquences Démultiplexage...
Alignement
Qualité
Recherche de variants
SNV
INDELS
Transloc.
CNV
UPD
Annotatio HGVS, dbSNP, 1 000 g...
Prédiction d'effets délétères
Analyse de cohortes couples, familles
Visualisation fichiers BAM

Logiciels associés aux séquenceurs
Torrent Browser
ICS illumina
GS Sequencer
GS AVA
Casava

Logiciels commerciaux
NextGene
SeqNext
CLC Bio Workbench
Ion Reporter
GeneSpring NGS
AVADIS
Golden helix SVS
Genome Browse
Ingenuity variant analysis
Alamut HT

Méthodes ou logiciels libres
SVDetect
IGV
Shrimp
Samtools
CNVSeq
Annovar
Bowtie
Control-FREEC
BWA
VarScan

Figure 31 : Liste non exhaustive des logiciels d'analyse de données issues du NGS et de leur champ d'application (Marceau-Renaut A, et al.)

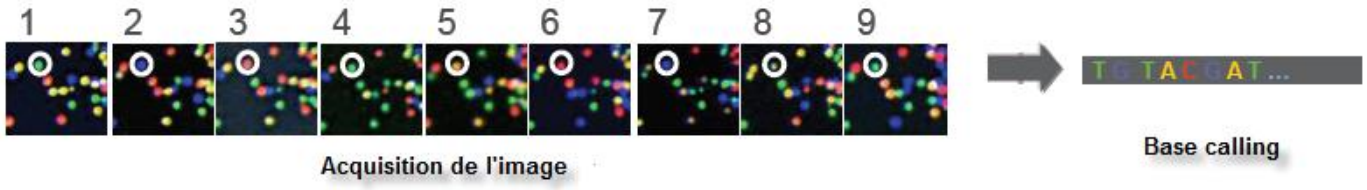


Figure 32 : Principe d'acquisition d'image (Illumina)

fichiers textes contenant les séquences nucléotidiques ainsi qu'un score de qualité de séquençage associé.

La deuxième étape est l'alignement des séquences obtenues sur un génome de référence, qui génère un nouveau fichier contenant les séquences classées par position d'alignement sur le génome ainsi qu'un score d'alignement associé.

3.3.2 Le logiciel de validation technique du run Sequencing Analysis Viewer

Ce logiciel permet, en chargeant les fichiers *RunInfo.xml* et *RunParameters.xml* générés par le *MiSeq*, de visualiser les caractéristiques générales du *run*.

Dans un premier temps, l'onglet « Analysis » **(Figure 33)** permet d'analyser les résultats suivants :

- L'intensité des bases : au cours d'un *run*, l'intensité de doit pas varier brusquement.
- La qualité des lectures pour chacun des cycles : il est normal de constater une baisse de la qualité des lectures au fur et à mesure des cycles.

  La baisse de qualité constatée lors de la lecture des index est normale et est due au fait que le *PhiX* qui sert de contrôle interne permettant de mesurer le score de qualité ne possède pas d'index.
- La distribution cumulée des « Q-scores » : elle indique la répartition du nombre de séquences en fonction de leur score de qualité.

  Illumina indique dans ses spécifications un Q-score supérieur ou égal à 30 pour plus de 80% des lectures en 2x100pb.
- Les données de *clusters* : la densité des *clusters* doit être comprise entre 600 et 900K/mm^2.
- La « flowcell chart » : cet outil permet de visualiser rapidement si un problème est survenu lors du *run* mais également de vérifier la répartition des *clusters* sur la *flowcell*.

Puis l'onglet « Summary » **(Figure 34)** nous présente un résumé des principaux paramètres du *run* :

- La densité des *clusters* ayant ou non passé les filtres de qualité *Illumina.*

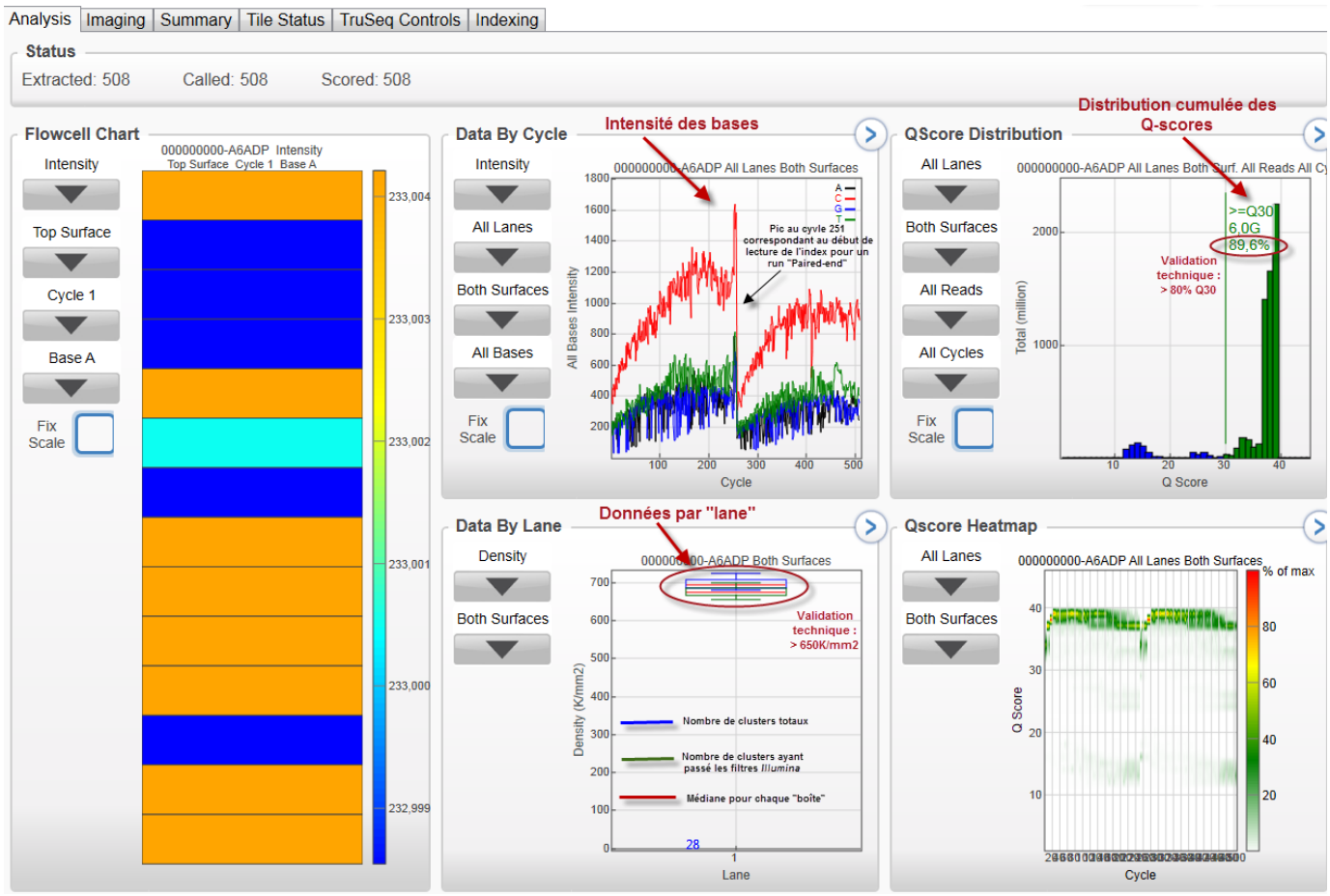


Figure 33 : Ecran principal « Analysis » du logiciel *SAV* d'*Illumina* - Exemple de notre « run »

Analysis | Imaging | Summary | Tile Status | TruSeq Controls | Indexing

Run Summary

Level	Yield Total (G)	Projected Total Yield (G)	Yield Perfect (G)	Yield <=3 errors (G)	Aligned (%)	% Perfect [Num Cycles]	% <=3 errors [Num Cycles]	Error Rate (%)	Intensity Cycle 1	% Intensity Cycle 20	% >= Q30
Read 1	3,3	3,3	0,3	0,3	10,68	72,6 [250]	96,1 [250]	0,99	214	95,1	90,2
Read 2 (I)	0,1	0,1	0,0	0,0	0,00	0,0 [5]	0,0 [5]	0,00	788	0,0	85,9
Read 3	3,3	3,3	0,2	0,3	10,53	67,8 [250]	94,5 [250]	1,12	124	196,0	89,1
Total	6,7	6,7	0,5	0,7	10,60	70,2	95,3	1,05	375	145,6	89,6

Read 1

Recommandation : 750-850K/mm2 — Recommandations : Taux < 0.5% — Recommandations : % > 80 — Recommandations : 0.8 < % alignement *PhiX* < 1.2

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Phas/Prephas (%)	Reads (M)	Reads PF (M)	% >= Q30	Yield (G)	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
1	28	699 +/- 13	96,98 +/- 0,37	0,190 / 0,138	13,69	13,27	90,2	3,3	250	10,7 +/- 0,1	0,99 +/- 0,03	0,20 +/- 0,04	0,20 +/- 0,03	0,25 +/- 0,03	214 +/- 20	95,1 +/- 2,5

Read 2 (I) = Lecture de l'index

Recommandations : Taux < 1.5%

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Phas/Prephas (%)	Reads (M)	Reads PF (M)	% >= Q30	Yield (G)	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
1	28	699 +/- 13	96,98 +/- 0,37	0,004 / 0,002	13,69	13,27	85,9	0,1	0	0,0 +/- 0,0	0,00 +/- 0,00	0,00 +/- 0,00	0,00 +/- 0,00	0,00 +/- 0,00	788 +/- 76	0,0 +/- 0,0

Read 3

Lane	Tiles	Density (K/mm2)	Cluster PF (%)	Phas/Prephas (%)	Reads (M)	Reads PF (M)	% >= Q30	Yield (G)	Cycles Err Rated	Aligned (%)	Error Rate (%)	Error Rate 35 cycle (%)	Error Rate 75 cycle (%)	Error Rate 100 cycle (%)	Intensity Cycle 1	% Intensity Cycle 20
1	28	699 +/- 13	96,98 +/- 0,37	0,198 / 0,213	13,69	13,27	89,1	3,3	250	10,5 +/- 0,0	1,12 +/- 0,04	0,32 +/- 0,03	0,39 +/- 0,02	0,43 +/- 0,02	124 +/- 10	196,0 +/- 5,1

Figure 34 : Ecran « Summary » du logiciel *SAV* d'*Illumina* - Exemple de notre « run »

Read 1 : Première lecture (sens)

Read 2 : Lecture de l'index

Read 3 : Seconde lecture (anti-sens)

- Le nombre total de séquences (*read*) obtenues et passant les filtres.
- Le taux d'erreur : il est calculé au cours du séquençage grâce à un alignement ; ce dernier doit être le plus faible possible et inférieur à 1,5%.

  Ce dernier rentre dans le calcul du *Q-score*.
- La déviation standard : elle renseigne pour chaque paramètre, sur l'homogénéité des données dans une ligne (entre les différents segments par exemple).

  Plus la déviation standard est faible et plus les données sont homogènes.

Une fois la validation technique effectuée, nous pourrons débuter l'analyse en elle-même.

3.3.3 Le logiciel d'analyse MiSeq Reporter

Il s'agit d'un logiciel que nous pouvons télécharger en ligne sur le site d'Illumina dans l'onglet « Support – Sequencing – Documentation » ; ce logiciel est gratuit et nous allons donc l'utiliser pour analyser notre run de séquençage.

Nous avons tout d'abord téléchargé, à partir de l'outil bio-informatique développé par l'UCSC (http://genome.ucsc.edu/) *[68]* les séquences de référence du génome humain, dans la version *hg19* en format « FASTA », transférable dans le logiciel *MiSeq Reporter*.

