HAL
open science

Identification et caractérisation de polymorphismes génétiques impliqués dans la réponse à l'imatinib dans la leucémie myéloïde chronique

Florence Lichou

To cite this version:

Florence Lichou. Identification et caractérisation de polymorphismes génétiques impliqués dans la réponse à l'imatinib dans la leucémie myéloïde chronique. Génétique. Université de Bordeaux, 2019. Français. NNT : 2019BORD0066 . tel-03228257

HAL Id: tel-03228257

https://theses.hal.science/tel-03228257v1

Submitted on 18 May 2021

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

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

Université de Bordeaux

Année 2019 Thèse n°84755

THÈSE

pour le

DOCTORAT DE L'UNIVERSITÉ DE BORDEAUX

Mention : Sciences de la Vie et de la Santé

Option : Génétique

Présentée et soutenue publiquement le 17 mai 2019

par

Florence Lichou

Née le 28 janvier 1993 à Châtellerault (86)

Identification et caractérisation de polymorphismes génétiques impliqués dans la réponse à l'imatinib dans la leucémie myéloïde chronique

---

Inserm U1218 ACTION, équipe « Oncogenèse mammaire et leucémique »

Membres du jury :

Pr. Stephen Manon, DR2 CNRS, Bordeaux	Président
Dr. Anne-Sophie Lia, MCU-PH, Limoges	Rapporteur
Dr. Stéphane Prost, Chercheur CEA, Paris	Rapporteur
Dr. Béatrice Turcq, CRHC CNRS, Bordeaux	Directeur de thèse

Identification et caractérisation de polymorphismes génétiques impliqués dans la réponse à l'imatinib dans la leucémie myéloïde chronique

La leucémie myéloïde chronique (LMC) est un syndrome myéloprolifératif rare traité par des inhibiteurs de tyrosine kinase, tels que l'imatinib. Malgré son efficacité, la résistance au traitement est un problème récurrent. Des variants génétiques responsables d'une altération de la mort cellulaire programmée (apoptose) pourraient notamment expliquer l'hétérogénéité de la réponse au traitement entre les patients. Dans un premier temps, l'objectif était de rechercher des variants candidats. Pour cela un panel de 45 gènes impliqués dans l'apoptose a été étudié par séquençage nouvelle génération chez 24 patients atteints de LMC, 12 répondeurs et 12 résistants au traitement par imatinib. A l'aide d'outils informatiques, 473 polymorphismes ont été détectés. Le nombre de patients étudiés étant limité, de nouvelles méthodes statistiques ont dû être développées pour analyser les résultats obtenus. Tout d'abord, les fréquences de survenue des variants chez les patients résistants et répondeurs ont été comparées aux fréquences observées dans la population générale et visualisées par une approche de statistiques descriptives. Cette stratégie a permis de réduire la liste à 95 polymorphismes pouvant être impliqués dans la résistance au traitement. Par la suite, les gènes ont été classés selon leur enrichissement en allèles variants. Au final, trois gènes candidats ont été choisis et séquencés chez 103 patients. Cette méthodologie, automatisée grâce à un algorithme informatique, a permis de mettre en évidence, un variant non-synonyme dans le gène *BCL-RAMBO*, retrouvé plus fréquemment chez les patients résistants de manière significative. Dans un second temps, l'objectif était de caractériser le rôle de ce variant dans la réponse à l'imatinib à l'aide de lignées cellulaires modifiées par la technologie CRISPR-Cas9. Des cellules n'exprimant plus la protéine ont été obtenues et ont permis de mettre en évidence le rôle majeur de la protéine BCL-RAMBO dans l'inhibition de l'apoptose. Des lignées cellulaires portant le variant candidat ont également été créées à l'aide d'une nouvelle technique utilisant CRISPR-Cas9 : l'exon entier contenant le nucléotide d'intérêt a été remplacé par un exon modifié. La modification d'un acide aminé induite par le variant a été associé à une perte de la sensibilité au traitement par imatinib dans ces lignées, comme suggéré après séquençage des patients. Ces données indiquent que BCL-RAMBO, facteur anti-apoptotique dans une lignée modèle de LMC, pourrait devenir une nouvelle cible thérapeutique afin de surmonter la résistance à l'imatinib.

Mots-clés : leucémie myéloïde chronique, résistance au traitement, apoptose, séquençage nouvelle génération, polymorphismes génétiques, BCL-RAMBO

Identification and characterization of genetic polymorphisms associated to imatinib sensitivity in chronic myeloid leukemia

Chronic myeloid leukemia (CML) is a rare myeloproliferative syndrome treated by tyrosine kinase inhibitors, such as imatinib. Despite its efficacy, resistance to treatment is a persistent clinical issue. Notably, genetic variants causing alterations in apoptosis may explain heterogeneity of imatinib sensitivity between patients. First, the goal was to look for candidate variants. For that purpose, a panel of 45 apoptotic genes was assessed by next-generation sequencing on 24 CML patients, 12 sensitive and 12 resistant to imatinib treatment. Using informatics tools, 473 polymorphisms were detected. As the number of sequenced samples was limited, novel statistical methods had to be developed to interpret the results. The variant frequency in resistant and sensitive patients was compared to variant frequency in the general population and visualized using descriptive statistics. This strategy allowed to obtain a restricted list of 95 polymorphisms that might be involved in resistance to the treatment. Then, genes were ranked according to variant allele enrichment. At the end, three candidate genes were chosen and sequenced for 103 CML patients. This methodology, automated using a computer algorithm, permitted to highlight a nonsynonymous variant in the *BCL-RAMBO* gene, significantly found more often in resistant patients. Second, the objective was to characterize the role of this variant in response to imatinib using model cell lines modified by CRISPR-Cas9 technology. *BCL-RAMBO* knock-out cells were obtained and allowed to demonstrate the major role of BCL-RAMBO protein in apoptosis inhibition. Additionally, cell lines carrying the variant were created using a new CRISPR-Cas9 mediated technique: the whole exon carrying the nucleotide of interest was replaced with a variant exon. The amino acid change induced by the identified polymorphism was associated with a loss of sensitivity to imatinib treatment in these cell lines as suggested after patient sequencing. These data indicate that BCL-RAMBO, anti-apoptotic factor in a CML cell line, could become a novel therapeutic target to overcome drug inefficacy for a subset of resistant patients.

Keywords: chronic myeloid leukemia, treatment resistance, apoptosis, next-generation sequencing, genetic polymorphisms, BCL-RAMBO

Remerciements

Je tiens tout d'abord à remercier le Pr. Stephen Manon d'avoir accepté de présider le jury. Je remercie également le Dr. Anne-Sophie Lia et le Dr. Stéphane Prost d'avoir accepté d'être rapporteurs de ce travail.

Je remercie le Pr. Pierre Soubeyran de m'avoir accueillie avec bienveillance au sein de l'unité Inserm 1218 ACTION. Je remercie également le Pr. François-Xavier Mahon de m'avoir accueillie au sein de l'équipe « Oncogenèse mammaire et leucémique ». Merci FX pour ton aide et tes conseils.

Un merci infini à Béa. Ces dernières années tu as été bien plus qu'une directrice de thèse. J'ai mis du temps à te tutoyer mais tu es aujourd'hui une amie et une confidente. Merci de m'avoir soutenue avant même que je n'arrive à Bordeaux dans nos toutes premières galères administratives. Cette aventure a été unique grâce à toi et j'ai appris énormément à tes côtés. Tu m'as permis de me surpasser et de saisir les opportunités qui se sont présentées pendant cette thèse. Merci pour ton investissement dans ce travail, tes conseils et ta patience qui a été mise à rude épreuve, surtout ces derniers jours !

J'ai travaillé ces dernières années dans la meilleure équipe de recherche. Merci à tous les membres, anciens et actuels, de l'équipe « Oncogenèse mammaire et leucémique ». C'était un plaisir de travailler à vos côtés. Merci à Achille, Aline, Amal, Amine, Ana, Bruno, Elodie B, Elodie R, Katie, Kelly, Julie, Olivier, Richard, Stéphane et Stéphanie. Merci à Francis, l'expert BIM, qui a débuté ce projet avec nous et partagé ses connaissances et ses idées. Merci Valérie PM de m'avoir fait découvrir l'outil incroyable CRISPR-Cas9 et merci Valérie L de m'avoir montré le monde obscur du western blot. Votre aide pour les manips a été précieuse. Merci à Elodie qui a participé au projet HIPER et multiplié les PCR pour mettre au point cette approche. Merci à Matthieu, mon binôme de thèse. J'ai eu la chance de vivre cette aventure avec toi, même si au départ, l'idée de passer 3 ans au côté d'un amateur de dinde enrobée de lard et de sauce à la menthe ne m'emballait pas ... Nous avons partagé beaucoup de choses ensemble : des iPad, des voyages aux Etats-Unis, des séances de muscu mais surtout beaucoup de fous rires. Je te souhaite plein de réussite dans tous tes futurs projets.

Je tiens également à remercier tous mes stagiaires, notamment Anaïs et Julie, qui ont activement participé au projet. Vos questions multiples m'ont permis de prendre du recul et m'ont donné de nouvelles idées.

Je souhaite également remercier les membres de l'équipe hémato U1035 avec qui nous avons partagé de nombreux déjeuners et pauses café : Amélie, Arnaud, Isabelle, Jean-Max, Pierre-Yves et Vanessa.

Un merci général à tous les résidents du dernier étage du bâtiment TP.

Merci au Dr. Michel Longy de m'avoir aidée à élucider les mystères des polymorphismes génétiques et merci au Dr. Gabriel Etienne de nous avoir consacré du temps pour mettre au point la cohorte de patients. Un grand merci également à toute l'équipe de séquençage de l'institut Bergonié, Delfine Lafon, Françoise Bonnet, Jennifer Chiron et au Pr. Nicolas Sévenet. Merci Christophe Hubert pour ton aide précieuse en NGS. Je remercie le Dr. Alain Monnereau pour son aide sur les approches statistiques. Un grand merci également Sébastien pour tout le temps que tu as consacré à nous expliquer les mystères des statistiques et ton investissement dans le projet.

Un grand merci aux copains de la meilleure promo de M2 génétique, qui m'ont accueillie à bras ouverts alors que je venais du grand Nord. Pauline et Laure merci pour les moments que nous avons partagés. Emilie, merci pour ton humour et ta bonne humeur. Je vous souhaite plein de bonheur et de réussite pour la suite.

Je remercie toute l'équipe d'enseignement de TP de génétique. Grâce à vous j'ai enfin découvert à quoi ressemblait une levure ! J'ai appris énormément à vos côtés.

Un grand merci d'ailleurs à mes élèves collégiens et lycéens ainsi que leurs familles qui ont également rythmé ces dernières années et m'ont confortée dans l'idée que je voulais enseigner.

Merci à ma famille. J'ai hâte de partager ce travail avec vous le 17 mai. Un merci tout particulier à ma maman qui a consacré des heures à me relire, me corriger et me conseiller. Et enfin, merci à Maël d'être à mes côtés depuis de longues années. J'ai hâte de vivre avec toi cette nouvelle aventure anglaise et toutes les aventures qui suivront.

Table des matières

Remerciements *5*

Table des matières *7*

Abréviations *13*

Liste des figures *19*

Liste des tableaux *25*

Avant-propos *27*

Introduction *29*

Partie I : Leucémie myéloïde chronique et résistance à l'imatinib 29

1. La leucémie myéloïde chronique 29
   1.1. Généralités 29
   1.2. Epidémiologie 30
   1.3. Evolution de la maladie 30
   1.4. Réarrangements chromosomiques 32
      1.4.1. Le chromosome Philadelphie 32
      1.4.2. Le gène de fusion *BCR-ABL1* 33
      1.4.3. Etiologie 34
   1.5. Anomalies moléculaires 34
      1.5.1. Le transcrit de fusion *BCR-ABL1* 34
      1.5.2. La protéine hybride BCR-ABL1 35
         1.5.2.1. La protéine normale ABL1 36
         1.5.2.2. La protéine normale BCR 37
         1.5.2.3. Fusion des protéines BCR et ABL1 38
   1.6. Autres syndromes Ph^+ et/ou *BCR-ABL1*$^+$ 39
2. Histoire du traitement de la LMC 40
   2.1. Les premiers traitements 40
   2.2. La découverte des inhibiteurs de tyrosine kinase 41
   2.3. La première thérapie ciblée : l'imatinib 41
   2.4. Les inhibiteurs de deuxième génération 43
      2.4.1. Le nilotinib 43
      2.4.2. Le dasatinib et le bosutinib 43
      2.4.3. Utilisations en clinique 44
   2.5. La troisième génération : le ponatinib 44
3. Les niveaux de réponse aux inhibiteurs de tyrosine kinase 45

3.1. Scores prédictifs....45
3.2. Critères d'évaluation de la réponse....45
3.2.1. Réponse hématologique....45
3.2.2. Réponse cytogénétique....46
3.2.3. Réponse moléculaire....46
3.3. Résistance à l'imatinib....46
4. Mécanismes de résistance à l'imatinib....48
4.1. Généralités....48
4.2. Mécanismes dépendants de BCR-ABL1....48
4.2.1. Mutations du domaine kinase de *BCR-ABL1*....48
4.2.2. Amplification et surexpression de BCR-ABL1....49
4.3. Mécanismes indépendants de BCR-ABL1....50
4.3.1. Mauvaise observance thérapeutique....50
4.3.2. Pharmacocinétique de l'imatinib....50
4.3.2.1. Diminution de la biodisponibilité orale....50
4.3.2.2. Augmentation de la liaison aux protéines plasmatiques....51
4.3.2.3. Modification du transport dans la cellule leucémique....51
4.3.3. Persistance d'une maladie résiduelle....51
4.3.3.1. Cellule souche résistante....52
4.3.3.2. Microenvironnement....52
4.3.3.3. Evolution clonale....52
4.3.4. Altération de voies cellulaires dans la cellule leucémique....53
Partie II : Mécanismes d'apoptose et régulation par la famille BCL2....54
1. L'apoptose....54
1.1. Historique....54
1.2. Modifications morphologiques....55
1.3. Voies alternatives de dégradation de la cellule....56
1.3.1. La nécrose....56
1.3.2. L'autophagie....56
1.4. Les voies majeures de l'apoptose....56
1.4.1. Les caspases....57
1.4.2. La voie extrinsèque....58
1.4.3. La voie intrinsèque....59
1.4.3.1. Initiation du signal apoptotique....59
1.4.3.2. Régulation par la famille BCL2....59
1.4.3.3. Exécution de l'apoptose....61
1.4.4. BID, à l'interface entre voies extrinsèque et intrinsèque....62
2. L'apoptose dans les cellules leucémiques BCR-ABL1$^+$....62
2.1. Inhibition de l'apoptose par BCR-ABL1....62
2.2. Altération des protéines de la famille BCL2 et résistance à l'imatinib....64

Partie III : Identification de variants génétiques par séquençage 67
1. Evolution des technologies de séquençage de l'ADN 67
1.1. Généralités 67
1.2. La première génération 67
1.2.1. Maxam et Gilbert 67
1.2.2. Sanger 68
1.3. La deuxième génération : le séquençage nouvelle génération 69
1.3.1. Une approche haut débit 69
1.3.2. Le séquençage nouvelle génération par synthèse 70
1.3.2.1. Pyroséquençage et séquençage par ligation 70
1.3.2.2. Les terminateurs de séquences réversibles 70
1.4. La troisième génération 72
1.4.1. Généralités 72
1.4.2. Séquençage en temps réel de molécules uniques 72
1.4.2.1. Détection par fluorescence 72
1.4.2.2. Détection par variations de courant électrique 73
1.5. Evolution des coûts de séquençage 74
2. La variabilité génétique inter-individuelle 75
2.1. Généralités 75
2.2. Distribution des variants selon le principe d'Hardy-Weinberg 76
2.3. Identifications de variants génétiques à grande échelle 77
2.3.1. Généralités 77
2.3.2. 1000 Genomes 77
2.3.3. ExAC et gnomAD 78
2.3.4. dbSNP 78
2.3.5. Actualisation du génome humain de référence 78
3. NGS, variants génétiques et réponse à l'imatinib 79

***Objectifs* 81**

***Matériel* 82**
1. Caractéristiques des patients et collecte d'échantillons 82
1.1. Cohorte initiale : séquençage NGS 82
1.2. Cohortes de validation 82
1.2.1. Cohorte pour le séquençage Sanger 82
1.2.2. Cohorte pour l'expression des gènes par PCRq en temps réel 83
2. Lignées cellulaires 83

***Méthodes* 85**
1. Séquençage nouvelle génération 85
1.1. Extraction de l'ADN génomique 85
1.2. Contrôle qualité de l'ADNg par électrophorèse sur gel d'agarose 85

1.3. Conception de panels de gènes 86
1.4. Conception des sondes de captures 87
1.5. Génération de la librairie de séquençage 93
1.6. Réaction de séquençage 97
2. Analyses informatiques des données de NGS 98
2.1. Données brutes issues du séquençage 98
2.2. Contrôle qualité 99
2.3. Alignement des séquences sur un génome de référence 99
2.4. Appel des variants 99
2.5. Annotation des variants 100
2.6. Tableau de contingence 101
3. Analyses statistiques à partir de la matrice d'occurrence 101
3.1. Statistiques descriptives : AFC et HCPC 101
3.2. Simulation de données, méthode non-paramétrique : rank products 102
3.3. Programme informatique 103
4. Validation des polymorphismes candidats 103
4.1. Extraction de l'ADNg et des ARN totaux 103
4.1.1. Extraction à partir de culots leucocytaires de patients 103
4.1.2. Dosage des acides nucléiques 104
4.2. Séquençage Sanger 104
4.2.1. Amplification par PCR 104
4.2.2. Electrophorèse sur gel d'agarose 105
4.2.3. Réaction de séquençage 106
4.3. PCRq en temps réel 106
4.3.1. DNase 106
4.3.2. Transcription inverse 106
4.3.3. Réaction de PCRq et analyse des résultats 106
5. Culture cellulaire 108
5.1. Milieux de culture 108
5.2. Sous-clonage cellulaire 108
5.3. Congélation des cellules en azote liquide 108
6. CRISPR-Cas9 109
6.1. Principe de la technologie 109
6.2. Conception et clonage des guides 110
6.3. Transformation bactérienne et extraction plasmidique 112
6.4. Conception d'une matrice de recombinaison homologue 113
6.5. Transfection 114
6.6. Crible des cellules modifiées 115
6.6.1. Amplification PCR 115
6.6.2. Test de clivage 116

6.6.3. Test de restriction enzymatique 117
7. Surexpression de BCL-RAMBO 117
7.1. Amplification de la séquence codante de *BCL-RAMBO* 117
7.2. Purification et digestion des produits PCR 118
7.3. Purification et digestion du vecteur de surexpression 119
7.4. Clonage de la séquence codante dans le vecteur de surexpression 119
7.5. Transformation bactérienne et purification du plasmide 119
7.6. Production de lentivirus 120
7.7. Transduction 120
7.8. PCR 121
7.9. RT-PCRq en temps réel 121
8. Apoptose évaluée par cytométrie en flux 122
8.1. Traitement des cellules à l'imatinib 122
8.2. Principe de la cytométrie en flux 122
8.3. $DiOC_6(3)$ 122
8.4. Annexine V-APC/DAPI ou IP 123
8.5. Calcul de l'apoptose induite 123
9. Apoptose évaluée par western blot 123
9.1. Traitement des cellules à l'imatinib 123
9.2. Isolement des mitochondries 124
9.3. Lysats protéiques 124
9.4. Dosage à l'acide bicinchoninique 124
9.5. Electrophorèse sur gel de polyacrylamide 125
9.5.1. Transfert sur membrane de PVDF 126
9.5.2. Détection des protéines d'intérêt à l'aide d'anticorps 126
9.5.3. Révélation des membranes 127

Résultats* *128

Partie I : Identification de variants génétiques par une approche NGS 128
1. Séquençage nouvelle génération de panels de gènes 128
2. Contrôle qualité de la réaction de séquençage 128
3. Détection et annotation des variants génétiques 131
4. La population générale : un groupe de référence 134

Partie II : Mise en évidence de polymorphismes candidats par de nouvelles approches statistiques 137
1. Analyser des données volumineuses de NGS obtenues à partir d'un nombre réduit d'échantillons. 137
2. Visualisation des données sur un graphique en deux dimensions 137
3. Identification de gènes candidats par l'approche des rank products 139
4. Gènes et polymorphismes candidats 141
4.1. *BCL-RAMBO* 141

4.2. *BCLAF1* 143
4.3. *BCL2A1* 144

Partie III : Etude des gènes candidats dans une cohorte de validation 147
1. Séquençage dans une cohorte de validation 147
2. Expression des gènes *BCL-RAMBO* et *BCL2A1* selon le génotype des patients 149
3. Courbes de survie 151

Partie IV : BCL-RAMBO régulation de l'apoptose et sensibilité à l'imatinib 153
1. Création de lignées cellulaires modifiées par CRISPR-Cas9 153
1.1. Lignées n'exprimant plus BCL-RAMBO (knock-out, KO) 153
1.2. Lignées n'exprimant plus les domaines BHNo et TM 156
1.3. Lignées surexprimant la protéine BCL-RAMBO 158
1.4. Lignées exprimant la protéine variante 160
1.4.1. Recombinaison homologue 160
1.4.2. HIPER 162
2. Induction de l'apoptose par l'imatinib 164
2.1. Détectée par cytométrie en flux 164
2.1.1. Lignées KO *BCL-RAMBO* 164
2.1.2. Délétion des domaines BHNo et TM 166
2.1.1. Surexpression de la protéine BCL-RAMBO 167
2.1.1. Changement d'acide aminé causé par le variant rs9306198 168
2.2. Détectée par isolement des mitochondries et western blot 169

***Discussion* 174**
1. Identification de polymorphismes génétiques par séquençage nouvelle génération 174
2. Création de lignées cellules modifiées grâce à la technologie CRISPR-Cas9 177
3. La régulation de l'apoptose par BCL-RAMBO impliquée dans la réponse à l'imatinib 178

***Bibliographie* 181**

***Annexes* 199**

Partie I : Annexe 1 199

Partie II : Annexe 2 223

Partie III : Annexe 3 228

Partie IV : Annexe 4 240

Partie V : Annexe 5 242

Abréviations

1000G	Base de données 1000 Genomes
ABCB1	= *ATP binding cassette subfamily B member 1*
	(ancien nom : MDR1, *multidrug resistance protein 1*)
ABL1	*ABL proto-oncogene 1, non-receptor tyrosine kinase*
	(ancien nom : v-*abl Abelson murine leukemia viral oncogene homolog 1)*
ADN	Acide désoxyribonucléique
ADNc	ADN complémentaire
ADNg	ADN génomique
AFC	Analyse factorielle des correspondances
AGP	*Alpha-1-acid glycoprotein*
A-MuLV	*Abelson-murine leukemia virus* ou virus de la leucémie murine Abelson
ANNOVAR	*Annotate variation*
APC	Allophycocianine
APS	Persulfate d'ammonium
AraC	Cytarabine
ARN	Acide ribonucléique
ARNg	ARN guide
ARNm	ARN messager
ASH	*American society of hematology*
ATG	Codon méthionine ou codon start
ATP	Adénosine triphosphate
AVIA	*Annotation, visualization, and impact analysis*
BAD	*BCL2 associated agonist of cell death*
BAM	*Binary alignment map*
BAP1	*BRCA1 associated protein 1*
BAX	*BCL2 associated X, apoptosis regulator*
BCA	Acide bicinchoninique
BCL2	*BCL2 apoptosis regulator* (ancien nom : *B-cell CLL/lymphoma 2*)
BCL2A1	*BCL2 related protein A1*
BCL-RAMBO	= *BCL2L13, BCL2 like 13*

BCR	*BCR activator of RhoGEF and GTPase* (ancien nom : *breakpoint cluster region*)
BET	Bromure d'éthidium
BH	*BCL2 homology domain* ou domaine d'homologie à BCL2
BIM	= *BCL2L11, BCL2 like 11*
BSA	*Bovine serum albumine* ou albumine de sérum bovin
BWA	*Burrows-wheeler aligner*
$CaCl_2$	Chlorure de calcium
CalB	*C2 calcium-dependent lipid-binding*
Cas9	*CRISPR associated protein 9*
CC	*Coiled coil* ou domaine en super hélice
CCA	*Clonal chromosomal abnormality* ou anomalie chromosomale clonale
CGH	*Comparative genomic hybridization* ou hybridation génomique comparative
CRISPR	*Clustered regurarly interspaced short palindromic repeats*
DAPI	4',6-diamidino-2-phénylindole
ddNTP	Didésoxyribonucléotide
$DiOC_6(3)$	3,3'-Dihexyloxacarbocyanine iodide
DISC	*Death-inducing signaling complex*
DMSO	Diméthylsulfoxyde
dNTP	Désoxyribonucléotide
ECL	*Enchanced chemiluminescence* ou chimioluninescence augmentée
EDTA	Éthylènediaminetétraacétique
ELN	*European leukemia net*
EUTOS	*European treatment and outcome study*
ExAC	*Exome aggregation consortium* ou consortium d'agrégation des exomes
FADD	*Fas associated via death domain*
FASTQ	Fichier de séquence nucléique fusionné à un fichier qualité
FC	*Fold-change*
FISH	*Fluorescence in situ hybridization* ou hybridation *in situ* en fluorescence
GATK	*Genome analysis toolkit*
GDP	Guanosine diphosphate
GnomAD	*Genome aggregation database*

GRB2	*Growth factor receptor bound protein 2*
GRC	*Genome reference consortium*
GTP	Guanosine triphosphate
HA	*Human influenza hemagglutinin*
HBS	*HEPES buffered saline* ou tampon HEPES salin
HCK	*HCK proto-oncogene, Src family tyrosine kinase*
HCPC	*Hierarchical clustering on principal components* ou classification hiérarchique sur composantes principales
HEPES	acide 4-(2-hydroxyéthyl)-1-pipérazine éthane sulfonique
HGNC	*HUGO gene nomenclature committee*
IAP	*Inhibitor of apoptosis*
IFN-α	Interféron alpha
IGV	*Integrative genomics viewer*
INDEL	Insertion et délétion
IP	Iodure de propidium
ITK	Inhibiteur de tyrosine kinase
IRES	*Internal ribosome entry site*
Kb	Kilobase
KCl	Chlorure de potassium
kDa	Kilodalton
KIT	*KIT proto-oncogene, receptor tyrosine kinase*
LAL	Leucémie aiguë lymphoblastique
LAM	Leucémie aiguë myéloïde
LB	*Lysogeny broth* ou bouillon de culture lysogène
LLC	Leucémie lymphoïde chronique
LLM	Leucémie de lignée mixte
LMC	Leucémie myéloïde chronique
LYN	*LYN proto-oncogene, Src family tyrosine kinase*
M-BCR	Région majeure de *BCR*
m-BCR	Région mineure de *BCR*
µ-MCR	Région micro de *BCR*
MAF	*Minor allele frequency* ou fréquence de l'allèle mineur
MAPK	*Mitogen-activated protein kinase*

Mb	Mégabases
$MgCl_2$	Chlorure de magnésium
$MgSO_4$	Sulfate de magnésium
MOI	*Multiplicty of infection*
MOMP	*Mitochondrial outer membrane permeabilization*
MPT	*Mitochondrial permeability transition*
NaCl	Chlorure de sodium
NaOH	Hydroxyde de sodium ou soude
NCBI	*National center for biotechnology information*
NES	*Nuclear export signal* ou signal d'export du noyau
NF-κB	*Nuclear factor-kappa B*
NGS	*Next-generation sequencing* ou séquençage nouvelle génération
NHEJ	*Non-homologous end-joining* ou jonction d'extrémités non homologues
NHGRI	*National human genome research institute*
NLS	*Nuclear localization signal* ou signal de localisation nucléaire
OMS	Organisation mondiale de la santé
ORF	*Open reading frame* ou cadre ouvert de lecture
PAM	*Protospacer adjacent motif*
Pb	Paire de bases
PBS	*Phosphate buffered saline* ou tampon phosphate salin
PCR	*Polymerase chain reaction* ou amplification en chaîne par polymérase
PCRq	PCR quantitative
PDGFR	*Platelet derived growth factor receptor beta*
Pfp	Pourcentage de fausse prédiction
Ph	Chromosome Philadelphie
pH	Potentiel hydrogène
PMSF	Fluorure de phénylméthylsulfonyle
PVDF	Polyfluorure de vinylidène
Rac-GAP	*Rac-GTPase-activating protein*
RCH	Réponse complète hématologique
RCyC	Réponse cytogénétique complète
Rho-GEF	*Rho/guanine nucleotide exchange factor*

RM4,5	Réponse moléculaire avec plus de 4,5 logs de réduction (*BCR-ABL1* ≤ 0,032%)
RMM	Réponse moléculaire majeure
RP	*Rank products*
RPMI	Milieu du Roswell park memorial institute
SDS	Dodécylsulfate de sodium
SFH	Société française d'hématologie
SH	*SRC homology domain* ou domaine d'homologie à SRC
shRNA	*Short hairpin RNA* ou petit ARN en épingle à cheveux
SLC22A1	*Solute carrier family 22 member 1* (ancien nom : OCT1, *organic cation transporter 1*)
SMD	Syndrome myélodisplasique
SMP	Syndrome myéloprolifératif
SMRT	*Single-molecule real-time*
SNP	*Single nucleotide polymorphism* ou polymorphisme d'un seul nucléotide
SOC	*Super optimal broth* ou bouillon de culture super optimal
SRC	*SRC proto-oncogene, non-receptor tyrosine kinase*
STIM 2	Essai clinique « *Stopping imatinib 2* »
SVF	Sérum de veau fœtal
TBE 0,5X	Tris 44,5 mM, Acide borique 44,5 mM, EDTA 1 mM
TEMED	Tetraméthyléthylènediamine
Tm	Température de fusion
TNF	*Tumor necrosis factor*
WT	*Wild-type* ou sauvage

Liste des figures

Figure 1 : Hématopoïèse simplifiée 29
Figure 2 : Hémogrammes de patients LMC en phases A) chronique et B) blastique.....31
Figure 3 : Origine du chromosome anormal Philadelphie.....32
Figure 4 : Gènes *ABL1* et *BCR* 33
Figure 5 : Principaux transcrits *BCR-ABL1* et noms des protéines associées 35
Figure 6 : Structure des domaines de la protéine ABL1 sauvage 36
Figure 7 : Conformations de la tyrosine kinase ABL1 à l'état inactif et actif.....37
Figure 8 : Structure des domaines de la protéine BCR sauvage 38
Figure 9 : Structure des domaines de la protéine de fusion $p210^{BCR-ABL1}$ 38
Figure 10 : Voies de signalisation constitutivement activées par BCR-ABL1 et mécanismes cellulaires impactés 39
Figure 11 : Chronologie des traitements utilisés dans la LMC.....40
Figure 12 : Inhibition de l'activité de BCR-ABL1 par l'imatinib 41
Figure 13 : Survie des patients atteints de LMC au cours du temps et de l'évolution des traitements.....43
Figure 14 : Détection du chromosome Philadelphie par deux approches 46
Figure 15 : Schéma des critères de suivi de la réponse en fonction des taux A) de cellules Ph^+ et B) de transcrits *BCR-ABL1* 47
Figure 16 : Suivi du taux de transcrits *BCR-ABL1* au cours du temps après traitement par l'imatinib par PCRq.....48
Figure 17 : Phénomène d'apoptose dans la formation des doigts d'une souris au cours du développement embryonnaire.....54
Figure 18 : Modifications morphologiques de la cellule en apoptose.....55
Figure 19 : Voies extrinsèque et intrinsèque d'induction de l'apoptose 57
Figure 20 : Structure d'une pro-caspase.....57
Figure 21 : Mécanisme d'activation des caspases initiatrices et effectrices.....58
Figure 22 : Structure des protéines des trois-sous familles de la famille BCL2 (21 membres)60
Figure 23 : Interactions des protéines de la famille BCL2 dans une cellule vivante 60
Figure 24 : Interactions des protéines de la famille BCL2 dans une cellule en apoptose 61
Figure 25 : Schéma simplifié de l'inhibition de l'apoptose intrinsèque par la voie PI3K/AKT 63

Figure 26 : Schéma simplifié de la régulation de l'apoptose par les voies des MAPK (sous-familles ERK, JNK et p38) 64
Figure 27 : Schéma simplifié de la régulation de l'apoptose par la régulation de l'expression des gènes notamment *via* le facteur de transcription TP53 64
Figure 28 : Principe du séquençage de Maxam-Gilbert 68
Figure 29 : Principe du séquençage par la méthode Sanger 69
Figure 30 : Liaison des produits de la librairie à la lame de séquençage et amplification des fragments 71
Figure 31 : Lecture des séquences fixées à la lame par ajout de dNTP fluorescents et capture d'images 71
Figure 32 : Principe du séquençage PacBio 73
Figure 33 : Principe du séquençage par nanopore 74
Figure 34 : Evolution du coût de séquençage pour un génome humain de 2001 à 2017 75
Figure 35 : Estimation des fréquences génotypiques selon les fréquences alléliques d'après le principe d'Hardy-Weinberg 76
Figure 36 : Estimation du nombre d'individus porteurs de l'allèle variant dans une population donnée en appliquant le principe d'Hardy-Weinberg à partir de la MAF 77
Figure 37 : Image d'une électrophorèse sur gel d'agarose révélée sous ultraviolets 86
Figure 38 : Exemple de profil de dosage obtenu au BioAnalyzer, après la première PCR de la création des librairies, représentant l'intensité d'absorbance du signal en fonction de la taille des fragments 94
Figure 39 : Schéma de deuxième PCR permettant l'ajout des index et des étiquettes universelles P5 et P7 95
Figure 40 : Exemple de profil de dosage obtenu au BioAnalyzer, après la deuxième PCR de la création des librairies, représentant l'intensité d'absorbance du signal en fonction de la taille des fragments 96
Figure 41 : Schéma du séquençage « paired-end » 97
Figure 42 : Schéma global des différentes étapes d'analyses des données NGS 98
Figure 43 : Exemple de fichier .fastq obtenu à l'issu d'un séquençage par MiSeq (Illumina). 98
Figure 44 : Exemple de graphique en trois dimensions illustrant le tableau de contingence. 101
Figure 45 : Simulation de Monte Carlo et calcul des rapports des cotes (ODDS ratios) 103
Figure 46 : Principe du système CRISPR-Cas9 109
Figure 47 : Carte du plasmide pX462 modifié 110

Figure 48 : Description du clonage de la séquence variable de l'ARNg dans le vecteur pX462 *via* la digestion du vecteur par l'enzyme *Bbs*I 112
Figure 49 : Profil de digestion attendu du vecteur pX462 après insertion de la séquence variable de l'ARNg de 20 nucléotides et digestion par les enzymes *Xba*I et *Eco*RI..... 113
Figure 50 : Principe de la modification de l'exon 11 de *BCL-RAMBO* par CRISPR-Cas9 et la recombinaison homologue pour l'insertion de l'allèle variant (T) du polymorphisme rs9306198 114
Figure 51 : Schéma des amorces sens et anti-sens permettant l'amplification de la séquence codante de *BCL-RAMBO* et le clonage dans le vecteur de surexpression..... 118
Figure 52 : Carte du plasmide lentiviral de surexpression p412-EF1-IRES-Puro..... 119
Figure 53 : Profil de digestion attendu du vecteur p412 après insertion ou non de la séquence codante de *BCL-RAMBO* et digestion par l'enzyme *Xho*I 120
Figure 54 : Score qualité Phred des bases identifiées par NGS..... 129
Figure 55 : Pourcentage de reads de haute qualité pour chaque patient (un patient = un index) 129
Figure 56 : Fraction de bases séquencées dans les régions d'intérêt en fonction de la profondeur de lecture..... 130
Figure 57 : Exemple de séquençage du gène *BIK* visualisé sur le logiciel IGV..... 130
Figure 58 : Exemple de polymorphismes (délétion et substitution) détectés dans le gène *BCL2A1* visualisés sur le logiciel IGV..... 131
Figure 59 : Nature des 480 polymorphismes identifiés (SNP, délétions ou insertions)..... 132
Figure 60 : Localisation des 480 polymorphismes identifiés 132
Figure 61 : Exemple de séquençage du gène *GRK2* visualisé sur le logiciel IGV 133
Figure 62 : Ajustement des génotypes observés par NGS en fonction de la MAF décrite dans la base de données 1000G 135
Figure 63 : Simulation d'un groupe de référence à partir des fréquences alléliques référencées dans 1000G ou estimées 136
Figure 64 : Représentation des AFC et HCPC réalisées sur les données de NGS..... 138
Figure 65 : Schéma de la séquence codante de *BCL-RAMBO* (ADNc) et de la protéine associée 142
Figure 66 : Expression de l'ARNm de *BCL-RAMBO* dans diverses maladies hématologiques 142
Figure 67 : Schéma du gène *BCL-RAMBO* et des polymorphismes identifiés par NGS..... 143
Figure 68 : Schéma du gène *BCLAF1* et des polymorphismes identifiés par NGS..... 144

Figure 69 : Schéma du gène *BCL2A1* et des polymorphismes identifiés par NGS. 145
Figure 70 : Répartition des individus porteurs ou non de l'allèle variant du polymorphisme rs9306198 du gène *BCL-RAMBO* après séquençage de 103 patients atteints de LMC, 78 répondeurs et 25 résistants au traitement à l'imatinib 147
Figure 71 : Répartition des patients porteurs ou non de l'allèle variant du polymorphisme rs139012879 du gène *BCLAF1* chez les patients atteints de LMC répondeurs et résistants au traitement à l'imatinib 148
Figure 72 : Répartition des patients porteurs ou non de l'allèle variant du polymorphisme rs3826007 du gène *BCL2A1* chez les patients atteints de LMC répondeurs et résistants au traitement à l'imatinib 148
Figure 73 : Expression relative du gène *BCL-RAMBO* selon A) le statut de réponse des patients à l'imatinib, B) le génotype du polymorphisme rs9306198 150
Figure 74 : Expression relative du gène *BCL2A1* selon A) le statut de réponse des patients, B) le génotype du polymorphisme rs3826007 150
Figure 75 : Evolution du pourcentage de patients atteignant la RCH au cours du temps en fonction du statut du polymorphisme rs9306198 du gène *BCL-RAMBO* pour 101 patients atteints de LMC 151
Figure 76 : Evolution du pourcentage de patients atteignant la RCH au cours du temps en fonction du statut du polymorphisme rs3826007 du gène *BCL2A1* pour 101 patients atteints de LMC 152
Figure 77 : Résultats d'une éléctrophorèse sur gel d'agarose des produits PCR obtenus à partir de l'amplification de l'ADNg (exon 3 de *BCL-RAMBO*) des cellules modifiées par CRISPR-Cas9 155
Figure 78 : Caryotype hypotriploïde de la lignée cellulaire K-562 155
Figure 79 : Expression protéique de la protéine BCL-RAMBO et de l'ACTINE β évaluée par western blot dans trois lignées cellulaires K-562 156
Figure 80 : Résultats d'une éléctrophorèse sur gel d'agarose des produits PCR obtenus à partir de l'amplification de l'ADNg du clone cellulaire modifié par CRISPR-Cas9 157
Figure 81 : Expression protéique de la protéine BCL-RAMBO et de l'ACTINE β dans deux lignées cellulaires K-562 157
Figure 82 : Schéma de la protéine BCL-RAMBO sous les formes WT et tronquée 158
Figure 83 : Résultats d'une électrophorèse sur gel d'agarose des produits PCR obtenus à partir de l'amplification du vecteur de surexpression transduit dans les cellules K-562 159