Le dossier du *run* à incrémenter au logiciel, doit comporter des fichiers générés par le séquenceur (notamment les fichiers « BCL » et les fichiers « LOCS » contenants les informations d'intensités pour chaque cycle et chaque segment de la *flowcell*) ainsi que la « sample sheet ». **(Figure 35)**

Lorsque le dossier du génome de référence ainsi que le dossier du *run* sont copiés dans les fichiers de données du logiciel, l'analyse se lance directement à l'ouverture du logiciel ; cette dernière peut durer plusieurs heures.

La première étape de l'analyse est le démultiplexage ; cette étape consiste, à l'aide de l'index spécifique accolé à chaque amplicon, à réorganiser les séquences par échantillon.

Les séquences triées sont ensuite alignées sur le génome en utilisant l'algorithme « Burrows Wheeler Aligner ».

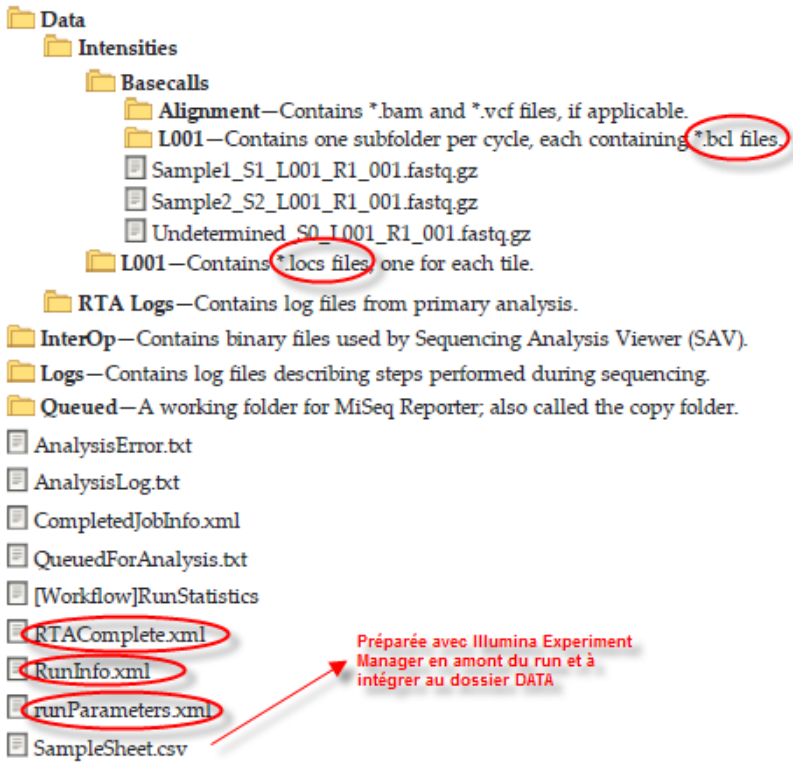


Figure 35 : Informations à intégrer au logiciel *MiSeq reporter* nécessaires à l'analyse des données (Illumina)

La dernière étape consiste en la détection des variants.

Nous récupérons différents types de fichiers lorsque l'analyse est terminée :

- Des fichiers « FASTQ » contenant l'enchainement des bases et le score de qualité pour chacune d'elles ;
- Des fichiers « BAM » correspondant à l'alignement des séquences sur le génome de référence ;
- Des fichiers « VCF » contenant des informations sur les différents variants détectés.

Il est ensuite nécessaire d'utiliser des outils de visualisation de ces fichiers générés.

3.3.4 La visualisation des données générées

Nous avons utilisé les fichiers « VCF » et « BAM » pour réaliser l'analyse des données.

Le logiciel *IGV – Integrative Genomics viewer (http://www.broadinstitute.org/software/igv/)* permet de visualiser les fichiers « BAM » et de réaliser un alignement avec le génome de référence.

En parallèle, manuellement, nous avons ouvert et analysé les fichiers « VCF » qui nous donnent des renseignements sur notamment :

- La profondeur avec laquelle chaque base a été lue ;
- La qualité de la séquence ;
- Les variations trouvées d'après les alignements réalisés par le *MiSeq* ;
- La fréquence de ces variations retrouvées.

La base de données *UCSC (http://cancer.sanger.ac.uk/cancergenome/projects/cosmic/)* permet de réaliser une dernière comparaison avec le génome, de vérifier si la mutation est avérée ou s'il s'agit d'un SNP que l'on pourra nommer si ce dernier est connu.

Ainsi nous avons pu comparer les résultats obtenus par le NGS avec ceux obtenus pas le séquençage Sanger et réaliser les premières statistiques.

VII RESULTATS

1 LA VALIDATION TECHNIQUE

L'étape préliminaire à toute analyse est la validation technique du *run* de séquençage, qui consiste à vérifier certains paramètres essentiels soulevés précédemment ***(Voir paragraphe 3.3.2 de la partie Matériel et Méthodes)***.

Nous avons donc recherché différentes données :

- La distribution cumulée des scores de qualité : nous obtenons 89,6% des lectures supérieures au Q30 (préconisé Q30 > 80%) ;
- La densité des *clusters* : nous relevons une moyenne de 699K/mm^2 (préconisé 600 K/mm^2 < densité *clusters* < 900 K/mm^2) ;
- Le nombre de *clusters* passant les filtres de qualité : nous en avons près de 97% ;
- Le taux d'erreur : inférieur à 1,12% (préconisé taux erreur < 1,5%).

En conclusion, il semble que les contrôles qualité obtenus soient très bons et nous permettent donc de débuter l'analyse à proprement parlé.

2 L'ANALYSE GENERALE DES SEQUENCES DES TEMOINS POLYCLONAUX

Deux fichiers « VCF » sont générés à la fin de l'analyse : « .vcf » et « .genome.vcf ».

Dans un premier temps, à partir du fichier « .genome.vcf », nous avons calculé la profondeur moyenne des différents amplicons, ce qui correspond au nombre de fois que ces séquences ont été lues ; pour certains amplicons d'un même gène, certains segments se chevauchant, nous avons calculé la profondeur moyenne de deux amplicons ensemble.

Nous avons obtenu des résultats que nous avons regroupés sous forme de tableau. Nous avons tout d'abord considéré les PBL dans des conditions que nous appellerons « standard » de PCR multiplex (56°C et 30 cycles d'élongation) pour tirer nos premières conclusions. **(Tableau XVIII)**

2.1 La couverture et la profondeur

La profondeur retrouvée est très hétérogène pour les différents amplicons ; ainsi nous avons une profondeur maximale de 127 944 lectures pour *TP53* exons 8 et 9

Tableau XVIII : Profondeur des différents amplicons pour les PBL en conditions standard

AMPLICON	CHR	POSITION DEBUT	POSITION FIN	PROFONDEUR MOYENNE PBL (conditions std)
DNMT3_X23	2	25457106	25457332	2
IDH1_X04	2	209112960	209113192	129
KIT_X17	4	55599055	55599380	67925
NPM1_X12	5	170837485	170837726	37534
FLT3_X20	13	28592540	28592819	30688
FLT3_X13	13	28608058	28608525	3487
TP53_X10	17	7573906	7574178	35820
TP53_X09	17	7576822	7577084	127944
TP53_X08	17	7576910	7577182	
TP53_X07	17	7577481	7577717	6709
TP53_X06	17	7578137	7578363	533
TP53_X05	17	7578299	7578582	110080
TP53_X04	17	7579267	7579517	62298
PP2cR	19	33792130	33792402	7171
PP2b	19	33792358	33792704	3
PP2Fa	19	33792669	33792906	12
PP1bR	19	33792775	33793095	
PP1Fa	19	33793049	33793350	11

Mix avec Enhancer
Plus faible valeur de profondeur
Plus forte valeur de profondeur
Conditions standard PCR retenues (56°C - 30 cycles)

mais une profondeur de 2 lectures pour *DNMT3* ou de 3 à 7171 lectures pour les amplicons du gène *CEBPα*.

Nous avons pu remarquer que le gène *DNMT3* n'était pas toujours retrouvé et que lorsqu'il était présent, sa couverture était très mauvaise et partielle.

Lors de la préparation de nos échantillons, nous avions réalisé un mélange « à parts égales », des deux PCR avec et sans *enhancer* pour chaque échantillon; mais ce mélange fut relativement subjectif puisque basé sur l'intensité des bandes sur gel et sur le dosage des produits PCR avant purification.

Etant donné que parmi les amplicons générés à partir du mix de PCR avec *enhancer* (*DNMT3A* et *CEBPα*), PP2cR a une profondeur nettement supérieure aux autres, cet amplicon a dû être amplifié préférentiellement par rapport aux autres lors de la PCR multiplex.

Ainsi, lorsque nous avons réalisé notre mélange volume à volume, nous avons sous-dosé les autres amplicons du mix avec *enhancer*, ce qui explique la très faible profondeur obtenue.

En revanche, pour les amplicons ayant une profondeur au-delà d'une centaine de lectures, la couverture des amplicons et l'homogénéité de la profondeur de lecture tout au long de l'amplicon sont très bonnes.

2.2 L'influence de la température et du nombre de cycles

Afin de connaître l'influence des conditions de PCR sur le résultat du séquençage, nous avions réalisé des PCR avec des conditions différentes, sur de l'ADN de PBL.

D'après les calculs de profondeur moyenne que nous avons réalisés, il s'avère que le nombre de cycles de PCR ait une influence différente en fonction du couple d'amorces **(Tableau XIX)**. Ainsi, pour *KIT* exon 17 ou *TP53* exon 6, l'augmentation du nombre de cycles de PCR diminue la profondeur, alors que pour IDH1 exon 4, FLT3 exon 20 ou TP53 exon 4 et 10, plus le nombre de cycles est augmenté et plus la profondeur est élevée ; enfin pour certains amplicons comme *NPM1* exon 12, l'influence n'est pas évidente.

De la même façon, nous avions testé différentes conditions de température pour lesquelles nous avons calculé la profondeur **(Tableau XX)**. Nous obtenons des conclusions semblables, l'influence de la température étant amplicon dépendante ; par

Tableau XIX : Profondeur des différents amplicons pour les PBL pour un nombre de cycles de PCR variable

AMPLICON	CHR	POSITION DEBUT	POSITION FIN	PROFONDEUR MOYENNE PBL 56°C - 27 cycles	PROFONDEUR MOYENNE PBL 56°C - 30 cycles	PROFONDEUR MOYENNE PBL 56°C - 33 cycles
DNMT3_X23	2	25457106	25457332	2	2	1
IDH1_X04	2	209112960	209113192	78	129	130
KIT_X17	4	55599055	55599380	153368	67925	47356
NPM1_X12	5	170837485	170837726	34685	37534	32435
FLT3_X20	13	28592540	28592819	24246	30688	35874
FLT3_X13	13	28608058	28608525	8152	3487	1930
IDH2_X04	15	90631676	90632027	115112	101488	116003
TP53_X10	17	7573906	7574178	17114	35820	40972
TP53_X09	17	7576822	7577084	138346	127944	90755
TP53_X08	17	7576910	7577182			
TP53_X07	17	7577481	7577717	5377	6709	6848
TP53_X06	17	7578137	7578363	874	533	611
TP53_X05	17	7578299	7578582	100481	110080	114833
TP53_X04	17	7579267	7579517	34987	62298	85082
PP2cR	19	33792130	33792402	24697	7171	13709
PP2b	19	33792358	33792704	couverture partielle 0 et <10	3	4
PP2Fa	19	33792669	33792906	21	12	30
PP1bR	19	33792775	33793095			
PP1Fa	19	33793049	33793350	18	11	18

Plus faible valeur de profondeur
Plus forte valeur de profondeur
Conditions standard PCR retenues (56°C - 30 cycles)

Tableau XX : Profondeur des différents amplicons pour les PBL pour une température de PCR variable

AMPLICON	CHR	POSITION DEBUT	POSITION FIN	PROFONDEUR MOYENNE PBL 30 cycles - 54°C	PROFONDEUR MOYENNE PBL 30 cycles - 56°C	PROFONDEUR MOYENNE PBL 30 cycles - 58°C	PROFONDEUR MOYENNE PBL 30 cycles - 60°C
DNMT3_X23	2	25457106	25457332	0	2	1	1
IDH1_X04	2	209112960	209113192	129	129	124	174
KIT_X17	4	55599055	55599380	69101	67925	69180	71138
NPM1_X12	5	170837485	170837726	38148	37534	33178	17915
FLT3_X20	13	28592540	28592819	35798	30688	24771	20262
FLT3_X13	13	28608058	28608525	3553	3487	2721	1943
TP53_X10	17	7573906	7574178	27817	35820	63698	125478
TP53_X09	17	7576822	7577084	135398	127944	117269	85811
TP53_X08	17	7576910	7577182				
TP53_X07	17	7577481	7577717	5976	6709	7975	9204
TP53_X06	17	7578137	7578363	557	533	400	393
TP53_X05	17	7578299	7578582	113290	110080	114433	118545
TP53_X04	17	7579267	7579517	60605	62298	65326	61651
PP2cR	19	33792130	33792402	8212	7171	18729	22010
PP2b	19	33792358	33792704	5	3	2	5
PP2Fa	19	33792669	33792906	16	12	30	40
PP1bR	19	33792775	33793095				
PP1Fa	19	33793049	33793350	25	11	15	25

Plus faible valeur de profondeur
Plus forte valeur de profondeur
Conditions standard PCR retenues (56°C - 30 cycles)

exemple pour *TP53* exons 7 et 10, l'augmentation de la température a une influence favorable sur la profondeur alors que pour *NPM1* exon 12 et *FLT3* exon 20, l'influence est défavorable.

Finalement, nous en avons conclu que les conditions de température que nous avions utilisées pour réaliser les PCR de nos patients, à savoir 30 cycles de PCR à une température d'élongation de 56°C, étaient convenables et aucun de ces tests n'est suffisamment concluant.

2.3 L'influence de la réalisation de mix communs

Afin de savoir s'il serait possible d'envisager de réaliser une PCR multiplex unique avec toutes les amorces, nous avions réalisé deux PCR multiplex contenant toutes les amorces, sur de l'ADN de PBL, en présence ou non de l'*enhancer*.

Les résultats compilés **(Tableau XXI)** montrent que globalement, les profondeurs seraient meilleures pour le mix commun avec l'*enhancer*. Cependant ce qui ressort, et qui était prévisible, est le fait que les amplicons qui semblaient, d'après les tests préalables avoir une meilleure efficacité avec l'*enhancer* (comme *CEBPα*) présentent une meilleure profondeur dans le mix avec l'*enhancer.* Inversement, les amplicons qui semblaient avoir une meilleure efficacité sans l'*enhancer* (comme *NPM1*) présentent effectivement une meilleure profondeur dans le mix sans *enhancer*.

3 L'ANALYSE DETAILLEE DES SEQUENCES DES PATIENTS

L'objectif de cette partie est maintenant d'explorer dans le détail les séquences obtenues pour les patients et notamment de savoir si les mutations attendues sont détectées. Pour cela nous avons analysé « manuellement » les fichiers « .vcf » générés lors de l'analyse.

3.1 La sélection des données

Dans un premier temps, nous avons décidé de ne prendre en compte les résultats que pour une profondeur de lecture supérieure à 1000 et un pourcentage de séquences contenant la variation supérieur à 10%. Cependant, certaines mutations retrouvées en technique classique après séquençage Sanger n'étaient pas retrouvées avec ces critères définis. Nous avons donc décidé, lorsque ces mutations étaient présentes en Sanger de récupérer les données de NGS si une des deux conditions au moins était remplie.

Tableau XXI : Profondeur des différents amplicons pour les mix communs avec et sans Enhancer

AMPLICON	CHR	POSITION DEBUT	POSITION FIN	PROFONDEUR MOYENNE PBL TOUT - Enhancer	PROFONDEUR MOYENNE PBL TOUT + Enhancer
DNMT3_X23	2	25457106	25457332	2	1
IDH1_X04	2	209112960	209113192	33	7
KIT_X17	4	55599055	55599380	2977	381
NPM1_X12	5	170837485	170837726	11058	330
FLT3_X20	13	28592540	28592819	17688	1136
FLT3_X13	13	28608058	28608525	268	60
TP53_X10	17	7573906	7574178	56922	58843
TP53_X09	17	7576822	7577084	114002	57801
TP53_X08	17	7576910	7577182		
TP53_X07	17	7577481	7577717	4410	13668
TP53_X06	17	7578137	7578363	374	1425
TP53_X05	17	7578299	7578582	125959	30666
TP53_X04	17	7579267	7579517	41311	14252
PP2cR	19	33792130	33792402	872	7202
PP2b	19	33792358	33792704	0	6
PP2Fa	19	33792669	33792906	0	61
PP1bR	19	33792775	33793095		
PP1Fa	19	33793049	33793350	2	11

Conditions standard PCR retenues (56°C - 30 cycles)
Couples d'amorces ayant une meilleure efficacité de PCR avec Enhancer
Couples d'amorces ayant une meilleure efficacité de PCR sans Enhancer
Mix avec Enhancer

Nous avons combiné les données de profondeur, les données relatives aux variants retrouvés ainsi que le pourcentage de ces variations dans un tableau **(Annexe 4)** qui nous a servi de base pour nos analyses.

3.2 L'analyse des données

Nous pouvons noter que toutes les mutations ponctuelles retrouvées par le séquençage Sanger le sont également par NGS.

En revanche, en ce qui concerne les insertions (comme pour le FLT3 ITD de DESMI) ou les délétions (comme le *TP53* exon 4 de CARMI), ces dernières ne sont détectées que si elles ne dépassent pas 6 pb (comme les mutations de *NPM1*). Il s'agit très probablement d'un problème dans l'algorithme de détection des variations que nous tenterons de résoudre avec l'aide du support technique d'*Illumina* ; le paramètre de détection des insertions / délétions semble par défaut configuré à 10pb maximum mais l'utilisation du kit *TruSeq Amplicon* permettrait, lors de l'analyse des données de modifier ce paramètre.

Ensuite, lorsqu'aucune mutation n'a été détectée en séquençage Sanger, nous n'en avons pas découvert de nouvelles en NGS (avec les paramètres définis à 1000 pour la profondeur de lecture et plus de 10% de variants).

L'analyse de *CEBPα* (basée sur le patient MOUCH) a tout de même posé des problèmes car la profondeur avec laquelle les bases ont été lues est très faible ; nous avons cependant essayé d'en tirer certaines informations ; ainsi nous retrouvons les mutations détectées en Sanger mais dans ces mêmes conditions d'analyse, d'autres variations coexistent sans pour autant avoir de valeur malgré un pourcentage de variants supérieur à 10%.

Il semble qu'une profondeur minimale de lecture de 100X soit nécessaire pour que les mutations retrouvées puissent être significatives.

De même l'analyse de DNMT3 s'avère impossible tant la couverture est souvent partielle et la profondeur infime.

Enfin, nous avons mis en évidence des artefacts lors de l'alignement de séquences présentant des répétitions de bases (T ou A) : il semble que cela soit dû à la polymérase lors de la PCR ou d'erreurs lors du *base-calling* aboutissant à de multiples possibilités de variants, comme c'est le cas par exemple pour le *TP53* exon 7 de DELCO ou le NPM1 exon 12 de MOUCH.

VIII DISCUSSION

1 L'ANALYSE FINANCIERE

La notion de coût de l'analyse doit être prise en compte dans le choix qui sera fait lors de l'établissement du *workflow* du NGS.

Nous avons voulu comparer au niveau du coût, la technique d'amplicons par PCR « maison » (mais en utilisant un kit commercial (*Nextera*^{XT}) permettant l'ajout des index et des adaptateurs ainsi que la normalisation de la librairie), avec la solution commerciale d'*Illumina* (*TruSeq Custom Amplicon)*, ceci afin de valider un protocole applicable en routine. Pour chacune de ces deux solutions nous avons calculé le prix de revient par échantillon dans plusieurs conditions (capacité de la flowcell et nombre d'échantillons multiplexés) qui influent sur la profondeur de lecture moyenne théorique. **(Annexe 5)**

Au niveau du coût du matériel, la technique de PCR « maison » semble la plus rentable lorsqu'elle est utilisée en routine. En revanche au niveau du coût de à la mise au point (et du temps passé à la mise au point), cette technique ne serait plus envisageable pour un nombre de gènes supérieur aux neuf gènes testés pendant ce travail. En effet la seule façon de vérifier en pratique l'efficacité des différents couplent d'amorces au sein 'une nouvelle PCR multiplex est de réaliser un *run* de NGS, ce qui serait bien sûr extrêmement couteux.Les simulations réalisées sur le *TruSeq Custom Amplicon* avec les versions 1 et 2 du panel de gènes semblent privilégier cette solution tant au niveau du gain de temps mais également au niveau d'une économie possible réalisée avec un nombre croissant de gènes.

2 LA MISE EN PLACE D'UN « WORKFLOW »

Nous avons tenté, tout au long de ce mémoire de répondre aux problématiques de l'intégration du NGS dans notre laboratoire de routine, et ce, afin de mettre en place un *workflow* viable au sein de notre équipe :

- Le choix de la technique d'enrichissement

  Il semble que nous devrons choisir une solution commerciale, comme par exemple le kit *TruSeq Custom Amplicon* pour réaliser en routine la recherche des mutations des gènes d'intérêt que nous avons choisi car notre expérience nous a montré que la technique par PCR sera trop lourde en terme de mise en

point et en pratique et ne sera pas applicable en routine. En effet notre expérience lors de ces tests nous a montré que l'utilisation de PCR multiplex pour la génération des amplicons à séquencer était trop hasardeuse en terme de couverture des différentes cibles lors du séquençage, et que la mise au point de PCR multiplex avec mille couples d'amorces (500 amplicons pour les 24 gènes) ne nous parrait pas faisable.

Nous n'avons malheureusement pas pu tester la solution commerciale en pratique pendant la réalisation de ce mémoire (pour des raisons budgétaires) et il conviendra donc de tester ce kit, après une vérification de la couverture par les bio informaticiens de la plateforme *Illumina,* et de s'assurer que les résultats sont interprétables pour tous les amplicons et que nous obtenons une profondeur de lecture effective suffisante pour chaque cible.

- Le choix de la plateforme de NGS

Suite à l'appel d'offre que nous avons passé au sein du CHU de Toulouse, la seule plateforme qui semble répondre à nos exigences en termes de coût, de taille, de praticité de l'interface et de prise en main est le *MiSeq* d'*Illumina* ; c'est pourquoi nous avons voulu réaliser nos tests sur cette plateforme.

En effet, le *MiSeq* permet d'accéder, à partir d'une faible quantité d'échantillon, à des données de séquençage de grande qualité avec un flux de travail simplifié. De plus, il est le seul séquenceur dernière génération à réunir amplification, séquençage et analyse des données sur un seul instrument de petite envergure.

- L'analyse bio-informatique et le traitement des données

Nous avons procédé à une analyse « manuelle » des données de NGS lors de ce test car nous avions peu de gènes, mais ce procédé ne serait plus viable si nous augmentons le nombre de gènes à étudier comme cela est prévu.

Nous devrons utiliser un logiciel adapté à la solution commerciale du *TruSeq* que nous préconisons. Ainsi le logiciel *Amplicon viewer* d'*Illumina* pourrait nous permettre de réaliser un alignement et de déduire les mutations détectées et ce de façon automatique.

Il serait alors possible de s'affranchir d'un bio informaticien à proprement parlé ; mais ceci ne pourra se confirmer que lorsque nous aurons testé cette solution.