Figure 84 : Expression relative du gène *BCL-RAMBO* normalisée sur l'expression du gène *GUS1* évaluée par PCRq dans trois lignées K-562 transduites avec le vecteur p412 : vide, portant la séquence codante de BCL-RAMBO de référence (Ref) ou variante (Var) (n=1).. 159
Figure 85 : Expression protéique de la protéine BCL-RAMBO et de l'ACTINE β dans trois lignées cellulaires K-562 transduites avec le vecteur p412 : vide, portant la séquence codante de BCL-RAMBO de référence (Ref) ou variante (Var) 160
Figure 86 : Séquences ciblées par les 4 ARNg à proximité du polymorphisme rs9306198 .. 160
Figure 87 : Résultats d'une éléctrophorèse sur gel d'agarose des produits PCR obtenus suite au test clivage réalisé à partir de l'ADNg de cellules WT ou ciblés par les ARNg 1, 2, 3 et 4.. 161
Figure 88 : Résultats d'une éléctrophorèse sur gel d'agarose des produits PCR, digérés par l'enzyme *Sma*I, obtenus à partir de l'ADNg de cellules WT ou transfectés avec quatre ARNg différents ainsi que la matrice de recombinaison de 999 pb..... 161
Figure 89 : Chromatogramme obtenu après le séquençage Sanger de la région modifiée de *BCL-RAMBO* d'un clone cellulaire 293T..... 162
Figure 90 : Schématisation de la stratégie HIPER (homology-independent precise and efficient replacement) permettant l'intégration de l'allèle variant du polymorphisme rs9306198 du gène *BCL-RAMBO* dans la lignée cellulaire K-562 163
Figure 91 : Modification de la séquence de l'exon 11 de *BCL-RAMBO* séquencée par NGS et visualisée sur le logiciel IGV..... 164
Figure 92 : Détection des cellules en apoptose par marquage $DiOC_6(3)$ en cytométrie en flux après 72h de traitement à 0,5 µM d'imatinib..... 165
Figure 93 : Pourcentage d'apoptose induite après 72h de traitement à 0,5 µM d'imatinib évalué par marquage $DiOC_6(3)$ pour la lignée WT et les deux lignées KO (n=4) 165
Figure 94 : A) Détection des cellules en apoptose par double marquage annexine V-APC/DAPI et cytométrie en flux après 72h de traitement à 0,5 µM d'imatinib, B) Pourcentage d'apoptose induite après 72h de traitement à 0,5 µM d'imatinib évalué par double-marquage annexine V-APC/DAPI pour la lignée WT et les deux lignées KO (n=3)..... 166
Figure 95 : Pourcentage d'apoptose induite après 72h de traitement à 0,5 µM d'imatinib évalué par marquage A) $DiOC_6(3)$ ou B) Annexine V-APC/DAPI pour la lignée WT et la lignée exprimant la protéine BCL-RAMBO tronquée (sans BHNo et TM) (n=3) 167
Figure 96 : Pourcentage d'apoptose induite après 72h de traitement à 0,5 µM d'imatinib évalué par marquage $DiOC_6(3)$ dans les lignées K-562 transduites avec le vecteur de surexpression vide (Vide), incluant la séquence codante de BCL-RAMBO de référence (Ref) ou variante (Var) vis-à-vis du polymorphisme rs9306198 (n=2)..... 168

Figure 97 : Pourcentage de cellules annexine V-APC$^+$ à 0,25 ; 0,5 ou 1 µM d'imatinib évalué par marquage $DiOC_6(3)$ pour la lignée WT et les deux lignées exprimant la protéine BCL-RAMBO sous la forme variante après 24h de traitement (n=1) 169
Figure 98 : Pourcentage d'apoptose induite après 48h de traitement à 0,25 ; 0,5 ou 1 µM d'imatinib évalué par marquage $DiOC_6(3)$ pour la lignée WT et les deux lignées exprimant la protéine BCL-RAMBO sous la forme variante (n=2) 169
Figure 99 : Représentation d'une cellule WT, en absence de traitement 170
Figure 100 : Représentation d'une cellule WT lorsque l'apoptose est induite notamment par le traitement à l'imatinib 171
Figure 101 : Expression des protéines caspase-3 et PARP1 (précurseur et protéine clivée) en fonction de l'expression de la protéine HSP60 dans les lignées K-562 WT, KO n°2 et ΔC-term après 24h de traitement à l'imatinib 0,5 µM (n=2) 171
Figure 102 : Expression des protéines cytochrome c et BIM dans la fraction mitochondriale et dans l'extrait total des lignées K-562 WT, KO n°2 et ΔC-term après 24h de traitement à l'imatinib 0,5 µM 172
Figure 103 : Expression de la protéine BCL-RAMBO dans la fraction mitochondriale et dans l'extrait total des lignées K-562 WT, KO n°2 et ΔC-term (ΔBHNo+TM) après 24h de traitement à l'imatinib 0,5 µM 172

Liste des tableaux

Tableau 1 : Description des trois phases de la LMC 30
Tableau 2 : Critères de réponse au traitement par imatinib définis par l'ELN en 2013 [87] ... 47
Tableau 3 : Liste des mutations du domaine kinase *d'ABL1* associées à une résistance aux différents ITK 49
Tableau 4 : Caractéristiques des 24 patients inclus dans la cohorte initiale pour le séquençage par NGS 82
Tableau 5 : Caractéristiques des 103 patients inclus dans la cohorte de validation pour le séquençage des polymorphismes candidats par approche Sanger 83
Tableau 6 : Caractéristiques des 78 patients inclus dans la cohorte de validation pour l'étude de l'expression des gènes candidats 83
Tableau 7 : Caractéristiques des 48 gènes inclus dans le panel test 88
Tableau 8 : Caractéristiques des 45 gènes associés à la régulation de l'apoptose 90
Tableau 9 : Caractéristiques des gènes *BCR* et *ABL1* 92
Tableau 10 : Mélange réactionnel et programme de la première PCR permettant l'amplification des fragments d'ADN avec adaptateurs 94
Tableau 11 : Mélange réactionnel et programme d'hybridation des sondes biotinylés aux fragments d'ADN 95
Tableau 12 : Mélange réactionnel et programme d'amplification de la deuxième PCR permettant l'ajout d'index aux fragments d'ADN 96
Tableau 13 : Liste non exhaustive des bases de données utilisées par l'outil AVIA 100
Tableau 14 : Exemple de tableau de contingence 101
Tableau 15 : Mélange réactionnel et programme pour l'amplification de fragments d'ADN à séquencer par Sanger 105
Tableau 16 : Description des amorces utilisées pour l'amplification de fragments des gènes *BCL-RAMBO*, *BCLAF1* et *BCL2A1*. 105
Tableau 17 : Description des amorces utilisées en PCRq pour étudier l'expression des gènes candidats dans une cohorte de validation 107
Tableau 18 : Caractéristiques des séquences cibles des ARNg et oligonucléotides commandés 111
Tableau 19 : Composition des milieux SOC et LB 112

Tableau 20 : Amorces utilisées pour l'amplification des fragments d'ADNg modifiés par CRISPR 116
Tableau 21 : Programme d'amplification PCR pour le crible des cellules modifiées par CRISPR 116
Tableau 22 : Mélange réactionnel et programme d'amplification PCR de la séquence codante de *BCL-RAMBO* 118
Tableau 23 : Séquences des amorces pour le séquençage du plasmide de surexpression du gène *BCL-RAMBO* 120
Tableau 24 : Mélange réactionnel et programme d'amplification PCR de la région de l'insert du vecteur p412 121
Tableau 25 : Composition du tampon de charge des lysats protéiques 125
Tableau 26 : Composition des gels de concentration et de séparation dilués dans l'eau 125
Tableau 27 : Composition des différents tampons de transfert 126
Tableau 28 : Description des anticorps primaires utilisés pour les expériences de western blot 126
Tableau 29 : Liste des cinq premiers gènes contenant le plus de variants altérant la protéine parmi les patients résistants selon la méthode RP 140
Tableau 30 : Caractéristiques du polymorphisme rs9306198 de *BCL-RAMBO* 143
Tableau 31 : Caractéristiques du polymorphisme rs139012879 de *BCLAF1* 144
Tableau 32 : Caractéristiques de six polymorphismes en déséquilibre de liaison dans le gène *BCL2A1* 145
Tableau 33 : Caractéristiques des trois polymorphismes mis en évidence dans la région promotrice du gène *BCL2A1* 149
Tableau 34 : Transcrits et isoformes associés codés par le gène *BCL-RAMBO* 154

Avant-propos

La notion de médecine de précision fait son apparition dans les années 2000 et a permis une évolution rapide des traitements anti-cancéreux. Le principe est simple : cibler les anomalies moléculaires jouant un rôle majeur dans le développement tumoral pour éliminer spécifiquement les cellules cancéreuses. Ces thérapies révolutionnaires ont amélioré grandement le pronostic des patients, cependant, ces inhibiteurs n'ont aucun effet chez certains malades. Cette hétérogénéité dans la réponse au traitement pourrait notamment être causée par la variabilité génétique existante entre les individus.

Le but de ce travail de thèse a été d'identifier, par une approche de séquençage nouvelle génération, des variants génétiques à l'origine de la résistance au traitement dans le cas de la leucémie myéloïde chronique (LMC).

La LMC est une maladie modèle particulièrement intéressante sous plusieurs aspects. D'une part, contrairement à d'autres cancers, une seule anomalie génétique est responsable du processus d'oncogenèse. Il s'agit d'une translocation réciproque équilibrée entre les chromosomes 9 et 22, à l'origine du gène de fusion *BCR-ABL1*. De plus, au cours du temps, très peu de mutations ou de remaniements chromosomiques s'accumulent dans les cellules leucémiques chez les patients. L'absence d'altérations génétiques aberrantes peut permettre d'identifier plus facilement les voies de signalisation impliquées dans le développement de la maladie ou dans la réponse au traitement.

D'autre part, il s'agit du premier modèle de maladie efficacement traitée par thérapies ciblées. Les inhibiteurs de tyrosine kinase, tels que l'imatinib, sont capables de cibler spécifiquement la protéine BCR-ABL1, ce qui entraîne l'élimination spécifique des cellules cancéreuses, sans affecter les cellules saines. Cependant, pour de nombreuses thérapeutiques ciblées, la résistance au traitement peut survenir : l'imatinib est inefficace pour environ 20% des patients atteints de LMC.

Les caractéristiques de la maladie, les critères d'évaluation de la réponse au traitement et les causes éventuelles de résistance au traitement sont présentés dans la première partie de l'introduction.

Une perturbation du processus de mort cellulaire programmée, l'apoptose, pourrait expliquer le manque d'efficacité de l'imatinib chez certains patients. En effet, dans de nombreux types de cancers, l'échec du traitement est associé à un défaut du programme apoptotique, essentiel à l'élimination des cellules tumorales. Des altérations de gènes

apoptotiques pourraient notamment expliquer ce phénomène. Un panel de 45 gènes candidats, a donc été établi et étudié au cours de la thèse. Ces gènes ont été séquencés afin de rechercher des variants génétiques susceptibles d'être associés à la résistance à l'imatinib dans la LMC.

Le principe de l'apoptose, les différentes voies moléculaires impliquées dans sa régulation et les gènes sélectionnés pour cette étude sont présentés dans la deuxième partie de l'introduction.

L'identification de variants génétiques est possible grâce aux différentes méthodes de séquençage développées depuis 1977. Pour ce travail de thèse, une approche de séquençage nouvelle génération a été choisie pour l'étude des gènes candidats. Grâce à cette technologie haut-débit, près d'une centaine de gènes ont pu être séquencés à partir d'échantillons de 24 patients atteints de LMC, répondeurs et résistants au traitement. Différents polymorphismes génétiques ont pu être mis en évidence en comparant les données de séquençage obtenues aux bases de données développées depuis les années 2000, rassemblant les connaissances sur la variabilité du génome humain. Les variants génétiques, identifiés chez certains patients atteints de LMC, pourraient être responsables d'une altération du processus apoptotique et donc responsables de la résistance à l'imatinib.

Un aperçu des différentes technologies de séquençage et des connaissances actuelles sur le génome humain est proposé dans la dernière partie de l'introduction.

Introduction

Partie I : Leucémie myéloïde chronique et résistance à l'imatinib

1. La leucémie myéloïde chronique

1.1. Généralités

La leucémie myéloïde chronique (LMC) est une hémopathie maligne rare, classée parmi les syndromes myéloprolifératifs (SMP). Ces pathologies sont caractérisées par une anomalie clonale de la cellule souche hématopoïétique à l'origine d'une production excessive de cellules de la lignée myéloïde (Figure 1).

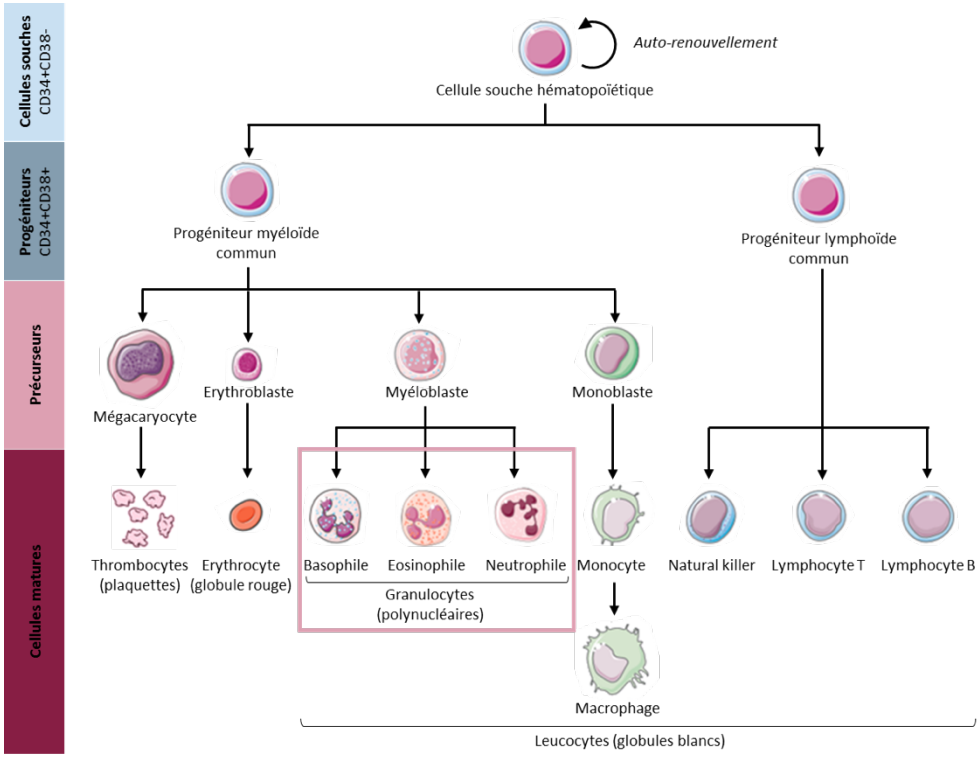


Figure 1 : Hématopoïèse simplifiée

Les cellules issues du progéniteur myéloïde commun sont altérées dans le cas de la LMC, en particulier les granulocytes, détectés en excès chez les patients.

Les cellules leucémiques présentent une prolifération élevée, une inhibition de la mort cellulaire programmée (apoptose) sans blocage de la maturation [1]. Bien qu'il s'agisse d'une altération de la cellule souche hématopoïétique, la nature des cellules affectées varie selon la cause de la maladie. Il existe différents SMP tels que la thrombocytémie essentielle, caractérisée

par un excès de thrombocytes (plaquettes sanguines), la polyglobulie de Vaquez, caractérisée par un excès d'érythrocytes (globules rouges), et la myélofibrose primitive, caractérisée par une fibrose de la moelle osseuse [1]. Dans le cas de la LMC, les cellules détectées en excès dans la moelle osseuse et le sang périphérique sont les granulocytes (basophiles, éosinophiles et neutrophiles). La LMC a été identifiée pour la première fois en 1845 par deux pathologistes, John Hughes Bennett [2] et Rudolf Virchow [3], mais ce n'est que près d'un siècle plus tard qu'un ensemble d'anomalies cellulaires, chromosomiques et moléculaires a été caractérisé.

1.2. Epidémiologie

La LMC peut survenir à tout âge mais touche principalement les adultes, à un âge médian de 65 ans. Environ 1 à 2 nouveaux cas/100.000 individus sont détectés chaque année. La LMC représente près de 15% des cas de leucémies nouvellement diagnostiquées chez l'adulte et la mortalité annuelle est actuellement de 1 à 2% [4,5]. Bien que cette maladie soit rare, la mortalité étant faible et l'espérance de vie globale en augmentation, le nombre d'individus atteints de LMC est en constante progression. Le nombre de cas recensés aux Etats-Unis était d'environ 30.000 en 2000, 100.000 en 2017 et pourrait atteindre 180.000 en 2030 [6].

1.3. Evolution de la maladie

Sans traitement, cette maladie évolue en trois phases (Tableau 1) [4].

Tableau 1 : Description des trois phases de la LMC

	Phase chronique	**Phase accélérée**	**Phase aiguë (crise blastique)**
Durée sans traitement	4 à 6 ans	12 à 18 mois	3 à 6 mois
Leucocytes dans le sang	$> 10.10^9$/L	En augmentation	Perte de la différenciation
Basophiles dans le sang	< 20%	≥ 20%	Perte de la différenciation
Blastes dans la moelle osseuse et le sang	≤ 10%	10 à 19%	≥ 20% Sortie de la moelle osseuse
Anomalies génétiques additionnelles	-	+	++

Les critères sont définis par l'organisation mondiale de la santé (OMS).

Dans un premier temps, la maladie est chronique pendant 4 à 6 ans en moyenne. La leucémie évolue lentement de manière graduelle. Pour 90 à 95% des patients, le diagnostic est établi durant cette phase [7]. Dans près de 50% des cas, la maladie est asymptomatique et par conséquent, le diagnostic est souvent fortuit, basé uniquement sur un hémogramme anormal après un test sanguin de routine. Un patient atteint de LMC en phase chronique présente notamment une forte proportion de leucocytes (hyperleucocytose) et l'apparition de cellules hématopoïétiques non différenciées (blastes) circulants dans le sang (Figure 2.A).

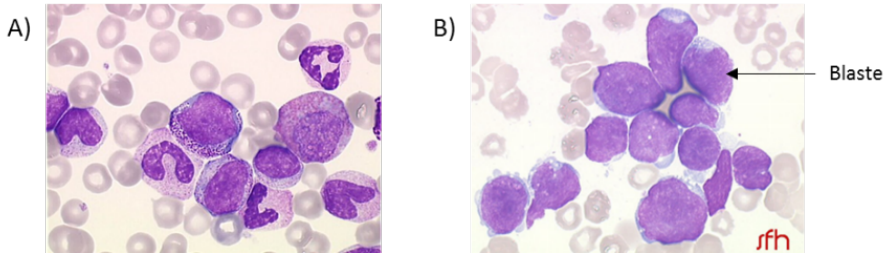


Figure 2 : Hémogrammes de patients LMC en phases A) chronique et B) blastique
Photographies de la Société française d'hématologie (SFH)

Une augmentation du volume de la rate (splénomégalie) est également observée dans la majorité des cas. Une diminution du taux d'hémoglobine (anémie) et/ou une augmentation du taux de plaquettes sont également détectées chez certains individus et peuvent causer un état de fatigue, une perte d'appétit, une perte de poids, une satiété rapide ou encore des douleurs abdominales.

La maladie peut ensuite progresser en phase accélérée pendant 12 à 18 mois. Cette accélération de la maladie conduit à une perte progressive de la capacité de différenciation terminale des cellules hématopoïétiques [8]. Cette phase se traduit notamment par une augmentation du nombre de blastes dans la moelle osseuse et le sang périphérique, et de nouveaux symptômes.

La dernière phase de la LMC est la phase d'acutisation (ou crise blastique). La maladie devient beaucoup plus agressive et des symptômes sévères sont décrits, similaires aux symptômes observés dans les cas de leucémies myéloïdes aiguës. En l'absence de traitement la maladie est mortelle en 3 à 6 mois. Dans 20% des cas, le patient progresse directement vers la dernière phase sans que la phase intermédiaire ne soit détectée. Durant cette période, les cellules perdent leur capacité à se différencier et à répondre aux facteurs de régulation, les blastes envahissent la moelle osseuse et s'infiltrent dans la circulation sanguine (Figure 2.B). La progression de la maladie en phase accélérée puis blastique est probablement due à

l'accumulation d'anomalies moléculaires qui conduisent graduellement à une perte de capacité à la différenciation terminale par les cellules leucémiques [8,9]. Ces anomalies surviendraient en raison d'une instabilité génomique croissante des cellules leucémiques [10]. Les altérations les plus fréquemment observées sont une trisomie du chromosome 8 ou encore des mutations et délétions de suppresseurs de tumeurs tels que *CDKN2A, TP53* ou *RB1* [11–14].

1.4. Réarrangements chromosomiques

1.4.1. Le chromosome Philadelphie

A la différence de la majorité des cancers, la LMC est caractérisée par la présence d'une anomalie génétique unique. Chez 95% des patients, un chromosome anormal est détecté dans les cellules hématopoïétiques : le chromosome Philadelphie. Ce chromosome tient son nom de la ville où il a été décrit pour la première fois en 1960, par Peter Nowell et David Hungerford [15]. Ces deux chercheurs ont observé un chromosome acrocentrique dans des cellules de patients atteints de LMC et émettent alors l'hypothèse que ce chromosome raccourci est issu d'une délétion chromosomique du chromosome 22. C'est la première fois qu'une anomalie chromosomique est associée à un cancer sans pour autant que son origine ou sa fonction ne soient connues. Par la suite, les techniques de marquages fluorescents des chromosomes ont évolué et permit à Janet Rowley, en 1973, de déterminer l'origine du chromosome Philadelphie [16]. Cette anomalie résulte d'une translocation réciproque équilibrée entre les bras longs des chromosomes 9 et 22 (t(9;22) (q34;q11.2)) (Figure 3). L'idée qu'une anomalie chromosomique puisse induire une maladie, être la cause et non pas la conséquence, commence alors à émerger.

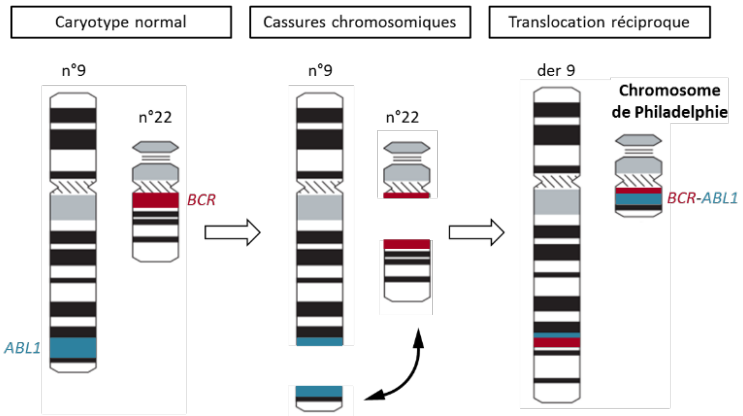


Figure 3 : Origine du chromosome anormal Philadelphie

Translocation réciproque équilibrée entre les chromosomes 9 et 22. Adapté de Apperley, Lancet, 2015 [17]

Le chromosome Philadelphie a été retrouvé dans différents types cellulaires des lignées myéloïde et lymphoïde : la LMC est donc définie comme une anomalie de la cellule souche hématopoïétique car l'ensemble des cellules sanguines peut être altéré [18].

1.4.2. Le gène de fusion *BCR-ABL1*

La cassure dans le chromosome 9 survient dans le gène aujourd'hui nommé *ABL1* (*ABL proto-oncogene 1, non-receptor tyrosine kinase*). L'histoire de sa découverte débute en 1970, lorsque Herbert Abelson et Louise Rabstein décrivent un nouveau rétrovirus : le virus de la leucémie murine Abelson (A-MulV) [19]. Ce virus est qualifié de transformant car il est capable d'induire un cancer chez la souris et ce grâce à un seul gène : le gène *v-Abl*. C'est à cette période que sont décrites les premières mutations ou altérations de gènes possédant des propriétés oncogéniques. L'homologue humain normal de *v-Abl*, *c-Abl* (nommé par la suite *ABL1*, après l'identification d'un second gène homologue, *ABL2* ou *ARG*), est retrouvé sur le bras long du chromosome 9 [20]. Différents points de clivage dans ce gène ont été décrits chez des patients atteints de LMC. Ces fractures dans l'ADN surviennent dans une région de plus de 300 kb en 5' du gène *ABL1*. Le gène contient 12 exons et les cassures sont détectées en amont de l'exon 1b (premier exon 1 alternatif), en aval de l'exon 1a (second exon 1 alternatif), ou le plus souvent entre les exons 1b et 1a (Figure 4) [21].

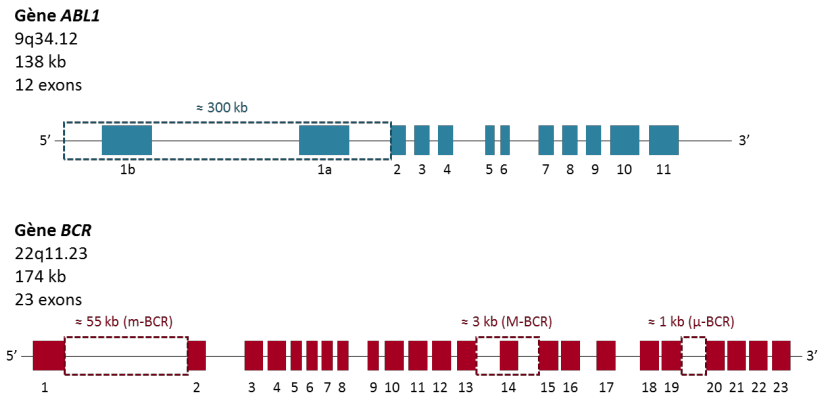


Figure 4 : Gènes *ABL1* et *BCR*

Les régions de cassures retrouvées dans la LMC sont indiquées en pointillés.

La cassure identifiée au niveau du chromosome 22 a initialement été détectée dans une région non décrite et relativement restreinte (5,8 kb) renommée *breakpoint cluster region* (*BCR*) [22,23]. Il apparait ensuite que *BCR* est un gène fonctionnel de 174 kb contenant 23 exons, cloné et cartographié par Heisterkamp *et al.* en 1988 (Figure 4) [24]. Chez les patients

atteints de LMC, il a tout d'abord été démontré que la cassure survenait soit dans l'intron 13 soit dans l'intron 14 (région d'environ 3 kb notée M-BCR, région majeure). Par la suite, deux autres sites de cassure ont été identifiés dans le gène *BCR*. Chez quelques rares patients atteints de LMC présentant un pourcentage de monocytes particulièrement élevé, des cassures dans l'intron 1 ont été détectées dans une région d'environ 55 kb (m-BCR, région mineure) [25]. Chez certains patients atteints d'un sous-type rare de LMC, appelé leucémie chronique à neutrophiles, des cassures sont présentes dans l'intron 19 dans une région d'environ 1000 pb (µ-BCR, région micro). La translocation réciproque entre ces deux chromosomes est donc à l'origine d'un nouveau gène de fusion, *BCR-ABL1*, porté par le chromosome 22 dérivé nommé chromosome Philadelphie (Figure 3).

1.4.3. Etiologie

L'origine de cette translocation réciproque reste à éclaircir. L'exposition aux radiations ionisantes a été identifiée à plusieurs reprises comme facteur de risque de survenue de la maladie *via* des études rétrospectives et des tests *in vitro* [26–28]. De plus, il est possible que cette translocation ne soit pas aléatoire et cible spécifiquement ces deux gènes en raison de leur proximité dans le noyau cellulaire. En effet, il a été montré que la distance physique entre les gènes *BCR* et *ABL1* dans les lymphocytes humains et des cellules positives pour le marqueur de surface CD34 (notées cellules $CD34^+$, cellules immatures de l'hématopoïèse) est plus courte que ce qui pouvait être attendu si leur répartition n'était due qu'au hasard [29,30].

1.5. Anomalies moléculaires

1.5.1. Le transcrit de fusion *BCR-ABL1*

Le gène hybride *BCR-ABL1* est exprimé dans les cellules leucémiques et, selon le point de cassure dans la région M-BCR du chromosome 22, l'ARN messager (ARNm) *BCR-ABL1* contient soit les 13 soit les 14 premiers exons de *BCR*. En revanche, quel que soit le point de cassure dans le chromosome 9, la partie *ABL1* de l'ARNm *BCR-ABL1* débute toujours par l'exon 2. L'exon 1 (1a ou 1b) d'*ABL1* est systématiquement épissé dans l'ARNm de *BCR-ABL1*. Les transcrits les plus fréquents sont notés b2a2 (ou e13a2, jonction de l'exon 13 de *BCR* et l'exon 2 d'*ABL1*) et b3a2 (ou e14a2, jonction de l'exon 14 de *BCR* et l'exon 2 *d'ABL1*). Ce sont deux ARNm d'environ 8,5 kb traduits en protéines d'environ 210 kDa

(p210$^{BCR-ABL1}$) (Figure 5). Les deux transcrits sont parfois exprimés chez un même patient mais le plus souvent une seule forme d'ARNm *BCR-ABL1* est détectée.

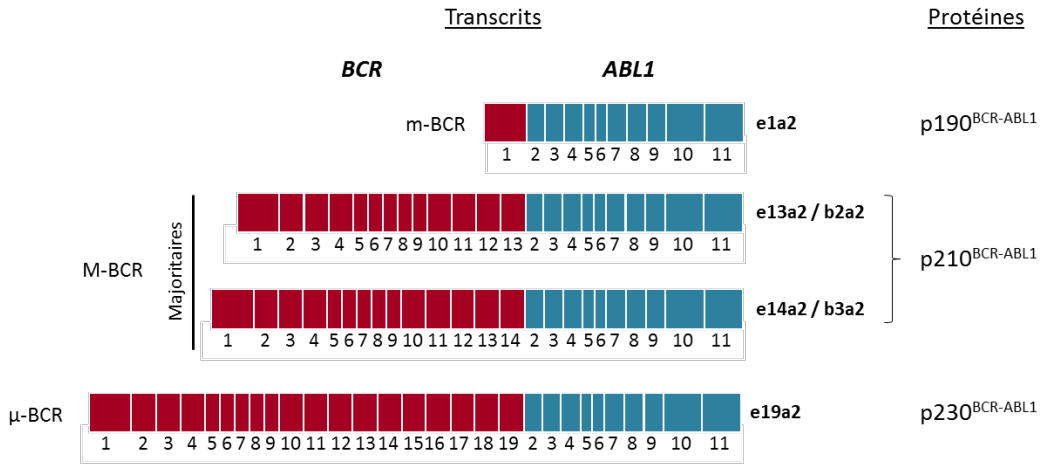


Figure 5 : Principaux transcrits *BCR-ABL1* et noms des protéines associées

Les transcrits e13a2/b2a2 et e14a2/b3a2, codant les protéines p210$^{BCR-ABL1}$, sont les plus fréquents dans la LMC.

Plus rarement, des fusions entre l'exon 1 de *BCR* (m-BCR) et l'exon 2 *d'ABL1* (e1a2, 7kb, 190 kDa, p190$^{BCR-ABL1}$) ou entre l'exon 19 de *BCR* (µ-BCR) et l'exon 2 *d'ABL1* (e19a2, 230 kDa, p230$^{BCR-ABL1}$) sont retrouvées chez les patients atteints de cas particuliers de LMC (Figure 5) [31]. La translocation étant réciproque, le chromosome dérivé 9 est porteur du gène *ABL1-BCR*. L'ARNm *ABL1-BCR* est exprimé chez environ 65% des patients atteints de LMC mais aucune protéine ABL1-BCR n'a été détectée [32]. De plus, les niveaux d'expression de ce gène ne semblent pas corrélés au pronostic de la maladie [33,34]. Par conséquent, le gène de fusion *ABL1-BCR* ne semble pas impliqué dans le processus d'oncogenèse. Pour confirmer le diagnostic de la LMC, le transcrit *BCR-ABL1* est détecté dans les cellules sanguines par une approche de RT-PCR (transcription inverse et amplification en chaîne par polymérase) en utilisant des amorces s'hybridant à l'ADN complémentaire (ADNc) de part et d'autre de la jonction.

1.5.2. La protéine hybride BCR-ABL1

La protéine BCR-ABL1 codée par le gène de fusion *BCR-ABL1* est à l'origine de l'activité tyrosine kinase dérégulée de ABL1 responsable de la leucémogenèse [35]. Avant de détailler la fonction de cette nouvelle protéine les propriétés des protéines normales BCR et ABL1 sont décrites ci-dessous.

1.5.2.1. La protéine normale ABL1

La protéine normale ABL1 est une protéine tyrosine kinase non récepteur qui navigue entre le noyau cellulaire et le cytoplasme. Il existe deux isoformes d'ABL1 (suite à l'épissage alternatif de l'exon 1), d'environ 145 kDa, exprimées de manière ubiquitaire dans l'organisme [36]. Cette kinase participe à la régulation de nombreux mécanismes essentiels dans la cellule tels que le cycle cellulaire, l'apoptose ou encore la différenciation [37–39]. Son activité est régulée en réponse à des facteurs de croissance, des protéines de l'adhésion cellulaire, aux dommages à l'ADN ou au stress oxydatif [40]. La protéine agit comme une plateforme de perception et de transmission de signaux intra- et extracellulaires [41]. La protéine ABL1 est divisée en plusieurs domaines (Figure 6).

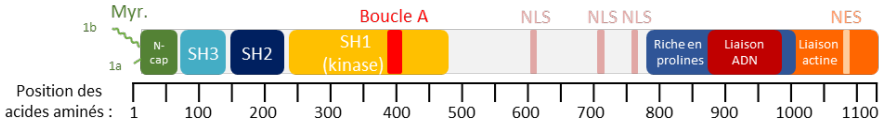


Figure 6 : Structure des domaines de la protéine ABL1 sauvage

L'extrémité N-terminale varie entre les deux isoformes. Une séquence de myristoylation (Myr) est codée par l'exon alternatif 1b. L'isoforme 1a est choisi comme séquence de référence (1130 acides aminés).

N-cap : coiffe N-terminale, SH : domaine d'homologie à SRC, Boucle A : boucle d'activation, NLS : signal de localisation nucléaire, NES : signal d'export nucléaire. UniProt : A0A024R8E2 et P00519

Dans la séquence protéique d'ABL1, trois domaines d'homologie à SRC (SH3, SH2 et SH1) sont identifiés. Les domaines SH2 et SH3 permettent les interactions d'ABL1 avec d'autres protéines. Le domaine SH1 quant à lui porte l'activité tyrosine kinase [42]. Au centre de la protéine, il existe 4 domaines riches en proline permettant la fixation d'autres protéines. Enfin, en C-terminal, la protéine ABL1 possèdent un domaine de liaison à l'ADN avec trois signaux de localisation nucléaire (NLS) et un domaine de liaison à l'actine avec un signal d'export nucléaire (NES) [37,43,44]. Ces derniers régulent la navette d'ABL1 entre le cytoplasme et le noyau. Le domaine kinase d'ABL1 présente une structure bilobée très conservée entre toutes les protéines ABL. L'adénosine triphosphate (ATP) se fixe dans le sillon formé entre les deux lobes du domaine (la poche de fixation de l'ATP) et forme des liaisons hydrogènes fortes avec plusieurs résidus de la kinase. Les kinases peuvent présenter deux formes majoritaires, active ou inactive, régulées selon leur conformation tridimensionnelle (Figure 7) [45].

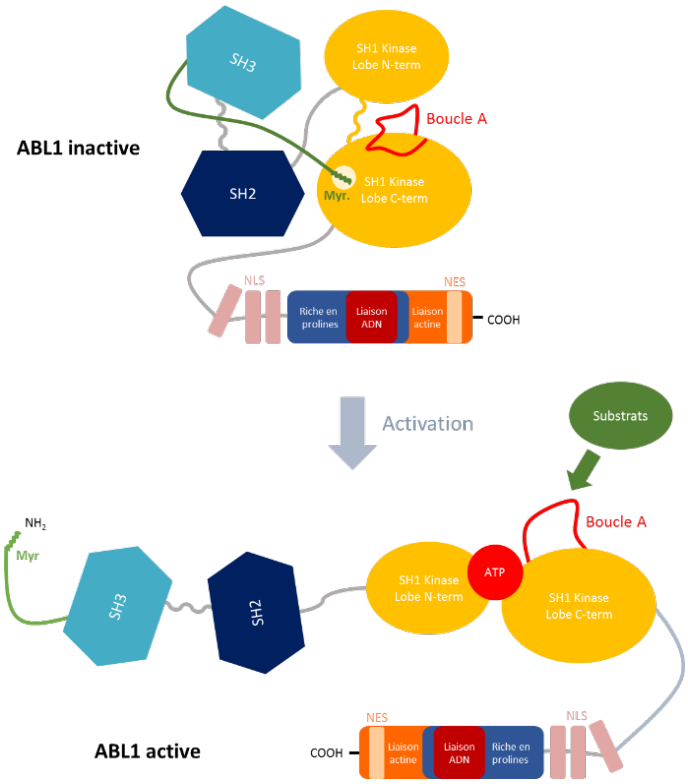


Figure 7 : Conformations de la tyrosine kinase ABL1 à l'état inactif et actif
Adapté de Woodring *et al.*, Journal of Cell Science, 2003 [46]

La boucle A au niveau du lobe C-terminal peut être dans une conformation étendue et agir comme plateforme de fixation du substrat ou elle peut être repliée et empêcher la liaison avec les protéines à phosphoryler. A la différence de la conformation inactive qui varie beaucoup d'une protéine à l'autre, la conformation active est très conservée entre toutes les kinases. De nombreux mécanismes régulent l'activation de la protéine, tels que la phosphorylation de certains résidus tyrosine essentiels ou l'interaction avec d'autres protéines activatrices. De plus, la protéine est maintenue dans un état auto-inhibé (isoforme 1b uniquement) grâce à une séquence de myristoylation présente en N-terminal de la protéine et liée au lobe C-terminal du domaine kinase pour maintenir la protéine repliée.

1.5.2.2. La protéine normale BCR

La protéine normale BCR de 1271 acides aminés (160 kDa) est exprimée de façon ubiquitaire dans le cytoplasme des cellules [36]. Cette protéine porte également une activité kinase mais à l'heure actuelle un seul substrat est connu (BRCA1 associated protein 1, BAP1) [47] et possiblement BCR lui-même [36]. La protéine BCR est divisée en plusieurs domaines (Figure 8).

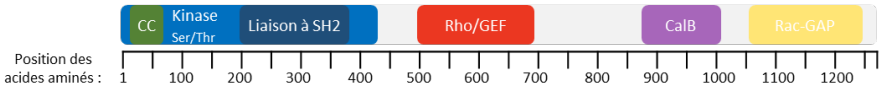


Figure 8 : Structure des domaines de la protéine BCR sauvage

CC = coiled coil: super-hélice, Rho/GEF = Rho/guanine nucleotide exchange factor, CalB = C2 calcium-dependent lipid-binding, Rac-GAP = Rac-GTPase-activating protein. UniProt : P11274

En N-terminal, il existe un domaine en super-hélice permettant la formation de dimères [48], suivi d'un domaine de liaison aux protéines contenant un domaine kinase sérine/thréonine et un domaine SH2. Au centre, deux régions stimulent le passage GTP à GDP de facteurs d'échange Rho guanidine (Rho/guanine nucleotide exchange factor, Rho/GEF) [49]. En C-terminal, BCR possède un domaine de fixation du calcium (C2 calcium-dependent lipid-binding, CalB) et un domaine à activité GTPase qui catalyse l'hydrolyse de GTP de protéines liées à Ras (Rac-GTPase-activating protein, Rac-GAP) [50].