- Le rendu des résultats aux cliniciens

Le premier problème est de déterminer à quelle fréquence nous devrons réaliser un *run* de NGS afin de rentabiliser chaque manipulation mais également de respecter nos exigences en termes de délais de rendu de résultats. Nous pourrions donc envisager de réaliser une manipulation tous les quinze jours afin de respecter les délais de rendu de résultats, en prenant les patients au diagnostic de LAM et MDS avec la version 1 du panel de gènes que nous avons établie (vingt-quatre gènes). Pour la version 2 (quatre-vingt-un gènes), il sera nécessaire de rediscuter de ces paramètres en intégrant la notion du temps nécessaire à l'analyse des données.

Afin que les résultats soient facilement interprétables par les cliniciens, il semble inconcevable de rendre une liste de vingt-quatre puis de quatre-vingt-un gènes avec le statut « muté » ou « non muté » ; il semblerait donc judicieux de ne rendre que les gènes « mutés » et de noter dans une partie du compte-rendu la totalité des gènes testés.

- La gestion des données de séquençage :

Chaque *run* de séquençage génère avec la version 2 de la *flowcell* 250, 7Gb de données brutes sans compter les données analysées. Une nouvelle version de la *flowcell* du MiSeq permettra très bientôt d'obtenir jusqu'à 15 Gb de données.

Il va donc être nécessaire de s'équiper de matériel informatique performant et surtout de systèmes de sauvegarde et de conservation des données évolutifs, ces dernières devant être sauvegardées pendant au minimum vingt ans.

Le NGS semble donc avoir toute sa place dans notre laboratoire de Génétique des Hémopathies ; néanmoins, si les progrès technologiques sont spectaculaires et permettent d'étudier un plus grand nombre de gènes pour un coût moindre, cela va

entrainer une modification des techniques de travail et il faudra donc prévoir des périodes de formations adaptées pour le personnel technique. Le recrutement de bio-informaticiens serait bien sûr un plus pour aider à la gestion et l'analyse de toutes ces données mais nous avons bon espoir d'y arriver sans à partir du moment où nous avons accès à des logiciels commerciaux permettant de rendre l'analyse plus abordable. Dans tous les cas, des investissements financiers mais aussi en termes de personnel seront très probablement nécessaires pour permettre l'intégration de cette technologie au sein du laboratoire.

3 L'ENJEU DE L'ACCREDITATION

Il ne faut pas oublier le fait que nous sommes engagés dans une politique d'accréditation COFRAC des laboratoires d'analyses médicales et que des normes sont à appliquer en ce qui concerne :

- La validation des techniques en termes de reproductibilité et de répétabilité ;
- La validation des résultats techniques en termes de fiabilité (profondeur, couverture) ;
- Les procédures à appliquer ;
- Le délai de rendu des résultats ;
- La conservation de ces résultats.

La dernière version de la norme ISO 15181 impose que 80% des analyses réalisées dans les laboratoires soient accréditées d'ici 2020 ; or pour des techniques de pointe comme le NGS ces étapes devront prendre encore du temps pour être au point, tant pour des raisons économiques que matérielles.

4 LES PERSPECTIVES

Nous avons encore beaucoup de travail à réaliser afin de boucler la mise au point et l'adaptation des techniques de NGS à notre laboratoire de routine.

Si nous avons conclu que la technique de design des amplicons « maison » n'était pas adaptée, il nous reste à tester la solution commerciale pour la version 1 de notre panel de gènes de LAM et MDS proposée par *Illumina*.

Un gros travail d'analyse des données reste à faire pour ce panel et nous allons devoir rédiger des procédures qui permettraient de passer cette technique en routine.

IX CONCLUSIONS

1 L'APPLICATION DU NGS A L'ACTIVITE DE ROUTINE

Ces expériences nous ont permis de conclure à la possibilité de mettre en place du NGS dans un laboratoire de routine du fait de ses nombreux avantages :

- La rapidité ; ainsi, le NGS permet d'étudier simultanément plus de patients pour un même gène ou plusieurs gènes pour un ou plusieurs patients dans des délais plus courts que le séquençage Sanger.
- L'étude plus exhaustive de certains gènes ; en effet, des gènes tels que *TP53*, *DNMT3A ou ASXL1*, de par leur taille et/ou l'existence de « hotspots » de mutations, ne sont jusqu'à présent séquencés que partiellement.
- L'économie du matériel biologique ; les nouvelles techniques de préparation des librairies ne nécessitent que quelques nano-grammes d'ADN et permettent l'étude d'un grand nombre de gènes.

Cependant il est important de respecter quelques précautions :

- L'utilisation de trousses commerciales afin de s'affranchir des difficultés inhérentes à la mise au point des PCR multiplex.
- La profondeur de séquençage : la grande difficulté réside dans la validation des résultats en particulier des sous-clones qui ne peuvent être validés par séquençage Sanger.
- La nécessité de valider la qualité de séquençage (couverture, profondeur).
- La robotisation ; quelle que soit la méthode de préparation des librairies, les étapes de purification sont nombreuses et longues, et donc source d'erreurs. Une automatisation de ces étapes permettrait d'augmenter la reproductibilité.
- La difficulté d'étudier de grandes insertions, type *FLT3-ITD,* qui sont mieux détectées en analyse de fragments.

D'autre part, certaines mutations sont situées dans des régions d'homopolymères ou dans des gènes très riches en GC, il conviendra de s'assurer que l'on obtient une couverture et une profondeur de lecture suffisante pour ces régions.

- L'analyse des résultats reste ardue même si elle devient de plus en plus aisée grâce aux nombreux logiciels disponibles.

2 L'EVALUATION POSSIBLE DE LA MALADIE RESIDUELLE

De par sa grande sensibilité, le NGS peut être un outil de choix pour le suivi de la maladie résiduelle en termes de mutations ou de réarrangements des immunoglobulines *[47]*.

Le suivi de la MRD (Minimal Residual Disease) dans les LAM est aujourd'hui réalisé en routine par RQ-PCR uniquement dans les cas où un transcrit de fusion a été détecté au diagnostic. La sensibilité du NGS pourrait permettre de réaliser un suivi des mutations détectées au diagnostic afin d'évaluer la réponse au traitement. Il convient cependant de s'assurer que la mutation suivie n'est pas une mutation secondaire présente dans un sous-clone.

Ainsi les duplications en tandem de *FLT3* font partie des mutations les plus fréquentes dans les LAM, mais le suivi de MRD utilisant ces mutations semble difficile à mettre en œuvre avec les techniques classiques.

En effet, la diversité des mutations de *FLT3* implique le séquençage de chaque mutation et le design d'amorces spécifiques *[69]*.

Cependant, ces difficultés pourraient être surmontées grâce au NGS, qui permet de déterminer à la fois le site d'insertion, la séquence dupliquée, le ratio allélique et, en cas de mutations multiples, la quantité relative de chaque clone.

Thol *et al*. *[70]* ont utilisé le pyro-séquençage massif parallèle pour suivre dix patients atteints de LAM FLT3-ITD positif. Dans cette étude, la sensibilité de détection était de 10^{-4} (10 000 *reads*/échantillon), mais cette sensibilité peut être améliorée en augmentant le nombre de *reads* (1 000 000 *reads*/échantillon, soit une sensibilité de 10^{-6}) ; cette analyse montre que les techniques de NGS sont applicable à tous les patients FLT3-ITD positif indépendamment du site d'insertion, de la longueur et de la séquence d'insertion, et permet d'évaluer les différents clones présents au cours de l'évolution de la maladie.

Concernant les mutations du gène *NPM1*, seuls les variants de types A, B et D sont aujourd'hui suivis en MRD par RQPCR or les nouvelles technologies à haut débit

permettraient de suivre toutes les mutations de *NPM1*, même les variants rares en un seul et même test *[70]*.

Pour les LAL, le suivi de la MRD utilise principalement la cytométrie en flux et la RQ-PCR avec des amorces spécifiques des réarrangements des gènes d'immunoglobulines ou du TCR pour chaque patient.

Une étude récente a évalué la MRD dans les LAL-T par NGS des régions variables du TCRB et TCRG *[71]*, sur une cohorte de quarante-trois patients, analysés au diagnostic et à J-29 du traitement. Les résultats montrent que le NGS permet une meilleure détection de la MRD que la CMF avec une sensibilité inférieure à 10^{-5} *[47]*.

Cependant, l'analyse des résultats de MRD par NGS reste pour l'instant très difficile car peu d'algorithmes ont été développés et il n'existe pas encore de solution logicielle clé en main.

X ANNEXES

Annexe 1 : Séquence des amorces utilisées dans les techniques classiques de recherche de mutations..... 51

Amorces PCR

FLT3_F	TCCTGTTTTGCTAATTCCATAAGCT
FLT3_R	TCCTGTTTTGCTAATTCCATAAGCT
NPM1_F	GAAGTGTTGTGGTTCCTTAAC
NPM1_R	TGGACAACACATTCTTGGCA
PP1F	TCGCCATGCCGGGAGAACTCTAAC
PP1R	CTGGTAAGGGAAGAGGCCGGCCAG
PP2F	CCGCTGGTGATCAAGCAGGA
PP2R	AATCTCCTAGTCCTGGCTCGC
DNMT3A_F	CTGGCCAGCACTCACCCT
DNMT3A_R	TGTTTAACTTTGTGTCGCTACCTCA

Amorces HRM

FLT3_X20_F2	TCACAGAGACCTGGCCGC
FLT3_X20_R1	GCCCCTGACAACATAGTTGG
TP53_X04_F3	CTGCACCAGCAGCTCCTACA
TP53_X04_R22	CAGGCATTGAAGTCTCATGGAA
TP53_X05_F12	CAACTCTGTCTCCTTCCTCTTCCTAC
TP53_X05_R22	GTCGTCTCTCCAGCCCCAGC
TP53_X06_F1	CCTCTGATTCCTCACTGATTGC
TP53_X06_R1	CTTAACCCCTCCTCCCAGAG
TP53_X07_F1	TTGGGCCTGTGTTATCTCCT
TP53_X07_R1	TGGCAAGTGGCTCCTGAC
TP53_X08_F1	TTGCTTCTCTTTTCCTATCCTGA
TP53_X08_R1	GCTTCTTGTCCTGCTTGCTT
TP53_X09_F1	CCTTTCCTTGCCTCTTTCCT
TP53_X09_R1	CCACTTGATAAGAGGTCCCAAG
TP53_X10_F2	GAACCATCTTTTAACTCAGGTACTGTGTAT
TP53_X10_R1	GAAGGGGCTGAGGTCACTC
KIT_X17_F3	CCTCCTTACTCATGGTCGGATC
KIT_X17_F3	CAAGCAGAGAATGGGTACTCACG
IDH1_F	CACGGTCTTCAGAGAAGCCAT
IDH1_R	GCATTTCTCAATTTCATACCTTGCTTA
IDH2_F	TGAAAGATGGCGGCTGCAGT
IDH2_R	GGGGTGAAGACCATTTTGAA