1.5.2.3. Fusion des protéines BCR et ABL1

La protéine $p210^{BCR\text{-}ABL1}$, codée par les transcrits majoritaires b2a2 et b3a2, possède le domaine N-terminal de BCR et C-terminal d'ABL1 (Figure 9).

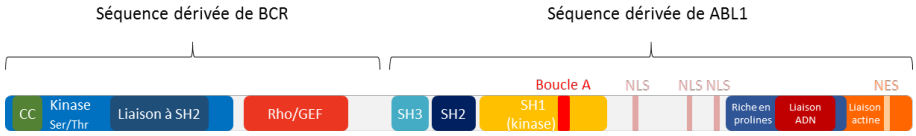


Figure 9 : Structure des domaines de la protéine de fusion $p210^{BCR\text{-}ABL1}$

Le domaine en super-hélice, le domaine de liaison à SH2, le domaine kinase et le domaine Rho/GEF de BCR sont présents. Tous les domaines d'ABL1 sont conservés à l'exception du domaine de myristoylation (codé par l'exon 1) : l'auto-inhibition par ABL1 est abolie. A la différence de la protéine ABL1 à l'état physiologique, la protéine hybride BCR-ABL1 est retenue dans le cytoplasme. La fusion de BCR avec ABL1 favorise la dimérisation grâce au domaine en super-hélice de BCR et ainsi l'autophosphorylation de nombreux résidus tyrosine de BCR-ABL1 [48]. Cet évènement génère des sites de fixation pour les domaines SH2 d'autres protéines et permet ainsi leur activation constitutive *via* l'activité tyrosine kinase de BCR-ABL1 [51,52]. Cette conformation entraîne notamment la phosphorylation du résidu tyrosine 177 de BCR. Ce dernier permet l'interaction avec la protéine adaptatrice GRB2, ce qui conduit à la sur-activation de plusieurs voies de signalisation

essentielles à la leucémogenèse [52–54]. Au final, l'activité catalytique de BCR-ABL1 est augmentée par son autophosphorylation et entraîne l'activation constitutive de voies kinases telles que RAS/MAPK, JAK/STAT ou PI3K/AKT par exemple (Figure 10).

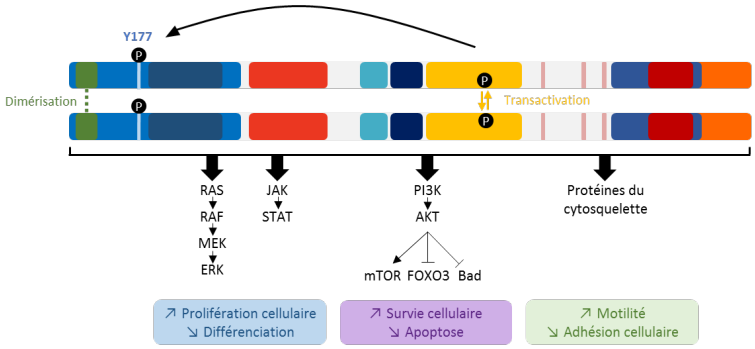


Figure 10 : Voies de signalisation constitutivement activées par BCR-ABL1 et mécanismes cellulaires impactés

Y : résidu tyrosine, P : groupement phosphate. Adapté de Frazer *et al.*, Ulster Medical Journal, 2007 [55]

Les cellules leucémiques présentent un phénotype de croissance anarchique et de résistance à l'apoptose. De plus, l'adhérence au microenvironnement médullaire est altérée, ce qui favorise la libération de cellules immatures (blastes) dans la circulation sanguine [56]. La régulation des voies cellulaires par l'environnement extracellulaire est perdue en présence de BCR-ABL1. La cellule leucémique devient indépendante aux facteurs de croissance extracellulaires tels que les cytokines : la cellule est dite « addict » à BCR-ABL1 [57].

1.6. Autres syndromes Ph$^+$ et/ou *BCR-ABL1*$^+$

Une minorité de patients pour lesquels le chromosome Philadelphie n'est pas détecté présente également un phénotype de LMC. Dans certains cas, de nouvelles translocations surviennent entre le chromosome 22 raccourci et d'autres chromosomes. Ces évènements peuvent « masquer » le chromosome Philadelphie : le chromosome 22 perd sa forme raccourcie caractéristique mais le transcrit *BCR-ABL1* est toujours détecté (LMC Ph$^-$ *BCR-ABL1*$^+$). Les patients porteurs de ces anomalies présentent des symptômes et une évolution de la maladie similaires aux autres patients atteints de LMC Ph$^+$ *BCR-ABL1*$^+$ [58]. Le chromosome Philadelphie est également retrouvé dans d'autres pathologies telles que la leucémie aiguë lymphoblastique (LAL- Ph$^+$) décrite la première fois en 1976 [59].

2. Histoire du traitement de la LMC

2.1. Les premiers traitements

Avant les années 2000, les solutions thérapeutiques proposées aux patients atteints de LMC étaient non spécifiques, toxiques et peu efficaces (Figure 11).

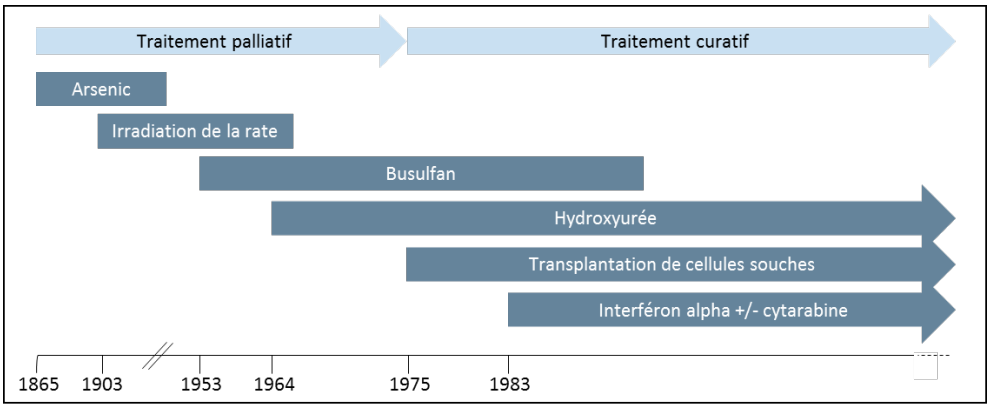


Figure 11 : Chronologie des traitements utilisés dans la LMC
Adapté de Hehlmann *et al.*, Lancet, 2007 [60]

A la fin des années 1800, les patients étaient traités par la solution Fowler composée d'arsénite de potassium, un dérivé de l'arsenic. Au début des années 1900, ce traitement a été remplacé par des irradiations de la rate puis par une succession de traitements de chimiothérapie. La prise de ces médicaments, comme le busulfan ou l'hydroxyurée, entraînait de nombreux effets secondaires et restait un traitement palliatif. Plus tard, l'interféron alpha (IFN-α) est utilisé seul ou en combinaison avec la cytarabine (AraC). Avec ces traitements, la maladie régresse et la survie est augmentée mais l'efficacité reste modeste et les effets secondaires sont sévères [61,62].

La seule thérapie réellement curative était la transplantation allogénique de cellules souches hématopoïétiques. Cette méthode nécessite un donneur compatible et entraîne une mortalité élevée. Le risque de cette thérapie augmente d'autant plus que la maladie est à un stade avancé [63]. Cette approche est toujours proposée aux patients atteints de LMC en phase chronique ayant échoué à au moins deux inhibiteurs de tyrosine kinase (voir ci-dessous) ou aux patients en phase accélérée ou blastique. De plus, à l'heure actuelle, il s'agit du seul traitement curatif de la LMC de l'enfant et de l'adolescent [64].

2.2. La découverte des inhibiteurs de tyrosine kinase

L'activité constitutive de la protéine BCR-ABL1 est l'évènement oncogénique critique dans le développement de la LMC. Cette constatation a fait naître une nouvelle idée de thérapie : inhiber spécifiquement BCR-ABL1 pour éliminer les cellules leucémiques. La société Ciba-Geigy (aujourd'hui Novartis) a donc initié un programme de recherche et développement d'inhibiteurs de kinases. Ils ont généré de grandes quantités de kinases afin de tester de larges banques d'inhibiteurs. Le taux de phosphorylation par BCR-ABL1 a été évalué en présence de chaque inhibiteur. Cette méthode leur a permis d'identifier en 1995 les composés de la classe des 2-phenylaminopyrimidine, inhibiteurs spécifiques des tyrosine kinases [65].

2.3. La première thérapie ciblée : l'imatinib

En 1996, l'inhibiteur CGP 57148B (ensuite nommé STI571 puis imatinib) a été caractérisé [66]. Il s'agit d'un inhibiteur compétitif de l'ATP qui cible spécifiquement la poche de fixation de l'ATP de la protéine BCR-ABL1 et bloque ainsi son autophosphorylation donc son activation (Figure 12).

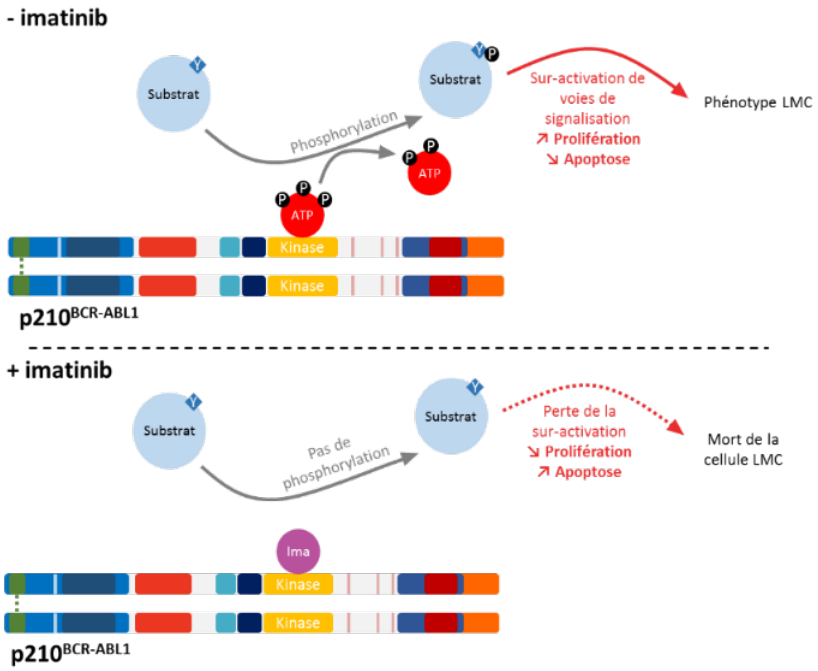


Figure 12 : Inhibition de l'activité de BCR-ABL1 par l'imatinib

Y : résidu tyrosine, P : groupement phosphate, Ima : imatinib

Adapté de Frazer *et al.*, Ulster Medical Journal, 2007 [55]

Cette molécule ne se fixe qu'à la forme inactive de BCR-ABL1. Les cellules étant devenues dépendantes à BCR-ABL1, cette inhibition entraîne la mort de la cellule leucémique *via* l'induction de l'apoptose. Par la suite, il a été déterminé que cet inhibiteur était hautement spécifique de BCR-ABL1 et ABL1 sauvage (inhibition minimale d'ABL2, PDGFR et KIT) et devient la molécule de choix pour traiter la LMC [67]. La grande efficacité de l'imatinib en comparaison avec les traitements précédents a permis une évolution rapide de la découverte du traitement à l'autorisation de mise sur le marché. Les premiers essais cliniques de phase I ont débuté en juin 1998 pour des patients atteints de LMC en phase chronique, réfractaires ou intolérants aux IFN-α, puis ont été étendus à des patients en phase blastique ou atteints de LAL Ph^+ [68]. L'imatinib était bien toléré par les patients qui ne présentaient que des effets secondaires courants tels que des nausées, œdèmes périorbitaires ou des crampes musculaires par exemple. Cet inhibiteur de tyrosine kinase (ITK) a démontré un bénéfice thérapeutique dès une prise quotidienne de 300 mg. Les résultats de cette phase I ont été présentés lors d'un congrès international organisé par la société américaine d'hématologie (American society of hematology, ASH) en décembre 1999. La phase II a montré des résultats similaires. Avec un suivi à long terme, la survie globale à 6 ans était de 76% [69]. Suite à ces résultats très favorables, l'imatinib (commercialisé Glivec® ou Gleevec®) a été approuvé par la « Food and Drug Administration » (Etats-Unis) en mai 2001, pour les patients réfractaires ou résistants aux IFN-α, soit à peine 3 ans après le début de la phase I des essais cliniques. En 2003, l'imatinib est autorisé pour le traitement de patients nouvellement diagnostiqués.

Le Glivec®/imatinib est une thérapie orale à la fois efficace et très bien tolérée par les patients atteints de LMC. Cette approche efficace et peu toxique a permis d'augmenter le taux de survie à 10 ans de moins de 10% à près de 90% (Figure 13) [66,70]. L'imatinib est ainsi devenu le traitement de première ligne majeur pour les patients nouvellement diagnostiqués. Les patients atteints de LMC traités par imatinib présentent aujourd'hui une espérance de vie comparable à la population générale [71].

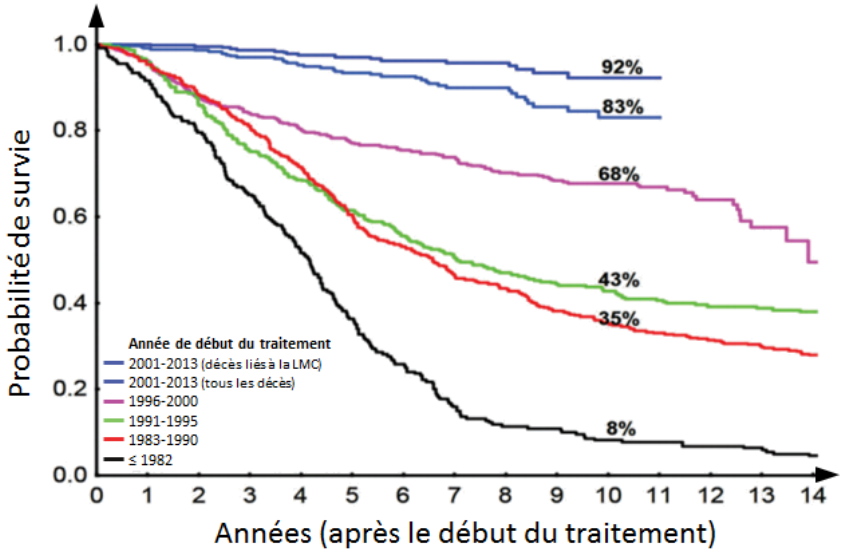


Figure 13 : Survie des patients atteints de LMC au cours du temps et de l'évolution des traitements
2035 patients ont été inclus dans cette étude entre 1993 et 2013.
Adapté de Mughal *et al.*, Haematologica, 2016 [72].

Malgré son efficacité, certains patients ne sont pas sensibles au traitement à l'imatinib (cf 3.3), c'est pourquoi, de nouveaux inhibiteurs de tyrosine kinase ont été développés.

2.4. Les inhibiteurs de deuxième génération

2.4.1. Le nilotinib

Dans un premier temps, l'AMN107 (nilotinib) a été développé par Novartis [73]. La structure de l'imatinib a été modifiée pour augmenter la force de la liaison avec ABL1. Le nilotinib est 30 à 50 fois plus « puissant » (*drug potency*) que l'imatinib (*in vitro*) et permet d'inhiber la plupart des protéines BCR-ABL1 mutées. Etant un dérivé de l'imatinib, le nilotinib possède le même profil d'inhibition en ciblant la conformation inactive et en affectant également KIT ou PDGFR.

2.4.2. Le dasatinib et le bosutinib

De nouveaux ITK ont également été développés en se basant sur le fait qu'ABL1 et la kinase SRC sont comparables. Leurs conformations actives sont proches mais leurs conformations inactives très différentes. Un inhibiteur de SRC peut donc être efficace pour inhiber la conformation active d'ABL1. Le composé BMS-354825 (dasatinib) développé par Bristol-Myers Squibb est un de ces inhibiteurs décrits initialement comme inhibiteur de SRC

mais également efficace contre ABL1 et BCR-ABL1 [74]. De manière intéressante, il est 350 fois plus puissant que l'imatinib (*in vitro*) et agit sur toutes les formes de BCR-ABL1 exceptées celles portant la mutation T315I [75]. Le SKI-606 (bosutinib) a été développé par Pfizer (anciennement Wyeth) en suivant la même approche [76].

2.4.3. Utilisations en clinique

Le dasatinib et le nilotinib ont rapidement montré des résultats concluants en recherche et ont été intégrés en clinique pour le traitement des patients résistants ou intolérants à l'imatinib. Ils sont aujourd'hui également proposés en traitement de première ligne. Le bosutinib est actuellement proposé pour les patients résistants aux autres ITK de première ligne. Malgré leur efficacité supérieure à l'imatinib, les effets secondaires de ces ITK de deuxième génération sont plus importants que pour l'imatinib. Il est intéressant de noter que les réponses atteintes avec ces ITK de deuxième génération sont plus profondes et plus rapides mais à long terme, aucune différence de survie globale n'est identifiée entre les patients traités avec des ITK de première ou de deuxième génération [77].

Une étude récente menée par le MD Anderson Cancer Center compare l'efficacité et les effets à long terme de l'imatinib, le nilotinib et le dasatinib [78]. 482 patients atteints de LMC en phase chronique ont été inclus de 2000 à 2013. L'étude a été conçue avec cinq bras différents (les doses indiquées sont quotidiennes) : imatinib 400 mg, imatinib 800 mg, dasatinib 2x50 mg, dasatinib 100 mg, nilotinib 2x400 mg. Les premiers résultats suggèrent une meilleure réponse moléculaire à n'importe quel temps et une survie à 5 ans sans évènement augmentée pour tous les bras comparés à l'imatinib 400 mg.

2.5. La troisième génération : le ponatinib

Un ITK de troisième génération est aujourd'hui disponible : l'AP24534 (ponatinib). Ce composé est le premier à pouvoir inhiber la protéine BCR-ABL1 avec la mutation T315I (Tableau 3). Il inhibe également les protéines VEGFR, FGFR, PGFR, LYN et FLT3 [79]. Au final, ces nouvelles thérapies sont efficaces et sélectives pour BCR-ABL1, cependant, leurs profils pharmacologiques et les taux de réponse sont relatifs à chaque patient traité.

3. Les niveaux de réponse aux inhibiteurs de tyrosine kinase

3.1. Scores prédictifs

Malgré le succès des ITK dans le traitement de la LMC, la sensibilité au traitement varie d'un individu à l'autre. Plusieurs scores ont été développés pour évaluer le pronostic des patients et aider les cliniciens au choix de traitement le plus adapté à chaque patient.

Le score de Sokal, décrit en 1984, est le plus utilisé pour catégoriser les patients atteints de LMC et prédire la survie [80]. Il a été initialement imaginé pour les patients traités au Busulfan mais il est toujours d'actualité pour les patients traités par ITK. Les patients sont répartis en trois groupes (risque faible, intermédiaire ou élevé) selon l'âge du patient, la taille de la rate, le taux de plaquettes sanguines et le pourcentage de blastes dans le sang. Le score de Hasford (ou Euro), publié en 1998, s'appuie sur les mêmes critères pour classer notamment les patients traités aux IFN-α [81]. En 2011, le score EUTOS (European treatment and outcome study) est proposé pour classer les patients traités aux ITK [82]. Ce score prend en compte la taille de la rate ainsi que le pourcentage de basophiles dans le sang avant traitement. Cette prédiction doit encore être testée à une plus grande échelle pour être utilisée couramment en clinique [83].

3.2. Critères d'évaluation de la réponse

La réponse au traitement peut également être suivie au cours du temps à l'aide de critères d'évaluation définis par un réseau de recherche financé par l'Union Européenne, l'European leukemia net (ELN) [63]. Ces indications s'appuient notamment sur l'évolution du pourcentage de cellules leucémiques chez les patients atteints de LMC.

3.2.1. Réponse hématologique

La réponse hématologique est définie d'après les résultats de l'hémogramme qui comporte une analyse quantitative et qualitative des cellules sanguines. Les valeurs de référence sont définies selon l'âge, le sexe et l'origine ethnique du patient. Une réponse complète hématologique (RCH) est caractérisée par un hémogramme normal qui se caractérise notamment par l'absence de blastes. De plus, si la rate avait grossi, elle reprend une taille normale.

3.2.2. Réponse cytogénétique

La réponse cytogénétique est définie en fonction du nombre de cellules porteuses du chromosome Philadelphie (cellules Ph^+). Une réponse cytogénétique complète (RCyC) est établie lorsqu'aucune cellule Ph^+ n'est détectée sur le caryotype des patients ou par hybridation *in situ* en fluorescence (FISH) (Figure 14) [84].

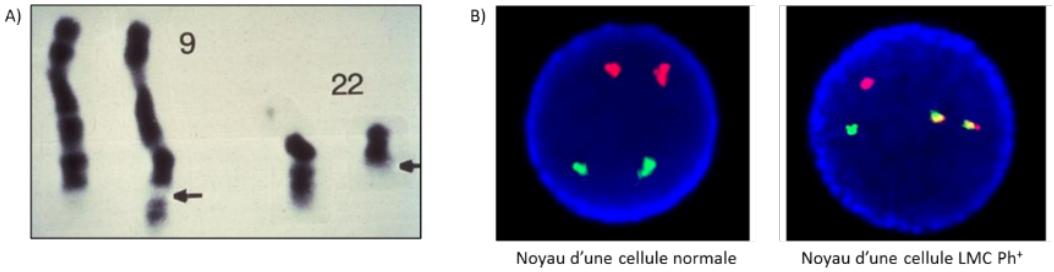


Figure 14 : Détection du chromosome Philadelphie par deux approches

A) Cytogénétique conventionnelle, B) FISH. Sonde *ABL1* en rouge, sonde *BCR* en vert : le gène *BCR-ABL1* apparaît en jaune.

3.2.3. Réponse moléculaire

La réponse moléculaire est définie en fonction du nombre de cellules exprimant le gène *BCR-ABL1*. Le taux de transcrits *BCR-ABL1* est mesuré par PCR quantitative (PCRq) en temps réel à partir d'échantillons sanguins en utilisant les mêmes amorces que pour la détection de transcrits au diagnostic. Une réponse moléculaire majeure (RMM) est définie par un niveau de transcrits *BCR-ABL1/ABL1*, détectés par PCRq, inférieur ou égal à 0,1% (en échelle internationale, SI). Un patient ayant atteint la RMM est considéré en rémission.

3.3. Résistance à l'imatinib

Bien que l'imatinib soit efficace, seulement 60% des patients sont maintenus avec une dose standard (400 mg par jour) après 6 ans de traitement [85]. Le traitement est stoppé ou la dose est modulée car ces patients sont intolérants ou résistants à la molécule. La réponse au traitement est suivie au cours du temps, à 3, 6 et 12 mois après le début du traitement, selon les critères définis par l'ELN. Les patients atteints de LMC traités par ITK en première ligne sont classés en trois catégories distinctes selon leurs niveaux de réponse hématologique, cytogénétique et moléculaire : optimale, en alerte ou en échec (Tableau 2 et Figure 15). Les premières indications ont été définies en 2009 [60] puis actualisées en 2013 [87].

Tableau 2 : Critères de réponse au traitement par imatinib définis par l'ELN en 2013 [87]

	Optimale	En alerte	En échec
Au diagnostic	NA	Haut risque ou anomalies chromosomiques clonales fréquentes dans des cellules Ph+ (CCA/Ph+)	NA
3 mois	BCR-ABL1 ≤ 10% et/ou Ph+ ≤ 35%	BCR-ABL1 > 10% et/ou 36% < Ph+ < 95%	Pas de RCH et/ou Ph+ > 95%
6 mois	BCR-ABL1 < 1% et/ou Ph+ = 0 (RCyC)	1% < BCR-ABL1 < 10% et/ou 1% < Ph+ < 35%	BCR-ABL1 > 10% et/ou Ph+ > 35%
12 mois	BCR-ABL1 ≤ 0,1% (RMM)	0,1% < BCR-ABL1 < 1%	BCR-ABL1 > 1% et/ou Ph+ > 0
> 12 mois	BCR-ABL1 ≤ 0,1% (RMM)	Anomalies chromosomiques clonales dans des cellules Ph- (CCA/Ph-, -7 ou 7q-)	Perte de la RCH Perte de la RCyC Perte confirmée de la RMM Mutations

Tous les pourcentages sont définis selon l'échelle internationale (SI).

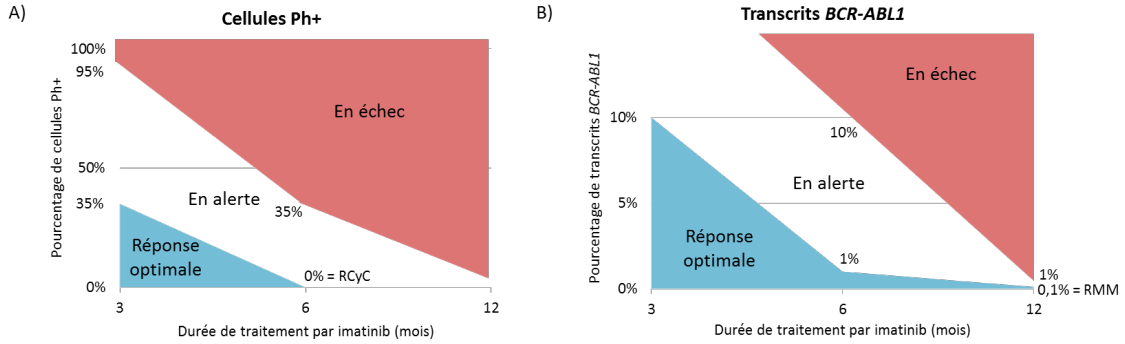


Figure 15 : Schéma des critères de suivi de la réponse en fonction des taux A) de cellules Ph$^+$ et B) de transcrits *BCR-ABL1*

Selon les recommandations de l'ELN, une fois le patient classé « en échec » il est maintenu dans cette catégorie peu importe l'évolution des taux de transcrits *BCR-ABL1* et du nombre de cellules Ph$^+$. En revanche un patient en réponse « optimale » peut être classé « en alerte » ou « en échec » dès que l'évolution de la réponse correspond à cette catégorie. Les patients en échec de réponse sont dits résistants au traitement par imatinib. Il existe deux grands types de résistance. On parle de résistance primaire (« refractoriness ») lorsque les cellules du patient ne sont pas sensibles au traitement et qu'il échoue à atteindre les critères de réponse dans un laps de temps défini (exemple en Figure 16). Le travail de thèse s'est focalisé sur cette l'identification de facteurs à l'origine de cette première forme de résistance.

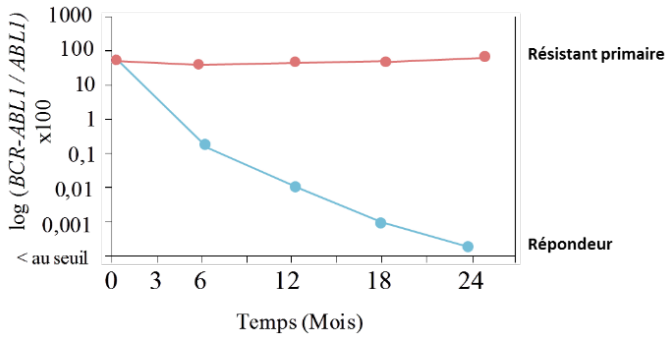


Figure 16 : Suivi du taux de transcrits *BCR-ABL1* au cours du temps après traitement par l'imatinib par PCRq

Deux cas sont représentés : un patient résistant primaire (en échec) et un patient répondeur (réponse optimale).

On parle ensuite de résistance secondaire (ou résistance acquise) lorsqu'une réponse favorable à l'imatinib est observée mais qu'elle est perdue au cours de temps. Ce second type de résistance survient chez environ 20% des patients ayant atteint la RCH et/ou RCyC [88]. Pour les patients résistants, la dose quotidienne d'imatinib peut être augmentée ou, un autre ITK, de deuxième ou troisième génération, est prescrit. Dans le cas d'une résistance persistante, une greffe de moelle osseuse allogénique peut également être envisagée.

4. Mécanismes de résistance à l'imatinib

4.1. Généralités

Le pourcentage de patients n'atteignant jamais la RCyC (résistants primaires) est estimé entre 20 et 25% [89]. Pour la majorité des patients, la résistance à l'imatinib est directement liée à BCR-ABL1. Pour les autres, plusieurs hypothèses sont proposées mais aucun mécanisme de résistance exact n'a été déterminé.

4.2. Mécanismes dépendants de BCR-ABL1

4.2.1. Mutations du domaine kinase de *BCR-ABL1*

Le premier cas de résistance à l'imatinib a été décrit en 2001 par Gorre et *al.* [90]. Le domaine kinase d'*ABL1* a été séquencé chez des patients en crise blastique, en rechute après une réponse initiale. Une mutation a été retrouvée chez 6 patients sur 9. Il s'agissait d'une substitution (changement d'une base azotée) non synonyme à l'origine d'un changement de la thréonine en position 315 en isoleucine (T315I). Ce résidu est nécessaire à la fixation de

l'imatinib pour la liaison à la poche de fixation de l'ATP et un changement d'acide aminé inhibe son activité. D'autres substitutions dans le domaine kinase d'*ABL1* ont par la suite été associées à la résistance aux différents ITK (Tableau 3) [90].

Tableau 3 : Liste des mutations du domaine kinase *d'ABL1* associées à une résistance aux différents ITK

Première génération			**Deuxième génération**			**Troisième génération**
imatinib			nilotinib	dasatinib	bosutinib	ponatinib
M237V	F311L	A380T	Y253F/H	V299L	E255K	T315M/L
M244V	**T315I**	F382L	E255K/V	**T315I**/A	V299L	?
L248R	F317L/V/I/C	L384M	**T315I**	F317L/V/I/C	**T315I**	
G250E/R	Y342H	L387M/F	F359V/I/C		?	
Q252R/H	M343T	M388L				
Y253F/H	A344V	Y393C				
E255K/V	M351T	H396R/P				
E258D	E355D/G	A397P				
L273M	F359V/I/C	S417F/Y				
E275K/Q	D363Y	I418S/V				
D276G	L364I	S438C				
T277A	A365V	E453G/K				
E279K	L370P	E459K/V				
V280A/I	V371A	P480L				
V289A	E373K	F486S				
V299L	V379I					

La mutation T315I, corrélée à une résistance aux ITK de 1ère et 2ème générations, est mise en évidence en gras. La mutation T315I ne cause pas de résistance au ponatinib en revanche, les mutations T315M et TT315L seraient impliquées [91,92]. Adapté de Soverini *et al.*, Clinical Lymphoma, Myeloma & Leukemia, 2015 [93]

Ces mutations sont classées en deux grandes catégories. Elles peuvent soit directement altérer la fixation de l'inhibiteur à la protéine en modifiant les points de contact entre BCR-ABL1 et l'imatinib [94], soit moduler la conformation tridimensionnelle de BCR-ABL1 en favorisant notamment la conformation active à laquelle l'imatinib ne peut pas se fixer [95]. Plus de 50 mutations ont été décrites. Quinze de ces mutations représentent près de 85% des mutations retrouvées chez les patients et sept résidus ont été identifiés comme essentiels pour la réponse à l'imatinib (G250, Y253, E255, T315, M351, F359, H396).

4.2.2. Amplification et surexpression de BCR-ABL1

Un deuxième mécanisme de résistance a été identifié, notamment dans notre laboratoire, en générant des lignées cellulaires résistantes par ajout de doses croissantes et graduelles d'imatinib [96,97]. Une amplification de *BCR-ABL1* a été identifiée dans ces lignées soit à cause de duplications du gène soit par la présence de chromosomes Philadelphie

supplémentaires [74]. Cependant, ce phénomène survient le plus souvent chez les patients en phases avancées de la maladie et ne concernerait qu'un nombre minoritaire de cas de résistance aux ITK.

4.3. Mécanismes indépendants de BCR-ABL1

4.3.1. Mauvaise observance thérapeutique

Le patient peut présenter des problèmes d'observance. La prise du médicament n'est pas régulière ou les doses ne sont pas respectées, intentionnellement ou par oubli. Une quantification de la concentration d'imatinib dans le sang est réalisée pour détecter ces variations dans la prise de médicament. Les patients présentant des problèmes d'observance (identifiés par le clinicien en charge de ces patients à l'hôpital) ont été exclus de notre cohorte d'étude.

4.3.2. Pharmacocinétique de l'imatinib

Pour qu'une thérapeutique soit efficace, elle doit parvenir jusqu'à sa cible. De nombreux facteurs peuvent moduler la disponibilité du médicament et donc son efficacité : on parle de facteurs pharmacocinétiques.

4.3.2.1. Diminution de la biodisponibilité orale

Selon les patients, il peut exister des variations dans l'absorption gastro-intestinale du médicament et le métabolisme de premier passage dans le foie. Ces différences affectent la biodisponibilité orale de l'imatinib [98]. En effet, les concentrations d'imatinib réelles détectées dans le sang peuvent grandement varier d'un individu à l'autre malgré une prise quotidienne identique [99,100]. L'imatinib est majoritairement métabolisé par les enzymes de la famille des cytochromes P450 (CYP) notamment CYP3A4 et CYP3A5 [101]. Plusieurs variants génétiques ont été identifiés dans ces deux gènes mais des résultats conflictuels entre les lignées cellulaires et les patients sont publiés [102,103].

4.3.2.2. Augmentation de la liaison aux protéines plasmatiques

La liaison de l'imatinib avec des protéines plasmatiques peut également être altérée. Une large majorité (89 à 96%) de l'imatinib dans le sang est liée à des protéines, notamment l'albumine et l'alpha-1-acid glycoprotein (AGP) [104]. Il n'y a que l'imatinib libre qui est disponible pour l'import dans les cellules. Une augmentation de la concentration de ces protéines plasmatiques pourrait donc augmenter la quantité de molécules d'imatinib non disponibles et ainsi induire une résistance au traitement.

4.3.2.3. Modification du transport dans la cellule leucémique

L'une des hypothèses les plus probables est que la résistance est associée à des anomalies dans l'influx ou l'efflux du médicament dans la cellule cible. Tout d'abord l'entrée de l'imatinib dans la cellule peut être diminuée. Par exemple, le transporteur OCT1 (organic cation transporter 1, ou SLC22A1) influe sur la disponibilité de l'imatinib dans les cellules leucémiques en inhibant son import à travers la membrane plasmique [105,106]. Plusieurs variants dans le gène *OCT1* ont été identifiés et corrélés à une mauvaise réponse au traitement par imatinib [103,107].

La sortie de l'imatinib hors de la cellule peut également être augmentée. Dans les années 70, l'efflux des chimiothérapies hors des cellules a été identifié comme mécanisme de résistance dans différentes pathologies [108]. La caractérisation de lignées cellulaires résistantes a notamment permis de mettre en évidence une surexpression d'ABCB1 (ATP binding cassette subfamily B member 1, aussi connu sous le nom MDR1, multidrug resistance protein 1) [97,109,110]. Ces études ont d'ailleurs montré qu'une inhibition des pompes à efflux à la membrane plasmique restaurait la sensibilité à l'imatinib.

4.3.3. Persistance d'une maladie résiduelle

Chez les patients atteints de LMC traités à l'imatinib, une maladie résiduelle est souvent détectée. Cela se traduit par la présence persistante de cellules BCR-ABL1^{+}. Ces cellules, non éliminées par l'imatinib, sont retrouvées chez près de 95% des patients en RCyC [111]. Ce mécanisme serait impliqué dans certains cas de résistance secondaire ou dans la rechute moléculaire après arrêt de traitement à l'imatinib. En effet, une perte de la réponse pour 55%

des patients a été observée au cours de l'essai clinique STIM 2 (*Stopping imatinib 2*) ayant pour objectif de stopper le traitement à l'imatinib pour les patients atteints de LMC répondeurs depuis au moins deux ans (ID de l'essai clinique : NCT01343173).

4.3.3.1. Cellule souche résistante

Les cellules souches sont résistantes à l'imatinib et pourraient être à l'origine d'une résistance secondaire ou d'une réapparition de la maladie si le traitement est arrêté [112]. Les cellules souches hématopoïétiques sont quiescentes et sont par conséquent, non dépendantes de l'activité de BCR-ABL1. Ces cellules en dormance sont donc résistantes au traitement par imatinib [113]. Bien que présentes en très faible proportion (0,5% des cellules $CD34^+$, cellules immatures de l'hématopoïèse), ces cellules forment un réservoir de cellules capables de produire de nouvelles cellules leucémiques. L'identification de vulnérabilités moléculaires spécifiques des cellules souches leucémiques est aujourd'hui un enjeu majeur pour surmonter la résistance et éliminer la maladie résiduelle [114].

4.3.3.2. Microenvironnement

Les cellules souches hématopoïétiques nichent dans le microenvironnement médullaire constitué en particulier de cellules stromales. Un ensemble de voies est activé afin de maintenir les cellules dans la moelle osseuse à l'état souche. Les cellules stromales sécrètent notamment CXCL12 reconnu par le récepteur CXCR4 à la surface des cellules souches. Ces signaux induisent une rétention des cellules dans la niche. Les cellules souches sécrètent quant à elles le facteur SCF qui favorise leur greffe au microenvironnement [115].

4.3.3.3. Evolution clonale

Au cours du temps, les cellules leucémiques peuvent acquérir des anomalies chromosomiques additionnelles favorables à leur persistance. Ces clones altérés peuvent alors proliférer en présence d'imatinib et participer à une diminution de la réponse au traitement et éventuellement à une progression de la maladie [9,116–118]. L'évolution clonale est plus susceptible d'expliquer la mauvaise réponse à l'imatinib que les mutations dans le domaine kinase d'*ABL1*. Cette constatation est d'autant plus vraie en crise blastique où 73% des patients démontrent une évolution clonale contre 30% qui portent des mutations dans le gène *ABL1* [119,120].

4.3.4. Altération de voies cellulaires dans la cellule leucémique

Enfin, la résistance pourrait également être due à des anomalies dans des voies alternatives ou régulées par BCR-ABL1. En effet, il se peut que l'imatinib inhibe efficacement BCR-ABL1 mais qu'en aval certaines voies cellulaires soient non fonctionnelles et bloquent l'élimination des cellules leucémiques. Les kinases de la famille SRC jouent un rôle crucial dans la signalisation en aval des récepteurs de surface des cellules hématopoïétiques. Une surexpression de certains de ces facteurs, tels que LYN ou HCK, a été associée à la résistance à l'imatinib [121]. Il semblerait que ces kinases agissent sur BCR-ABL1 en stabilisant la protéine en conformation active, ne pouvant ainsi pas être ciblée par l'imatinib [122].

De plus, l'apoptose, inhibée par BCR-ABL1 est essentielle à l'élimination des cellules leucémiques en réponse à l'imatinib. Ce mécanisme serait altéré chez certains patients non répondeurs [123]. Cette voie de résistance, qui fait l'objet d'étude de ce travail de thèse, est détaillée dans la deuxième partie de l'introduction.

Partie II : Mécanismes d'apoptose et régulation par la famille BCL2

1. L'apoptose

L'apoptose est le phénomène de mort programmée par la cellule, aussi appelé suicide cellulaire. La cellule provoque son « autodestruction » en réponse à des signaux intra- ou extracellulaires. C'est un processus essentiel à l'état physiologique qui permet notamment de modeler les organismes durant leur développement (exemple de la formation des doigts, Figure 17). Par la suite, durant la vie de l'organisme, l'apoptose intervient pour maintenir l'homéostasie cellulaire en parallèle de la prolifération cellulaire ou encore participer à la réaction immunitaire. Une apoptose dérégulée intervient dans différentes pathologies (ex : maladies neurodégénératives telles que la maladie d'Alzheimer, cancer ...).

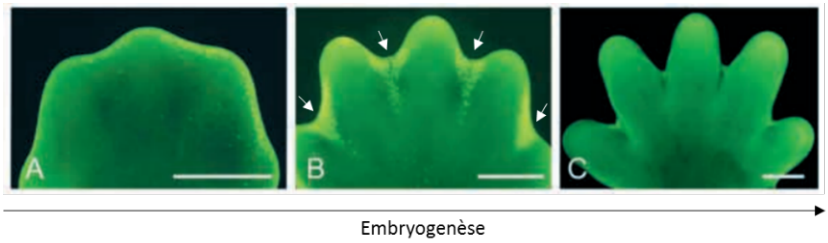


Figure 17 : Phénomène d'apoptose dans la formation des doigts d'une souris au cours du développement embryonnaire

Au cours du temps les tissus interdigitaux sont éliminés par apoptose (marquage des cellules en apoptose en vert clair). Adapté de Wood *et al*., Development, 2000 [124]

1.1. Historique

Ce phénomène a été décrit pour la première fois en 1842 par le chercheur Carl Vogt et, 30 ans plus tard, le terme « apoptose » a été donné par John Kerr et ses collaborateurs [125]. En 2002, les chercheurs Sydney Brenner, Robert Horvitz et John Sulston reçoivent le prix Nobel de médecine et physiologie pour leurs « découvertes sur la régulation génétique du développement des organes et de la mort programmée des cellules » dans le modèle du nématode *Caenorhabditis elegans* (*C. elegans*) [126,127]. L'organisme *C. elegans* adulte est formé à partir de 1090 cellules somatiques dont 131 entrent en apoptose. Ces 131 cellules sont éliminées au cours du développement de manière hautement régulée et constante d'un individu à l'autre ce qui en a fait un modèle d'étude idéal pour identifier le déterminisme génétique et les voies moléculaires impliquées dans la régulation de l'apoptose.

1.2. Modifications morphologiques

L'apoptose se déroule en trois phases majeures caractérisées par plusieurs changements morphologiques (Figure 18).

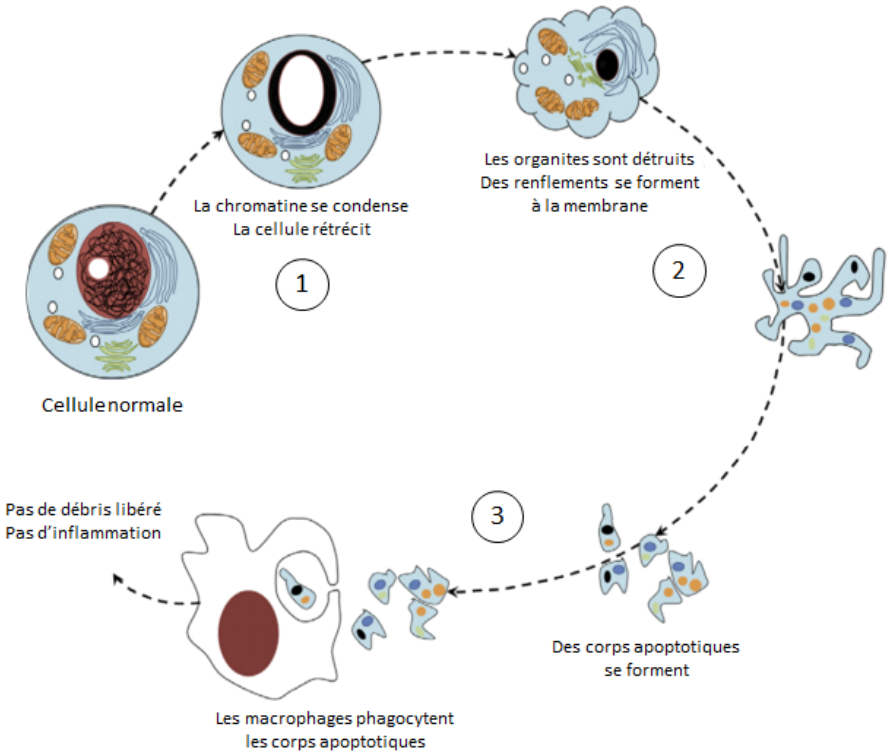


Figure 18 : Modifications morphologiques de la cellule en apoptose

Adapté de Abou-Ghali et Stiban, Saudi Journal of Biological Sciences, 2015 [128]

Dans un premier temps, la chromatine se condense de manière irréversible conduisant à une réduction du volume nucléaire et à un rétrécissement global de la cellule. Vient ensuite l'étape du bourgeonnement. Le noyau est détruit par fragmentation et, par altération du cytosquelette, les débris et les organites sont englobés dans des renflements à la surface de la membrane plasmique. Ces bourgeonnements donnent une forme très caractéristique aux cellules en cours d'apoptose. Enfin, ces bulles forment les corps apoptotiques, libérés dans le milieu extracellulaire. Les vésicules sont dégradées par phagocytose : la cellule est entièrement détruite sans que des résidus ne soient libérés dans le milieu extracellulaire. De par l'absence de débris cellulaires libres dans l'environnement extracellulaire et l'élimination rapide des corps apoptotiques par les macrophages, aucune réaction inflammatoire ne survient. C'est un mécanisme d'élimination de la cellule hautement régulé. Ces changements morphologiques sont induits par des cascades d'évènements moléculaires nécessitant une consommation d'énergie.

1.3. Voies alternatives de dégradation de la cellule

1.3.1. La nécrose

L'apoptose se différencie de la nécrose qui correspond au phénomène de mort cellulaire non programmée. Lorsque des cellules entrent en nécrose, la chromatine est entièrement détruite par l'action d'endonucléases et le noyau est détruit. A l'opposé de l'apoptose, le volume cellulaire augmente et les cellules éclatent en libérant les débris dans le milieu extracellulaire. La nécrose est un processus non régulé par la cellule et toxique. La libération de débris dans l'environnement conduit au déclenchement de réactions inflammatoires. A la différence de l'apoptose, c'est un mécanisme passif sans consommation d'énergie.

1.3.2. L'autophagie

L'autophagie est un processus hautement conservé permettant la dégradation et le recyclage spécifique de composés cellulaires en réponse à différents stimuli [129]. Les éléments cytoplasmiques à dégrader sont englobés dans des vésicules nommées autophagosomes fusionnant avec les lysosomes contenant des enzymes de dégradation. Ce mécanisme intervient notamment dans la survie cellulaire *via* le contrôle qualité des cellules en évitant l'accumulation de protéines non fonctionnelles ou mal repliées. Il peut également permettre la dégradation d'organites endommagés tels que la mitochondrie et ainsi éviter l'apoptose. A l'inverse, l'autophagie peut induire la mort cellulaire. Par exemple, la protéine d'autophagie ATG12 peut éliminer certains facteurs anti-apoptotiques et ainsi provoquer la mort cellulaire [130].

1.4. Les voies majeures de l'apoptose

Au cours du XIXème siècle, les recherches permettent d'identifier progressivement les mécanismes moléculaires qui interviennent dans ces changements morphologiques. Deux voies majeures d'induction de l'apoptose sont identifiées : la voie extrinsèque, induite par l'activation des récepteurs « de mort », et la voie intrinsèque, induite par des stress environnementaux, active *via* la modification du potentiel membranaire mitochondrial (Figure 19).

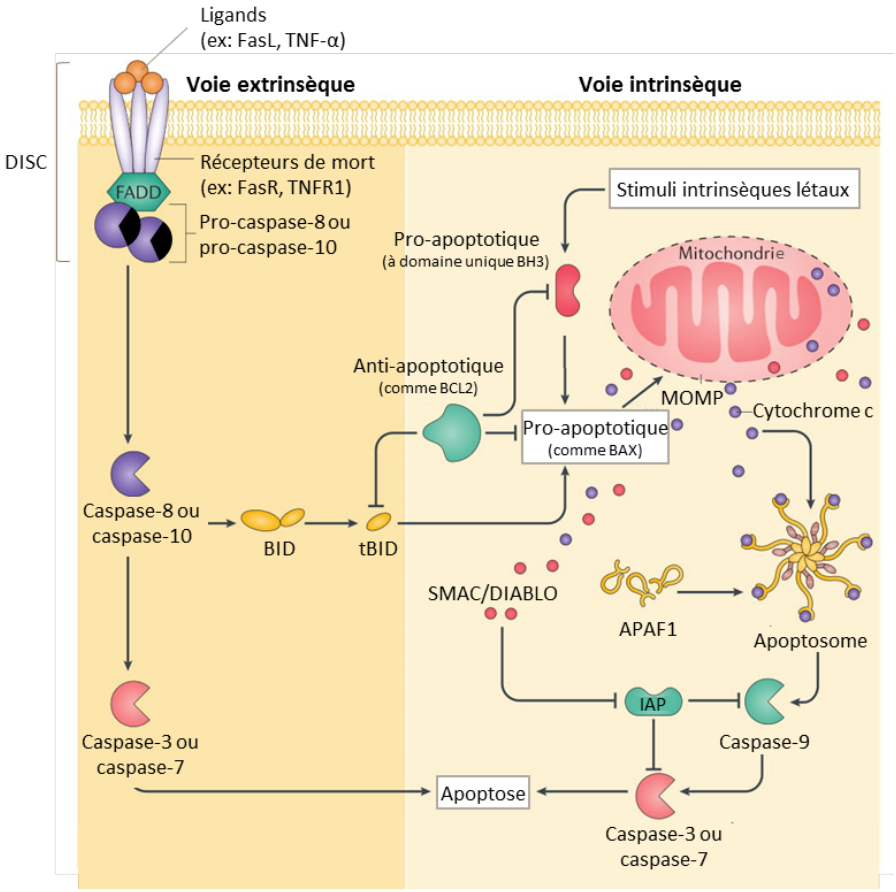


Figure 19 : Voies extrinsèque et intrinsèque d'induction de l'apoptose

Adapté de Ichim et Tait, Nature Reviews Cancer, 2016 [131]

1.4.1. Les caspases

L'induction de l'apoptose par ces deux voies passe par l'activation des caspases (cysteinyl aspartate-specific proteases). Il s'agit de protéases très conservées exprimées sous forme de pro-enzymes inactives dans la plupart des types cellulaires. Les caspases sont activées par clivage de leur domaine N-terminal (pro-domaine) et la dissociation de la grande et de la petite sous-unité. Les coupures sont induites au niveau de résidus aspartate (Figure 20) [132].

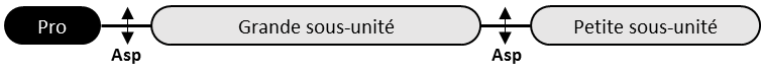


Figure 20 : Structure d'une pro-caspase

Les deux sous-domaines et le pro-domaine (Pro) sont dissociés suite au clivage de la pro-caspase au niveau des résidus asparatate (Asp).

Lorsque les caspases sont activées, elles peuvent à leur tour activer d'autres caspases *via* leur activité protéolytique : on parle alors d'une cascade de protéases. Cette voie permet la transmission dans la cellule du signal apoptotique qui entraîne la mort rapide de la cellule. Dix caspases principales ont été identifiées et sont classées en trois groupes : les caspases initiatrices (caspase-2, -8, -9 et -10), les caspases effectrices (caspase-3, -6 et -7) et les caspases inflammatoires (caspase-1, -4 et -5). Les caspases initiatrices s'auto-activent en formant des dimères, puis, elles activent les caspases effectrices qui déclenchent le signal apoptotique (Figure 21). Les caspases inflammatoires sont impliquées dans une autre forme de mort cellulaire programmée à l'origine d'une réponse immunitaire [132].

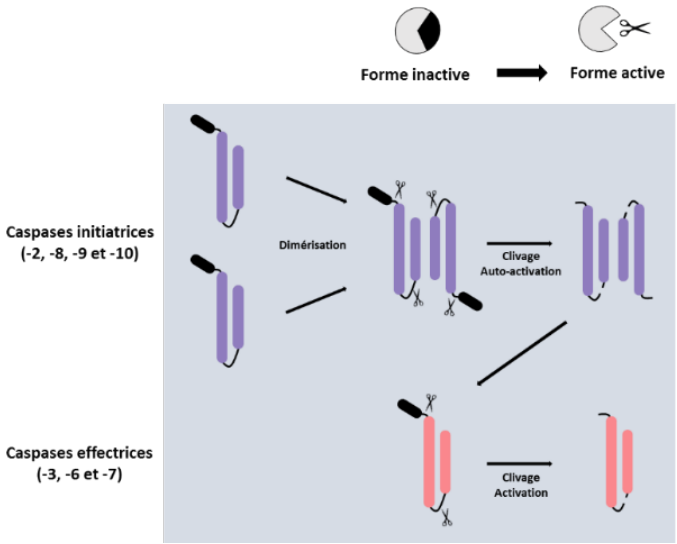


Figure 21 : Mécanisme d'activation des caspases initiatrices et effectrices

1.4.2. La voie extrinsèque

La voie extrinsèque est médiée par l'activation de récepteurs transmembranaires tels que les récepteurs de mort de la superfamille des tumor necrosis factor (TNF) [133]. Ces récepteurs sont composés d'un domaine extracellulaire conservé, riche en cystéine, et d'un domaine cytoplasmique appelé domaine de mort [134]. Ces récepteurs permettent la transmission d'un signal de mort du milieu extra au milieu intra-cellulaire. Une série de couples ligand/récepteur a été identifiée tels que FasL/FasR ou TNF-α/TNFR1 [135,136]. Les récepteurs forment des amas à la surface des cellules et sont activités par des ligands trimériques. A la suite de l'activation de ces récepteurs, des protéines adaptatrices cytoplasmiques sont recrutées et se lient aux domaines de mort. Par exemple après l'activation de FasR par FasL, la protéine adaptatrice Fas associated via death domain (FADD) est recrutée

[137]. Cette molécule s'associe à son tour aux caspases initiatrices 8 ou 10 sous forme de pro-enzymes. Ce complexe de protéines est appelé death-inducing signaling complex (DISC). L'ensemble de ces interactions permet l'activation des caspases initiatrices qui vont à leur tour activer les caspases effectrices 3 ou 7 et induire la phase d'exécution de l'apoptose (Figure 19) [138].

1.4.3. La voie intrinsèque

1.4.3.1. Initiation du signal apoptotique

Il s'agit de la voie principale d'induction de l'apoptose. La mort cellulaire est induite par un ensemble de stimuli non liés à des récepteurs spécifiques. Certains signaux d'activation sont dits négatifs tels que l'absence de facteurs de croissance, d'hormones et de cytokines. La cellule étant dans un environnement défavorable à sa croissance, la mort cellulaire est initiée. D'autres stimuli sont dits positifs tels que la présence de radicaux libres ou la survenue d'infections virales.

1.4.3.2. Régulation par la famille BCL2

Tous ces signaux agissent sur la membrane mitochondriale interne *via* la régulation des membres de la famille BCL2. Il s'agit d'un ensemble de facteurs interagissant entre eux pour maintenir l'intégrité de la membrane mitochondriale ou au contraire induire sa dégradation. Cette famille englobe au moins 21 protéines réparties en trois sous-familles : les inhibiteurs de l'apoptose comme BCL2, les effecteurs de l'apoptose tels que BCL2 associated X (BAX) eux-mêmes activés par les protéines « à domaine BH3 unique ». Au cours de ce travail, les 21 gènes de la famille *BCL2* ont été séquencés chez des patients atteints de LMC. Des variations de séquence de ces gènes pourraient être associées à une altération du processus d'apoptose et par conséquence entraîner la résistance à l'imatinib dans la LMC.

Les protéines de la famille BCL2 partagent une structure tridimensionnelle similaire et possèdent au moins un des domaines d'homologie à BCL2 (BH1-4) en hélices alpha. Toutes les protéines de la famille BCL2 possèdent au minimum le domaine BH3 LXXXGD avec X représentant n'importe quel acide aminé (Figure 22). L'interaction entre les facteurs permettant la régulation de l'apoptose se fait le plus souvent *via* ce domaine BH3 [139]. La plupart des protéines possèdent également un domaine transmembranaire C-terminal, composé d'une série

de résidus hydrophobes, nécessaires à l'ancrage à la membrane mitochondriale externe (Figure 22) [140].

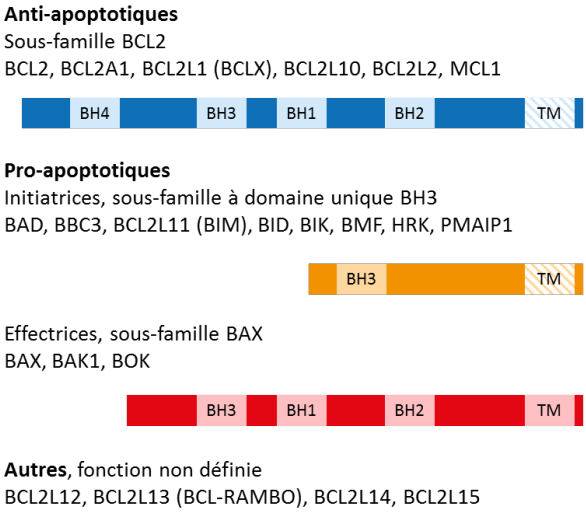


Figure 22 : Structure des protéines des trois-sous familles de la famille BCL2 (21 membres)
BH : domaine d'homologie à BCL2, TM : domaine transmembranaire. TM n'est pas présent dans toutes les protéines anti-apoptotiques et de la sous-famille des protéines pro-apoptotiques initiatrices. Pour quatre protéines, la fonction exacte n'est pas encore définie. Adapté de Cory et Adams, Nature Reviews, 2002 [141]

Dans une cellule vivante, les facteurs anti-apoptotiques, en concentrations élevées, inhibent les facteurs pro-apoptotiques effecteurs, notamment BAX, présents sous formes monomériques dans la membrane mitochondriale externe (Figure 23).

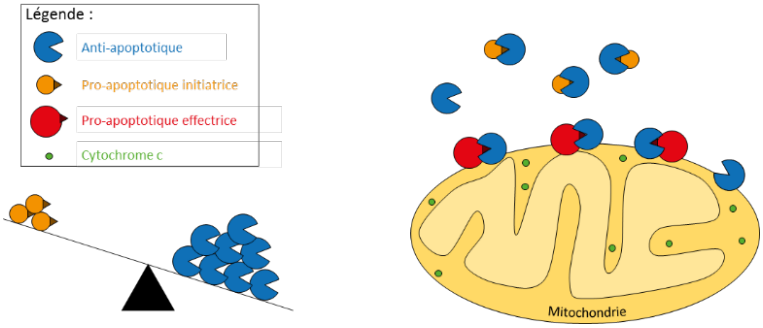


Figure 23 : Interactions des protéines de la famille BCL2 dans une cellule vivante
La concentration de protéines anti-apoptotiques est supérieure à la concentration de protéines pro-apoptotiques initiatrices : l'intégrité de la membrane mitochondriale est maintenue.

Lorsqu'un signal apoptotique est induit dans la cellule, l'expression des protéines pro-apoptotiques initiatrices à domaine unique BH3 est augmentée. Les facteurs anti-apoptotiques comme BCL2 sont séquestrés, libérant ainsi les protéines pro-apoptotiques effectrices comme BAX. De plus, les effecteurs sont également directement activés par les protéines initiatrices (Figure 24).

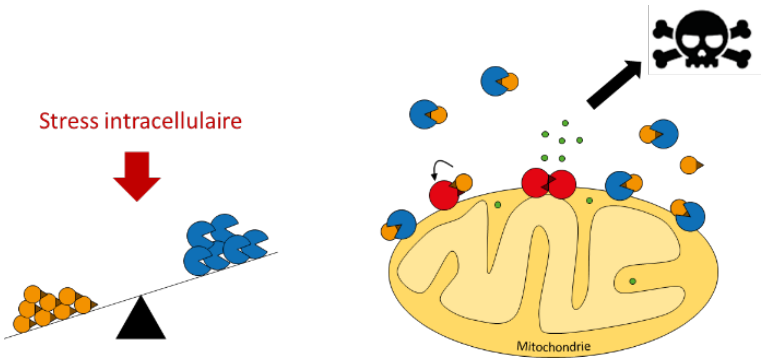


Figure 24 : Interactions des protéines de la famille BCL2 dans une cellule en apoptose

La concentration de protéines pro-apoptotiques initiatrices est augmentée par différents stress intracellulaires. La concentration de protéines anti-apoptotiques est alors inférieure, les protéines pro-apoptotiques effectrices sont libérées, permettant l'induction de l'apoptose.

Les protéines de type BAX forment des oligomères dans la membrane mitochondriale et créent ainsi des pores de transition de perméabilité mitochondriale (*mitochondrial permeability transition*, MPT). L'ouverture de ces canaux entraîne une perte de potentiel membranaire mitochondrial (*mitochondrial outer membrane permeabilization*, MOMP) et la libération de protéines pro-apoptotiques dans le cytoplasme jusqu'alors séquestrées dans l'espace inter-membranaire telles que le cytochrome c, le complexe Smac/DIABLO ou encore la protéase à sérine HtrA2/Omi [142].

1.4.3.3. Exécution de l'apoptose

Ces protéines libérées dans le cytoplasme entraînent une cascade d'activation de caspases à l'origine de la phase d'exécution de l'apoptose. Le cytochrome c se lie à APAF1 et à la pro-caspase-9 formant un complexe : l'apoptosome. La formation de cette structure permet le clivage de la pro-caspase-9 et par conséquent son activation. Cette dernière active à son tour la caspase-3 et provoque l'apoptose. Les protéines Smac/DIABLO et HtrA2/Omi quant à elles semblent induire l'apoptose par l'inhibition d'inhibiteurs de l'apoptose désignés IAP (inhibitor of apoptosis) (Figure 19).

1.4.4. BID, à l'interface entre voies extrinsèque et intrinsèque

BID, protéine de la famille BCL2 à domaine unique BH3, est exprimée dans le cytoplasme à l'état inactif. Cette protéine est activée par la caspase-8, elle-même activée dans la voie extrinsèque. La forme tronquée (tBID) agit comme un facteur pro-apoptotique activant les facteurs de type BAX responsables du changement de potentiel membranaire mitochondrial dans la voie intrinsèque (Figure 19) [143,144].

2. L'apoptose dans les cellules leucémiques BCR-ABL1^{+}

2.1. Inhibition de l'apoptose par BCR-ABL1

Dans la cellule leucémique, l'apoptose est bloquée par BCR-ABL1 *via* différentes voies de signalisation telles que la voie PI3K/AKT. La sur-activation de cette voie par BCR-ABL1 entraîne notamment une augmentation de la phosphorylation de la protéine pro-apoptotique BCL2 associated agonist of cell death (BAD) [145]. Les protéines anti-apoptotiques BCL2 et BCL2-like 1 (BCL2L1 ou BCLX), jusqu'alors séquestrées par BAD, sont libérées dans le cytoplasme [146]. La sur-activation de la voie PI3K/AKT favorise également l'inhibition de l'expression du gène *BIM*, codant un facteur pro-apoptotique essentiel à l'induction de l'apoptose par la voie intrinsèque, *via* l'inhibition du facteur de transcription FOXO3 [147]. L'expression de la protéine anti-apoptotique BCLX est quant à elle augmentée par BCR-ABL1 *via* l'activation du facteur de transcription STAT5 [148]. L'ensemble de ces mécanismes, induits par l'activité tyrosine kinase constitutive de BCR-ABL1, favorise la survie cellulaire.

Au cours de ce travail de thèse, huit gènes codant des kinases ou des facteurs de transcription de la voie PI3K/AKT, impliqués dans la régulation des protéines de la famille BCL2, ont été étudiés. Le rôle des huit protéines codées par ces gènes est résumé dans la Figure 25.

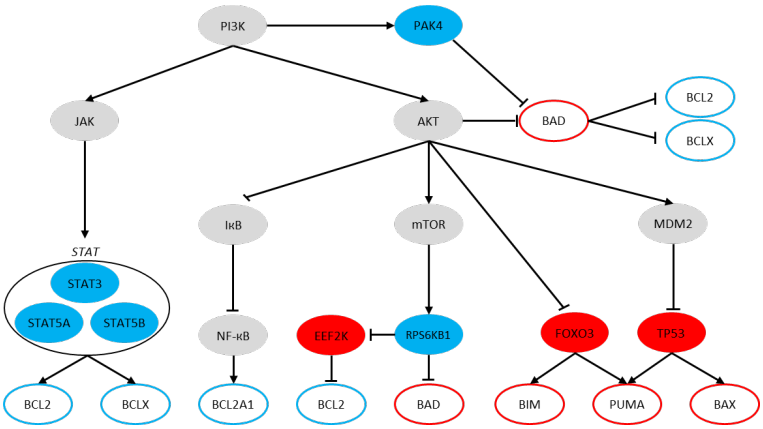


Figure 25 : Schéma simplifié de l'inhibition de l'apoptose intrinsèque par la voie PI3K/AKT

Les protéines de la famille BCL2 sont entourées. Les kinases et facteurs de transcription impliqués dans la régulation des membres de la famille BCL2, et inclus dans notre étude, sont colorés. Bleu : anti-apoptotique, rouge : pro-apoptotique. Les flèches indiquent une régulation positive (↓) ou négative (⊥).

Les interactions entre les différentes protéines sont estimées à partir de diverses études [149–159].

Les voies des MAPK sont également essentielles dans la régulation de l'apoptose. La sur-activation de la voie RAS/MAPK favorise l'inhibition de l'apoptose *via* l'inactivation de BIM. En effet, la protéine est phosphorylée par ERK ce qui favorise sa dégradation par le protéasome [160]. L'apoptose peut également être induite par la voie RAS/MAPK notamment par les deux sous-familles de MAPK, JNK et p38. L'apoptose est notamment induite en réponse aux ITK *via* l'activation du facteur TP53. De manière intéressante, le gène *TP53* est le gène le plus fréquemment muté dans les cancers et pourrait donc être associé à une absence de réponse à l'imatinib dans la LMC [161].

Pour ce projet de thèse, 16 gènes supplémentaires, responsables de la régulation de la famille BCL2 par les voies des MAPK ou de la régulation de TP53, ont été inclus dans le panel à séquencer. Le rôle de ces protéines est résumé dans les Figure 26 et Figure 27. Au total, un groupe de 45 gènes, 21 gènes de la famille *BCL2* et 24 régulateurs codant des kinases et des facteurs de transcription, a été étudié pour ce projet de thèse cherchant à identifier des marqueurs de la résistance à l'imatinib dans la LMC.

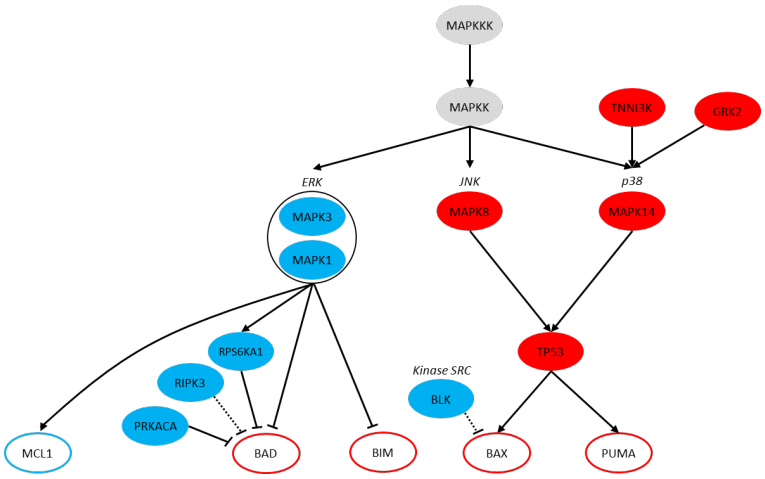


Figure 26 : Schéma simplifié de la régulation de l'apoptose par les voies des MAPK (sous-familles ERK, JNK et p38)

Les protéines de la famille BCL2 sont entourées. Les kinases et facteurs de transcription impliqués dans la régulation des membres de la famille BCL2, et inclus dans notre étude, sont colorés. Bleu : anti-apoptotique, rouge : pro-apoptotique. Les flèches indiquent une régulation positive (↓) ou négative (⊥).

Les interactions entre les différentes protéines sont estimées à partir de diverses études [162–168].

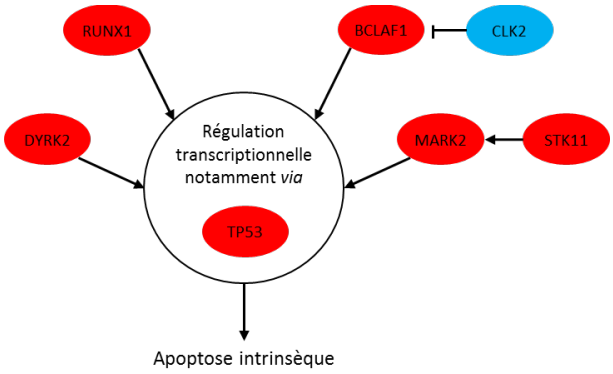


Figure 27 : Schéma simplifié de la régulation de l'apoptose par la régulation de l'expression des gènes notamment *via* le facteur de transcription TP53

Les kinases et facteurs de transcription impliqués dans la régulation des membres de la famille BCL2, et inclus dans notre étude, sont colorés. Bleu : anti-apoptotique, rouge : pro-apoptotique. Les flèches indiquent une régulation positive (↓) ou négative (⊥).

Les interactions entre les différentes protéines sont estimées à partir de diverses études [169–173].

2.2. Altération des protéines de la famille BCL2 et résistance à l'imatinib

Plusieurs mécanismes de résistance à l'imatinib ont été associés à des modifications d'expression des membres de la famille BCL2. Par exemple, une surexpression du facteur anti-

apoptotique BCLX entraîne une résistance à l'imatinib dans des lignées cellulaires modèles de LMC [174]. Dans des lignées cellulaires résistantes à l'imatinib, une augmentation de l'expression du facteur anti-apoptotique BCL2 a été mise en évidence [175]. De plus, la surexpression de BCL2 a été largement décrite dans différentes pathologies et associée à l'oncogenèse, la progression tumorale et la résistance aux thérapies [176–178]. C'est d'ailleurs ainsi que ce gène a été découvert, BCL2 étant surexprimé dans des lymphomes B (B-cell lymphoma, BCL) suite à une translocation entre les chromosomes 14 et 18 [179,180].

Enfin la protéine pro-apoptotique BIM, a également été reportée dans plusieurs études comme étant importante dans l'inefficacité de l'imatinib. Une diminution de l'expression de *BIM* par ARN interférence confère une résistance à l'imatinib dans des lignées cellulaires de LMC [181]. De plus, deux variants génétiques, retrouvés chez des patients atteints de LMC, ont été associés à une résistance au traitement. Ng *et al.* ont montré qu'une délétion de 2,9 kb dans l'intron 2 du gène *BIM* entraînait un épissage alternatif à l'origine de la perte du domaine BH3 rendant la protéine BIM non fonctionnelle [182]. Ce variant a été retrouvé chez 59% de patients résistants contre 7,4% de répondeurs dans une cohorte de 203 individus (présent à 12,3% dans la population générale). A noter cependant que cette délétion n'est présente que dans la population asiatique. De plus, d'autres études ont par la suite démontré que l'implication de ce polymorphisme dans la résistance ne serait pas significative [183–185].

Au laboratoire, Augis *et al.* ont mis en évidence un polymorphisme silencieux (rs724710, c29201 C>T) dans le domaine BH3, associé à une résistance plus élevée chez les patients porteurs de l'allèle T [186]. Dans cette étude, ce polymorphisme a été séquencé par une approche Sanger chez 72 patients atteints de LMC classés en deux catégories en fonction de leur réponse à l'imatinib, d'après les critères définis en 2009 par l'ELN. Trente-deux patients étaient répondeurs, en réponse optimale, et 40 non répondeurs, en réponse sub-optimale ou en échec selon l'ELN. L'allèle T était présent à 44% chez les répondeurs, à 65% chez les non répondeurs et à 31% chez des individus sains de la région bordelaise qu'ils ont séquencés afin d'obtenir une population régionale de référence. Cet allèle variant a donc été identifié comme sur-représenté chez les patients non répondeurs au traitement par rapport à la population de référence. Dans cette étude, l'allèle T a été associé à une diminution de la transcription de *BIM*, chez des témoins sains. Ce polymorphisme a également été mis en évidence par Gagné *et al.* qui ont étudié le rôle de polymorphismes du gène *BIM* dans la résistance des cellules leucémiques aux corticostéroïdes dans des cas de LAL pédiatriques [187].

L'objectif du projet de thèse a été d'identifier et caractériser de nouveaux variants dans 45 gènes associés à la régulation de l'apoptose (21 gènes de la famille *BCL2*, 24 gènes codant des kinases et facteurs de transcription) pouvant expliquer les cas de patients atteints de LMC résistants à l'imatinib. Pour cela, une approche de séquençage nouvelle génération permettant l'analyse simultanée de plusieurs gènes à partir de l'ADNg extrait de 24 patients a été choisie.

Partie III : Identification de variants génétiques par séquençage

1. Evolution des technologies de séquençage de l'ADN

1.1. Généralités

Le séquençage consiste à déterminer l'enchaînement des nucléotides d'un fragment d'ADN ou d'un génome entier. C'est en 1977, que les deux premières technologies sont développées par Maxam-Gilbert et Sanger. Bien qu'elles soient efficaces, ces méthodes sont limitées par la faible quantité de données générées. Malgré cela, l'approche Sanger a permis le séquençage du tout premier génome humain publié en 2001. Dans les années 2000, les approches haut-débit de deuxième génération font leur apparition. Ces nouvelles technologies, appelées séquençages de nouvelle génération, révolutionnent le domaine des biotechnologies et permettent l'identification de génomes de milliers d'espèces. La limite de ces approches est la taille des fragments séquencés, restreinte à quelques centaines de nucléotides. Depuis 2010, des techniques de séquençage de troisième génération sont développées et permettent la caractérisation de longs fragments d'ADN de plusieurs milliers de bases en une seule réaction.

1.2. La première génération

1.2.1. Maxam et Gilbert

La première méthode de séquençage, décrite par Allan Maxam et Walter Gilbert, est basée sur la dégradation chimique sélective [188]. Dans un premier temps, le fragment d'ADN à séquencer est marqué par un isotope radioactif en 5' puis différentes réactions chimiques indépendantes sont réalisées. Chaque réaction permet la modification spécifique d'un ou plusieurs nucléotide(s) donné(s). Dans un deuxième temps, les fragments marqués sont clivés au niveau des bases modifiées à l'aide d'un composé organique : la pipéridine. Enfin, les fragments sont séparés par électrophorèse sur gel de polyacrylamide (résolution d'une pb) et, selon la taille des fragments et le réactif chimique utilisé, la séquence d'ADN peut être déduite (Figure 28).

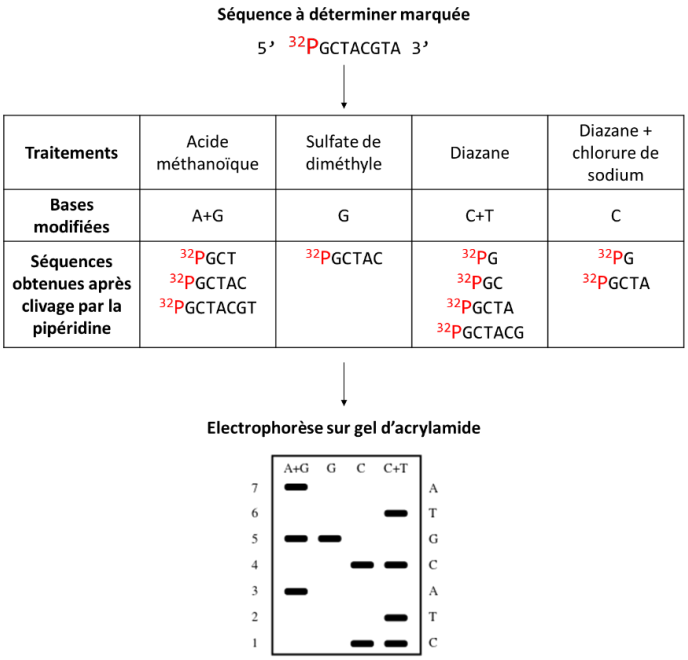


Figure 28 : Principe du séquençage de Maxam-Gilbert

1.2.2. Sanger

La deuxième méthode a été décrite la même année par Frederick Sanger et reste, à l'heure actuelle, couramment utilisée dans les laboratoires [189]. Le séquençage Sanger est basé sur une synthèse enzymatique sélective. Cette stratégie permet d'identifier jusqu'à environ 1000 pb d'un fragment d'ADN. L'ADN double brin à séquencer peut être fragmenté et cloné dans des vecteurs de séquences connues. La séquence à identifier peut également être amplifiée par PCR. Une polymérisation de l'ADN est initiée à l'aide d'une amorce complémentaire (oligonucléotide d'environ 20 nt) au vecteur ou au début de la séquence d'intérêt. Comme pour une réaction de PCR, la séquence complémentaire est générée par incorporation de désoxyribonucléotides (dNTP) présents dans le milieu réactionnel. La subtilité par rapport à la PCR est l'ajout à faibles concentrations de didésoxyribonucléotides (ddNTP) qui sont incorporés aléatoirement lors de l'élongation. Ces éléments agissent comme des terminateurs de chaîne et la polymérisation est stoppée pour le fragment d'ADN correspondant (Figure 29).

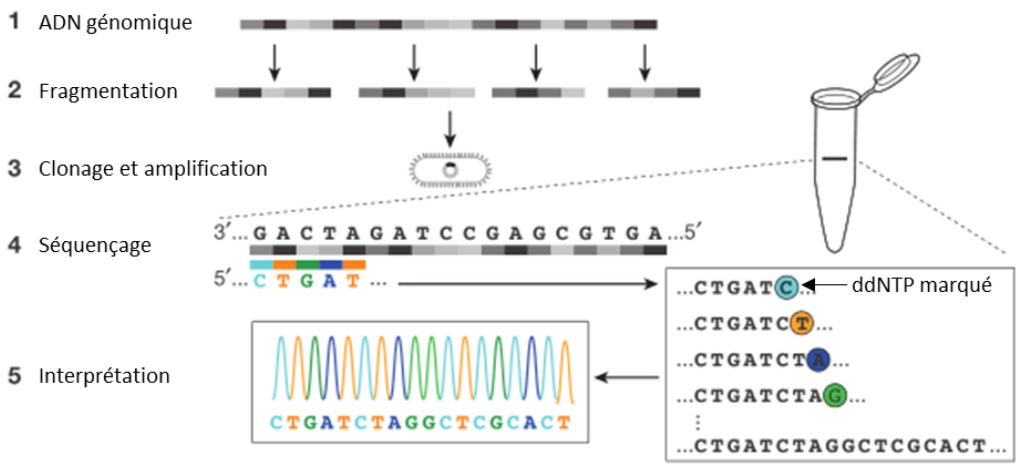


Figure 29 : Principe du séquençage par la méthode Sanger

Adapté de Shendure *et al.*, Nature, 2017 [190]

Initialement chaque ddNTP était incorporé indépendamment dans quatre réactions différentes puis les produits obtenus étaient séparés par une électrophorèse en gel de polyacrylamide à haute résolution et visualisés *via* un marquage radioactif. Aujourd'hui, les ddNTP sont marqués par des fluorophores ce qui permet de ne réaliser qu'une seule réaction avec les quatre ddNTP associés à quatre fluorescences différentes. Cette technologie a permis à Frédérick Sanger de publier la toute première séquence d'un génome entier, celui du bactériophage ΦX174 comprenant 5375 nucléotides [191]. L'usage de produits toxiques et de radio-isotopes est moindre dans le séquençage Sanger, ce dernier devient donc la méthode de choix pour l'identification de séquences d'ADN.