Annexe 2 : Séquence des amorces de NGS utilisées..... 58

NOM	SEQUENCE	SEQUENCE + QUEUE	TAILLE AMPLICON
FLT3_X20_NGS_F1	GGTACCTCCTACTGAAGTTGAGTCTAGAA	CTTTCCCTACACGACGCTCTTCCGATCTGGTACCTCCTACTGAAGTTGAGTCTAGAA	280
FLT3_X20_NGS_R1	TTACCATGATAACGACACAACACAA	GGAGTTCAGACGTGTGCTCTTCCGATCTTTACCATGATAACGACACAACACAA	/
NPM1_X12_F1	GAAGTGTTGTGGTTCCTTAAC	CTTTCCCTACACGACGCTCTTCCGATCTGAAGTGTTGTGGTTCCTTAAC	243
NPM1_X12_R1	TGGACAACACATTCTTGGCA	GGAGTTCAGACGTGTGCTCTTCCGATCTTGGACAACACATTCTTGGCA	/
FLT3_NGS_F	TCCTGTTTTGCTAATTCCATAAGCT	CTTTCCCTACACGACGCTCTTCCGATCTTCCTGTTTTGCTAATTCCATAAGCT	468
FLT3_NGS_R	AGACAACATCTCATTCTATGCAACAAT	GGAGTTCAGACGTGTGCTCTTCCGATCTAGACAACATCTCATTCTATGCAACAAT	/
PP1b_NGS_F	GAGTTCCTGGCCGACCTGTTCCA	CTTTCCCTACACGACGCTCTTCCGATCTGAGTTCCTGGCCGACCTGTTCCA	321
PP1_R	CTGGTAAGGGAAGAGGCCGGCCAG	GGAGTTCAGACGTGTGCTCTTCCGATCTCTGGTAAGGGAAGAGGCCGGCCAG	/
PP1_F	TCGCCATGCCGGGAGAACTCTAAC	CTTTCCCTACACGACGCTCTTCCGATCTTCGCCATGCCGGGAGAACTCTAAC	301
PP1a_NGS_R	GCCTTCTCCTGCTGCCGGCTGT	GGAGTTCAGACGTGTGCTCTTCCGATCTGCCTTCTCCTGCTGCCGGCTGT	/
PP2c_NGS_F	AACGTGGAGACGCAGCAGAA	CTTTCCCTACACGACGCTCTTCCGATCTAACGTGGAGACGCAGCAGAA	273
PP2_R	AATCTCCTAGTCCTGGCTCGC	GGAGTTCAGACGTGTGCTCTTCCGATCTAATCTCCTAGTCCTGGCTCGC	/
PP2_bis_F	CTGGACGGCAGGCTGGA	CTTTCCCTACACGACGCTCTTCCGATCTCTGGACGGCAGGCTGGA	238
PP2a_bis_R	TGGTGGTCTGGCCGCA	GGAGTTCAGACGTGTGCTCTTCCGATCTTGGTGGTCTGGCCGCA	/
PP2b_NGS_F	TGCAGTTCCAGATCGCGCA	CTTTCCCTACACGACGCTCTTCCGATCTTGCAGTTCCAGATCGCGCA	347
PP2b_NGS_R	ATTGTCACTGGTCAGCTCCAGCA	GGAGTTCAGACGTGTGCTCTTCCGATCTATTGTCACTGGTCAGCTCCAGCA	/
IDH1_X4_F2	CACGGTCTTCAGAGAAGCCAT	CTTTCCCTACACGACGCTCTTCCGATCTCACGGTCTTCAGAGAAGCCAT	235
IDH1_X4_R2	GCATTTCTCAATTTCATACCTTGCTTA	GGAGTTCAGACGTGTGCTCTTCCGATCTGCATTTCTCAATTTCATACCTTGCTTA	/
IDH2_F1	TGAAAGATGGCGGCTGCAGT	CTTTCCCTACACGACGCTCTTCCGATCTTGAAAGATGGCGGCTGCAGT	353
IDH2_R2	GGGGTGAAGACCATTTTGAA	GGAGTTCAGACGTGTGCTCTTCCGATCTGGGGTGAAGACCATTTTGAA	/
DNMT3A_X23_F1	CTGGCCAGCACTCACCCT	CTTTCCCTACACGACGCTCTTCCGATCTCTGGCCAGCACTCACCCT	228
DNMT3A_X23_R1	TGTTTAACTTTGTGTCGCTACCTCA	GGAGTTCAGACGTGTGCTCTTCCGATCTTGTTTAACTTTGTGTCGCTACCTCA	/
KIT _X17_NGS_F1	ATTATGTGAACATCATTCAAGGCGTA	CTTTCCCTACACGACGCTCTTCCGATCTATTATGTGAACATCATTCAAGGCGTA	326
KIT_X17_R3	CAAGCAGAGAATGGGTACTCACG	GGAGTTCAGACGTGTGCTCTTCCGATCTCAAGCAGAGAATGGGTACTCACG	/
TP53_X04_NGS_bis_F1	ACCCAGGTCCAGATGAAGCT	CTTTCCCTACACGACGCTCTTCCGATCTACCCAGGTCCAGATGAAGCT	251
TP53_X04_R22	CAGGCATTGAAGTCTCATGGAA	GGAGTTCAGACGTGTGCTCTTCCGATCTCAGGCATTGAAGTCTCATGGAA	/

TP53_X05_F12	CAACTCTGTCTCCTTCCTCTTCCTAC	CTTTCCCTACACGACGCTCTTCCGATCTCAACTCTGTCTCCTTCCTCTTCCTAC	284
TP53_X05_NGS_R	ATCAGTGAGGAATCAGAGGCCT	GGAGTTCAGACGTGTGCTCTTCCGATCTATCAGTGAGGAATCAGAGGCCT	/
TP53_X06_NGS_F1	GCTGGGGCTGGAGAGACGAC	CTTTCCCTACACGACGCTCTTCCGATCTGCTGGGGCTGGAGAGACGAC	228
TP53_X06_R1	CTTAACCCCTCCTCCCAGAG	GGAGTTCAGACGTGTGCTCTTCCGATCTCTTAACCCCTCCTCCCAGAG	/
TP53_X07_NGS_F1	GCGACAGAGCGAGATTCCAT	CTTTCCCTACACGACGCTCTTCCGATCTGCGACAGAGCGAGATTCCAT	237
TP53_X07_R1	TGGCAAGTGGCTCCTGAC	GGAGTTCAGACGTGTGCTCTTCCGATCTTGGCAAGTGGCTCCTGAC	/
TP53_X08_F1	TTGCTTCTCTTTTCCTATCCTGA	CTTTCCCTACACGACGCTCTTCCGATCTTTGCTTCTCTTTTCCTATCCTGA	273
TP53_X08_NGS_bis_R1	GGTGTTGTTGGGCAGTGCT	GGAGTTCAGACGTGTGCTCTTCCGATCTGGTGTTGTTGGGCAGTGCT	/
TP53_X09_NGS_F1	AGGAAGAGAATCTCCGCAAGA	CTTTCCCTACACGACGCTCTTCCGATCTAGGAAGAGAATCTCCGCAAGA	263
TP53_X09_R1	CCACTTGATAAGAGGTCCCAAG	GGAGTTCAGACGTGTGCTCTTCCGATCTCCACTTGATAAGAGGTCCCAAG	/
TP53_X10_NGS_F1	CAGCTGTATAGGTACTTGAAGTGCA	CTTTCCCTACACGACGCTCTTCCGATCTCAGCTGTATAGGTACTTGAAGTGCA	273
TP53_X10_R1	GAAGGGGCTGAGGTCACTC	GGAGTTCAGACGTGTGCTCTTCCGATCTGAAGGGGCTGAGGTCACTC	/

Annexe 3 : Protocoles de PCR utilisés pour la PCR 1 du NGS 58

CEBPa / DNMT3a				
	LOT N°	Concentration	n=1(µL)	
Nbre de tubes	/	/	1	15
EAU	/	/	5	75
ENHANCER 360 GC		5X	10	150
AMPLITAQ GOLD 360 GC		2X	25	375
PP1b_NGS_F	/	10µM	0,5	7,5
PP1_R	/	10µM	0,5	7,5
PP1_F	/	10µM	0,5	7,5
PP1a_NGS_R	/	10µM	0,5	7,5
PP2c_NGS_F	/	10µM	0,5	7,5
PP2_R	/	10µM	0,5	7,5
PP2_bis_F	/	10µM	0,5	7,5
PP2a_bis_R	/	10µM	0,5	7,5
PP2b_NGS_F	/	10µM	0,5	7,5
PP2b_NGS_R	/	10µM	0,5	7,5
DNMT3A_X23_F1	/	10µM	0,5	7,5
DNMT3A_X23_R1	/	10µM	0,5	7,5
TOTAL	/		46	690

ADN	/	**100ng/µl**	4	

Volume final	50

1 PCR

Test : 4 patients LAM + 3 patients LLC + PBL = 8 (MOU.CH. / DEL.CO. / DES.MI. / GER.JE.)
Test : 54/58/60°C = 3 PBL (CAR .MI. / LIO.MI. / LAG.AN.)
Test : 27 / 33 cycles = 2 PBL

FLT3_TKD et ITD / NPM1 / IDH1 / IDH2 / TP53 / KIT_X17				
	LOT N°	Concentration	n=1(µL)	
Nbre de tubes	/	/	1	15
EAU	/	/	8	120
AMPLITAQ GOLD 360 GC		2X	25	375
FLT3_X20_NGS_F1	/	10µM	0,5	7,5
FLT3_X20_NGS_R1	/	10µM	0,5	7,5
NPM1_X12_F1	/	10µM	0,5	7,5
NPM1_X12_R1	/	10µM	0,5	7,5
FLT3_NGS_F	/	10µM	0,5	7,5
FLT3_NGS_R	/	10µM	0,5	7,5
IDH1_X4_F2	/	10µM	0,5	7,5
IDH1_X4_R2	/	10µM	0,5	7,5
IDH2_F1	/	10µM	0,5	7,5
IDH2_R2	/	10µM	0,5	7,5
KIT _X17_NGS_F1	/	10µM	0,5	7,5
KIT_X17_R3	/	10µM	0,5	7,5
TP53_X04_NGS_bis_F1	/	10µM	0,5	7,5
TP53_X04_R22	/	10µM	0,5	7,5
TP53_X05_F12	/	10µM	0,5	7,5
TP53_X05_NGS_R	/	10µM	0,5	7,5
TP53_X06_NGS_F1	/	10µM	0,5	7,5
TP53_X06_R1	/	10µM	0,5	7,5
TP53_X07_NGS_F1	/	10µM	0,5	7,5
TP53_X07_R1	/	10µM	0,5	7,5
TP53_X08_F1	/	10µM	0,5	7,5
TP53_X08_NGS_bis_R1	/	10µM	0,5	7,5
TP53_X09_NGS_F1	/	10µM	0,5	7,5
TP53_X09_R1	/	10µM	0,5	7,5
TP53_X10_NGS_F1	/	10µM	0,5	7,5
TP53_X10_R1	/	10µM	0,5	7,5
TOTAL	/		46	690

ADN	/	**100ng/µl**	4	

Volume final	50

1 PCR
Test : 4 patients LAM + 3 patients LLC + PBL = 8 (MOU.CH. / DEL.CO. / DES.MI. / GER.JE.)
Test : 54/58/60°C = 3 PBL (CAR .MI. / LIO.MI. / LAG.AN.)
Test : 27 / 33 cycles = 2 PBL