1.3. La deuxième génération : le séquençage nouvelle génération

1.3.1. Une approche haut débit

L'utilisation du séquençage Sanger pour la caractérisation de nouveaux génomes présente plusieurs limites telles que le nombre de réactions simultanées réduit. Le séquençage nouvelle génération (*next-generation sequencing*, NGS) fut alors développé pour répondre à cette problématique. Cette nouvelle technologie apporte deux améliorations majeures par rapport au séquençage Sanger. Tout d'abord, au lieu de centaines, des milliers voire des millions de réactions peuvent être initiées en parallèle. Ensuite, les données générées par l'appareil sont directement détectables et analysables, il n'est plus nécessaire de réaliser des électrophorèses.

1.3.2. Le séquençage nouvelle génération par synthèse

1.3.2.1. Pyroséquençage et séquençage par ligation

Une première technologie appelée pyroséquençage a été développée en 2005 par la société 454 Life Sciences (racheté par Roche) [192]. Cette approche est reprise en 2010 par Ion Torrent (racheté par Thermo Fisher Scientific). Les fragments d'ADN sont liés à des billes individualisées puis amplifiés par PCR en émulsion. La séquence nucléotidique est ensuite détectée grâce à une réaction de chimioluminescence survenant chaque fois qu'une base est incorporée. Cette approche permet de séquencer des fragments jusqu'à 700 pb ce qui présente un réel avantage pour la caractérisation de génomes *de novo.* En revanche, l'inconvénient majeur de cette approche est le taux d'erreur de séquençage élevé. Le manque d'exactitude est notamment causé par les répétitions de plusieurs nucléotides identiques. Pour pallier cette limite, chaque fragment d'ADN doit être détecté de très nombreuses fois, et par conséquent, le coût de séquençage par échantillon est élevé.

En 2007, la méthode Sequencing by Oligo Ligation Detection (SOLiD) est proposée par Applied Biosystems (aujourd'hui Thermo Fisher Scientific) [193]. Cette approche de séquençage par ligation se base sur l'hybridation de sondes de séquences connues (contenant deux bases spécifiques) à partir d'une amorce complémentaire du fragment d'ADN à identifier. Chaque fois qu'une sonde est correctement hybridée à la séquence d'intérêt, un fluorophore est libéré et sa fluorescence émise est détectée. En utilisant un panel d'amorces et de sondes, l'ensemble du fragment peut être caractérisé. La complémentarité de deux bases simultanées augmente grandement la fiabilité des résultats comparée aux méthodes précédentes. En revanche, la taille des séquences lues est réduite (35 pb) et le coût par réaction de séquençage reste élevé.

1.3.2.2. Les terminateurs de séquences réversibles

Un autre système a été commercialisé en 2006 par Solexa (racheté par Illumina). L'appareil *Genome analyzer* est basé sur la synthèse de fragments complémentaires et l'utilisation de terminateurs de séquence fluorescents réversibles. L'ADN est tout d'abord fragmenté et des adaptateurs sont ajoutés de part et d'autre de chaque séquence. Ces fragments sont ensuite liés à la surface d'une « lame » (*flow-cell*) par complémentarité des adaptateurs avec des séquences universelles fixées à la lame et un cycle de PCR est réalisé. Après dénaturation des fragments, des « ponts » sont formés : chaque brin d'ADN est lié par ses deux

extrémités à la matrice. Par la suite, des réactions de PCR classiques permettent l'amplification de chaque fragment qui reste toujours fixé sur la lame. Ces étapes sont répétées de nombreuses fois jusqu'à la formation de groupes denses de fragments (clusters), chaque groupement correspondant à un fragment d'ADN initial unique (Figure 30).

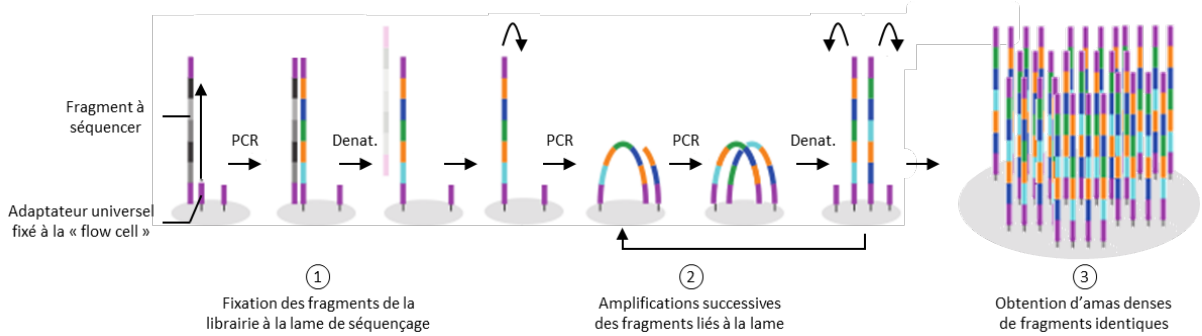


Figure 30 : Liaison des produits de la librairie à la lame de séquençage et amplification des fragments
Adapté de Shendure *et al.*, Nature, 2017 [190]

La séquence de chaque fragment est par la suite identifiée. Des amorces, complémentaires aux adaptateurs, et des dNTP associés à un terminateur marqué sont ajoutés au milieu. Par réaction de polymérisation, un dNTP est incorporé au fragment d'ADN et, après excitation au laser, la fluorescence émise par chaque cluster est capturée par informatique enfin, le terminateur est éliminé (Figure 31). Ces étapes sont répétées 75 à 300 fois selon les appareils et les réactifs utilisés.

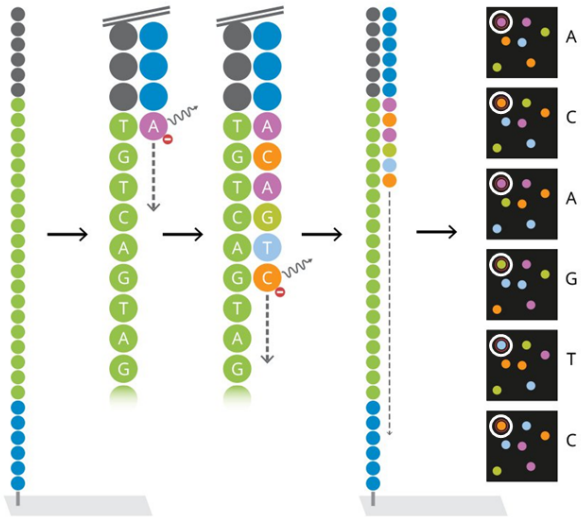


Figure 31 : Lecture des séquences fixées à la lame par ajout de dNTP fluorescents et capture d'images

Grâce aux derniers appareils développés, cette technologie permet d'obtenir jusqu'à 400 millions de lectures de 250 pb en une seule réaction. De plus, les résultats obtenus sont d'une grande fiabilité avec un taux d'erreur estimé à 10^{-3} [194]. Enfin, cette technologie est l'une des moins coûteuses (<$0,10/Mb) [195]. C'est donc la méthode qui a été choisie pour le séquençage des 45 gènes associés à la régulation de l'apoptose à partir d'échantillons de patients atteints de LMC.

1.4. La troisième génération

1.4.1. Généralités

Les méthodes de séquençage de première et deuxième génération ne permettent de séquencer que des fragments de longueurs restreintes. La détection de régions riches en répétitions ou de grandes variations structurales est donc limitée par ces approches. A partir de 2010, de nouvelles méthodes de séquençage de molécules uniques ont été développées et permis la lecture de fragments de plus de 100.000 pb.

1.4.2. Séquençage en temps réel de molécules uniques

1.4.2.1. Détection par fluorescence

La méthode la plus utilisée est la technologie *single-molecule real-time* (SMRT) commercialisée en 2010 par Pacific Biosciences et souvent nommée le séquençage PacBio [196,197]. Brièvement, des supports (SMRT cells) contenant des dizaines de milliers de puits nanoscopiques sont utilisés. Une polymérase liée à une molécule d'ADN est fixée au fond de chaque cavité. Le séquençage est ensuite réalisé en temps réel selon un principe similaire à l'approche Illumina. Le fragment d'ADN est répliqué par polymérisation en incorporant des dNTP marqués. A chaque dNTP intégré, le fluorophore est excité par de la lumière émise en dessous de chaque puits. La taille miniature de cette structure limite la propagation de la lumière émise au fond du puits et empêche ainsi la détection de fluorescences non spécifiques. Par la suite, le fluorophore est clivé ce qui permet l'incorporation du dNTP suivant (Figure 32).

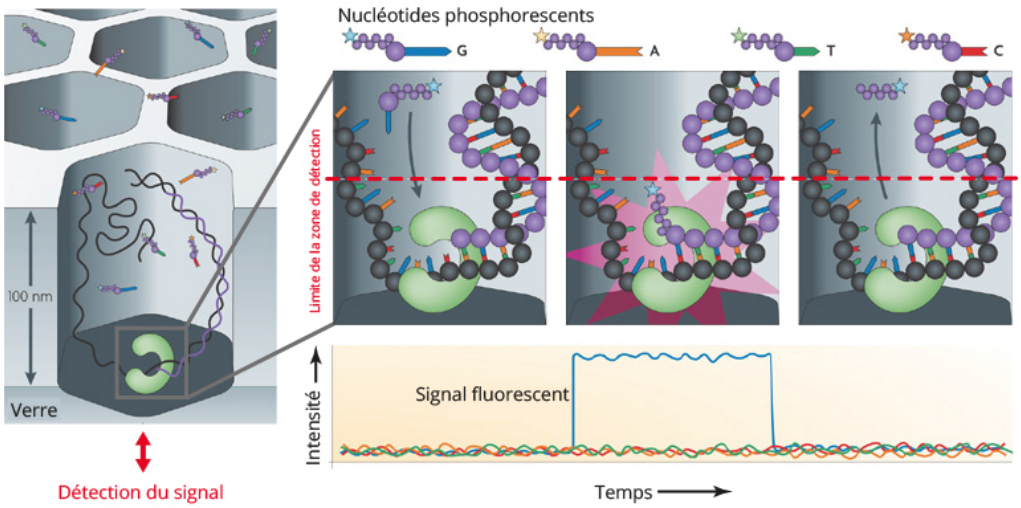


Figure 32 : Principe du séquençage PacBio

Adapté de Metzker *et al.*, Nature Reviews Genetics, 2010 [198]

Le taux d'erreur de 10 à 15% peut être contrecarré par une profondeur de lecture suffisante et l'utilisation de nouveaux algorithmes. De plus, chaque molécule d'ADN englobe plusieurs fois la même séquence grâce à une amplification PCR préalable du fragment d'ADN circularisé ce qui augmente la profondeur de lecture. Le coût de cette technologie récente reste élevé par rapport au séquençage de deuxième génération. De plus la concentration et la qualité des échantillons doivent être élevées pour obtenir des résultats concluants.

1.4.2.2. Détection par variations de courant électrique

En 2014, Oxford Nanopore Technologies propose un nouvel appareil de séquençage atypique : le MinION. Il s'agit d'un séquenceur miniature (de la taille d'une grosse clé USB et pesant moins de 100g) s'utilisant branché par port USB à un ordinateur. Malgré son petit format il permet le séquençage *de novo* de génomes entiers et ce, sans aucune phase de polymérisation de l'ADN [199]. L'appareil contient une membrane recouverte de centaines de nanopores. Un courant électrique à voltage constant est induit à travers la membrane et traverse les nanopores. Chaque fois qu'une base azotée traverse l'orifice, une altération du courant est observée. Chaque perturbation est caractéristique de l'élément spécifique qui l'a induit. La longue molécule d'ADN à séquencer est associée à une enzyme qui permet la séparation des deux brins d'ADN. Ce complexe se lie au sommet de l'orifice et, lorsque l'enzyme dissocie la double hélice, l'un des deux brins traverse l'ouverture. Le courant ionique émis est détecté au cours du temps et permet de déduire la séquence du brin d'ADN (Figure 33).

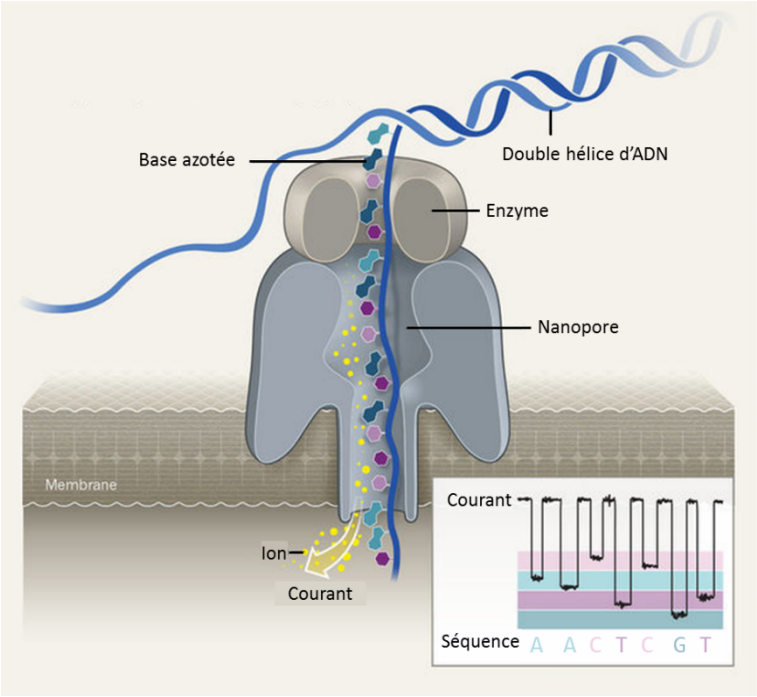


Figure 33 : Principe du séquençage par nanopore
Adapté de Perkel, Blog NaturesJobs, 2017 [200]

A l'heure actuelle, cette technologie permet d'identifier des fragments de tailles et de qualités similaires à l'approche PacBio mais le débit de données est faible. Cependant, le développement de nouveaux algorithmes permet d'augmenter la fiabilité des résultats. De plus, la quantité de données générées peut être augmentée en utilisant plusieurs appareils MinION. C'est le cas du GridION qui contient 5 séquenceurs ou encore le PromethION contenant près de 300 fois plus de pores que le MinION.

1.5. Evolution des coûts de séquençage

L'approche Sanger a permis le premier séquençage de l'Homme par le projet collaboratif « Génome Humain ». La première version a été publiée en 2001 [201] puis une version complète de 3400 Mb, englobant 20.500 gènes, a été dévoilée en 2004 [202] après plus de 10 ans de recherche par 20 équipes internationales et un budget de 3 milliards de dollars.

En parallèle, le chercheur controversé Graig Venter a fondé une compagnie privée, Celera Genomics, ayant également pour but de séquencer le génome humain mais en seulement trois ans. La séquence « brute » a été publiée en même temps que celle du consortium public international [203] puis la séquence complète a été dévoilée en 2007 [204].

En 2008, le *national human genome research institute* (NHGRI) avait pour objectif de réduire les coûts de séquençage d'un génome humain à 1000 dollars en 10 ans grâce aux

approches de séquençage de nouvelles générations [205]. Cet objectif est aujourd'hui atteint, dans certaines conditions (séquençages simultanés de plusieurs génomes par des appareils de grande capacité). De manière intéressante, cette évolution fut bien plus rapide que ce qui pouvait être estimé par des modèles mathématiques comme la Loi de Moore qui permet de prédire l'évolution du coût des nouvelles technologies au cours du temps (Figure 34).

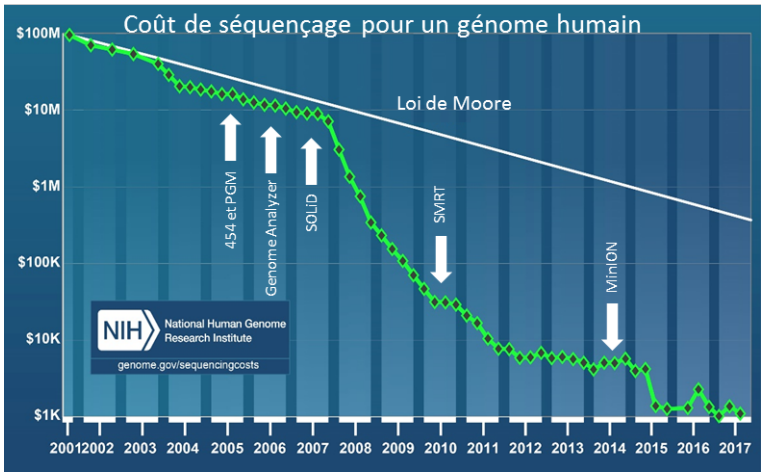


Figure 34 : Evolution du coût de séquençage pour un génome humain de 2001 à 2017

Ces approches innovantes de séquençage de l'ADN ont permis l'exploration du génome humain à la recherche de variants génétiques pouvant être impliqués dans la progression tumorale ou la réponse aux traitements anti-cancéreux.

2. La variabilité génétique inter-individuelle

2.1. Généralités

L'identification de différents génomes et l'amélioration des approches de séquençage ont permis l'identification de variations génétiques entre les individus à l'origine de la variabilité phénotypique. Ces différences peuvent être des substitutions (modifications d'une paire de bases) aussi appelées *single nucleotide polymorphisms* (SNP), des insertions ou délétions d'une ou plusieurs paires de bases (indels), des courtes répétitions de séquences ou l'insertion d'éléments transposables. Des variations plus importantes sont également décrites telles que des délétions ou des duplications de gènes entiers.

Si ces modifications surviennent dans la séquence codante d'un gène, plusieurs changements peuvent être observés dans la séquence protéique correspondante. Il peut s'agir de variations synonymes (ou silencieuses) : le codon est modifié mais l'acide aminé associé est

identique. Au contraire, il peut s'agir de variations non-synonymes associées à un changement d'acide aminé (faux-sens) voire à l'apparition d'un codon stop (non-sens). Dans ce dernier cas, le cadre ouvert de lecture est interrompu : la protéine n'est plus exprimée ou tronquée. Des insertions ou délétions peuvent également être à l'origine d'un décalage du cadre de lecture générant une séquence protéique différente.

La probabilité de survenue de chaque variant peut varier d'une population à une autre. En identifiant les variations présentes dans un groupe d'individus, la fréquence allélique peut être déterminée. La fréquence d'un allèle variant est notée fréquence de l'allèle mineur (*minor allele frequency*, MAF).

2.2. Distribution des variants selon le principe d'Hardy-Weinberg

La loi de distribution génotypique d'Hardy-Weinberg a été développée par Godfrey Hardy et Wilhem Weinberg en 1908. Ce modèle mathématique permet, dans le cas d'une population « modèle » diploïde, le calcul des fréquences génotypiques à partir des fréquences alléliques pour un locus donné (Figure 35).

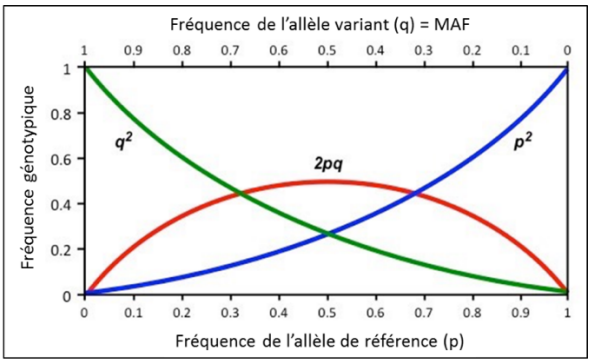


Figure 35 : Estimation des fréquences génotypiques selon les fréquences alléliques d'après le principe d'Hardy-Weinberg

Cette approche permet donc d'estimer la proportion d'individus porteurs de l'allèle variant dans une population donnée à partir des fréquences alléliques d'un polymorphisme (exemple en Figure 36). La fréquence de l'allèle variant (notée q) équivaut à la MAF. La fréquence de l'allèle de référence peut être déduite par le calcul dans le cas d'un polymorphisme biallélique (p = 1-q). Les fréquences génotypiques sont ensuite déterminées en estimant que la répartition des gamètes chez les individus respecte les lois de Mendel. Enfin, la proportion d'individus porteurs ou non de l'allèle variant peut être déduite à partir des fréquences calculées pour chaque génotype possible.

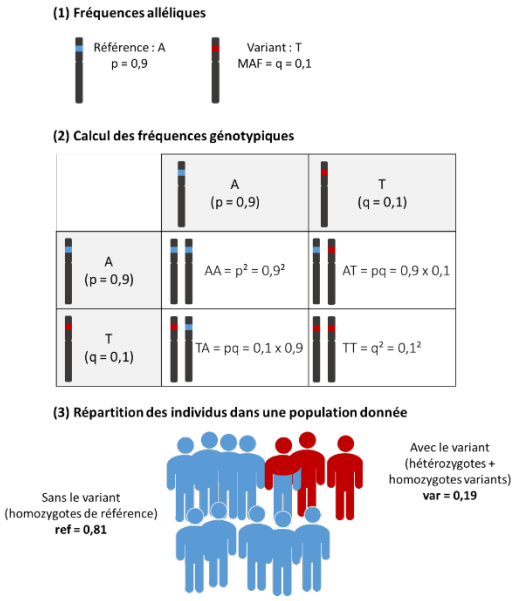


Figure 36 : Estimation du nombre d'individus porteurs de l'allèle variant dans une population donnée en appliquant le principe d'Hardy-Weinberg à partir de la MAF

2.3. Identifications de variants génétiques à grande échelle

2.3.1. Généralités

Différents projets d'identification et de caractérisation des polymorphismes génétiques chez l'Homme ont été initiés tels que le projet 1000 Genomes (1000G) ou le consortium d'agrégation des exomes (ExAC). L'ensemble des résultats obtenus sont actuellement en libre accès dans différentes bases de données disponibles en ligne telles que dbSNP.

2.3.2. 1000 Genomes

Le projet 1000G s'est déroulé entre 2008 et 2015 [206,207]. Le but était de créer le catalogue de variants et génotypes le plus complet possible. Pour cela, plusieurs milliers d'individus ont été séquencés par une approche de NGS. Les variations par rapport au génome de référence ont été détectées et les fréquences alléliques et génotypiques ont pu être déterminées. Cette initiative a permis de répertorier les MAF de 84,4 millions de variants identifiés chez 2504 individus répartis dans 26 populations. Les MAF référencées dans cette base de données ont été utilisées pour l'analyse des données au cours de ce projet de thèse.

2.3.3. ExAC et gnomAD

Le projet ExAC a pour but de regrouper et d'harmoniser les données de séquençage d'exomes publiées [208]. Ce projet a inclus 60.706 individus. Plus récemment le projet *genome aggregation database* (gnomAD) a été initié. Il suit le même principe qu'ExAC mais les régions non codantes ont également été incluses. Cette seconde base de données englobe les variations issues de 125.748 exomes et 15.708 génomes entiers.

2.3.4. dbSNP

En 1998, le centre américain d'information en biotechnologies (*national center for biotechnology information*, NCBI) met en ligne la base de données dbSNP. Elle regroupe aujourd'hui une grande partie des données obtenues par 1000G, ExAC ou gnomAD. Près de 400 millions de variations de séquences d'ADN chez l'Homme (SNP, indels et répétitions courtes en tandem) sont aujourd'hui incluses.

2.3.5. Actualisation du génome humain de référence

Le génome de référence utilisé à l'origine a été identifié par le séquençage d'un individu lambda. Cette séquence contient donc, à de nombreux locus, soit la version de référence d'un gène soit la version variante d'un gène or, une séquence modèle pour identifier de nouveaux polymorphismes ne doit inclure que les séquences les plus fréquentes dans la population étudiée. Les initiatives de séquençage à grande échelle permettent aujourd'hui au *genome reference consortium* (GRC) d'actualiser régulièrement ce génome de référence dans le but, entre autres, de n'inclure que les séquences les plus fréquemment retrouvées dans la population générale. La dernière version du génome humain, publiée en décembre 2013, est la 20ème (GRCh38, hg38). Pour ce projet de thèse, c'est la 19ème version qui a été utilisée (GRCh37, hg19). A l'heure actuelle, certains polymorphismes détectés dans les projets de séquençage présentent toujours une MAF supérieure à 50%. Par exemple, les chercheurs Koko *et al.* ont analysé la base de données ExAC et constaté que 26.537 variants étaient associés à une fréquence de l'allèle de référence <0,5 [209].

3. NGS, variants génétiques et réponse à l'imatinib

Plusieurs études ont mis en évidence des polymorphismes génétiques pouvant être associés à la résistance à l'imatinib dans la LMC par deux approches de NGS. La première approche consiste à réaliser des séquençages du génome ou de l'exome complet permettant la recherche de variants sans *a priori*. Cependant, ces méthodes présentent un coût élevé et peuvent limiter le nombre d'échantillons étudiés. De plus, des approches informatiques et statistiques robustes doivent être développées afin d'identifier les quelques polymorphismes d'intérêt parmi des milliers de variants non impliqués dans la réponse à l'imatinib. En 2016, Lavrov *et al.* séquencent les exomes complets d'échantillons sanguins de 8 patients atteints de LMC [210]. La moitié des patients sont en échec de réponse, les autres patients présentent une réponse optimale au traitement à l'imatinib selon les critères définis par l'ELN en 2013. Sept variants dans des gènes associés à la progression cancéreuse ont été mis en évidence. Récemment, cette même équipe a réalisé une nouvelle analyse de l'exome ainsi qu'une étude transcriptomique d'ARN et miARN chez des patients atteints de LMC [211]. Cette fois-ci, ils n'ont identifié aucun marqueur moléculaire associé à l'efficacité des ITK chez les patients atteints de LMC. Les effectifs étaient de 62 patients pour le séquençage, 33 pour l'analyse ARN et 12 pour l'analyse des miARN. En 2017, Kim *et al.* séquencent les exomes entiers de 13 patients atteints de LMC et suggèrent l'importance des gènes *DNMT3A* et *ASXL1* dans la réponse au traitement [212]. Plus récemment, en 2018, Shokeen *et al.* ont séquencé l'exome de 4800 gènes chez 18 patients atteints de LMC traités à l'imatinib (10 répondeurs, 8 résistants) ainsi que 5 individus contrôles [213]. Cette large étude a permis de mettre en évidence un variant dans le gène *MFGE8*, marqueur de la résistance à l'imatinib, présent chez 4 patients en échec de réponse et aucun répondeur.

La deuxième méthode consiste à ne séquencer qu'un panel de gènes. Cette technique est moins coûteuse et permet d'augmenter la cohorte d'étude, néanmoins, cette approche peut créer un biais dans l'identification des variants en limitant le nombre de gènes candidats. Cette approche a été choisie notamment par Zheng *et al.* en 2016. Ils ont séquencé 8 polymorphismes dans 4 gènes impliqués dans la régulation de l'apoptose (*FAS*, *FASLG*, *CASP10*, *APAF1*) chez 187 patients [214]. Les patients porteurs de l'allèle variant du polymorphisme rs2234767 du gène *FAS* présentent une probabilité inférieure d'atteindre une réponse moléculaire profonde (MR4.5).

Au cours de ce travail de thèse, un panel de 45 gènes associés à la régulation de l'apoptose a été séquencé par NGS à partir d'échantillons de 24 patients atteints de LMC. Un ensemble de variants génétiques a été mis en évidence en comparant les données de séquençage obtenues aux bases de données telles que dbSNP et 1000G. L'identification des polymorphismes génétiques les plus susceptibles d'être associés à la résistance à l'imatinib et la caractérisation de leur impact sur la régulation de l'apoptose ont fait l'objet de ce travail de thèse.

Objectifs

L'objectif principal de ce travail de thèse était d'identifier de nouveaux polymorphismes génétiques impliqués dans la sensibilité au traitement à l'imatinib de patients atteints de LMC. Le travail s'est focalisé sur 45 gènes impliqués dans la régulation de l'apoptose intrinsèque (21 gènes de la famille *BCL2* et 24 gènes codant des facteurs de transcription et kinases régulant les protéines de la famille BCL2). La mort cellulaire programmée étant un processus nécessaire à l'élimination des cellules cancéreuses, la présence de variants de gènes apoptotiques peut altérer l'efficacité de cette voie et causer la résistance au traitement à l'imatinib dans la LMC.

Dans un premier temps, le but de ce projet était d'identifier des variants présents parmi ces 45 gènes chez 24 patients atteints de LMC, soit répondeurs soit résistants à l'imatinib, à l'aide d'une expérience de NGS. Les données issues de ce séquençage devaient être analysées à l'aide d'approches informatiques et de nouvelles méthodes statistiques. Le but étant notamment de comparer la fréquence de survenue de l'allèle variant pour chaque polymorphisme identifié entre les patients répondeurs, et les patients résistants.

Dans un deuxième temps, l'objectif était de valider les gènes et variants candidats, plus fréquents chez les patients en échec de réponse et donc pouvant être impliqués dans la résistance au traitement. Pour cela, le génotype, l'expression des gènes et l'association avec l'évolution de la réponse à l'imatinib devait être étudié dans une plus grande cohorte de 103 patients.

Dans un troisième temps, le but était de caractériser le rôle du ou des variants candidats dans la régulation de l'apoptose et la réponse au traitement à l'aide de modèles cellulaires modifiés grâce à l'outil d'édition du génome, CRISPR-Cas9. Tout d'abord, afin de recréer spécifiquement le variant dans une lignée cellulaire modèle de LMC, l'approche de recombinaison homologue pouvait être utilisée. Ce mécanisme de réparation étant minoritaire dans les cellules humaines, l'objectif était également de développer une nouvelle stratégie, utilisant l'approche CRISPR-Cas9, afin d'induire un changement de séquences nucléotidiques de manière plus efficace. Par la suite, le rôle du ou des gènes dans la sensibilité des cellules à l'imatinib pouvait être également être étudié grâce à des lignées cellulaire n'exprimant plus, ou surexprimant, la protéine d'intérêt.

Matériel

1. Caractéristiques des patients et collecte d'échantillons

1.1. Cohorte initiale : séquençage NGS

Vingt-quatre patients atteints de LMC Ph^+ nouvellement diagnostiqués ont été inclus dans la cohorte initiale. Ces patients ont été traités à l'imatinib *de novo* 400 mg/jour. Douze patients ont été classés dans la catégorie « répondeurs », en réponse optimale selon les critères définis par l'ELN en 2013, et douze patients dans la catégorie « résistants primaires », en échec de réponse selon les mêmes critères. Aucun patient ne présentait d'altérations du domaine kinase de *BCR-ABL1* au diagnostic. Tous les patients inclus dans cette étude ont fourni un consentement éclairé selon la déclaration d'Helsinki. Le sexe ratio, l'âge médian au diagnostic et le score de Sokal sont indiqués dans le Tableau 4.

Tableau 4 : Caractéristiques des 24 patients inclus dans la cohorte initiale pour le séquençage par NGS

Caractéristiques des patients	Tous les patients	En réponse optimale	En échec
Nombre de patients (%)	24 (100)	12 (50)	12 (50)
Sexe			
Homme	15	8	7
Femme	9	4	5
Age médian au diagnostic (gamme)	59 (20-78)	61 (36-75)	57 (20-78)
Risque selon le score de Sokal : Faible/Intermédiaire/Elevé (%)	8/7/9 (33/29/38)	6/2/4 (50/17/33)	2/5/5 (17/42/42)

Des culots leucocytaires contenant 10 millions de cellules prélevées au diagnostic, en grande majorité des cellules leucémiques Ph^+, ont été obtenus pour chaque patient. Ces échantillons ont été fournis par la tumorothèque de l'hôpital Haut-Lévêque (Pessac, Bordeaux, France).

1.2. Cohortes de validation

1.2.1. Cohorte pour le séquençage Sanger

Un deuxième groupe de 103 patients (incluant les patients de la cohorte initiale) a été sélectionné pour valider les résultats obtenus par NGS. Les patients ont été sélectionnés selon les mêmes critères que la cohorte initiale (Tableau 5).

Tableau 5 : Caractéristiques des 103 patients inclus dans la cohorte de validation pour le séquençage des polymorphismes candidats par approche Sanger

ND : score de Sokal non déterminé

Caractéristiques des patients	Tous les patients	En réponse optimale	En échec
Nombre de patients (%)	103 (100)	78 (76)	25 (24)
Sexe			
Homme	61	45	16
Femme	42	33	9
Age médian au diagnostic (gamme)	59 (18-85)	61 (18-85)	56 (19-78)
Risque selon le score de Sokal : Faible/Intermédiaire/Elevé/ND (%)	34/40/21/8 (33/39/20/8)	32/29/13/4 (41/37/17/5)	2/11/8/4 (8/44/32/16)

1.2.2. Cohorte pour l'expression des gènes par PCRq en temps réel

L'expression de gènes candidats a été évaluée pour un groupe de 78 patients atteints de LMC pour lesquels des échantillons d'ARN au diagnostic étaient disponibles. Le génotype des polymorphismes d'intérêt sont connus pour ces individus (inclus dans la cohorte de 103 patients pour le séquençage Sanger).

Tableau 6 : Caractéristiques des 78 patients inclus dans la cohorte de validation pour l'étude de l'expression des gènes candidats

ND : score de Sokal non déterminé

Caractéristiques des patients	Tous les patients	En réponse optimale	En échec
Nombre de patients (%)	78 (100)	60 (77)	18 (23)
Sexe			
Homme	44	34	10
Femme	34	26	8
Age médian au diagnostic (gamme)	61 (18-85)	62 (18-83)	57 (33-85)
Risque selon le score de Sokal : Faible/Intermédiaire/Elevé/ND (%)	26/30/15/7 (33/38/19/9)	24/22/11/3 (40/37/18/5)	2/8/4/4 (11/45/22/22)

2. Lignées cellulaires

La lignée cellulaire 293T (ATCC® CRL-3216™) a été utilisée pour la mise au point de plusieurs expériences. Cette lignée a été établie à partir de cellules embryonnaires de rein puis modifiée pour permettre la surexpression d'une version mutante de l'antigène « SV40 large T » afin de faciliter les transfections. La lignée cellulaire K-562 (ATCC® CCL-243™) a été choisie comme modèle de la LMC. Cette lignée a été établie à partir de cellules issues d'un prélèvement

médullaire d'une patiente de 53 ans atteinte de LMC en crise blastique. Ces cellules portent donc le chromosome Ph et expriment le transcrit *BCR-ABL1*.

Méthodes

1. Séquençage nouvelle génération

1.1. Extraction de l'ADN génomique

L'ADN génomique (ADNg) pour l'approche NGS a été extrait suivant le protocole du kit DNA extraction (Agilent Technologies). Les cellules sont lysées dans une solution contenant du Triton X-100 et du sucrose et incubées 2 min dans la glace. Après centrifugation 15 min à 350 g à 4°C, le surnageant est éliminé. Le culot est repris dans une solution contenant du dodécylsulfate de sodium (SDS) et une pronase (100 µg/mL) afin de digérer les protéines en acides aminés. Le mélange est incubé toute la nuit à 37°C sous agitation. Le lendemain, les échantillons sont placés 10 min dans la glace puis une solution contenant du chlorure de sodium (NaCl) est ajoutée afin de précipiter les protéines. Le tout est incubé 5 min dans la glace puis centrifuger 15 min à 2000 g à 4°C. Le surnageant contenant les acides nucléiques est donc récupéré. Par la suite, une RNase à 20 µg/mL est ajouté afin d'éliminer les ARN. Le mélange est incubé 15 min à 37°C. L'ADNg obtenu est précipité par ajout de deux volumes d'éthanol absolu et le filament d'ADN est récupéré à l'aide d'un cône de pipette. L'ADN est lavé à l'éthanol 70% puis séché à température ambiante et repris dans 50 µL d'eau distillée UltraPure™ sans DNase ni RNase (eau distillée ultra pure, Thermo Fisher Scientific).

1.2. Contrôle qualité de l'ADNg par électrophorèse sur gel d'agarose

L'ADNg extrait pour le NGS ne doit pas être fragmenté. Une électrophorèse sur gel d'agarose faiblement concentré est réalisée pour contrôler l'intégrité des molécules extraites. Pour cela, de l'agarose en poudre est dissoute dans du tampon TBE 0,5X (Tris 44,5 mM, Acide borique 44,5 mM, EDTA 1 mM) (agarose 0,5% m/v). Le tout est chauffé jusqu'à l'obtention d'un mélange homogène. Après refroidissement, du bromure d'éthidium (BET), un intercalant de l'ADN (0,5 µg/mL) est ajouté à la solution. Le gel est coulé dans un support adapté puis, après polymérisation de l'agarose, le gel est placé dans une cuve remplie de tampon TBE 0,5X.

Les échantillons sont mélangés à un tampon de charge (bleu de bromophénol 0,025 % m/v, glycérol 50 % v/v, EDTA 1 mM) puis déposés dans les puits du gel. Un contrôle positif contenant de l'ADN de phage λ non dégradé et de poids moléculaire connu (50 kb) a également été déposé. Le tout migre sous un champ électrique de 100 volts pendant environ 30 min. Enfin, la fluorescence émise par le BET est détectée à l'aide d'un transilluminateur à ultraviolets (Gel logic 200 imaging system, Kodak) (Figure 37).

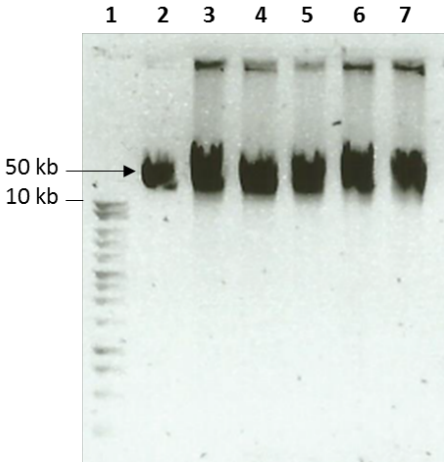


Figure 37 : Image d'une électrophorèse sur gel d'agarose révélée sous ultraviolets

Puits 1 : échelle de taille de 250 pb à 10 kb (intervalles de 1 kb), puits 2 : ADNg de phage λ non fragmenté de 50 kb, puits 3 à 7 : ADNg extrait à partir de cellules prélevées au diagnostic chez cinq patients différents

Tous les signaux présentent un haut poids moléculaire et aucune fragmentation de l'ADNg n'est détectée (absence de signaux de poids moléculaires plus faibles). Cette première analyse confirme que les ADNg sont de qualité suffisante pour la suite de l'expérience.

1.3. Conception de panels de gènes

Un total de 95 gènes, répartis en trois groupes, a été séquencé par NGS afin d'identifier de nouveaux variants génétiques pouvant être impliqués dans la résistance à l'imatinib. Un premier panel de 48 gènes a tout d'abord été établi afin de tester de nouvelles approches statistiques développées au cours de cette thèse pour l'analyse des résultats de NGS (ce travail a fait l'objet d'une publication soumise à Human Genomics, Annexe 1). Il s'agit de gènes impliqués dans la pharmacocinétique (n=22), dans la réparation de l'ADN et la régulation du cycle cellulaire (n=15), dans les voies des cytokines (n=6) et des gènes codant des kinases ou phosphatases (n=5) (Tableau 7).

Un deuxième panel de 45 gènes associés à la régulation de l'apoptose a été conçu. Ce dernier inclut 21 gènes de la famille *BCL2* et 24 gènes, codant des kinases ou facteurs de transcription, associés à la régulation des facteurs de la famille BCL2 (

Tableau 8). Le travail de thèse s'est focalisé sur l'étude de ces 45 gènes. En complément, les gènes *BCR* et *ABL1* ont également été séquencés par NGS (Tableau 9).