CEBPa / DNMT3a / FLT3_TKD et ITD / NPM1 / IDH1 / IDH2 / TP53 / KIT_X17				
	LOT N°	Concentration	n=1(µL)	
Nbre de tubes	/	/	1	5
EAU	/	/	0 ou 10	0
+/- ENHANCER 360 GC		5X	10	50
AMPLITAQ GOLD 360 GC		2X	25	125
PP1b_NGS_F	/	10µM	0,3	1,5
PP1_R	/	10µM	0,3	1,5
PP1_F	/	10µM	0,3	1,5
PP1a_NGS_R	/	10µM	0,3	1,5
PP2c_NGS_F	/	10µM	0,3	1,5
PP2_R	/	10µM	0,3	1,5
PP2_bis_F	/	10µM	0,3	1,5
PP2a_bis_R	/	10µM	0,3	1,5
PP2b_NGS_F	/	10µM	0,3	1,5
PP2b_NGS_R	/	10µM	0,3	1,5
DNMT3A_X23_F1	/	10µM	0,3	1,5
DNMT3A_X23_R1	/	10µM	0,3	1,5
FLT3_X20_NGS_F1	/	10µM	0,3	1,5
FLT3_X20_NGS_R1	/	10µM	0,3	1,5
NPM1_X12_F1	/	10µM	0,3	1,5
NPM1_X12_R1	/	10µM	0,3	1,5
FLT3_NGS_F	/	10µM	0,3	1,5
FLT3_NGS_R	/	10µM	0,3	1,5
IDH1_X4_F2	/	10µM	0,3	1,5
IDH1_X4_R2	/	10µM	0,3	1,5
IDH2_F1	/	10µM	0,3	1,5
IDH2_R2	/	10µM	0,3	1,5
KIT _X17_NGS_F1	/	10µM	0,3	1,5
KIT_X17_R3	/	10µM	0,3	1,5
TP53_X04_NGS_bis_F1	/	10µM	0,3	1,5
TP53_X04_R22	/	10µM	0,3	1,5
TP53_X05_F12	/	10µM	0,3	1,5
TP53_X05_NGS_R	/	10µM	0,3	1,5
TP53_X06_NGS_F1	/	10µM	0,3	1,5
TP53_X06_R1	/	10µM	0,3	1,5
TP53_X07_NGS_F1	/	10µM	0,3	1,5
TP53_X07_R1	/	10µM	0,3	1,5
TP53_X08_F1	/	10µM	0,3	1,5
TP53_X08_NGS_bis_R1	/	10µM	0,3	1,5
TP53_X09_NGS_F1	/	10µM	0,3	1,5
TP53_X09_R1	/	10µM	0,3	1,5
TP53_X10_NGS_F1	/	10µM	0,3	1,5
TP53_X10_R1	/	10µM	0,3	1,5
TOTAL	/		46,4	232
ADN	/	**100ng/µl**	4	

Annexe 4 : Analyse des variants retrouvés pour chaque patient 73

IDENTITE PATIENT	GENE	CHR	POSITION SUR LE GENOME	BASE DE REFERENCE	BASE LUE	QUALITE	PROFONDEUR	FREQUENCE DU VARIANT	FREQUENCE DU VARIANT (%)	RESULTATS ANALYSE NGS
MOUCH	TP53 X7	chr17	7577698	A	T	100	1583	0,1049	10%	Variation dans l'intron 6-7 (voir ligne suivante : deux possibilités de subsitution ?)
		chr17	7577698	A	G	100	1583	0,0398	4%	
		chr17	7577678	CT	C	100	6921	0,1298	13%	Variation dans l'intron 6-7 (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr17	7577678	CTTT	C	100	6921	0,0228	2%	
		chr17	7577678	C	CT	100	6921	0,0328	3%	
		chr17	7577678	CTT	C	100	6921	0,0472	5%	
		chr17	7577696	A	G	100	1491	0,1388	14%	Variation dans l'intron 6-7 (voir ligne suivante : deux possibilités de subsitution ?)
		chr17	7577696	A	T	91	1491	0,0295	3%	
	TP53 X4	chr17	7579472	G	C	100	64866	0,4883	49%	**SNP rs1042522**
	NPM1 X12	chr5	170837543	C	CTCTG	100	22820	0,2491	25%	**Mutation NPM1 type A**
		chr5	170837513	CT	C	100	20327	0,3459	35%	Variation dans l'intron (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr5	170837513	C	CT	100	20327	0,0366	4%	
		chr5	170837513	CTT	C	100	20327	0,054	5%	
	IDH1 X4	chr2	209113048	G	GA	82	73	0,137	14%	Variation dans l'intron
		chr2	209113112	C	T	100	75	0,2	20%	**Mutation R132H**
	FLT3 X13	chr13								**Couverture entre 3000 et 5000 mais pas de mutation détectée.**
	CEBPa	chr19	33792595	A	C	41	8	0,375	38%	**Variation non connue dans l'exon ? Artefact dû à la faible profondeur ?**
		chr19	33792596	C	A	41	8	0,375	38%	
		chr19	33792597	C	G	39	9	0,3333	33%	
		chr19	33792599	A	T	41	8	0,375	38%	
		chr19	33792609	C	G	45	6	0,5	50%	
		chr19	33792610	G	C	45	6	0,5	50%	
		chr19	33792622	GGGCA	G	45	6	0,5	50%	**NM_004364.3(CEBPA_v001):c.695_698del ; NM_004364.3(CEBPA_i001):p.(Val232Alafs*85)**
		chr19	33792629	G	GGT	45	6	0,5	50%	**NM_004364.3(CEBPA_v001):c.691**

										_692insAC ; NM_004364.3(CEBPA_i001):p.(Pro231Hisfs*88)
		chr19	33792633	T	A	45	6	0,5	50%	**Variation non connue dans l'exon ? Artefact dû à la faible profondeur ?**
		chr19	33792656	G	C	29	5	0,4	40%	
		chr19	33792657	G	T	60	7	0,5714	57%	
		chr19	33792660	G	T	76	8	0,625	63%	
		chr19	33792662	A	C	60	7	0,5714	57%	
		chr19	33792664	G	A	43	7	0,4286	43%	
		chr19	33792665	TG	T	25	8	0,25	25%	
		chr19	33792668	A	T	60	7	0,5714	57%	
		chr19	33792731	G	GGCGGGT	24	9	0,2222	22%	**NM_004364.3(CEBPA_v001):c.584_589dup ; NM_004364.3(CEBPA_i001):p.(His195_Pro196dup)**
		chr19	33792845	A	C	46	16	0,25	25%	**Variation non connue dans l'exon ?**
DELCO	TP53 X7	chr17	7577698	A	T	100	2193	0,1008	10%	Variation dans l'intron 6-7 (voir ligne suivante : deux possibilités de subsitution ?)
		chr17	7577698	A	G	75	2193	0,0237	2%	
		chr17	7577678	CT	C	100	9574	0,1014	10%	Variation dans l'intron 6-7 (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr17	7577678	CTTT	C	60	9574	0,0152	2%	
		chr17	7577678	CTT	C	100	9574	0,0346	3%	
		chr17	7577678	C	CT	100	9574	0,0873	9%	
		chr17	7577696	A	G	100	2002	0,1139	11%	Variation dans l'intron 6-7 (voir ligne suivante : deux possibilités de subsitution ?)
		chr17	7577696	A	T	100	2002	0,034	3%	
	TP53 X4	chr17	7579472	G	C	100	54389	0,9966	100%	**SNP rs1042522**
	NPM1 X12	chr5	170837544	T	TCTGC	100	35078	0,2439	24%	**Mutation NPM1 type D**
		chr5	170837513	CT	C	100	31539	0,3439	34%	Variation dans l'intron (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr5	170837513	C	CT	100	31539	0,0417	4%	
		chr5	170837513	CTT	C	100	31539	0,0555	6%	
	KIT X17	chr4	55599129	T	A	100	60886	0,3156	32%	**SNP rs 11935331**
		chr4	55599159	G	A	100	119664	0,3193	32%	**SNP rs 4864921**
	FLT3 X20	chr13	28592642	C	A	100	26963	0,5128	51%	**Mutation D835Y**
		chr13	28608459	T	C	100	5263	0,631	63%	**SNP rs34374211**
	IDH1 X4	chr2	209113048	GA	G	20	83	0,0482	5%	Variation dans l'intron : répétition de A sur la séquence donc erreur de lecture ?
		chr2	209113048	G	GA	100	83	0,1928	19%	
		chr2	20911311	C	T	100	91	0,2308	23%	**Mutation R132H**

			2							
	DNMT3	chr2								**L'amplicon n'a pas été détecté au séquençage**
DESMI	NPM1 X12	chr5	170837543	C	CTCTG	100	16482	0,2678	27%	**Mutation NPM1 type A**
		chr5	170837513	CT	C	100	14661	0,3232	32%	Variation dans l'intron (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr5	170837513	C	CT	100	14661	0,0439	4%	
		chr5	170837513	CTT	C	100	14661	0,0561	6%	
	TP53 X7	chr17	7577678	CT	C	100	4750	0,1347	13%	Variation dans l'intron 6-7 (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr17	7577678	CTTT	C	100	4750	0,0255	3%	
		chr17	7577678	C	CT	100	4750	0,0326	3%	
		chr17	7577678	CTT	C	100	4750	0,0619	6%	
		chr17	7577696	A	G	100	1242	0,1151	12%	Variation dans l'intron 6-7 (voir ligne suivante : deux possibilités de subsitution ?)
		chr17	7577696	A	T	94	1242	0,0322	3%	
	TP53 X4	chr17	7579472	G	C	100	31108	0,9977	100%	**SNP rs1042522**
	IDH2 X4	chr15								**Les 312 bases de l'amplicon ont été lues entre 44000 et 90000 fois et aucune variation n'est détectée.**
	FLT3 X13	chr13								**Couverture entre 3000 et 5000 mais pas de mutation détectée.**
GERJE	TP53 X7	chr17	7577698	A	T	100	1722	0,101	10%	Variation dans l'intron 6-7 (voir ligne suivante : deux possibilités de subsitution ?)
		chr17	7577698	A	G	100	1722	0,0465	5%	
		chr17	7577678	CT	C	100	7391	0,1276	13%	Variation dans l'intron 6-7 (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr17	7577678	CTTT	C	86	7391	0,0176	2%	
		chr17	7577678	C	CT	100	7391	0,0314	3%	
		chr17	7577678	CTT	C	100	7391	0,0457	5%	
		chr17	7577696	A	G	100	1575	0,1384	14%	Variation dans l'intron 6-7 (voir ligne suivante : deux possibilités de subsitution ?)
		chr17	7577696	A	T	100	1575	0,033	3%	
	TP53 X4	chr17	7579472	G	C	100	64998	0,4928	49%	**SNP rs1042522**
	KIT X17	chr4	55599321	A	T	100	49806	0,1458	15%	**Mutation D816V**
	NPM1 X12	chr5	170837513	CT	C	100	40866	0,4461	45%	Variation dans l'intron (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr5	170837513	C	CT	100	40866	0,0493	5%	
		chr5	170837513	CTT	C	100	40866	0,0648	6%	

	FLT3 X13	chr13								**Couverture entre 3000 et 5000 mais pas de mutation détectée ??**
LAGAN	TP53 X7	chr17	7577696	A	G	100	1199	0,1259	13%	Variation dans l'intron 6-7 (voir ligne suivante : deux possibilités de subsitution ?)
		chr17	7577696	A	T	68	1199	0,0284	3%	
		chr17	7577678	CT	C	100	5391	0,1482	15%	Variation dans l'intron 6-7 (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr17	7577678	CTTT	C	100	5391	0,0237	2%	
		chr17	7577678	C	CT	100	5391	0,0304	3%	
		chr17	7577678	CTT	C	100	5391	0,0473	5%	
	TP53 X5	chr17	7578534	C	G	100	135465	0,2263	23%	**Mutation K132N**
		chr17	7578541	A	C	100	132130	0,2678	27%	**Mutation L130R**
	TP53 X4	chr17	7579472	G	C	100	63669	0,4955	50%	**SNP rs1042522**
	NPM1 X12	chr5	170837513	C	CT	100	40165	0,1655	17%	Variation dans l'intron (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr5	170837513	CT	C	100	40165	0,4219	42%	
		chr5	170837513	C	CTT	100	40165	0,0311	3%	
		chr5	170837513	CTT	C	100	40165	0,0903	9%	
	KIT X17	chr4	55599129	T	A	100	16476	0,2747	27%	**SNP rs 11935331**
		chr4	55599159	G	A	100	32392	0,2771	28%	**SNP rs 4864921**
LIOMI	IDH1 X4	chr2	209113048	G	GA	100	145	0,1448	14%	Variation dans l'intron.
	TP53 X7	chr17	7577678	CT	C	100	5942	0,1415	14%	Variation dans l'intron 6-7 (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr17	7577678	CTTT	C	100	5942	0,0266	3%	
		chr17	7577678	C	CT	100	5942	0,0311	3%	
		chr17	7577678	CTT	C	100	5942	0,0512	5%	
		chr17	7577696	A	G	100	1295	0,1598	16%	Variation dans l'intron 6-7 (voir ligne suivante : deux possibilités de subsitution ?)
		chr17	7577696	A	T	100	1295	0,0355	4%	
	TP53 X4	chr17	7577121	G	A	100	54543	0,1693	17%	**Mutation R273C**
	KIT X17	chr4	55599159	G	A	100	47490	0,3484	35%	**SNP rs 4864921**
	NPM1 X12	chr5	170837513	CT	C	100	55204	0,4526	45%	Variation dans l'intron (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr5	170837513	C	CT	100	55204	0,0457	5%	
		chr5	170837513	CTT	C	100	55204	0,0769	8%	
CARMI	TP53 X7	chr17	7577678	CT	C	100	6403	0,1396	14%	Variation dans l'intron 6-7 (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr17	7577678	CTTT	C	100	6403	0,0237	2%	
		chr17	7577678	C	CT	100	6403	0,0339	3%	
		chr17	7577678	CTT	C	100	6403	0,0487	5%	
		chr17	7577696	A	G	100	1460	0,1452	15%	Variation dans

		chr17	7577696	A	T	62	1460	0,0253	3%	l'intron 6-7 (voir ligne suivante : deux possibilités de subsitution ?)
	NPM1 X12	chr5	170837513	CT	C	100	27211	0,4467	45%	Variation dans l'intron (voir lignes suivantes : répétition de T sur la séquence = erreur de lecture ?)
		chr5	170837513	C	CT	100	27211	0,0513	5%	
		chr5	170837513	CTT	C	100	27211	0,06	6%	
	TP53 X4	chr17	7579472	G	C	100	6034	0,9654	97%	**SNP rs1042522**
		chr17								**Couverture à 6000, la délétion de 12 pb n'est pas détectée.**

<u>Critères d'analyse définis</u> : Profondeur > 1000 et Fréquence des mutations >10%

Mutations analysées mais dont les critères de profondeur <u>ou</u> de fréquence définis ne sont pas respectés

Concordance avec le séquençage Sanger

Discordance avec le séquençage Sanger

SNP

Annexe 5 : Analyse financière comparative du coût par patient du NGS 74

EVALUATION FINANCIERE NGS - 24 gènes (NGS_V1) Amplicons 250pb

MATERIEL

TruSeq Custom Amplicon (*Illumina*)
15 600 euros / 96 échantillons
162,50 euros / échantillon

Index (*Illumina*)
210 euros / 24 échantillons x 4 runs
52,50 euros / 24 échantillons

Nextera X^{T} (*Illumina*)
2 433 euros / 96 échantillons
25 euros / échantillon

Index (*Illumina*)
210 euros / 24 échantillons x 4 runs
2 euros / échantillons

AmpliTaq Gold 360 Master Mix (*Life Technologies*)
250 euros / kit
20 euros / échantillon

Amorces + Tag (*Eurofin*)
10 euros / amplicon (2 amorces/amplicon)
10 euros / échantillon

Nombre amplicons 250pb : 500

		PROFONDEUR (X)	PRIX / ECH. (euros)	PRIX / ECH. (euros)
Flowcell Standard 2x250 14Millions Clusters 7Gb Output 1000 euros	*24 échantillons*	1166*	206**	99****
	12 échantillons	2332	248	140
	6 échantillons	4666	331	223
Nano Flowcell 2x250 1Million Clusters 500Mb Output 520 euros	*24 échantillons*	83	186***	79****
	12 échantillons	166	208	100
	6 échantillons	332	251	143
Flowcell V3 2x300 25Millions Clusters 15Gb Output 1100 euros	*24 échantillons*	2080	211	103
	12 échantillons	4160	256	149
	6 échantillons	8320	348	240

*(=(14millions/500/24)
**(=(162,5x24+52,5+1000)/24)
***(=(162,5x24+52,5+520)/24)

****(=10+20+2+25+(1000/24))

*****(=10+20+2+25+(520/24))

Evaluation sur 2012 : *LAM = 140 échantillons*
MDS = 266 échantillons

Prévisions sur 2014 : *LAM seules è 12 échantillons / mois*
LAM + MDS è 34 échantillons / mois

XI REFERENCES BIBLIOGRAPHIQUES

1. Patel J.P, et *al*., Prognostic relevance of integrated genetic profiling in acute myeloid leukemia. N Engl J Med, 2012. 366(12): p. 1079-89
2. Bejar R., et *al*., Validation of a prognostic model and the impact of mutations in patients with lower-risk myelodysplastic syndromes. J Clin Oncol, 2012. 30(27): p. 3376-82
3. Deschler B, et *al.,* Acute myeloid leukemia: epidemiology and etiology. Cancer, 2006; 107 (9): 2099-107
4. Stone R. M., et *al*., Acute myeloid leukemia. Hematology Am Soc Hematol Educ Program, (2004) 98-117
5. Pabst T, et *al*., Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. Nat Genet. 2001 Mar;27(3):263-70
6. Pasquet M, et *al*., High frequency of GATA2 mutations in patients with mild chronic neutropenia evolving to MonoMac syndrome, myelodysplasia, and acute myeloid leukemia., Blood. 2013 Jan 31;121(5):822-9. doi: 10.1182/blood-2012-08-447367. Epub 2012 Dec 6
7. Welch, J.S., Genomics of AML : Clinical Applications of Next-Generation Sequencing – Am Soc Hematol 2011:30-5
8. Swerdlow S. H., et *al*., WHO Classification of Tumors of Haematopoietic and Lymphoid Tissues - International Agency for Research on cancer - (2008) - 4th Edition
9. Grimace D, et *al*., The importance of diagnostic cytogenetics on outcome in AML: analysis of 1,612 patients entered into the MRC AML 10 trial. The Medical Research Council Adult and Children's Leukaemia Working Parties. Blood, 1998; 92 (7): 2322-33
10. Melnick, A., et *al.,* Deconstructing a disease: RARalpha, its fusion partners, and their roles in the pathogenesis of acute promyelocytic leukemia. (1999), Blood 93, 3167-3215
11. Speck N. A., et *al*., Core-binding factor: a central player in hematopoiesis and leukemia. Cancer (1999),.Res 59, 1789s-1793s
12. Smith ML, et *al.,* Independent prognostic variables in acute myeloid leukaemia. - Blood Rev. 2011 Jan;25(1):39-51. doi: 10.1016/j.blre.2010.10.002. Epub 2010 Nov 27. Review
13. Krzysztof M, et *al*., Chromosome Aberrations, Gene Mutations and Expression Changes, and Prognosis in Adult Acute Myeloid Leukemia – ASH _ Hematology 2006
14. Tsimberidou, A. M., et *al*., The role of gemtuzumab ozogamicin in acute leukaemia therapy. Br J Haematol 132 (2006).
15. Nakao M, et *al*., Internal tandem duplication of the flt3 gene found in acute myeloid leukemia. Leukemia - 1996 Dec;10(12):1911-8I
16. Schnittger S, et *al*., Analysis of FLT3 length mutations in 1003 patients with acute myeloid leukemia: correlation to cytogenetics, FAB subtype, and prognosis in the AMLCG study and usefulness as a marker for the detection of minimal residual disease.- Blood. 2002 Jul 1;100(1):59-66
17. Fröhling S, et *al*., Prognostic significance of activating FLT3 mutations in younger adults (16 to 60 years) with acute myeloid leukemia and normal cytogenetics: a study of the AML Study Group Ulm. - Blood. 2002 Dec 15;100(13):4372-80. Epub 2002 Aug 8

18. Schnittger S, et *al*., Nucleophosmin gene mutations are predictors of favorable prognosis in acute myelogenous leukemia with a normal karyotype. Blood, 2005; 106 (12): 3733-9

19. Dohner K, et *al*., Mutant nucleophosmin (NPM1) predicts favorable prognosis in younger adults with acute myeloid leukemia and normal cytogenetics: interaction with other gene mutations. Blood, 2005; 106 (12): 3740-6

20. Thiede C, et *al*., Prevalence and prognostic impact of NPM1 mutations in 1485 adult patients with acute myeloid leukemia (AML). Blood, 2006;1070 :4011-20

21. Pabst T, et *al*., Dominant-negative mutations of CEBPA, encoding CCAAT/enhancer binding protein-alpha (C/EBPalpha), in acute myeloid leukemia. Nat Genet. 2001 Mar;27(3):263-70}

22. Wouters BJ, et *al*., Double CEBPA mutations, but not single CEBPA mutations, define a subgroup of acute myeloid leukemia with a distinctive gene expression profile that is uniquely associated with a favorable outcome. Blood 2009 ; 113 : 3088-91

23. Pabst T, et *al*., Heterogeneity within AML with CEBPA mutations; only CEBPA double mutations, but not single CEBPA mutations are associated with favourable prognosis. Br J Cancer 2009 ; 100 : 1343-6

24. Mardis ER, et *al*., Recurring mutations found by sequencing an acute myeloid leukemia genome. N Engl J Med 2009;361:1058-66

25. Dang L, et *al*., Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. Nature 2009 ; 462 : 739-44

26. Ley TJ, et *al*., DNMT3A mutations in acute myeloid leukemia. N Engl J Med 2010;363:2424-33

27. Bennett JM, et *al*., Proposals for the classification of the acute leukaemias. French-American-British (FAB) co-operative group., Haematol., 1976 Aug;33(4):451-8

28. Yoshida K, et *al*., Frequent pathway mutations of splicing machinery in myelodysplasia. Nature 2011 ; 478 : 64-9

29. Greenberg P, et *al*., International scoring system for evaluating prognosis in myelodysplastic syndromes. Blood 1997; 89 : 2079-88

30. Maxam A, M, et *al.,* A new method for sequencing DNA. 1977. Biotechnology. 1992;24:99-103

31. Sanger F, Nicklen S, et al., DNA sequencing with chain-terminating inhibitors. 1977. - Coulson AR. - Biotechnology. 1992;24:104-8

32. Metzker, M.L., Sequencing technologies - the next generation (2010) - Nat Rev Genet, 11(1)31-46

33. Mardis, E.R. Anticipating the 1,000 dollar genome. (2006) Genome Biol, 7, 112

34. Sboner, A., et *al*., The real cost of sequencing: higher than you think! (2011) Genome Biol, 12, 125

35. Glenn, T.C., Field guide to next-generation DNA sequencers. (2011) Mol Ecol Resour, 11, 759- 769

36. Misale, S., et *al*., Emergence of KRAS mutations and acquired resistance to anti-EGFR therapy in colorectal cancer (2012) – Nature 486(7404):532-6.

37. Dessai, A.N., et *al.,* Next-generation sequencing : ready for the clinics? – (2012) – Clin Genet 81:503-10

38. Wong, K.M., Hudson, T.J., et al., Unraveling the Genetics of Cancer : Genome Sequencing and Beyond – (2011) – Annu. Rev Genomics Hum Genet 12:407-30

39. Vivien Marx -Technology Feature – The genome jigsaw – Nature vol 501 – 263

40. Rowley JD. Chromosomal translocations: revisited yet again. Blood 2008;112: 2183-9.2

41. Mrózek K, et *al.,* Cytogenetics in acute leukemia. Blood Rev 2004;18:115-36

42. Dohner H, et *al*., Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. Blood 2010;115:453-74

43. Mrózek K, et *al*., Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. J Clin Oncol 2012;30:4515-23.

44. Mrozek, K., et *al.,* Chromosome aberrations, gene mutations and expression changes, and prognosis in adult acute myeloid leukemia (2006). Hematology Am Soc Hematol Educ Program, 169-177

45. Delhommeau F, et *al*., Mutation in TET2 in myeloid cancers. N Engl J Med, 2009. 360(22): p. 2289-301

46. Genomic and Epigenomic Landscapes of Adult De Novo Acute Myeloid Leukemia The Cancer Genome Atlas Research Network May 1, 2013, at NEJM.org

47. Marceau-Renaut A, et *al*., .Le séquençage de nouvelle generation (NGS) a-t-il deja sa place dans nos laboratoires d'oncohematologie ? Hematologie 2013 ; 19 : 112-122.