1.4. Conception des sondes de captures

Afin de séquencer spécifiquement ces gènes par NGS, des sondes de capture ont été conçues à l'aide du logiciel SureDesign (Agilent Technologies). Il s'agit d'oligonucléotides de 120 pb, complémentaires des exons codants, des jonctions intron-exon et des régions promotrices des gènes d'intérêt. Chaque séquence à capturer a été élargie de 10 pb de part et d'autre en 5' et 3'.
Au final, 2236 sondes ont été conçues pour capturer les régions cibles des 48 gènes du panel test (Tableau 7), 2017 sondes pour le panel de gènes impliqués dans la régulation de l'apoptose (

Tableau 8) et 279 sondes pour cibler les gènes *BCR* et *ABL1* (Tableau 9). Dans le panel de gènes apoptotiques, le pourcentage de couverture moyen estimé était de 98,5%. Les deux gènes les moins bien couverts (estimation) étaient *BIM/BCL2L11* avec un taux de capture estimé à 86,9% et *BCL2L13/BCL-RAMBO* avec un taux de 88%. Ces sondes sont biotinylées pour permettre la capture des fragments hybridés par des colonnes streptavidine. Les séquences ont été conçues à partir des données disponibles en ligne en date du 22 octobre 2014.

Tableau 7 : Caractéristiques des 48 gènes inclus dans le panel test

Symbole du gène	Nom détaillé du gène	Autres symboles	Localisation chromosomique	Intervalle du gène (hg19)	Nombre de régions ciblées	Taille des régions ciblées (pb)	Couverture attendue (%)	Régions couverture élevée	Régions couverture faible
Protéines plasmatiques, transporteurs membranaires et régulateurs (n=10)									
ABCB1	ATP binding cassette subfamily B member 1	*MDR1, PGY1, CLCS*	7q21.12	chr7:87133549-87342649	32	5021	98,6	30	2
ABCC2	ATP binding cassette subfamily C member 2	*CMOAT*	10q24.2	chr10:101542453-101611398	32	5587	100	32	0
ABCG2	ATP binding cassette subfamily G member 2 (Junior blood group)	*MXR, BCRP, ABCP, CD338*	4q22.1	chr4:89013376-89152484	18	3238	100	18	0
HFE	homeostatic iron regulator	*HLA-H*	6p22.2	chr6:26087499-26096835	8	1855	100	8	0
HIF1A	hypoxia inducible factor 1 subunit alpha	*MPO1, PASD8, bHLHE78*	14q23.2	chr14:62162109-62213813	16	3541	100	16	0
ORM1	orosomucoid 1		9q32	chr9:117085293-117088647	5	1331	100	5	0
SLC22A1	solute carrier family 22 member 1	*OCT1*	6q25.3	chr6:160542795-160579760	11	2213	100	11	0
SLC22A4	solute carrier family 22 member 4	*OCTN1, DFNB60*	5q31.1	chr5:131630126-131679538	13	2549	100	13	0
SLCO1A2	solute carrier organic anion transporter family member 1A2	*SLC21A3, OATP*	12p12.1	chr12:21422472-21572528	21	4194	87,2	18	3
SLCO1B1	solute carrier organic anion transporter family member 1B1	*SLC21A6, OATP-C, LST-1*	12p12.1	chr12:21284118-21392133	15	2480	100	15	0
Enzymes du métabolisme et régulateurs (n=12)									
CYP1A1	cytochrome P450 family 1 subfamily A member 1	*CYP1, P450DX, CP11, P1-450*	15q24.1	chr15:75012820-75017961	7	1878	100	7	0
CYP1A2	cytochrome P450 family 1 subfamily A member 2	*CP12, P3-450*	15q24.1	chr15:75041174-75047439	7	1755	100	7	0
CYP2C19	cytochrome P450 family 2 subfamily C member 19	*CYP2C, CPCJ*	10q23.33	chr10:96447901-96612840	15	3856	100	15	0
CYP2C8	cytochrome P450 family 2 subfamily C member 8	*CPC8*	10q23.33	chr10:96796875-96829264	10	2072	97,4	9	1
CYP2C9	cytochrome P450 family 2 subfamily C member 9	*CYP2C10*	10q23.33	chr10:96698405-96749320	10	2959	99,7	10	0
CYP2D6	cytochrome P450 family 2 subfamily D member 6	*CPD6, P450-DB1, CYP2D*	22q13.2	chr22:42522566-42540482	11	1930	100	11	0
CYP3A4	cytochrome P450 family 3 subfamily A member 4	*CYP3A3*	7q22.1	chr7:99355746-99381898	14	2082	100	14	0
CYP3A5	cytochrome P450 family 3 subfamily A member 5	*PCN3, CP35*	7q22.1	chr7:99245918-99277631	15	2276	100	15	0
NR1I2	nuclear receptor subfamily 1 group I member 2	*ONR1, PXR, BXR, SXR, PAR2*	3q13.33	chr3:119499321-119537341	10	5580	100	10	0
NR1I3	nuclear receptor subfamily 1 group I member 3	*MB67, CAR1*	1q23.3	chr1:161199462-161208102	10	1836	100	10	0
UGT1A1	UDP glucuronosyltransferase family 1 member A1	*UGT1A, GNT1*	2q37.1	chr2:234668884-234681215	6	1793	100	6	0
UGT1A9	UDP glucuronosyltransferase family 1 member A9	*UGT1AI, HLUGP4, LUGP4*	2q37.1	chr2:234580489-234681215	5	1775	100	5	0
Cycle cellulaire et prolifération (n=5)									
CCND1	cyclin D1	*BCL1, PRAD1, U21B31*	11q13.3	chr11:69455845-69466185	5	1377	100	5	0
PPP2R2A	protein phosphatase 2 regulatory subunit Balpha	*PR52A, B55A*	8p21.2	chr8:26148997-26228655	17	6615	98,4	16	1
RPA1	replication protein A1	*REPA1, RPA70, HSSB, RF-A, RP-A*	17p13.3	chr17:1733256-1800479	17	2313	100	17	0
RPA2	replication protein A2		1p35.3	chr1:28218664-28241267	8	1540	100	8	0
RPA3	replication protein A3	*REPA3*	7p21.3	chr7:7676621-7758248	9	1840	100	9	0

Réparation de l'ADN (n=10)									
ERCC2	ERCC excision repair 2, TFIIH core complex helicase subunit	*XPD, MAG, EM9, TFIIH*	19q13.32	chr19:45854877-45873886	23	3177	97,7	22	1
ERCC3	ERCC excision repair 3, TFIIH core complex helicase subunit	*XPB, BTF2, RAD25, TFIIH, GTF2H*	2q14.3	chr2:128015162-128051762	15	3023	100	15	0
ERCC4	ERCC excision repair 4, endonuclease catalytic subunit	*XPF, RAD1, FANCQ*	16p13.12	chr16:14014004-14042214	13	3164	98,8	12	1
ERCC5	ERCC excision repair 5, endonuclease	*XPGC, ERCM2*	13q33.1	chr13:103459695-103528263	27	8572	97,1	26	1
ERCC6	ERCC excision repair 6, chromatin remodeling factor	*CKN2, CSB, RAD26, ARMD5*	10q11.23	chr10:50666851-50747594	26	7377	100	26	0
ERCC8	ERCC excision repair 8, CSA ubiquitin ligase complex subunit	*CKN1, CSA*	5q12.1	chr5:60170432-60240915	14	1801	100	14	0
LIG1	DNA ligase 1		19q13.33	chr19:48618896-48673611	29	3599	100	29	0
RAD23B	RAD23 homolog B. nucleotide excision repair protein	*HHR23B, P58*	9p31.2	chr9:110045408-110094013	13	2391	100	13	0
XPA	XPA, DNA damage recognition and repair factor	*XPAC, XP1*	9p22.33	chr9:100437711-100459701	6	1059	100	6	0
XPC	XPC complex subunit, DNA damage recognition and repair factor	*XPCC, RAD4*	3p25.1	chr3:14187431-14220293	16	3358	100	16	0
Voies des cytokines (n=6)									
CXCL8	C-X-C motif chemokine ligand 8	*IL8*	4q13.3	chr4:74606213-74608191	3	788	100	3	0
IFNG	interferon gamma		12q15	chr12:68549123-68553537	4	713	100	4	0
IFNGR1	interferon gamma receptor 1	*IFNGR, CD119*	6q23.3	chr6:137519158-137540596	9	1992	100	9	0
IFNGR2	interferon gamma receptor 2	*IFNGT1, AF-1*	21q22.11	chr21:34757289-34851120	12	2252	100	12	0
SOCS1	suppressor of cytokine signaling 1	*SSI-1, JAB, TIP3, Cish1*	16p13.13	chr16:11348690-11350049	2	780	100	2	0
SOCS2	suppressor of cytokine signaling 2	*SSI-2, STATI2, Cish2*	12q	chr12:93963588-93969230	6	2740	100	6	0
Kinases et phosphatases (n=5)									
AKT1	AKT serine/threonine kinase 1	*RAC, PKB, PRKBA, AKT*	14q32.33	chr14:105236668-105262090	16	3614	92,3	15	1
ULK3	unc-51 like kinase 3	*FLJ90566*	15q24.1	chr15:75129559-75135562	17	2003	100	17	0
PTPN1	protein tyrosine phosphatase non-receptor type 1	*PTP1B*	20q13.13	chr20:49126848-49199262	10	1715	100	10	0
PTPN2	protein tyrosine phosphatase non-receptor type 2	*TCPTP*	18p11.21	chr18:12785812-12884347	12	1867	100	12	0
PTPN22	protein tyrosine phosphatase non-receptor type 22	*PTPN8, Lyp*	1p13.2	chr1:114357517-114414391	21	3032	100	21	0

Le symbole, le nom détaillé, les autres symboles et la localisation chromosomique de chaque gène sont extraits du comité de nomenclature des gènes de l'organisation du génome humain (HGNC) [215]. La version du génome humain de référence utilisée est hg19. L'intervalle du gène, le nombre de régions ciblées et leur taille ainsi que les couvertures attendues ont été déterminés à l'aide des bases de données Gencode, RefSeq et VEGA.

Une couverture élevée correspond >=90% des positions complémentaires d'une sonde de capture et à l'inverse, une couverture faible correspond à <90% des positions capturées.

Tableau 8 : Caractéristiques des 45 gènes associés à la régulation de l'apoptose

Symbole du gène	Nom détaillé du gène	Autres symboles	Localisation chromosomique	Intervalle du gène (hg19)	Nombre régions ciblées	Taille régions ciblées (pb)	Couverture attendee (%)	Régions couverture élevée	Régions couverture faible
Famille *BCL2* (n=21)									
***BCL2*-like (anti-apoptotiques) (n=6)**									
BCL2	BCL2 apoptosis regulator	*PPP1R50*	18q21.33	chr18:60795848-60987371	2	2255	100,0	2	0
BCL2A1	BCL2 related protein A1	*BCL2L5, A1, GRS, BFL1, ACC1, ACC2*	15q25.1	chr15:80253399-80263798	3	971	100,0	3	0
BCL2L1	BCL2 like 1	*BCLX, BCL2L, PPP1R52*	20q11.21	chr20:30253742-30311802	3	1897	100,0	3	0
BCL2L10	BCL2 like 10	*BCL-B, Boo, Diva*	15q21.2	chr15:52402035-52404982	2	704	100,0	2	0
BCL2L2	BCL2 like 2	*BCL-W, KIAA0271, PPP1R51*	14q11.2	chr14:23775961-23778184	4	944	100,0	4	0
MCL1	MCL1 apoptosis regulator, BCL2 family member	*BCL2L3*	1q21.2	chr1:150549830-150552224	3	1332	100,0	3	0
BH3-only (pro-apoptotiques, initiatrices) (n=8)									
BAD	BCL2 associated agonist of cell death	*BCL2L8, BBC2*	11q13.1	chr11:64037398-64052186	5	1254	100,0	5	0
BBC3	BCL2 binding component 3	*PUMA, JFY1*	19q13.32	chr19:47724948-47736033	5	1316	95,1	4	1
BCL2L11	BCL2 like 11	*BIM, BOD*	2q13	chr2:111876945-111921818	14	3020	87,0	11	3
BID	BH3 interacting domain death agonist		22q11.21	chr22:18218336-18257441	9	1495	100,0	9	0
BIK	BCL2 interacting killer	*NBK*	22q13.2	chr22:43506744-43525321	5	646	100,0	5	0
BMF	Bcl2 modifying factor	*FLJ00065*	15q15.1	chr15:40383978-40401095	6	1011	100,0	6	0
HRK	harakiri, BCL2 interacting protein	*DP5*	12q24.22	chr12:117318827-117319122	1	296	100,0	1	0
PMAIP1	phorbol-12-myristate-13-acetate-induced protein 1	*NOXA, APR*	18q21.32	chr18:57567170-57570090	3	701	100,0	3	0
***BAX*-like (pro-apoptotiques, effectrices) (n=3)**									
BAX	BCL2 associated X protein, apoptosis regulator	*BCL2L4*	19q13.33	chr19:49458062-49465065	4	2464	95,2	3	1
BAK1	BCL2 antagonist/killer 1	*BCL2L7, BAK*	6p21.31	chr6:33541570-33548080	8	1227	100,0	8	0
BOK	BCL2 family apoptosis regulator BOK	*BCL2L9, BOKL, MGC4631*	2q37.3	chr2:242498126-242511847	5	1041	100,0	5	0
Autres (fonctions non définies) (n=4)									
BCL2L12	BCL2 like 12		19q13.33	chr19:50168389-50177183	7	2044	100,0	7	0
BCL2L13	BCL2 like 13	*BCL-RAMBO, MIL1*	22q11.21	chr22:18111611-18213631	12	8176	88,1	9	3
BCL2L14	BCL2 like 14	*BCL-G*	12p13.2	chr12:12202788-12291899	15	4043	93,6	14	1
BCL2L15	BCL2 like 15	*Bfk, FLJ22588*	1p13.2	chr1:114423709-114430179	4	770	100,0	4	0

Régulateurs de la survie cellulaire (n=24)									
Kinases (n=17)									
BLK	BLK proto-oncogen, Src family tyrosine kinase	*MGC10442*	8p23.1	chr8:11351500-11422118	15	4441	99,1	15	0
CLK2	CDC like kinase 2		1q22	chr1:155232999-155247511	14	2269	100,0	14	0
DYRK2	dual specificity tyrosine phosphorylation regulated kinase 2		12q15	chr12:68042491-68056454	4	6450	100,0	4	0
EEF2K	eukaryotic elongation factor 2 kinase	*CaMKIII*	16p12.2	chr16:22217582-22295327	18	3023	100,0	18	0
GRK2	G protein-coupled receptor kinase 2	*ADRBK1, BARK1*	11q13.2	chr11:67033871-67054005	21	3003	93,7	20	1
MAPK1	mitogen-activated protein kinase 1	*PRKM1, PRKM2, ERK, ERK2, MAPK2, p41mapk*	22q11.22	chr22:22123483-22221980	8	1483	100,0	8	0
MAPK14	mitogen-activated protein kinase 14	*PRKM14, PRKM15, Mxi2, p38*	6p21.31	chr6:35995444-36076234	15	4371	100,0	15	0
MAPK3	mitogen-activated protein kinase 3	*PRKM3, ERK1, p44erk1, p44mapk*	16p11.2	chr16:30127979-30134640	8	2419	98,8	8	0
MAPK8	mitogen-activated protein kinase 8	*PRKM8, JNK, JNK1, SAPK1*	10q11.22	chr10:49514672-49644493	14	6989	96,7	13	1
MARK2	microtubule affinity regulating kinase 2	*EMK1, PAR-1, PAR-1B*	11q13.1	chr11:63606390-63676719	21	3900	98,8	20	1
PAK4	p21 (RAC1) activated kinase 4		19q13.2	chr19:39616410-39669229	11	2474	95,7	10	1
PRKACA	Protein kinase cAMP-activated catalytic subunit alpha	*PKACa*	19p13.1	chr19:14203914-14228584	12	2075	100,0	12	0
RIPK3	receptor interacting serine/threonine kinase 3	*RIP3*	14q12	chr14:24805371-24809252	10	1967	100,0	10	0
RPS6KA1	ribosomal protein S6 kinase A1	*RSK1, HU-1, p90Rsk*	1p36.11	chr1:26856239-26900944	28	4328	97,6	27	1
RPS6KB1	ribosomal protein S6 kinase B1	*S6K1, p70(S6K)-alpha, PS6K*	17q23.1	chr17:57970397-58024973	19	3723	97,9	18	1
STK11	serine/threonine kinase 11	*PJS, LKB1*	19p13.3	chr19:1205788-1226656	9	2597	100,0	9	0
TNNI3K	TNNI3 interacting kinase	*CARK*	1p31.1	chr1:74663912-75010122	33	6174	100,0	33	0
Régulateurs de la transcription (n=7)									
BCLAF1	BCL2 associated transcription factor 1	*BTF, KIAA0164*	6q23.3	chr6:136582238-136610999	14	3366	100,0	14	0
FOXO3	forkhead box O3	*AF6q21, FOXO2*	6q21	chr6:108881016-108986102	5	2995	100,0	5	0
RUNX1	runt related transcription factor 1	*AML1, CBFA2, PEBP2A2, AMLCR1*	21q22.12	chr21:36164266-36421651	12	3998	100,0	12	0
STAT3	signal transducer and activator of transcription 3	*APRF*	17q21.2	chr17:40467753-40540523	24	3033	100,0	24	0
STAT5A	signal transducer and activator of transcription 5A	*MGF*	17q21.2	chr17:40439555-40462697	21	4085	99,8	21	0
STAT5B	signal transducer and activator of transcription 5B		17q21.2	chr17:40353746-40428735	20	3013	100,0	20	0
TP53	tumor protein 53	*p53, LFS1*	17p13.1	chr17:7565247-7590878	15	2403	95,7	14	1

Tableau 9 : Caractéristiques des gènes *BCR* et *ABL1*

Symbole du gène	Nom détaillé du gène	Autres symboles	Localisation chromosomique	Intervalle du gène (hg19)	Nombre de régions ciblées	Taille des régions ciblées (pb)	Couverture attendue (%)	Régions couverture élevée	Régions couverture faible
ABL1	ABL proto-oncogene 1, non-receptor tyrosine kinase	*ABL, c-ABL, p150, JTK7*	9p34.12	chr9:133589258-133761080	14	4650	100	14	0
BCR	BCR activator of RhoGEF and GTPase	*CML, PHL, ALL, D22S662*	22q11.23	chr22:23521881-23660231	30	12989	99,1	30	0

1.5. Génération de la librairie de séquençage

La librairie de séquençage a été réalisée en collaboration avec la plateforme génome et transcriptome de Bordeaux (PGTB) suivant le protocole du kit SureSelect^{QXT} Target Enrichment for Illumina Multiplexed Sequencing (Agilent Technologies).

La première étape est la tagmentation, c'est-à-dire la fragmentation de l'ADNg et l'ajout simultané d'adaptateurs universels aux extrémités des fragments d'ADNg à l'aide d'une transposase. Deux microlitres de chaque échantillon d'ADNg, dilué à 25 ng/µL, sont mélangés à 17 µL de tampon (SureSelect QXT Buffer) et 2 µL d'enzyme complexée aux adaptateurs (SureSelect QXT Enzyme ILM). La réaction de tagmentation est réalisée dans un thermocycleur pendant 10 min à 45°C suivie d'une incubation d'une minute à 4°C. La fragmentation est stoppée par l'ajout d'une solution de dénaturation de l'enzyme contenant notamment du thiocyanate de guanidium (SureSelect QXT Stop Solution). Les fragments d'ADN sont purifiés à l'aide de billes magnétiques recouvertes de groupements carboxyl. Brièvement, les fragments d'ADN sont mélangés à des billes magnétiques (Agencourt AMPure XP beads, Beckman Coulter) à un ratio 1:1 et incubés 5 min à température ambiante. Les tubes sont placés sur un support magnétique. Après quelques minutes, lorsque la solution est claire, le surnageant est éliminé. Les billes sont ensuite lavées deux fois, toujours sur le support magnétique, en ajoutant 200 µL d'éthanol 70% dans chaque tube. Les traces d'éthanol sont éliminées en incubant les échantillons au moins 1 min à 37°C. L'ADN est élué en ajoutant 11 µL d'eau dépourvue de nucléases pour chaque échantillon. Après mélange et 2 min d'incubation à température ambiante, les tubes sont de nouveau placés sur le support magnétique. Après quelques minutes, le surnageant contenant les fragments d'ADN purifiés est récupéré.

Les fragments purifiés sont ensuite amplifiés par une réaction de PCR à l'aide de l'ADN polymérase Herculase II Fusion (Agilent Technologies) en présence de diméthylsulfoxyde (DMSO) (Tableau 10). Les amorces reconnaissent spécifiquement les adaptateurs ajoutés lors de la première étape.

Tableau 10 : Mélange réactionnel et programme de la première PCR permettant l'amplification des fragments d'ADN avec adaptateurs

Réactifs	Volume pour une réaction (µL)
Tampon Herculase II (5X)	10
Mélange équimolaire des 4 dNTP (25 mM chaque)	0,5
DMSO	2,5
Amorces SureSelect QXT (sens, anti-sens)	1
ADN polymérase Herculase II Fusion	1
ADN purifié	10
Eau dépourvue de nucléases	Qsp 50 µL

Nombre de cycles	Température (°C)	Durée
1	68	2 min
1	98	2 min
8	98	30 sec
	57	30 sec
	72	1 min
1	72	5 min

Les produits PCR sont purifiés à l'aide de billes magnétiques (Agencourt AMPure XP beads, Beckman Coulter) et élués dans 13 µL d'eau dépourvue de nucléases. La taille des produits PCR purifiés, la concentration et la qualité des échantillons sont évaluées par électrophorèse à l'aide d'une puce DNA 1000 dans un appareil 2100 Bioanalyzer (Agilent Technologies) (exemple en Figure 38). La majorité des amplicons doivent être d'environ 300 pb. Les 24 échantillons sont dilués afin d'homogénéiser les concentrations (500 ng/12 µL par échantillon).

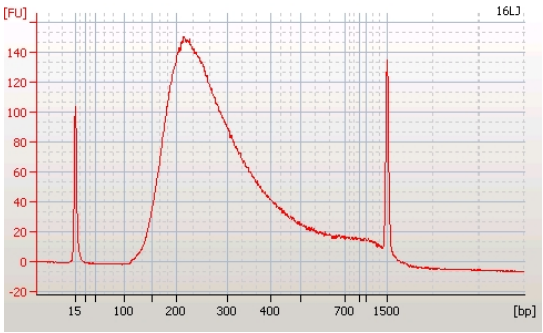


Figure 38 : Exemple de profil de dosage obtenu au BioAnalyzer, après la première PCR de la création des librairies, représentant l'intensité d'absorbance du signal en fonction de la taille des fragments

Pour éliminer une contamination éventuelle des amorces issues de la réaction de PCR, un réactif est ajouté (SureSelect QXT Fast Blocker Mix). Les fragments d'ADN sont ensuite dénaturés 5 min à 95°C puis incubés pendant 10 min à 65°C. Pendant ce temps, les sondes de capture biotinylées en présence des réactifs d'hybridation sont ajoutées au mélange. Le programme d'hybridation est réalisé dans un thermocycleur (Tableau 11).

Tableau 11 : Mélange réactionnel et programme d'hybridation des sondes biotinylés aux fragments d'ADN

Réactifs	Volume pour une réaction (µL)
Solution RNase Block (25%)	2
SureSelectXT Custom Capture Library	2
Tampon SureSelect QXT Fast Hybridization	6
ADN + Fast Blocker	17
Eau dépourvue de nucléases	qsp 30 µL

Nombre de cycles	Température (°C)	Durée
1	65	1 min
60	65	1 min
	37	3 sec
1	65	infini

Les fragments d'intérêt sont capturés *via* l'interaction forte des sondes biotinylées à la streptavidine fixée sur des billes magnétiques (Dynabeads MyOne Streptavidin T1 magnetic beads, Thermo Fisher Scientific). Les produits de PCR-sondes biotinylées sont mélangés aux billes, à un ratio 1:8 et incubés 30 min à température ambiante sous agitation. Les tubes sont placés sur un support magnétique et, après quelques minutes, le surnageant est éliminé. Le tout est ensuite lavé avec 200 µL de tampon (Wash Buffer 1). Après mélange, les tubes sont de nouveau placés sur un support magnétique puis le surnageant est éliminé. Les échantillons sont alors lavés trois fois en ajoutant 200 µL d'un deuxième tampon maintenu à 65°C (Wash Buffer 2) puis incubés 10 min à 65°C. Chaque échantillon est élué dans 23 µl d'eau dépourvue de nucléases. Les billes sont toujours associées aux sondes et donc aux fragments d'ADN capturés.

Enfin, des couples d'index, différents pour chaque échantillon, sont ajoutés aux fragments d'ADN par PCR. Pour cela des amorces divisées en trois parties sont utilisées. Elles sont composées de séquences étiquettes universelles (P7 et P5) utilisées pour la liaison des produits PCR à la *flow-cell*, des index et de séquences complémentaires aux adaptateurs universels ajoutés aux fragments d'ADNg à l'étape de tagmentation (Figure 39 et Tableau 12).

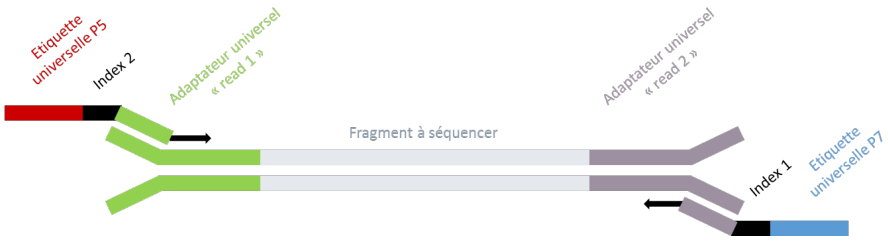


Figure 39 : Schéma de deuxième PCR permettant l'ajout des index et des étiquettes universelles P5 et P7
Les flèches indiquent le sens d'élongation par la polymérase.

Tableau 12 : Mélange réactionnel et programme d'amplification de la deuxième PCR permettant l'ajout d'index aux fragments d'ADN

Réactifs	Volume pour une réaction (µL)
Tampon Herculase II (5X)	10
Mélange équimolaire des 4 dNTP (25 mM chaque)	0,5
ADN polymérase Herculase II Fusion	1
ADN purifié + billes	23
Amorce P7	1
Amorce P5	1
Eau dépourvue de nucléases	qsp 50 µL

Nombre de cycles	Température (°C)	Durée
1	98	2 min
14	98	30 sec
	58	30 sec
	72	1 min
1	72	5 min

Pendant la PCR, les sondes biotinylées associées aux billes magnétiques sont dissociées des fragments d'ADN grâce aux étapes de dénaturation à 98°C. Les billes magnétiques streptavidine sont ensuite isolées en plaçant les échantillons sur un support magnétique à l'issue de la PCR et en récupérant le surnageant. Par la suite, un nouveau cycle de purification des échantillons à l'aide de billes magnétiques (Agencourt AMPure XP beads, Beckman Coulter) est réalisé. Les échantillons sont ensuite élués dans 25 µl d'eau dépourvue de nucléases. Les produits de PCR (librairies) sont dosés à l'aide d'une puce DNA High Sensitivity 1000 dans un appareil 2100 Bioanalyzer (Agilent Technologies) (exemple en Figure 40), puis mélangés à concentration égale avant séquençage à une concentration finale de 4 nM. La majorité des fragments doivent être d'environ 350-400 pb.

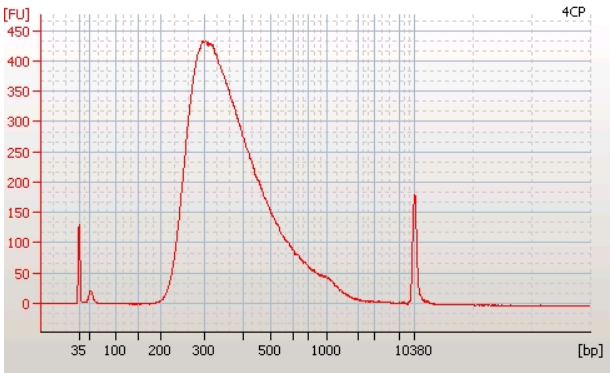


Figure 40 : Exemple de profil de dosage obtenu au BioAnalyzer, après la deuxième PCR de la création des librairies, représentant l'intensité d'absorbance du signal en fonction de la taille des fragments

Juste avant le séquençage, les fragments d'ADN (2 nM) sont dénaturés par ajout de soude (NaOH) à 0,1 N et incubation 5 min à température ambiante. La réaction de dénaturation est stoppée par ajout du tampon HT1 froid (tampon d'hybridation, Illumina). L'ajout de tampon permet également la dilution des librairies à 10 pM. Enfin, les librairies sont mélangées à de l'ADN de PhiX 10 pM (Illumina) (90% librairie, 10% PhiX).

1.6. Réaction de séquençage

Les fragments d'ADN sont séquencés à partir des deux extrémités (*paired-end*), sur 2x150 pb, par un séquenceur MiSeq (Illumina), avec une profondeur de couverture moyenne estimée à 100X. Deux cassettes de séquençage MiSeq Micro utilisant la chimie v2 (Illumina) ont été utilisées.

Brièvement, les fragments des librairies sont liés à la lame (*flow-cell*) par complémentarité de leurs étiquettes universelles P5 et P7. Dans un premier temps, après formation des clusters par amplification en pont (Introduction, Partie III, 1.3.2.2), les séquences d'ADN liées par leurs adaptateurs P5 sont dissociées de la lame. Ainsi, tous les fragments d'un même cluster sont issus d'un même brin d'ADN. Les séquences d'ADN des fragments sont lues à partir d'une amorce complémentaire de l'adaptateur universel « read 1 » : une première séquence de 150 pb, le « read 1 », est obtenue. L'index 1 de 8 pb, spécifique à chaque patient, est également détecté à partir d'une amorce s'hybridant à l'adaptateur universel « read 2 ». (Figure 41). Dans un second temps, une nouvelle étape d'amplification en pont et de dissociation des fragments liés par leurs adaptateurs P7 permet la lecture de chaque élément dans le sens inverse selon un principe similaire (Figure 41).

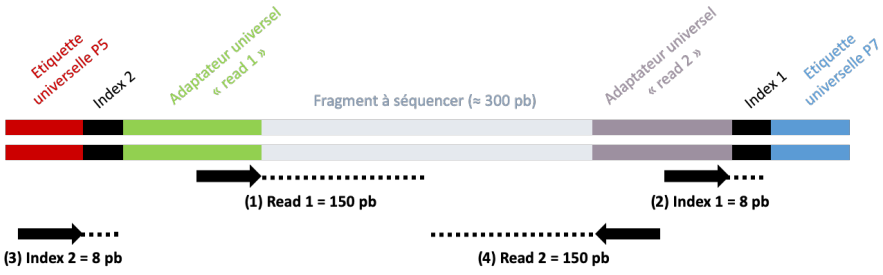


Figure 41 : Schéma du séquençage « paired-end »

Quatre étapes de lecture se succèdent afin d'obtenir la séquence du fragment à partir des deux extrémités ainsi que les index spécifiques de chaque échantillon.

2. Analyses informatiques des données de NGS

Une analyse des séquences obtenues par NGS a été réalisée en coopération avec le Centre de Bio-informatique de Bordeaux (CBiB) puis grâce à l'utilisation d'outils en ligne (usegalaxy.org [216] et avia-abcc.ncifcrf.gov [217]) ainsi que du logiciel R (Figure 42).

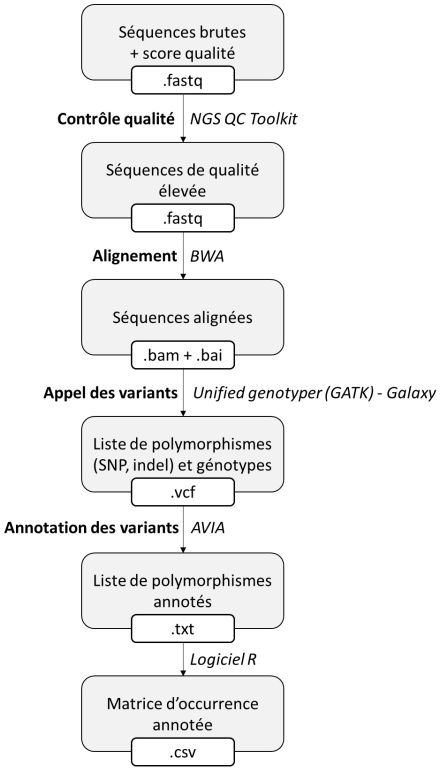


Figure 42 : Schéma global des différentes étapes d'analyses des données NGS

2.1. Données brutes issues du séquençage

A l'issu du séquençage des listes de séquences de 150 pb (reads) sont obtenues. Pour chaque base lue par l'appareil, un score de confiance est attribué (score Phred = qualité de la base). Ces données sont regroupées dans des fichiers .fastq (Exemple en Figure 43).

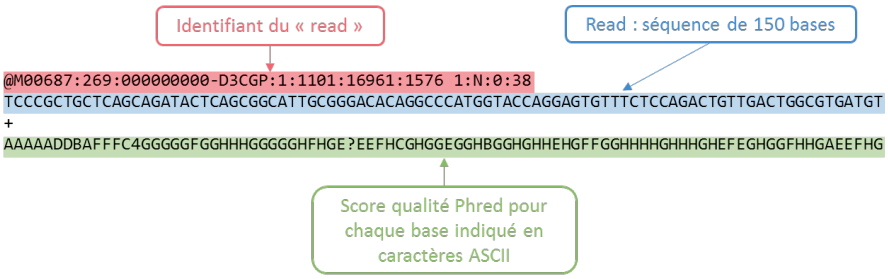


Figure 43 : Exemple de fichier .fastq obtenu à l'issu d'un séquençage par MiSeq (Illumina)

La première étape consiste à attribuer chaque séquence au patient correspondant : cette étape est appelée démultiplexage. L'appareil détecte les index dans chaque séquence et génère deux fichiers .fastq pour chaque patient (lectures des brins sens et anti-sens, read 1 et read 2).

2.2. Contrôle qualité

L'étape suivante consiste à évaluer la qualité du séquençage et éliminer les données « non fiables ». L'outil NGS QC Toolkit v.2.3.3 [218] est utilisé. Les séquences conservées présentent au minimum 70% des bases avec un score qualité Phred supérieur à 20 (1 erreur/100 bases lues).

2.3. Alignement des séquences sur un génome de référence

Les séquences conservées après le filtre de qualité sont comparées à un génome de référence à l'aide de l'outil Burrows-Wheeler Aligner (BWA) [219]. La version du génome de référence utilisée pour cette étude est *Homo sapiens* b37 (GRCh37, hg19, hg_g1k_v37).

De nouveau, certaines séquences peuvent être éliminées si la qualité de l'alignement est trop faible (<70% de bases complémentaires au génome de référence, par read). De plus, les reads sont alignés par paire (read 1 et read 2) pour chaque échantillon séquencé. A l'issue de cette étape, des fichiers .bam et .bai, correspondant respectivement aux séquences individuelles alignées sur un génome de référence et aux annotations des fichiers, sont obtenus (48 fichiers au total, 2 par patient). Les séquences alignées peuvent être visualisées à l'aide du logiciel *integrative genomics viewer* (IGV, Broad Institute) [220].

2.4. Appel des variants

Vient ensuite l'étape de l'appel des variants qui consiste à identifier les différences entre les séquences obtenues par NGS et le génome de référence. L'outil *unified genotyper* de la suite *genome analysis toolkit* (GATK, Broad Institute) est utilisée *via* la plateforme use.galaxy.org [216,221]. Le score de qualité Phred seuil est 20. A l'issue de cette étape des fichiers au format .vcf sont récupérés pour chaque patient séquencé. Chaque entrée du fichier correspond à une position donnée pour laquelle une variation de la séquence a été identifiée.

Plusieurs informations sont indiquées à chaque position telles que :

- CHROM : chromosome
- POS : position sur le chromosome (hg19)
- REF : base(s) de référence
- ALT : base(s) variante(s)
- DP : profondeur de lecture
- GQ : fiabilité du génotype (selon Phred)
- GT : génotype (0/1 = hétérozygote, 1/1 = homozygote variant)

Les seuls variants conservés sont ceux pour lesquels, pour au moins un patient, 30 reads au minimum sont alignés sur le génome (DP≥30).

2.5. Annotation des variants

Pour chaque patient une liste de positions dans le génome et de variants associés est collectée. Des dizaines de bases de données sont interrogées *via* l'outil d'annotation en ligne AVIA (avia-abcc.ncifcrf.gov) (Tableau 13) [217].

Tableau 13 : Liste non exhaustive des bases de données utilisées par l'outil AVIA

Nature du variant	
ANNOVAR	Localisation
dbSNP v147	Identifiant du polymorphisme (rs...)
Fréquence de survenue du variant	
1000 genomes (1000G)	MAF à partir de 2504 individus
ExAC v3	MAF à partir de 60.706 (inclus 1851 échantillons de 1000G)
Conséquences du variant	
Combination Annotation Dependent Depletion (CADD)	Prédit si le variant est délétère (cause d'une pathologie) Exemple de scores : 10 : parmi les 10% de variants les plus délétères du génome 20 : 1% 30 : 0,1% ...
Mutation Assessor	Prédit l'impact fonctionnel des SNP selon la conservation au cours de l'évolution MEDIUM ou HIGH : pas de conséquence NEUTRAL ou LOW : protéine non fonctionnelle
Mutation Taster	Prédit si le variant est associé à une pathologie
Polymorphism Phenotyping v2 (PolyPhen-2)	Prédit l'impact de SNP non-synonyme sur la structure et la fonction de la protéine BENIGN < Probably DAMAGING < DAMAGING
Sorting Intolerant From Tolerant (SIFT)	Prédit l'impact du variant sur la structure de la protéine selon la redondance du code génétique. Score de 0 à 1 DAMAGING : altère la structure ≤ 0,05 TOLERATED : ne modifie pas la structure > 0,05

2.6. Tableau de contingence

La dernière étape consiste à fusionner les données obtenues pour chaque patient dans une matrice d'occurrence reprenant l'ensemble des polymorphismes et génotypes identifiés pour les 24 patients atteints de LMC séquencés. Les homozygotes de référence sont notés 0_0, les hétérozygotes 0_1 et les homozygotes variant 1_1. Pour certains polymorphismes, les MAF sont >0,5. Dans ces cas-là, les génotypes observés sont inversés. Par la suite, la fréquence de survenue du variant (hétérozygote + homozygotes variant) est estimée pour les patients répondeurs, résistants mais également pour la population générale en fonction de la MAF.

3. Analyses statistiques à partir de la matrice d'occurrence

3.1. Statistiques descriptives : AFC et HCPC

Plusieurs analyses statistiques ont été développées afin d'identifier les variants les plus susceptibles d'être associés à la résistance à l'imatinib parmi tous les variants identifiés par NGS. Tout d'abord, une analyse factorielle des correspondances (AFC) est réalisée afin de visualiser les données obtenues sur un graphique en deux dimensions [222]. Par cette approche multivariée, les données d'une matrice d'occurrence sont tout d'abord représentées dans un graphique multidimensionnel (exemple Tableau 14 et Figure 44).

Tableau 14 : Exemple de tableau de contingence

Polymorphismes	Répondeurs	Résistants	Population générale
1	0	0,7	0,2
2	0,4	0,6	0,1
3	0,8	0,5	0,8

Les fréquences de survenue de l'allèle variant sont indiquées pour chaque groupe d'individus (patients résistants, patients répondeurs, population générale) et chaque polymorphisme (1, 2 et 3).

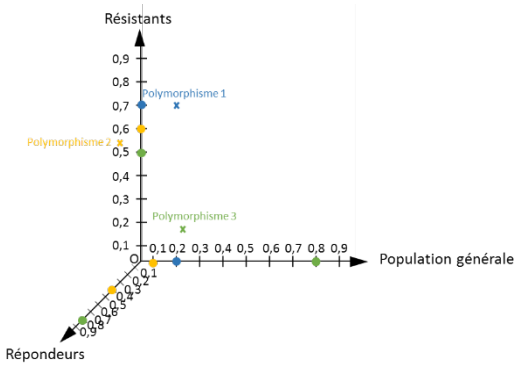


Figure 44 : Exemple de graphique en trois dimensions illustrant le tableau de contingence

Chaque point représente un polymorphisme.

Ces données sont ensuite projetées dans un graphique en deux dimensions tout en conservant la structure globale du graphique en trois dimensions représentative du tableau de contingence. La localisation des points sur le graphique est corrélée avec les différences de fréquences de survenue du variant entre les trois groupes. Par la suite, une classification hiérarchique sur composantes principales (*hierarchical clustering on principal components*, HCPC) est effectuée. Une analyse non supervisée est réalisée et les clusters sont définis. Ensuite, un partitionnement en K-moyennes permet d'optimiser les groupes générés. Le centroïde de chaque cluster est alors déplacé selon la répartition des points et une nouvelle classification est réalisée en se basant sur la nouvelle valeur du centroïde.

3.2. Simulation de données, méthode non-paramétrique : rank products

La méthode des rank products (RP) est une approche statistique non-paramétrique permettant de comparer deux conditions et de classer les données selon la différence entre ces deux conditions [223,224]. Pour cette étude, les gènes séquencés sont classés selon leur enrichissement en variants causant des altérations de la séquence protéique et les différences entre les patients répondeurs et les patients résistants.

Cette approche nécessite des réplicas expérimentaux. Des données artificielles sont donc générées à partir des données observées *via* une simulation de Monte Carlo. La méthode du *naïve bootstrap* est choisie car elle s'applique à des données estimées non Gaussiennes. Huit expériences simulées ont été générées (4 à partir des données de patients répondeurs, 4 à partir des données de patients résistants) avec 1000 échantillonnages sans remise de type *bootstrap* de 12 patients à chaque mesure. A partir de ces réplicas, les fréquences alléliques sont déterminées et comparées aux fréquences déterminées dans la population générale (MAF selon 1000G) en calculant le risque de survenue du variant (ODDS ratio) pour chaque expérience (8 fois) (Figure 45).

$$ODDS_i = \frac{\sum_{k=1}^{n} \frac{q_{ki} \times (1 - p_{ki})}{p_{ki} \times (1 - q_{ki})}}{999}$$

$\Sigma ODDS_i$ est la somme des ODDSi de chaque groupe de patients atteints de LMC pour k=999 nouveaux échantillonnages sans remise « bootstrap », q est la fréquence de survenue du variant pour chaque gène i dans un groupe de patients, p est la fréquence de survenue du variant pour chaque gène i dans la population générale.

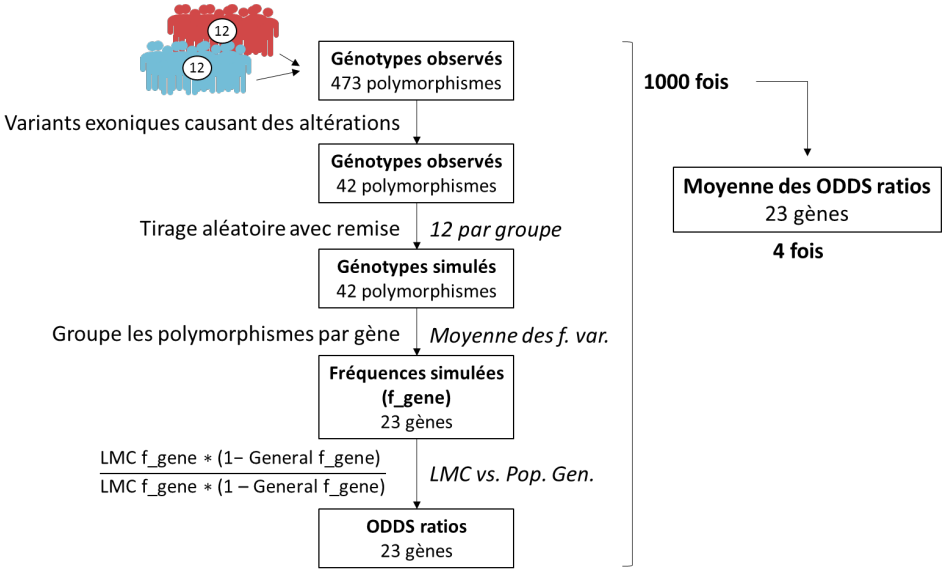


Figure 45 : Simulation de Monte Carlo et calcul des rapports des cotes (ODDS ratios)

La méthode des RP est ensuite appliquée. Les ODDS ratios pour chaque gène sont comparés entre les patients atteints de LMC répondeurs et résistants au traitement.

3.3. Programme informatique

Afin de réaliser l'ensemble de ces analyses informatiques et statistiques un programme et une interface graphique à utiliser avec le logiciel de statistiques R (v3.5) ont été développés (script en détails dans les données supplémentaires de l'article en Annexe 1, description de l'interface en Annexe 2).

4. Validation des polymorphismes candidats

4.1. Extraction de l'ADNg et des ARN totaux

4.1.1. Extraction à partir de culots leucocytaires de patients

L'ADNg et les ARN totaux sont extraits à partir d'une méthode basée sur le mélange phénol-chloroforme. Chaque culot de 10 millions de cellules est repris dans 1 mL de mélange TRIzol™ (Thermo Fisher Scientific) contenant notamment du phénol et du thiocyanate de guanidium. Les cellules sont entièrement lysées grâce à un appareil Disruptor Genie (Scientific Industries). Du chloroforme:alcool isoamylique 24:1 (200 µL) (Sigma-Aldrich) est ajouté au mélange cellulaire. Le tout est incubé 3 min à température ambiante puis centrifugé 15 min à 11.300 g. Une interphase blanche contenant l'ADNg apparait entre le mélange TRIzol™ et la phase aqueuse supérieure contenant l'ARN. Les phases contenant l'ADNg et l'ARN sont

collectées indépendamment. L'interface contenant l'ADNg est mélangée à 300 µL d'éthanol absolu. Après 3 min d'incubation à température ambiante les filaments d'ADNg sont visibles. Ils sont prélevés à l'aide d'un cône puis lavés à l'éthanol 75%. Après séchage à température ambiante, les filaments sont repris dans de l'eau distillée ultra pure. La phase aqueuse contenant l'ARN est mélangée de nouveau à 200 µL de chloroforme:alcool isoamylique 24:1 (Sigma-Aldrich) et incubée 3 min à température ambiante puis centrifugée 15 min à 11.300 g. La phase supérieure contenant l'ARN est précipitée par ajout de 500 µL d'isopropanol (Sigma-Aldrich) puis centrifugée 15 min à 11.300 g. Le surnageant est éliminé et le culot contenant l'ARN est lavé en présence d'éthanol 75 %. Après séchage à température ambiante, le culot est repris dans de l'eau distillée ultra pure.

4.1.2. Dosage des acides nucléiques

La concentration et la qualité des solutions d'ADNg et d'ARN sont évaluées par spectrophotométrie en utilisant l'appareil Nanodrop 2000 (Thermo Fisher Scientific). Les acides nucléiques absorbent la lumière à 260 nm. Les absorbances à 230 nm et 280 nm sont respectivement associées aux solvants et aux protéines (acides aminés aromatiques). La pureté des solutions d'acides nucléiques est considérée comme acceptable lorsque le ratio A260nm/A280nm est ≈1,8/2 pour l'ADN et ≈2/2,2 pour l'ARN, et lorsque le ratio A260nm/A230nm est compris entre 2 et 2,2. La concentration en ng/µL est évaluée sachant qu'une unité d'absorbance à 260 nm correspondant à 50 µg/mL pour l'ADN double brin et 40 µg/mL pour l'ARN simple brin.

4.2. Séquençage Sanger

4.2.1. Amplification par PCR

Les polymorphismes d'intérêt sont identifiés dans une plus grande cohorte de patients par approche Sanger. Pour cela, les fragments d'ADN à séquencer sont tout d'abord amplifiés par PCR à l'aide du kit KOD Hot Start DNA Polymerase (Merck) à partir d'ADNg extrait d'échantillons de patients. Le mélange réactionnel et le programme d'amplification sont présentés dans le Tableau 15. La réaction de PCR est réalisée dans un thermocycleur (T100 Thermal Cycler, Bio-Rad).

Tableau 15 : Mélange réactionnel et programme pour l'amplification de fragments d'ADN à séquencer par Sanger

Réactifs	Volume pour une reaction (µL)	Concentration finale
Tampon	2,5	1X
$MgSO_4$	1,5	1,5 mM
dNTP	2,5	0,2 mM
Amorce sens	0,75	300 nM
Amorce anti-sens	0,75	300 nM
ADN polymérase KOD Hot Start	0,25	0,01 unité/µL
ADNg	0,5	100 ng
Eau	qsp 25 µL	

Nombre de cycles	Température (°C)	Durée
1	95	2 min
40	95	20 sec
	60	10 sec
	70	15 sec

Quatre PCR différentes sont réalisées afin de séquencer les différents polymorphismes identifiés par NGS dans les gènes *BCL-RAMBO, BCLAF1* et *BCL2A1*. De plus, une PCR permettant l'amplification d'une séquence de 996 pb du promoteur de *BCL2A1*, non séquencée par NGS, a été réalisée (Tableau 16).

Tableau 16 : Description des amorces utilisées pour l'amplification de fragments des gènes *BCL-RAMBO*, *BCLAF1* et *BCL2A1*.

Gène	Localisation	Polymorphismes	Sens	Anti-sens	Taille attendue (pb)
BCL-RAMBO	Exon 11	rs9306198	GCACGTGGAGAAAGAAGAGG	GACCACCTCCTCCACTTCAG	274
BCLAF1	Exon 4	rs139012879	GTCGTTCACGTTCCAGGAGT	TTCGGTTCCTCCTGTGATTT	241
BCL2A1	Exon 1	rs1138357 rs1138358 rs3826007	TCTCAGCACATTGCCTCAAC	TCATTATGAACTCCGCAACAA	450
BCL2A1	Exon 2, intron 2	rs8026803 rs16971619 rs146807490	GGCATGGAAAACTGTTATCATCT	CTGGGATTACAGGCATGAGC	298
BCL2A1	Promoteur	?	CCCCGTGGATGGCGTTTAAG	AAGCCACTTCCTCATCCTGTG	996

4.2.2. Electrophorèse sur gel d'agarose

L'efficacité de l'amplification et la taille des amplicons PCR sont évaluées par électrophorèse sur gel d'agarose 2% m/v. En plus des échantillons, un marqueur de taille (échelle de poids moléculaire d'intervalle de 100 pb, de 100 pb à 3000 pb) est déposé afin d'estimer la taille des produits PCR obtenus (100 bp DNA ladder, Promega).

4.2.3. Réaction de séquençage

Les produits PCR sont séquencés en Sanger par la société GATC Biotech, devenue par la suite, Eurofins Genomics. Les amorces utilisées pour le séquençage sont les mêmes que celles utilisées pour l'amplification PCR. Les chromatogrammes sont visualisés grâce au logiciel FinchTV (Geospiza).

4.3. PCRq en temps réel

4.3.1. DNase

L'expression des gènes candidats est évaluée par RT-PCRq en temps réel. Dans un premier temps, les éventuelles contaminations de l'ARN par de l'ADNg sont éliminées en utilisant le kit DNA-*free*™ DNA Removal (Thermo Fisher Scientific). Les échantillons contenant 10 µg d'ARN totaux sont mélangés à de la DNase I à 0,1 unité/µL dans un tampon spécifique 1X pour un volume total de 20 µL puis incubés 30 min à 37°C. La réaction est stoppée par ajout de 2 µL d'un réactif d'inactivation (DNase inactivation reagent) contenant notamment de l'acide citrique. Après une incubation de 2 min à température ambiante, les échantillons sont centrifugés 1 min à 10.000 g et le surnageant contenant l'ARN est récupéré.

4.3.2. Transcription inverse

L'ADNc est obtenu à partir de l'ARN traité à la DNase en utilisant le kit Transcriptor First Strand cDNA Synthesis Kit (Roche). Les échantillons contenant 1 µg d'ARN totaux sont mélangés à un mélange de dNTP à 1 mM chacun, un inhibiteur de RNase à 1 unité/µL, des amorces hexamères aléatoires à 60 µM, une reverse transcriptase à 0,5 unité/µL dans un tampon adapté 1X. L'amplification du brin complémentaire est réalisée dans un thermocycleur (T100 Thermal Cycler, Bio-Rad) et comprend plusieurs étapes successives : 10 min à 25°C, 1h à 50°C et 5 min à 85°C.

4.3.3. Réaction de PCRq et analyse des résultats

L'expression des gènes candidats est évaluée par PCRq en temps réel à l'aide du kit Brilliant III UltraFast SYBR Green PCRq Master Mix (Agilent Technologies). Le mélange réactionnel comprend, dans un volume total de 20 µL : 2 µL d'ADNc dilué au 10ème, le Master Mix SYBR Green à 50% v/v, un colorant de référence à 0,03 µM, les amorces sens et anti-sens

à 300 nM chacune. Les échantillons sont déposés en duplicatas en microplaques de 96 puits. L'amplification et la lecture de la fluorescence sont réalisées par l'appareil Mx3005P (Agilent Technologies). Plusieurs étapes successives sont définies : 3 min à 95°C, 40 cycles incluant 15 sec à 95°C et 22 sec à 60° (la fluorescence est détectée à chaque cycle), puis une étape finale avec augmentation de la température de 60°C à 95°C par paliers de 0,1°C et en mesurant la fluorescence en continu. Cette dernière étape permet d'estimer la température de fusion des amplicons.

Une première paire d'amorces a été conçue pour détecter l'expression du transcrit majoritaire de *BCL2A1* (jonction exons 1 et 3). Une deuxième paire d'amorces permet de détecter les 13 transcrits de *BCL-RAMBO* comprenant tous l'exon 11. L'expression du gène *ACTB* (codant l'actine β) est également évaluée afin de normaliser les résultats (Tableau 17). En effet il s'agit d'un gène de ménage, exprimé de manière ubiquitaire dans tous les types cellulaires.

Tableau 17 : Description des amorces utilisées en PCRq pour étudier l'expression des gènes candidats dans une cohorte de validation

	Amorce sens		Amorce anti-sens		
Gène	**Localisation**	**Séquence**	**Localisation**	**Séquence**	**Taille attendue (pb)**
BCL2A1	Exon 1	AAATTGCCCCGGATGTGGAT	Jonction exon 1 et 3	ACAAAGCCATTTTCCCAGCC	115
BCL-RAMBO	Exon 11	CTTCTGTGGTCTTGCCAGCG	Exon 11	GTAGCAGGGCTGGATTTCTCA	148
ACTB	3'-UTR	AGCATCGGGTGATGTTCATT	3'-UTR	ATTACAAGCATGCGTCACCA	107

Les variations d'expression des gènes sont étudiées selon la méthode des ΔΔCt, le Ct correspondant au nombre de cycles nécessaires pour que l'échantillon soit détecté (cycle seuil). Le Ct moyen est calculé pour chaque échantillon (moyenne des duplicatas) puis les ΔCt, ΔΔCt et l'expression relative du gène pour chaque échantillon sont estimés. L'échantillon (patient répondeur) avec le ΔCt médian devient l'échantillon de référence.

$$\Delta\mathrm{Ct}_{\text{échantillon}} = \mathrm{Ct}_{\text{moyen}} du\ gène\ cible - Ct_{moyen}\ du\ gène\ de\ ménage$$

$$\Delta\Delta\mathrm{Ct}_{\text{échantillon}} = \Delta\mathrm{Ct}_{\text{échantillon}} - \Delta\mathrm{Ct}_{\text{échantillon de référence}}$$

$$Expession\ relative = 2^{-\Delta\Delta\mathrm{Ct}_{\text{échantillon}}}$$

5. Culture cellulaire

Après validation du polymorphisme rs9306198 du gène *BCL-RAMBO*, les rôles du gène et du variant dans la réponse à l'imatinib ont été évalués par différentes analyses fonctionnelles à partir de lignées cellulaires modèles.

5.1. Milieux de culture

Les cellules 293T sont cultivées en milieu Dulbecco's Modified Eagle's (DMEM, Thermo Fisher Scientific). Les cellules K-562 sont cultivées en milieu Roswell Park Memorial Institute (RPMI) 1640 (Thermo Fisher Scientific). Les deux milieux sont supplémentés avec 10% de sérum de veau fœtal (SVF, Eurobio), 1 mM de L-glutamine, 100 unités/mL de pénicilline, 0,1 mg/mL de streptomycine (Thermo Fisher Scientific). Les cellules sont cultivées à 37°C en étuve en atmosphère humide 5% CO_2 à une concentration de 200.000 cellules/mL.

5.2. Sous-clonage cellulaire

Les cellules d'un mélange cellulaire sont individualisées par sous-clonage. Deux approches sont utilisées selon le type cellulaire. Les cellules contrôles 293T sont clonées par la méthode de dilution limite en plaque 96 puits. Les cellules K-562 quant à elles ont la capacité de former des colonies en milieu semi-solide. Cette propriété est utilisée pour individualiser les cellules et obtenir des lignées clonales. Une dilution au 1:6 de 800 cellules en milieu supplémenté est réalisée dans un milieu contenant de la methylcellulose (MethoCult™ H4534 classique sans EPO, Stemcell technologies). Les cellules sont maintenues 10 jours en atmosphère humide à 37°C à 5% CO_2. Lorsque les colonies sont visibles, elles sont prélevées et ensemencées dans une plaque 96 puits contenant 100 µL de milieu.

5.3. Congélation des cellules en azote liquide

Les cellules sont régulièrement congelées et conservées en azote liquide. Pour cela, 5 à 10 millions de cellules sont mélangés dans un cryotube avec du milieu de culture (RPMI ou DMEM selon le type cellulaire) contenant 10% de DMSO, 50% de SVF. La température des échantillons est lentement diminuée par stockage 24h dans des boîtes CoolCell© (Thermo Fisher Scientific) à -80°C avant leur transfert dans de l'azote liquide (-195,79°C).

6. CRISPR-Cas9

6.1. Principe de la technologie

Au cours de ce projet, des lignées cellulaires modifiées ont été générées grâce à la technologie d'édition du génome CRISPR-Cas9 (Clustered Regularly Interspaced Short Palindromic Repeats-CRISPR associated protein 9) basée sur l'utilisation d'une endonucléase, la Cas9, complexée à un court ARN simple brin court, l'ARN guide (ARNg). Cette ARN contient une séquence de 20 nucléotides complémentaires de la région génique que l'on souhaite modifier, suivie d'une séquence *scaffold* qui permet à l'ARNg de former un complexe avec la Cas9. Dans le génome, la séquence complémentaire doit être suivie d'une courte séquence nommée *protospacer adjacent motif* (PAM). Pour cette étude, la Cas9 utilisée est issue de *Streptococcus pyogenes* et sa séquence PAM est NGG, avec N équivalent à n'importe quel nucléotide. Le complexe CRISPR-Cas9, exprimé dans le noyau cellulaire, parcourt l'ADNg jusqu'à ce que la séquence complémentaire de l'ARNg, suivie du PAM, soit détectée. L'enzyme génère alors une coupure double brin de l'ADNg, le plus souvent à 3 pb du PAM (Figure 46) [225].

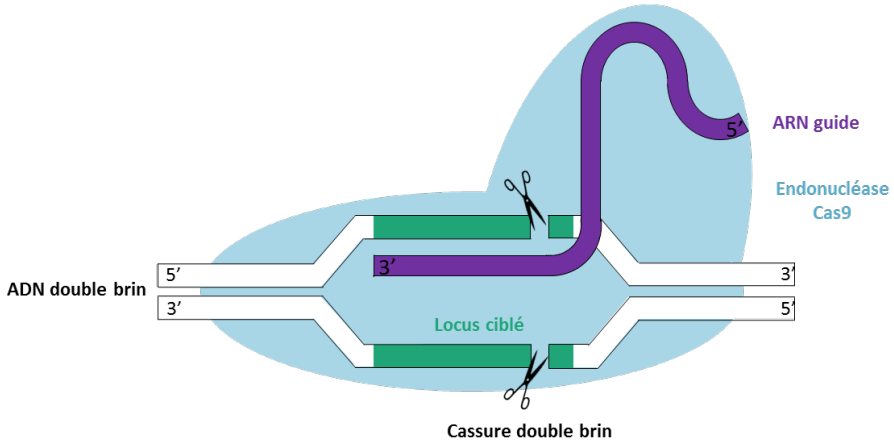


Figure 46 : Principe du système CRISPR-Cas9

Cette cassure double brin peut être réparée par deux mécanismes cellulaires différents. Le mécanisme de réparation NHEJ (*non-homologous end joining*) est majoritaire dans les cellules mammifères. Le plus souvent, la réparation est correcte : le complexe CRISPR reconnaît donc de nouveau la séquence et coupe une nouvelle fois la double hélice. Après des coupures à répétitions, le mécanisme de réparation peut générer des erreurs et créer de courtes insertions ou délétions aléatoires au niveau du site de cassure. Cette propriété est utilisée afin

de générer des décalages de cadre de lecture au début des séquences codantes et ainsi éteindre l'expression des gènes.

La recombinaison homologue, bien plus rare que la NHEJ dans les cellules humaines, permet également la réparation des lésions de l'ADN en présence d'une séquence modèle : la matrice de recombinaison. La séquence de la matrice est intégrée dans la séquence d'ADNg et permet ainsi la substitution spécifique d'une ou plusieurs bases ou même l'intégration de fragments exogènes dans l'ADNg.

Si deux ARNg différents, ciblant deux régions proches, sont exprimés simultanément dans la cellule, la réparation peut se faire en éliminant le fragment d'ADN présent entre les deux sites de cassure. Cette méthode facilite le crible des cellules modifiées puisque le fragment d'ADN édité sera raccourci par rapport au fragment d'ADN non modifié.

6.2. Conception et clonage des guides

Les séquences des ARNg sont conçues à l'aide de l'outil en ligne CHOPCHOP v2 (http://chopchop.cbu.uib.no) [226]. Chaque guide est cloné individuellement dans un vecteur PX462 modifié (fourni par le laboratoire de Feng Zhang, Addgene #48141). Dans ce plasmide, la nickase Cas9 D10A a été remplacée par la nucléase Cas9 à spécificité améliorée issue de *Streptococcus pyogenes* (eSpCas9(1.1)). La résistance à la puromycine a été améliorée en clonant la séquence Puro-lowCpG. Ce plasmide contient également un promoteur U6 suivi d'un site de clonage en amont de la séquence 3' de l'ARNg (chimeric guide RNA scaffold) (Figure 47). Cette conformation permet l'incorporation des 20 pb de l'ARNg variables selon la séquence ciblée.

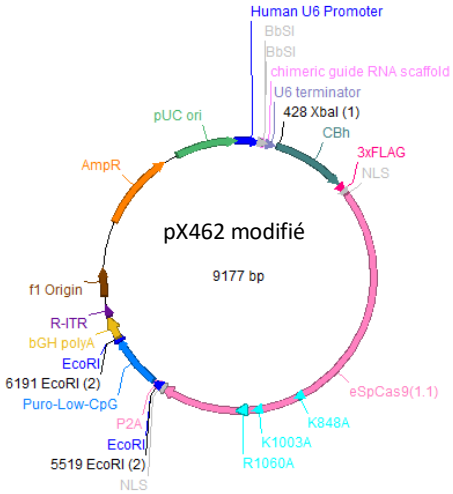


Figure 47 : Carte du plasmide pX462 modifié

Pour introduire la séquence de 20 nucléotides de l'ARNg dans le vecteur, l'approche d'hybridation d'oligonucléotides phosphorylés précédemment décrite est utilisée [227]. Pour chaque ARNg, deux oligonucléotides sont commandés (Eurogentec). Le premier contient la séquence reconnue par l'ARNg précédée de CACC, le deuxième englobe la séquence complémentaire précédée de AAAC. Si la séquence ne débute pas par une guanine (G), ou termine par une cytosine (C) pour la séquence complémentaire, la base doit être ajoutée aux oligonucléotides (nécessaire à la transcription) (Tableau 18).

Tableau 18 : Caractéristiques des séquences cibles des ARNg et oligonucléotides commandés

Lignées	Cible	Délétion attendue (pb)	N° guide	Séquences cibles	PAM	Oligonucléotide 1	Oligonucléotide 2
KO	Exon 2	99	1	cggataccgttccggatgtc	AGG	**caccG**cggataccgttccggatgtc	**aaac**gacatccggaacggtatccg**C**
			2	ggtagattagggccgcgggt	CGG	**cacc**ggtagattagggccgcgggt	**aaac**acccgcggccctaatctacc
KO	Exon 3	66	1	agaagaggacgccattgaat	TGG	**caccG**agaagaggacgccattgaat	**aaac**attcaatggcgtcctcttct**C**
			2	gtatgttgttctcagctact	TGG	**cacc**gtatgttgttctcagctact	**aaac**agtagctgagaacaacatac
ΔBHNo+TM	Exon 11	627	1	tatcttctgcagtgatttcca	GGG	**caccG**tatcttctgcagtgattcca	**aaac**tggaatcactgcagaagata**C**
			2	aagcccatgccgctgtctga	GGG	**caccG**aagcccatgccgctgtctga	**aaac**tcagacagcggcatgggctt**C**
RH	Exon 11	-	1	actgtgatcacttctgtgtc	AGG	**caccG**actgtgatcacttctgtgtc	**aaac**gacacagaagtgatcacagt**C**
			2	aggttcccttgtccccagca	GGG	**caccG**aggttcccttgtccccagca	**aaac**tgctggggacaagggaacct**C**
			3	agggaggtggcagtgatatg	TGG	**caccG**agggaggtggcagtgatatg	**aaac**catatcactgccacctccct**C**
			4	gcagtgatatgtggaagcaa	GGG	**cacc**gcagtgatatgtggaagcaa	**aaac**ttgcttccacatatcactgc

Les extrémités de ces fragments d'ADN (5 µM chacun) sont phosphorylées par incubation 30 min à 37°C en présence de polynucléotide kinase T4 (New England Biolabs) et d'ATP (1 mM) dans un tampon approprié. Ces oligonucléotides complémentaires phosphorylés sont hybridés par une incubation de 5 min à 95°C dans un bloc chauffant, ce dernier est ensuite éteint pour permettre une diminution lente de la température des échantillons pendant environ 2h. Au final, un fragment partiellement double brin est obtenu avec les 20 pb complémentaires et, de part et d'autre, les séquences simple brin CACC et CAAA. Ces extrémités cohésives permettent un clonage dans le vecteur pX462 modifié précédemment (50 ng/µL) digéré par l'enzyme de restriction *BbS*I (0,05 unité/µL, FastDigest, Thermo Fisher Scientific) (Figure 48).

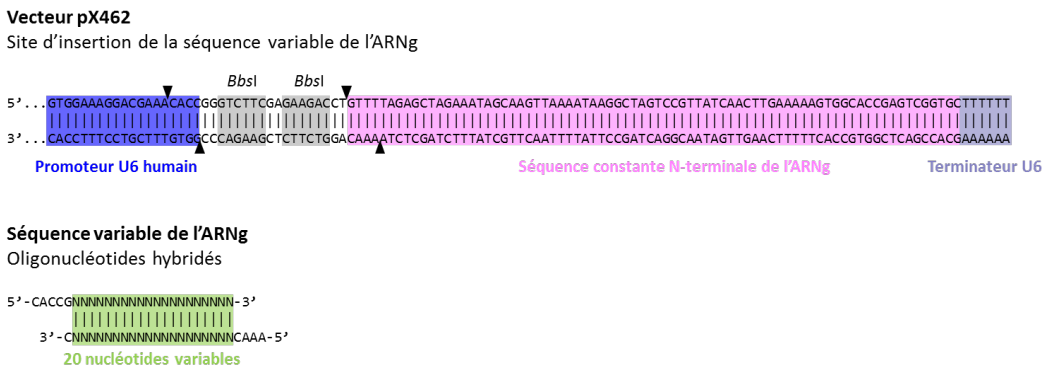


Figure 48 : Description du clonage de la séquence variable de l'ARNg dans le vecteur pX462 *via* la digestion du vecteur par l'enzyme *Bbs*I

Le kit Quick Ligation™ est choisi pour le clonage (New England Biolabs). Les inserts à 25 nM chacun sont mélangés au vecteur digéré à 5 ng/µL et une Quick ligase à 60 unités/µL dans un tampon approprié pour un volume final de 10 µL. Les échantillons sont incubés 15 min à température ambiante.

6.3. Transformation bactérienne et extraction plasmidique

Les plasmides sont amplifiés dans des bactéries compétentes *Escherichia coli* TOP10 (Thermo Fisher Scientific). Tout d'abord, 2,5 µL de produit de ligation sont mélangés à 25 µL de bactéries compétentes et incubés 20 min dans la glace. Les vecteurs sont introduits dans les bactéries par un choc thermique de 30 sec à 42°C puis refroidissement rapide à 4°C. Par la suite, 200 µL de milieu SOC (*super optimal broth*, Tableau 19) sont ajoutés et le mélange est incubé 30 min à 37°C afin de permettre l'expression des gènes de résistance aux antibiotiques et incubés 30 min à 37°C. Enfin, 150 µL du mélange sont étalés sur une boîte contenant du milieu de culture LB (lysogeny broth, Tableau 19) agar solide contenant 100 µg/mL d'ampicilline. Le tout est incubé toute la nuit à 37°C.

Tableau 19 : Composition des milieux SOC et LB

Milieu SOC	
Réactifs	**Concentration**
Tryptone	2% m/v
Extrait de levure	0,5 m/v
NaCl	10 mM
KCl	2,5 mM
$MgCl_2$	10 mM
$MgSO_4$	10 mM
Glucose	20 mM

Milieu LB	
Réactifs	**Concentration**
Peptone	10 g/L
Extrait de levure	5 g/L
NaCl	5 g/L
+/- Agar	12 g/L

Le lendemain, plusieurs colonies, résistantes à l'ampicilline sont visibles sur la boîte. A l'aide d'une anse, une colonie est prélevée et mélangée à 50 mL de milieu LB liquide contenant 100 µg/mL d'ampicilline. Après plusieurs heures d'incubation à 37°C sous agitation (220 rpm), les vecteurs plasmidiques sont extraits à l'aide du kit midiprep (NucleoBond® Xtra Midi, Macherey Nagel). L'intégrité des plasmides est évaluée par restriction enzymatique. Pour cela, 50 ng/µL de vecteur sont mélangés à 0,05 unité/µL d'enzyme *Xba*I et *Eco*RI dans un tampon adapté 1X (FastDigest, Thermo Fisher Scientific) pour un volume final de 20 µL. Le profil attendu après digestion est indiqué en Figure 49. Après 30 min d'incubation à 37°C, une électrophorèse sur gel d'agarose 0,7% est réalisée. De plus, l'insertion correcte de l'ARNg est validée par séquençage Sanger à l'aide d'une amorce sens s'hybridant sur le promoteur U6 en amont de la séquence codante (5' CAAGGCTGTTAGAGAGATAATTGGA 3').

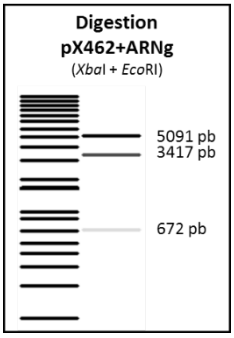


Figure 49 : Profil de digestion attendu du vecteur pX462 après insertion de la séquence variable de l'ARNg de 20 nucléotides et digestion par les enzymes *Xba*I et *Eco*RI

6.4. Conception d'une matrice de recombinaison homologue

Une matrice de recombinaison double brin de 999 pb est testée pour l'insertion du variant par recombinaison homologue (gBlock gene fragment, Integrated DNA Technologies). Cette matrice est complémentaire d'une séquence de 999 pb de l'exon 11 de *BCL-RAMBO* contenant le polymorphisme rs9306198. Plusieurs modifications sont introduites dans ce fragment modèle par rapport à la séquence endogène. Tout d'abord, le variant (allèle T) a été intégré au lieu de l'allèle de référence (C) au milieu de la matrice. De plus, à seulement 5 bases de ce variant, deux bases CC ont été inclues au lieu du motif TT présent dans la séquence sauvage (wild-type, WT). Ces changements synonymes permettent de créer un nouveau site de restriction enzymatique (*Sma*I) facilitant les étapes de crible des cellules modifiées (Figure 50).

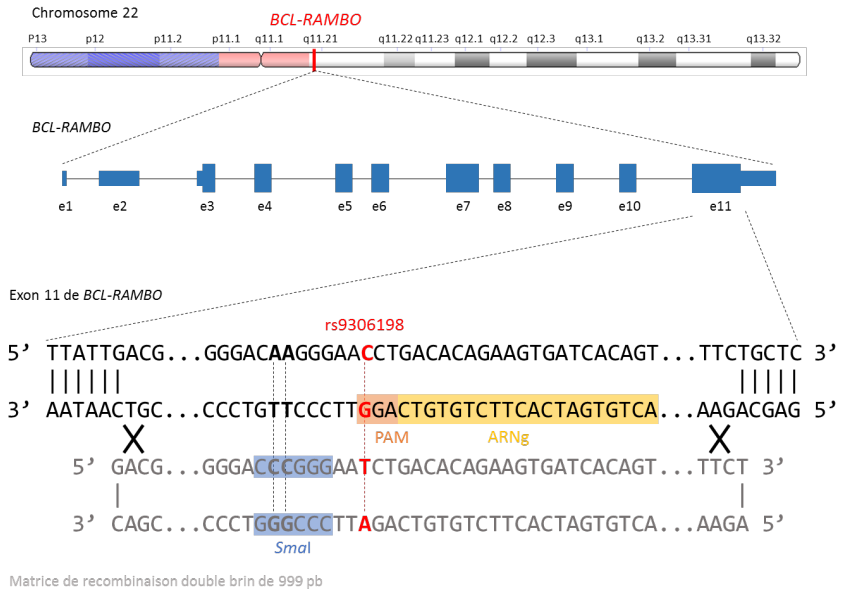


Figure 50 : Principe de la modification de l'exon 11 de *BCL-RAMBO* par CRISPR-Cas9 et la recombinaison homologue pour l'insertion de l'allèle variant (T) du polymorphisme rs9306198

6.5. Transfection

Afin d'exprimer le système CRISPR-Cas9 dans les lignées 293T et K-562, les différents vecteurs ont été introduits dans les cellules par transfection. La méthode de co-précipitation avec du phosphate de calcium est choisie pour introduire les plasmides dans les cellules 293T. La veille de la transfection, 400.000 cellules/condition dans 2 mL de milieu, sont déposées dans une plaque de 6 puits. Le jour de la transfection, l'ADN à introduire dans les cellules est mélangé à 140 mM de $CaCl_2$ et 1X de tampon HBS (*HEPES buffered saline*). Après une phase d'incubation de 5 min à température ambiante, la solution est déposée dans chaque puits. Après 6 à 8 h d'incubation en atmosphère humide à 37°C à 5% CO_2, les cellules sont « lavées » en remplaçant le milieu de culture.

La nucléofection est choisie pour transfecter les cellules K-562. Deux appareils différents ont été utilisés. Les premières expériences ont été réalisées avec l'appareil Amaxa® Nucleofector® II (Lonza). Pour chaque transfection, 1 million de cellules/condition, en phase exponentielle de croissance, sont transfectées. Les cellules sont reprises dans 100 µL de solution de transfection Ingenio® (Mirus, Euromedex) et l'ADN est ajouté à la suspension cellulaire. La solution est placée dans une cuvette à électroporation (Euromedex) et l'électroporation est réalisée dans l'appareil de nucléofection en suivant le programme T-016. Enfin, les cellules sont ensemencées dans 1 mL de milieu et maintenues en atmosphère humide à 37°C à 5% CO_2. Plus récemment, le système 4D-Nucleofector™ (Lonza), plus efficace et moins toxique pour les cellules, a été utilisé. Le protocole est similaire. Les cellules sont reprises dans 100 µL de

tampon composé de 82 µL de solution de nucléofection et 18 µL de supplément (kit SF Cell Line 4D-Nucleofector™ X, Lonza) puis l'ADN est ajouté. Après transfert dans des cuvettes d'éléctroporation (Nucleocuvette™, Lonza), le programme FF-120 du nucleofector est utilisé. Les cellules K-562 sont ensuite transférées dans 2 mL de milieu et maintenues en atmosphère humide à 37°C à 5% CO_2.

Les plasmides transfectés codent un gène de résistance à la puromycine. Les cellules transfectées sont donc traitées, 24h après la transfection, par ajout de 3,5 µg/mL de puromycine. Les cellules sont ensuite sélectionnées après 48h de traitement. Pour chaque transfection, 2 µg de vecteur, exprimant à la fois la Cas9 et l'ARNg, sont transfectés. De plus, pour l'expérience de recombinaison homologue 2,5 µg de matrice sont co-transfectés. Comme contrôle positif, 2 µg de plasmide exprimant la GFP sont utilisés (pmaxGFP™ Vector, Lonza).

6.6. Crible des cellules modifiées

6.6.1. Amplification PCR

Afin de déterminer l'efficacité du sysème CRISPR-Cas9 et les modifications induites, les séquences d'ADN ciblées sont amplifiées par PCR. Le kit Phire Tissue Direct PCR Master Mix (Thermo Fisher Scientific) est utilisé pour extraire l'ADN et réaliser la réaction de PCR. Une petite quantité de cellules est prélevée (env. 1000) puis lysée dans 10 µL de tampon de dilution et 0,25 µL de DNARelease Additive contenant notamment de la protéinase K et du glycérol. Le mélange est incubé 5 min à température ambiante puis 2 min à 98°C.

La PCR est réalisée dans le tampon Master Mix du kit 1X et 500 nM de chaque amorce (Tableau 20) auquel 0,5 µL de lysat cellulaire est ajouté pour un volume final de 10 µL. Le programme PCR est indiqué dans le Tableau 21. Les amplicons sont visualisés par électrophorèse sur gel d'agarose 2%.

Tableau 20 : Amorces utilisées pour l'amplification des fragments d'ADNg modifiés par CRISPR

Lignées	Cible	Amorce sens	Amorce anti-sens	Produit complet (pb)	Produit avec délétion (pb)
KO	Exon 2	aggtcacatctggacggaga	tctgggcactttcctaccca	339	240
KO	Exon 3	tgcacaagacattttggtgtga	accaccataagccaaagataatacc	279	213
ΔBHNo+TM	Exon 11	gacgcttgtccttcgtttctag	catgccactgccatgttgag	1020	303
RH	Exon 11	acggagctggagagaagagt	acctcaagccctaacatgcc	747	

Tableau 21 : Programme d'amplification PCR pour le crible des cellules modifiées par CRISPR

Nombre de cycles	Température (°C)	Durée
1	98	5 min
35	98	5 sec
	60	5 sec
	72	20 sec
1	72	1 min

Lorsque deux ARNg sont utilisés en simultané, un crible par PCR est suffisant pour déterminer si des coupures ont été générées dans l'ADN. Pour évaluer l'efficacité d'un seul ARNg un test de clivage doit être réalisé.

6.6.2. Test de clivage

Un test de clivage (surveyor assay) est réalisé à l'aide du kit Guide-it^{TM} Mutation Detection (Clontech) afin d'évaluer l'efficacité de la coupure. Brièvement, si l'ARNg est efficace et que des coupures doubles brins sont générées dans l'ADNg, des insertions ou délétions aléatoires sont générées. Ces fragments sont amplifiés par PCR, dénaturés puis hybridés par diminution lente de la température. Si des altérations de la séquence ont été créées, des mésappariements se forment dans ces nouveaux fragments doubles brins. Une enzyme, la resolvase, a la propriété de reconnaître et cliver ces boucles de mésappariement.