48. Bejar R, et *al*., Clinical effect of point mutations in myelodysplastic syndromes. N Engl J Med 2011 ; 364 : 2496-506

49. Jadersten M, et *al*., TP53 mutations in low-risk myelodysplastic syndromes with del(5q) predict disease progression. J Clin Oncol 2011 ; 29 : 1971-9

50. Schnittger S, et *al*., ASXL1 exon 12 mutations are frequent in AML with intermediate risk karyotype and are independently associated with an adverse outcome. Leukemia. 2013 Jan;27(1):82-91. doi: 10.1038/leu.2012.262. Epub 2012 Sep 11.

51. Thol F, et *al*., Incidence and Prognostic Influence of *DNMT3A* Mutations in Acute Myeloid Leukemia, JCO.2011.35.4894.

52. Langemeijer SMC, et *al*., Acquired mutations in TET2 are common in myelodysplastic syndromes.Nat Genet 2009 ; 41 : 838-42

53. Kosmider O, et *al*., TET2 mutation is an independent favorable prognostic factor in myelodysplastic syndromes (MDSs). Blood 2009 ; 114 : 3285-91

54. Nibourel O, et *al*., Association of TET2 Alterations with NPM1 Mutations and Prognostic Value in De Novo Acute Myeloid Leukemia (AML). ASH Annual Meeting Abstracts 2009; 114 : 163.

55. Rücker FG, et *al*., TP53 alterations in acute myeloid leukemia with complex karyotype correlate with specific copy number alterations, monosomal karyotype, and dismal outcome. Blood 2012 ; 119 : 2114-21

56. Schnittger S, et *al*., KIT-D816 mutations in AML1-ETO-positive AML are associated with impaired event-free and overall survival. Blood 2006; 107: 1791-9

57. Cairoli R, et *al*., Prognostic impact of c-KIT mutations in core binding factor leukemias: An Italian retrospective study. Blood 2006 ; 107 : 3463-8

58. Neubauer, A., et *al*., Patients with acute myeloid leukemia and RAS mutations benefit most from postremission high-dose cytarabine: a Cancer and Leukemia Group B study. J Clin Oncol, 2008. 26(28): p. 4603-9

59. Neubauer, A., et *al*., Mutations in the ras proto-oncogenes in patients with myelodysplastic syndromes. Leukemia, 1994. 8(4): p. 638-41

60. Jeromin, S., et *al*., High frequencies of SF3B1 and JAK2 mutations in refractory anemia with ring sideroblasts associated with marked thrombocytosis strengthen the assignment to the category of myelodysplastic/myeloproliferative neoplasms. Haematologica, 2013. 98(2): p. e15-7

61. Thompson, C.B., Metabolic enzymes as oncogenes or tumor suppressors. N Engl J Med, 2009. 360(8): p. 813-5.

62. Abbas, S., et *al*., Acquired mutations in the genes encoding IDH1 and IDH2 both are recurrent aberrations in acute myeloid leukemia: prevalence and prognostic value. Blood, 2010. 116(12): p. 2122-6

63. Papaemmanuil E, et *al*., Somatic SF3B1 mutation in myelodysplasia with ring sideroblasts. N Engl J Med. 2011;365(15): n1384-1395

64. Thol F, Kade S, et *al*., Frequency and prognostic impact of mutations in SRSF2, U2AF1, and ZRSR2 in patients with myelodysplastic syndromes. Blood 2012; 119: 3578–3584

65. Thol F, et *al*., SETBP1 mutation analysis in 944 patients with MDS and AML Leukemia advance online publication, 21 May 2013;doi:10.1038/leu.2013.145

66. Schnittger, S., et *al*., Use of CBL exon 8 and 9 mutations in diagnosis of myeloproliferative neoplasms and myelodysplastic/myeloproliferative disorders: an analysis of 636 cases. Haematologica, 2012. 97(12): p. 1890-4

67. Bacher, U., et *al*., Mutations of the TET2 and CBL genes: novel molecular markers in myeloid malignancies. Ann Hematol, 2010. 89(7): p. 643-52.

68. Dreszer, T.R., et *al*., The UCSC Genome Browser database: extensions and updates 2011. Nucleic Acids Res, 40, D918-923

69. Abdelhamid E, et *al*., Minimal residual disease monitoring based on FLT3 internal tandem duplication in adult acute myeloid leukemia. Leuk Res 2012 ; 36 : 316-23.

70. Thol F, et *al*., Next-generation sequencing for minimal residual disease monitoring in acute myeloid leukemia patients with FLT3-ITD or NPM1 mutations. Genes Chromosomes Cancer 2012 ; 51 : 689-95

71. Wu D, et *al*., High-throughput sequencing detects minimal residual disease in acute T lymphoblastic leukemia. Sci Transl Med 2012 ; 4 : 134ra63

XIII LISTE DES FIGURES

Figure 1 : Différenciation des cellules souches hématopoïétiques 13

Figure 2 : Action des facteurs multipotents sur les CSH 14

Figure 3 : Action des facteurs restreints sur les CSH 14

Figure 4 : Dérégulation de la différenciation cellulaire 15

Figure 5 : Blastes typique d'une LAM5 visibles sur le frottis sanguin secondaire 18

Figure 6 : Blastes présentant des corps d'Auer typiques de LAM et visibles sur le frottis médullaire 18

Figure 7 : Anomalies des cellules typiques de SMD visibles sur le frottis médullaire 25

Figure 8 : Principe du séquençage enzymatique Sanger 28

Figure 9 : Principe du HRM – Dénaturation et courbe de fusion 28

Figure 9 bis : Principe du HRM – Mutation 28

Figure 10 : Amélioration des techniques de séquençage de nouvelle génération . 29

Figure 11 : Evolutions du prix du séquençage 29

Figure 12 : Principe de la technologie du séquenceur *454* 31

Figure 13 : Principe de la technologie *SOLID* 31

Figure 14 : Principe de la technologie *Ion Torrent* 32

Figure 15 : Exemple de la structure d'une *flowcell* utilisée sur le *HiSeq* d'*Illumina* 33

Figure 16 : Principe de la technologie du séquenceur *Illumina* 33

Figure 17 : Principe de l'émission de fluorescence lors du séquençage *Illumina* ... 33

Figure 18 : Gamme de séquenceurs *Illumina* 33

Figure 19 : Principe de la technologie des *Nanopores* 34

Figure 20 : Principe de la technologie *PacBio RS* 34

Figure 21 : Proportion des gènes mutés dans une cohorte de 200 patients 36

Figure 22 : Regroupement des types de mutations dans une cohorte de 200 patients et correspondances 36

Figure 23 : Interconnexion des mutations de gènes et groupes de risques cytogénétiques 37
Figure 24 : Complexité des mutations retrouvées dans les LAM 37
Figure 25 : Etapes du kit *TruSeq* 47
Figure 26 : Composition des adaptateurs *Illumina* 50
Figure 27 : Visualisation d'une mutation sur le logiciel d'analyse « Mutation Surveyor » 54
Figure 28 : Taille des fragments obtenus pour la PCR 1 de NGS sur gel d'acrylamide 58
Figure 29 : *Workflow* de la préparation de la librairie en technique d'enrichissement par PCR mulitplex 60
Figure 30 : Résultats obtenus d'après la puce Agilent sur la PCR 2 de NGS 61
Figure 31 : Liste non exhaustive des logiciels d'analyse de données issues du NGS et de leur champ d'application 66
Figure 32 : Principe d'acquisition d'image 66
Figure 33 : Ecran principal « Analysis » du logiciel « SAV » d'*Illumina* - Exemple de notre *run* 67
Figure 34 : Ecran « Summary » du logiciel « SAV » d'*Illumina* - Exemple de notre *run* 67
Figure 35 : Informations à intégrer au logiciel « MiSeq Reporter » nécessaires à l'analyse des données 68

XIV LISTE DES TABLEAUX

Tableau I Comparaison de la fréquence relative des mutations dans les LAM et SMD 11

Tableau II Classifications cytogénétiques pronostiques anglaise et française des LAM 19

Tableau III Classification FAB des LAM 24

Tableau IV Révision 2008 de la classification OMS des leucémies aigües et néoplasies associées 24

Tableau V Classification OMS – 2008 des SMD 26

Tableau VI Score IPSS 27

Tableau VII Caractéristiques des technologies de séquençage 30

Tableau VIII Q-score et précision 35

Tableau IX Impact pronostic des mutations retrouvées dans les SMD 37

Tableau X Evolution de la liste de gènes à une V2 comprenant 81 gènes 49

Tableau XI Résultats des recherches de mutations par les techniques « classiques » sur notre cohorte de patients tests 55

Tableau XII Valeurs de Ct obtenues pour le test des couples d'amorces en *SyBrGreen* avec et sans *Enhancer* 57

Tableau XIII Dosage des produits post PCR 1 pour chaque échantillon 59

Tableau XIV Dosage des produits de PCR1 purifiée à ajouter dans le mix de PCR2 et concentration initiale de PCR2 60

Tableau XV Taille des amplicons attendue avec queue flottante et avec adaptateurs 62

Tableau XVI Volume de PCR 2 à mélanger de façon équimolaire pour la PCR quantitative 62

Tableau XVII Comparaison des différentes *Flow-cells* disponibles 63

Tableau XVIII Profondeur des différents amplicons pour les PBL en conditions standard 70

Tableau XIX Profondeur des différents amplicons pour les PBL pour un nombre de cycles de PCR variable 71
Tableau XX Profondeur des différents amplicons pour les PBL pour une température de PCR variable 71
Tableau XXI Profondeur des différents amplicons pour les mix communs avec et sans *Enhancer* 72

L'intégration du séquençage de nouvelle génération dans le diagnostic médical : application aux leucémies aiguës myéloïdes et syndromes myélodysplasiques.

The integration of the new generation sequencing in the medical diagnosis : application to acute myeloïd leukemias and myelodysplasic syndromes.

Mémoire d'Ingénieur C.N.A.M., Toulouse, 2013.

RESUME

Le séquençage de nouvelle génération (NGS) représente un véritable saut technologique. Si les premières applications ont surtout été axées sur le séquençage de génome entier dans le but d'identifier de nouvelles mutations géniques, il fait aujourd'hui son apparition dans les laboratoires hospitaliers. En Onco-Hématologie, les applications hospitalières permettraient de mettre en évidence des mutations de plus en plus nombreuses avec un intérêt dans la prise en charge thérapeutique du patient, ce avec de nombreux avantages mais également certaines contraintes. Nous comparerons les techniques de PCR et HRM classiques en association avec le séquençage Sanger aux résultats obtenus par NGS afin de savoir si le NGS pourrait passer en routine.

Mots clés : NGS, leucémies, syndromes myélodysplasiques, mutations, bio-informatique

---

SUMMARY

New-generation sequencing (NGS) is a true technological leap forward. Although the first applications have mainly focused on whole-genome sequencing intended to identify new gene mutations, it is appearing today in hospital laboratories. Hospital applications in hematologic oncology would allow to highlight increasingly many mutations with an interest in the therapeutic assumption of the patient, and thus with numerous advantages but also some constraints. We will compare the techniques of PCR and classical HRM in partnership with the Sanger sequencing, with the results obtained with NGS in order to know if NGS could become a technique of routine.

Key words : NGS, leukemias, myelodysplasic syndromes, mutations, bioinformatics