Les amplicons PCR sont donc incubés 5 min à 95°C afin de dénaturer les structures doubles brins. Par la suite, la température est diminuée progressivement, de -2°C/sec jusqu'à 85°C puis de -0,1°C jusqu'à 25°C. Cette étape permet l'hybridation aléatoire des brins d'ADN même si la complémentarité n'est pas totale (formation de mésappariements). La resolvase est ensuite ajoutée à la solution. L'action de l'enzyme se fait grâce à une incubation de 15 min à 37°C. Les amplicons résultants sont visualisés par une électrophorèse sur gel d'agarose 2%.

6.6.3. Test de restriction enzymatique

Si la recombinaison homologue a été efficace, un nouveau site de restriction enzymatique *Sma*I a été créé à 5 pb en amont du polymorphisme rs9306198 de *BCL-RAMBO.* Les produits PCR (20 µL) sont donc digérés par l'enzyme *Sma*I (0,025 unité/µL) dans un tampon adapté 1X adapté (FastDigest, Thermo Fisher Scientific) pour un volume final de 60 µL. Un contrôle négatif est réalisé en mélangeant tous les réactifs sans l'enzyme de restriction. Les produits PCR sont ensuite visualisés à l'aide d'une électrophorèse sur gel d'agarose 2%.

7. Surexpression de BCL-RAMBO

Le gène *BCL-RAMBO* a été surexprimé dans les cellules K-562 *via* la transduction d'un vecteur de surexpression.

7.1. Amplification de la séquence codante de *BCL-RAMBO*

Dans un premier temps, la séquence codante de *BCL-RAMBO* a été amplifiée à partir d'ADNc de deux patients, porteur ou non de l'allèle variant du polymorphisme rs9306198, afin d'obtenir les deux versions du gène *BCL-RAMBO*. Les amorces sens et anti-sens sont complémentaires respectivement de l'exon 2 en amont du codon start et de l'exon 11 en aval du codon stop. Le transcrit majoritaire de *BCL-RAMBO* n'est pas connu mais 10 des 11 transcrits codants contiennent à la fois l'exon 2 et l'exon 11. Plusieurs éléments sont ajoutés à chaque amorce pour permettre le clonage dans le vecteur de surexpression, la traduction et la détection de la protéine exprimée.

Tout d'abord une séquence correspondant au site de restriction *Eco*RI précédé de deux bases (AT) est ajoutée à l'amorce sens. Ces deux bases supplémentaires favorisent la stabilité de l'enzyme au niveau du site de restriction et donc son efficacité. Un site de restriction *Mlu*I, précédé également des bases AT, est ajouté à l'amorce anti-sens. En 5' de la séquence codante un codon start (ATG) suivi d'une séquence étiquette HA (*human influenza hemagglutinin*) est intégrée à l'amorce sens (Figure 51).

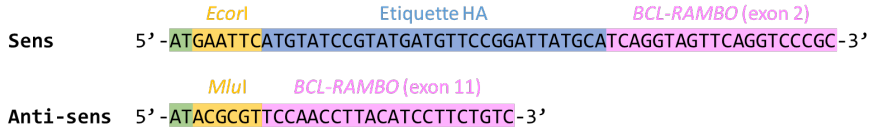


Figure 51 : Schéma des amorces sens et anti-sens permettant l'amplification de la séquence codante de *BCL-RAMBO* et le clonage dans le vecteur de surexpression

L'amplification PCR est réalisée à l'aide du kit KOD Hot Start DNA Polymerase (Merck) (Tableau 22). Les produits PCR sont analysés par électrophorèse sur gel d'agarose 1%.

Tableau 22 : Mélange réactionnel et programme d'amplification PCR de la séquence codante de *BCL-RAMBO*

Réactifs	Volume pour une reaction (µL)	Concentration finale
Tampon	2,5	1X
$MgSO_4$	1,5	1,5 mM
dNTP	2,5	0,2 mM
Amorce sens	1	400 nM
Amorce anti-sens	1	400 nM
ADN polymérase KOD Hot Start	0,5	0,02 unité/µL
ADNc	1	env. 50 ng
Eau	qsp 25 µL	

Nombre de cycles	Température (°C)	Durée
1	95	2 min
35	95	20 sec
	60	30 sec
	70	2 min

7.2. Purification et digestion des produits PCR

Les produits PCR obtenus sont purifiés à l'aide du kit NucleoSpin® Gel and PCR Clean-up (Macherey-Nagel). Les amorces, dNTP, sels et autres impuretés sont éliminés grâce à une colonne de Silice. Les contaminations sont éliminées par un lavage avec un tampon à l'éthanol. Les amplicons purifiés sont élués dans 20 µL de tampon d'élution. Les échantillons sont par la suite dosés au Nanodrop 2000 puis, les produits PCR à 10 ng/µL sont digérés avec les enzymes de restriction *Eco*RI et *Mlu*I (0,025 unité/µL de chaque) dans un tampon adapté 1X (FastDigest, Thermo Fisher Scientific) pour un volume final de 20 µL. Les échantillons sont incubés 15 min à 37°C.

7.3. Purification et digestion du vecteur de surexpression

Le vecteur utilisé est le p412-EF1-IRES-Puro contenant notamment un promoteur EF1α, un gène de résistance à l'ampicilline et un gène de résistance à la puromycine ainsi que les séquences virales (Figure 52).

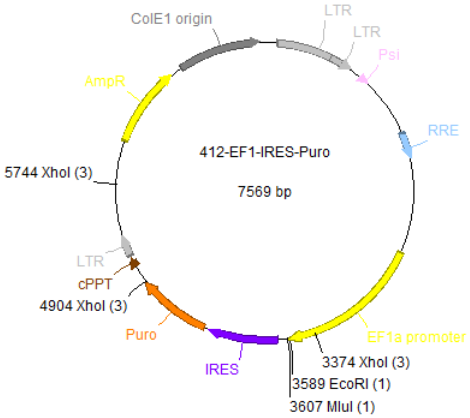


Figure 52 : Carte du plasmide lentiviral de surexpression p412-EF1-IRES-Puro

Le vecteur (50 ng/µL) est linéarisé par digestion avec les mêmes enzymes à 0,025 unité/µL (*Eco*RI/*Mlu*I) dans un tampon adapté 1X (FastDigest, Thermo Fisher Scientific) pour un volume final de 40 µL. Le tout est incubé pendant 15 min à 37°C. Le produit est visualisé par électrophorèse sur gel d'agarose 0,7%. Le vecteur digéré est purifié à l'aide du kit SmartPure (Eurogentec) et dosé au Nanodrop.

7.4. Clonage de la séquence codante dans le vecteur de surexpression

Les produits PCR purifiés et digérés sont clonés dans le vecteur également digéré et purifié à un ratio 2:1 (2 fois plus de vecteur que d'insert). Ce ratio a été défini après optimisation du protocole. Le kit Quick Ligation™ est choisi pour le clonage (New England Biolabs). Les produits PCR digérés et purifiés à 2,5 ng/µL sont mélangés au vecteur digéré à 5 ng/µL et une Quick ligase à 60 unités/µL dans un tampon approprié pour un volume final de 10 µL. Les échantillons sont incubés 15 min à température ambiante.

7.5. Transformation bactérienne et purification du plasmide

Des bactéries TOP10 sont transformées avec les mélanges de ligation suivant le même protocole qu'indiqué précédemment (cf. 6.3). Les plasmides sont extraits et purifiés en utilisant le kit NucleoBond™ Xtra Midi EF (endotoxin free = dépourvu d'entoxine, Macherey-Nagel).

Les vecteurs purifiés à 50 ng/µL sont digérés avec l'enzyme *Xho*I à 0,025 unité/µL(FastDigest, Thermo Fisher Scientific) dans un tampon approprié 1X pour un volume final de 20 µL. Cette étape permet de vérifier l' intégrité du vecteur et l'insertion correcte de la séquence codante de *BCL-RAMBO.*

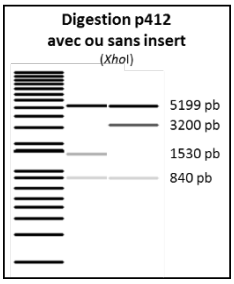


Figure 53 : Profil de digestion attendu du vecteur p412 après insertion ou non de la séquence codante de *BCL-RAMBO* et digestion par l'enzyme *Xho*I

Les plasmides sont également séquencés par Sanger à l'aide de différentes amorces, complémentaires du vecteur ou de la séquence codante de *BCL-RAMBO* (Tableau 23). Les séquences codantes clonées dans le vecteur correspondent au transcrit 1 de *BCL-RAMBO* (exons 2, 3, 5, 7, 8, 9 et 11).

Tableau 23 : Séquences des amorces pour le séquençage du plasmide de surexpression du gène *BCL-RAMBO*

Séquence ciblée	Sens/anti-sens	Séquence amorce
Promoteur EF1α	Sens	TCAAGCCTCAGACAGTGGTTC
Exon 2	Sens	AGGTCACATCTGGACGGAGA
Exon 11	Sens	ACGGAGCTGGAGAGAAGAGT
Exon 11	Anti-sens	GACCACCTCCTCCACTTCAG
3'-UTR	Anti-sens	AAGCAGCAAGTAGCCGCTTC
IRES	Anti-sens	CCGGGCCCTCACATTGC

7.6. Production de lentivirus

Les trois vecteurs (sans insert, avec la séquence codante de référence, avec la séquence codante variante) ont été empaquetés dans des lentivirus par la plateforme de vectorologie de l'unité 1218 ACTION.

7.7. Transduction

Le vecteur de surexpression est intégré au génome des cellules K-562 *via* une transduction de particules virales. Les productions lentivirales contiennent environ 10^6 particules/mL. Des cellules K-562 (1 million de cellules) sont transduites à une MOI

(*multiplicity of infection*) de 1. Après environ 12h d'incubation, les virus sont éliminés par centrifugation des cellules et élimination du surnageant. Les cellules transduites sont alors sélectionnées part traitement à la puromycine (3,5 µg/mL) pendant 48h.

7.8. PCR

L'insertion du vecteur dans l'ADNg des cellules transduites est vérifiée par amplification PCR à l'aide du kit GoTaq® G2 Hot Start Polymerase (Promega) (Tableau 24). L'ADN est extrait à l'aide du kit Phire Tissue Direct PCR Master Mix (Thermo Fisher Scientific) à partir d'environ 100.000 cellules prélevées 5 jours après la transduction. Les amorces utilisées sont complémentaires du promoteur EF1α (5'-TCAAGCCTCAGACAGTGGTTC-3') et de la séquence IRES (5'-CCGGGCCCTCACAATTGC-3'), de part et d'autre de l'insert, sont utilisées. Les produits PCR sont analysés par électrophorèse sur gel d'agarose 1,5%.

Tableau 24 : Mélange réactionnel et programme d'amplification PCR de la région de l'insert du vecteur p412

Réactifs	Volume pour une réaction (µL)	Concentration finale
Tampon (sans $MgCl_2$)	5	1X
$MgCl_2$	1,5	1,5 mM
dNTP	0,5	200 nM
Amorce sens	1	400 nM
Amorce anti-sens	1	400 nM
Polymérase G2 Hot Start	0,125	0,025 unité/µL
Lysat cellulaire	2	
Eau	qsp 25 µL	

Nombre de cycles	Température (°C)	Durée
1	94	3 min
35	94	30 sec
	58	30 sec
	72	2 min 10
1	72	5 min

7.9. RT-PCRq en temps réel

L'expression du gène *BCL-RAMBO* dans les lignées transduites est évaluée par RT-PCRq en temps réel. Les éventuelles contaminations de l'ARN par de l'ADNg sont éliminées selon le même protocole que décrit précédemment (cf 4.3.1). La transcription inverse est réalisée à l'aide du kit LunaScript® RT SuperMix (New England Biolabs). L'ARN traité à la DNase (1 µg) est mélangé au SuperMix 1X dans un volume final de 20 µL. Les échantillons

sont incubés 2 min à 25°C, 10 min à 55°C puis 1 min à 95°C. L'expression du gène *BCL-RAMBO* est normalisée sur l'expression du gène de ménage *GUSB* codant la glucuronidase beta.

8. Apoptose évaluée par cytométrie en flux

8.1. Traitement des cellules à l'imatinib

Les différentes lignées cellulaires modifiées, issues de la lignée K-562, ont été traitées par différentes concentrations d'imatinib (STI571, Gleevec®, Novartis) pendant plusieurs jours. Pour cela, 100.000 cellules dans 500 µL de milieu ont été déposées dans des plaques de 24 puits. Puis 500 µL de milieu contenant différentes concentrations d'imatinib ont été ajoutés (concentrations finales : 0 ; 0,25 ; 0,5 et 1 µM).

8.2. Principe de la cytométrie en flux

La cytométrie en flux permet l'analyse individuelle et multi-paramétrée de cellules. Avant chaque passage au cytomètre, les cellules sont incubées avec des marqueurs spécifiques. La fluorescence émise par les composés fluorescents est capturée et visualisée sur une interface informatique. Les premières analyses (lignées KO, ΔC-term=ΔBHNo+TM) ont été réalisées à l'aide d'un cytomètre BD FACSCanto II (BD Biosciences). Les dernières analyses (lignées HIPER, surexpression de *BCL-RAMBO*) ont été réalisées à l'aide du cytomètre BD Accuri™ C6 (BD Biosciences). Cet appareil ne possédant pas le laser d'excitation à 407 nm, le marquage au DAPI a été remplacé par l'iodure de propidium (IP, excitation à 488 nm).

8.3. $DiOC_6(3)$

Le $DiOC_6(3)$ (3,3'-Dihexyloxacarbocyanine Iodide, Thermo Fisher Scientific) est un marqueur lipophile fluorescent. Cette molécule se lie à la membrane mitochondriale (mais également à la membrane du réticulum endoplasmique) *via* ses groupements hydrophiles.Ce marqueur permet de discriminer les cellules vivantes des cellules en cours d'apoptose qui perdent la fluorescence suite à l'altération du potentiel membranaire mitochondrial.

8.4. Annexine V-APC/DAPI ou IP

L'annexine V est un marqueur spécifique des phosphatidylsérines exposées à l'extérieur de la membrane plasmique lorsque l'apoptose est induite (flip-flop de la membrane). Cette molécule est couplée à l'allophycocianine (APC) pour permettre sa détection par cytométrie en flux. Afin de discriminer les cellules en cours d'apoptose des cellules mortes, un deuxième marquage au DAPI ou IP est réalisé. Ces marqueurs du noyau cellulaire ne pénètrent que dans les cellules perméables donc les cellules mortes.

8.5. Calcul de l'apoptose induite

Le niveau d'apoptose induite est calculé pour chaque lignée cellulaire suivant la formule suivante :

$$\%\ apoptose\ induite = \frac{(\%\ apoptose_{cellules\ traitées} - \%\ apoptose_{cellules\ non\ traitées}\ ) * 100}{(100 - \%\ apoptose_{cellules\ non\ traitées})}$$

9. Apoptose évaluée par western blot

9.1. Traitement des cellules à l'imatinib

L'induction de l'apoptose a également été évaluée *via* l'expression ou l'activation de différentes protéines mais également *via* leur localisation cellulaire. En effet, après traitement à l'imatinib, les cellules ont été prélevées et les mitochondries ont été isolées des autres compartiments cellulaires. Une grande quantité de cellules est nécessaire pour réaliser cette extraction. C'est pour cela que 20 millions de cellules/condition ont été traitées avec 0,5 µM d'imatinib (STI571, Gleevec®, Novartis) pendant 24h. Comme contrôle, 20 millions de cellules ont également été maintenues en culture pendant 24h sans imatinib. Après 24h de traitement, les cellules sont prélevées et comptées. Huit millions de cellules sont isolées pour permettre l'extraction des protéines totales. Le reste des cellules est utilisé pour l'isolement des mitochondries.

9.2. Isolement des mitochondries

Le kit Mitochondria Isolation for Cultured Cells (Thermo Fisher Scientific) a été utilisé afin d'isoler les mitochondries des cellules traitées ou non à l'imatinib. Les différentes fractions cellulaires sont purifiées grâce à leurs différences de densité par gradients discontinus. Le culot cellulaire est tout d'abord mélangé à 800 µL de réactif A (+ inhibiteur de protéases 1X), vortexé pendant 5 secondes puis maintenu dans la glace pendant 2 min. Par la suite, 10 µL de réactif B est ajouté, le tout est de nouveau vortexé à forte vitesse pendant 5 secondes puis maintenu dans la glace pendant 5 min. Ces étapes permettent la lyse des cellules et la libération des organites cellulaires. Le réactif C (+ inhibiteur de protéases 1X) (800 µL) est ensuite ajouté et le mélange est centrifugé 10 min à 700 g à 4°C. Les gros débris cellulaires forment un culot et le surnageant, contenant notamment les mitochondries, est récupéré puis de nouveau centrifugé 15 min à 12.000 g à 4°C. Le surnageant contenant le cytoplasme est isolé et le culot contenant les mitochondries est mélangé à 500 µL de réactif C puis centrifugé 5 min à 12.000 g. Le culot contenant les mitochondries est conservé avant l'extraction des protéines.

9.3. Lysats protéiques

Les lysats totaux sont réalisés en mélangeant 8 millions de cellules à 320 µL d'un mélange de tampon Cell Lysis 1X (Cell Signaling Technology) et de fluorure de phénylméthylsulfonyle 1 mM (PMSF). Après 10 min d'incubation dans la glace, les échantillons sont centrifugés 30 min à 12.000 g à 4°C. Le surnageant contenant les protéines extraites est récupéré. Les lysats mitochondriaux sont réalisés en mélangeant le culot obtenu avec 50 µl de tampon de lyse (SDS 4% m/v, glycérol 20% v/v, Tris-HCl 0,5 M à pH 6,8 10% v/v).

9.4. Dosage à l'acide bicinchoninique

La concentration des échantillons est déterminée en utilisant le kit de dosage Pierce™ BCA Protein Assay (Thermo Fisher Scientific). Chaque échantillon est dilué au 100ème dans un mélange de réactifs contenant notamment de l'acide bicinchoninique (BCA) et des ions Cu^{2+} à pH basique. Les protéines réduisent les ions Cu^{2+} en ions Cu^{+} qui forment un complexe avec le BCA, de couleur violette. Les solutions sont déposées dans une plaque de 96 puits et incubées 45 min à 37°C. L'absorbance de chaque échantillon, à une longueur d'onde d'émission de 595 nm, est déterminée à l'aide d'un spectrophotomètre associé à un lecteur de plaques (iMark,

Bio-Rad). L'absorbance est proportionnelle à la concentration protéique. En parallèle des échantillons, une gamme de concentrations connues d'albumine de sérum bovin (*bovine serum albumin*, BSA) est déposée afin de déterminer précisément la quantité de protéines pour chaque condition. Les lysats protéiques sont dilués à 1 µg/µL dans 20% v/v de tampon de charge (Tableau 25). Les protéines sont dénaturées 5 min à 95°C.

Tableau 25 : Composition du tampon de charge des lysats protéiques

Réactifs	Concentration finale
SDS	10% m/v
Glycérol	50% v/v
Tris-HCl 0,5 M (pH = 6,8)	25% v/v
β-mercaptoéthanol	25% v/v
Bleu de bromophénol	0,013% m/v

9.5. Electrophorèse sur gel de polyacrylamide

Une électrophorèse sur gel de polyacrylamide est réalisée afin de permettre la séparation des protéines, des différents échantillons, selon leur masse moléculaire.

Les gels d'électrophorèse sont divisés en deux parties : le gel de concentration permettant de concentrer tous les échantillons déposés, puis le gel de séparation permettant la migration des protéines selon leur masse. Le gel de séparation est coulé dans un support entre deux plaques de verre et, après polymérisation, le gel de concentration est à son tour coulé dans le support. La composition des gels est indiquée dans le Tableau 26. Selon la taille de la protéine étudiée, le pourcentage d'acrylamide peut varier de 10 à 15%.

Tableau 26 : Composition des gels de concentration et de séparation dilués dans l'eau

Réactifs	Séparation	Concentration
30% Acrylamide/Bis acrylmaide, 37,5:1	10% à 15%	3,75%
Resolving gel buffer 3 M	375 mM	-
Stacking gel buffer 0,5 M	-	125 mM
SDS 20 %	0,1%	0,1%
TEMED	0,06%	0,08%
APS 1,5%	0,075%	0,075%

Les gels sont déposés dans une cuve contenant un tampon de migration (SDS 0,1%, Glycine 192 mM, Tris base 25 mM). Quinze microgrammes de chaque échantillon sont déposés dans les puits formés dans le gel de concentration puis migrent sous un champ électrique de 50 à 75 volts pendant plusieurs heures. Un marqueur de taille (3 µL) est également déposé (Page Ruler, Thermo Fisher Scientific).

9.5.1. Transfert sur membrane de PVDF

Lorsque le front de migration atteint le bas du gel, la migration est stoppée et les protéines intégrées dans le gel sont transférées sur une membrane de polyfluorure de vinylidène (PVDF) préalablement activée au méthanol. Le gel et la membrane sont déposés dans un appareil de transfert entre plusieurs papiers absorbants (papier Whatman, Dutscher) imbibés de différents tampons dilués dans l'eau (Tableau 27). Dans l'ordre, 6 papiers Cathode sont déposés puis le gel d'acrylamide, la membrane de PVDF activée, 2 papiers Anode II et 4 papiers Anode I. Le transfert se fait en 2h sous un champ électrique de 15 volts.

Tableau 27 : Composition des différents tampons de transfert

Réactifs	Tampon Anode I (pH = 10,4)	Tampon Anode 2 (pH = 10,4)	Tampon Cathode
Tris	0,3 M	25 mM	25 mM
Méthanol	20% v/v	20% v/v	20% v/v
Acide 6-aminohexaonique	-	-	40 mM

Les trois tampons sont préparés dans l'eau

9.5.2. Détection des protéines d'intérêt à l'aide d'anticorps

Les sites libres de la membrane sont saturés avec une solution de TBS-T 1X (Tris base 20 mM, NaCl 137 mM, Tween 20 0,1%), avec 5% de lait en poudre, pendant 1h sous agitation à température ambiante. Les anticorps primaires, dilués soit dans une solution de TBS-T lait 5% ou soit dans une solution de TBS-T BSA 10% et azide 0,1%, sont ajoutés et les échantillons sont incubés toute la nuit sous agitation à 4°C (Tableau 28). Le lendemain, les anticorps non fixés sont éliminés avec 4 lavages de 10 min avec du TBS-T 1X puis les anticorps secondaires sont ajoutés (dilués au 1/10.000 dans TBS-T 1X lait 5%). Après 1h30 d'incubation sous agitation à température ambiante, 4 lavages de 10 min sont de nouveau réalisés.

Tableau 28 : Description des anticorps primaires utilisés pour les expériences de western blot

Protéine	Taille (kDa)	Ref. anticorps	Espèce	Solution de dilution	Dilution	Gel
ACTINE β	42	AC-74	Souris	TBS-T Lait 5%	1/10.000	12,5%
BCL-RAMBO	83	6D161	Souris	TBS-T Lait 5%	1/1000	10%
BIM	22	B7929	Lapin	TBS-T Lait 5%	1/2000	15%
Cytochrome c	15	A-8	Souris	TBS-T Lait 5%	1/1000	15%
Caspase-3	32 (clivée : 20+17+11)	E-8	Souris	TBS-T Lait 5%	1/1000	15%
HSP60	60	K-19	Chèvre	TBS-T BSA 10%, azide 0,1%	1/10/000	12,5%
PARP1	116 (clivée : 89+24)	B-10	Souris	TBS-T Lait 5%	1/250	12,5%

9.5.3. Révélation des membranes

La position des anticorps sur la membrane de PVDF est révélée par chimioluminescence. Les membranes sont incubées 5 min avec des réactifs de chimioluminescence (*enhanced chemiluminescence*, ECL, Bio-Rad). Ces molécules réagissent avec la peroxydase de raifort liée aux anticorps secondaires. La position des anticorps secondaires et donc l'emplacement des protéines d'intérêt est révélé selon un principe de révélation photographique en chambre noire.

Résultats

Partie I : Identification de variants génétiques par une approche NGS

1. Séquençage nouvelle génération de panels de gènes

Au cours de ce projet de thèse un total de 95 gènes, susceptibles d'être associés à la résistance à l'imatinib, a été séquencé par NGS à partir d'échantillons au diagnostic de 24 patients atteints de LMC, 12 répondeurs et 12 résistants.

Un premier groupe de 48 gènes codant des facteurs impliqués dans la pharmacocinétique, le cycle cellulaire, la réparation de l'ADN ou encore les voies des cytokines, a été étudié. Au cours de la thèse, des analyses informatiques et statistiques ont été développées pour analyser les résultats de NGS obtenus à partir d'un petit nombre d'échantillons. Ces nouvelles approches ont été testées et mises au point grâce à ce premier groupe de 48 gènes. Cette étude fait l'objet d'une première publication soumise à Human Genomics (Annexe 1). L'analyse des variants génétiques mis en évidence par le séquençage de ces 48 gènes ne sera pas détaillée dans la partie Résultats. De même, les gènes *BCR* et *ABL1* ont été séquencés. Aucun variant n'a été identifié dans le domaine kinase d'*ABL1*. L'analyse détaillée du séquençage de ces deux gènes ne sera pas détaillée dans la partie Résultats.

Le travail de thèse s'est concentré sur un groupe de 45 gènes impliqués dans la régulation de l'apoptose. Les résultats obtenus pour ce panel de gènes sont détaillés ci-dessous.

2. Contrôle qualité de la réaction de séquençage

L'objectif de ce projet de thèse était d'identifier des variants génétiques pouvant être associés à la résistance à l'imatinib dans la LMC. Pour cela, une expérience de NGS à partir d'échantillons au diagnostic de 24 patients atteints de LMC, 12 résistants et 12 répondeurs au traitement, a été réalisée. Le séquençage a été restreint aux exons codants, jonctions intron-exon et régions promotrices de 45 gènes impliqués dans la régulation de l'apoptose (21 gènes de la famille *BCL2* et 24 gènes codant des kinases et facteurs de transcription). Cette méthode est plus abordable que le séquençage de génomes ou d'exomes entiers et permet donc d'analyser un nombre plus important de patients en une seule réaction de séquençage. Ces régions spécifiques sont capturées à l'aide de sondes complémentaires des séquences d'intérêt au cours de la création des librairies de séquençage.

A l'issue du NGS, 24 listes de séquences de 150 pb (reads) sont obtenues et la qualité de lecture des fragments d'ADNg est évaluée. Un pourcentage important de séquences de qualité élevée est obtenu avec près de 610 Mb (soit environ quatre millions de reads) présentant un score de qualité (score Phred) >30 (moins d'une erreur pour 1000 bases lues). Ces séquences représentent 95% de l'ensemble des bases détectées (Figure 54). De plus, la répartition entre les 24 patients est relativement uniforme, les résultats pour l'ensemble des patients peuvent donc être comparés (Figure 55). Le séquençage a été efficace et les résultats sont fiables. Les différences de séquence entre les patients et le génome de référence peuvent donc être détectées.

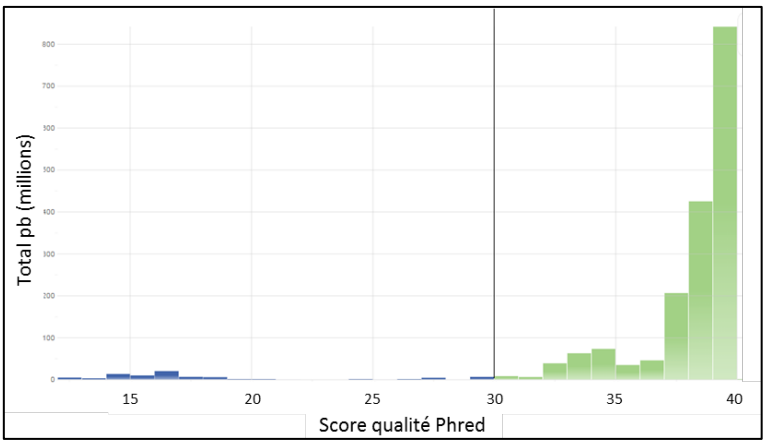


Figure 54 : Score qualité Phred des bases identifiées par NGS

Le seuil est placé à 30 (1 erreur pour 1000 bases lues).

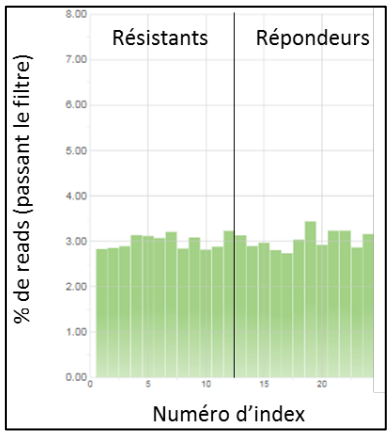


Figure 55 : Pourcentage de reads de haute qualité pour chaque patient (un patient = un index)

Index 1 à 12, patients résistants ; index 13 à 24, patients répondeurs

Pour cela, les reads sont alignés au génome humain de référence et un contrôle qualité est de nouveau réalisé. La profondeur de lecture, c'est-à-dire le nombre de fois où une même base a été séquencée, est supérieure à 35X pour plus de 90% des positions ciblées. La

profondeur de lecture moyenne est d'environ 100X (exemple du séquençage des patients résistants, Figure 56) cependant, pour certaines régions de taille réduite, couverte par une sonde unique, la profondeur de lecture était bien inférieure (exemple du gène *BIK*, Figure 57).

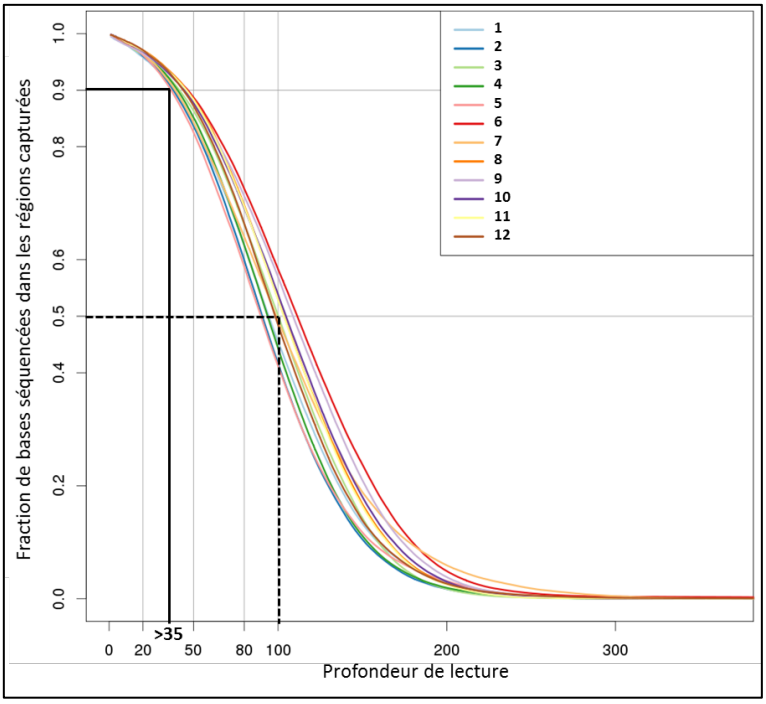


Figure 56 : Fraction de bases séquencées dans les régions d'intérêt en fonction de la profondeur de lecture

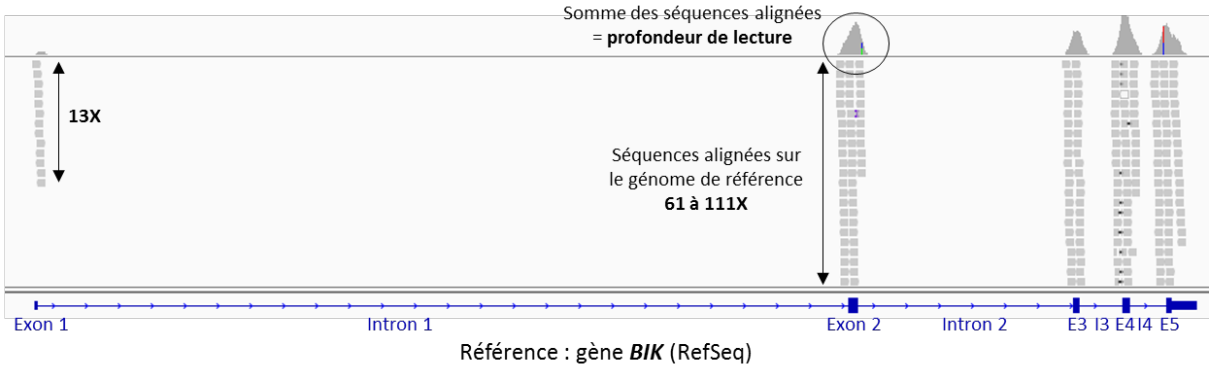


Figure 57 : Exemple de séquençage du gène *BIK* visualisé sur le logiciel IGV

Chaque rectangle gris correspond à un read de 150 pb. Cinq régions ont été capturées et séquencées et correspondent aux 5 exons du gène *BIK*. La profondeur de lecture de l'exon 1 est inférieure aux profondeurs de lecture des exons suivants.

Au final, un panel de 45 gènes associés à la régulation de l'apoptose a été séquencé par NGS à partir de cellules leucocytaires prélevées au diagnostic pour 24 patients atteints de LMC, répondeurs et résistants au traitement à l'imatinib. L'extraction de l'ADNg et la génération des

librairies étant de qualités élevées, peu d'erreurs de lecture ont été détectées par l'appareil de séquençage. De plus, la plupart des régions d'intérêt ont été séquencées de nombreuses fois (en moyenne >100X) ce qui augmente la fiabilité des résultats. Les différences de séquence entre les patients et le génome de référence peuvent donc être détectées avec une confiance importante dans les résultats.

3. Détection et annotation des variants génétiques

A partir des échantillons d'ADNg obtenus pour 24 patients atteints de LMC, 810 polymorphismes ont été mis en évidence. Pour 330 de ces polymorphismes la profondeur de lecture était inférieure à 30X pour tous les patients. Les variations identifiées à ces positions pourraient être des erreurs de séquençage et c'est pourquoi elles ont été exclues de l'analyse. Au final, 480 polymorphismes ont donc été conservés. Ces variations de séquence peuvent être visualisées facilement par le logiciel IGV (exemple du gène *BCL2A1*, Figure 58).

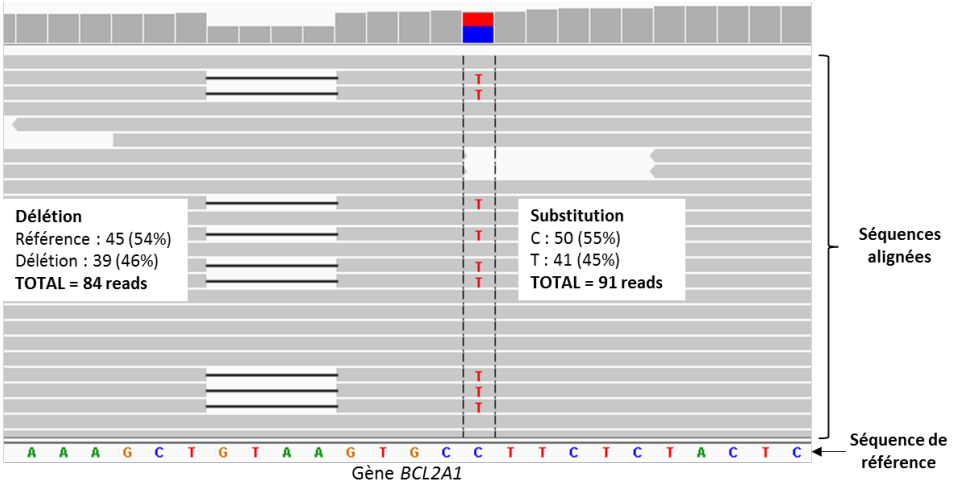


Figure 58 : Exemple de polymorphismes (délétion et substitution) détectés dans le gène *BCL2A1* visualisés sur le logiciel IGV

Près de 88% des variants sont des substitutions d'une paire de bases (SNP). Les autres polymorphismes sont des petites insertions (5,8%) ou délétions (6,3%) (Figure 59). La plus grande délétion identifiée par cette approche de séquençage est de 18 pb et la plus grande insertion est de 14 pb. Les polymorphismes sont retrouvés à 39,6% dans les exons, 0,4% au niveau de sites d'épissage et 3,1% dans les promoteurs. Les 56,9% restants correspondent à des polymorphismes retrouvés dans des régions introniques bien que les sondes n'aient pas été conçues pour capturer ces régions non codantes (Figure 60).

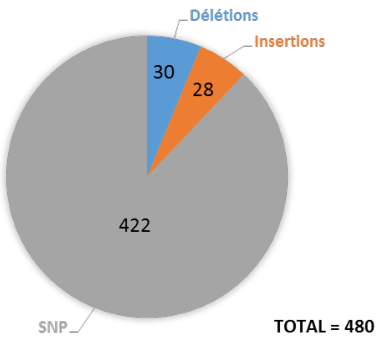
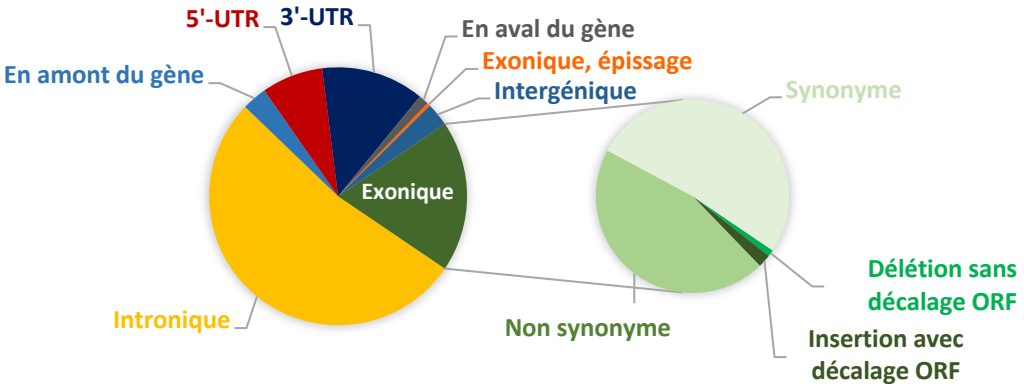


Figure 59 : Nature des 480 polymorphismes identifiés (SNP, délétions ou insertions)



Localisation des polymorphismes	Nombre	%
Intronique	253	52,7
En amont du gène	15	3,1
5'-UTR	37	7,7
3'-UTR	62	12,9
En aval du gène	6	1,3
Exonique, épissage	2	0,4
Intergénique	14	2,9
Exonique	91	19,0
Délétion sans décalage ORF	1	0,2
Insertion avec décalage ORF	2	0,4
Non synonyme	41	8,5
Synonyme	47	9,8
TOTAL	480	100,0

Figure 60 : Localisation des 480 polymorphismes identifiés

L'impact des variants exoniques sur la séquence protéique est également détaillé (non-synonyme, synonyme, indels avec ou sans décalage du cadre de lecture (*open reading frame*, ORF)).

La capture de variants dans les régions non ciblées est due à un biais dans l'étape de la capture des séquences d'intérêt. En effet, la librairie de séquençage est constituée de fragments d'environ 300 pb. Il est donc certain qu'un grand nombre de fragments comprenait à la fois des régions codantes et des régions non codantes, capturées par les sondes biotinylées de 120 pb.

Ceci se vérifie facilement grâce à la visualisation de l'alignement des séquences sur le logiciel IGV (exemple du gène *GRK2*, les reads recouvrent sans interruption toute la région de l'exon 7 à l'exon 15 du gène, Figure 61).

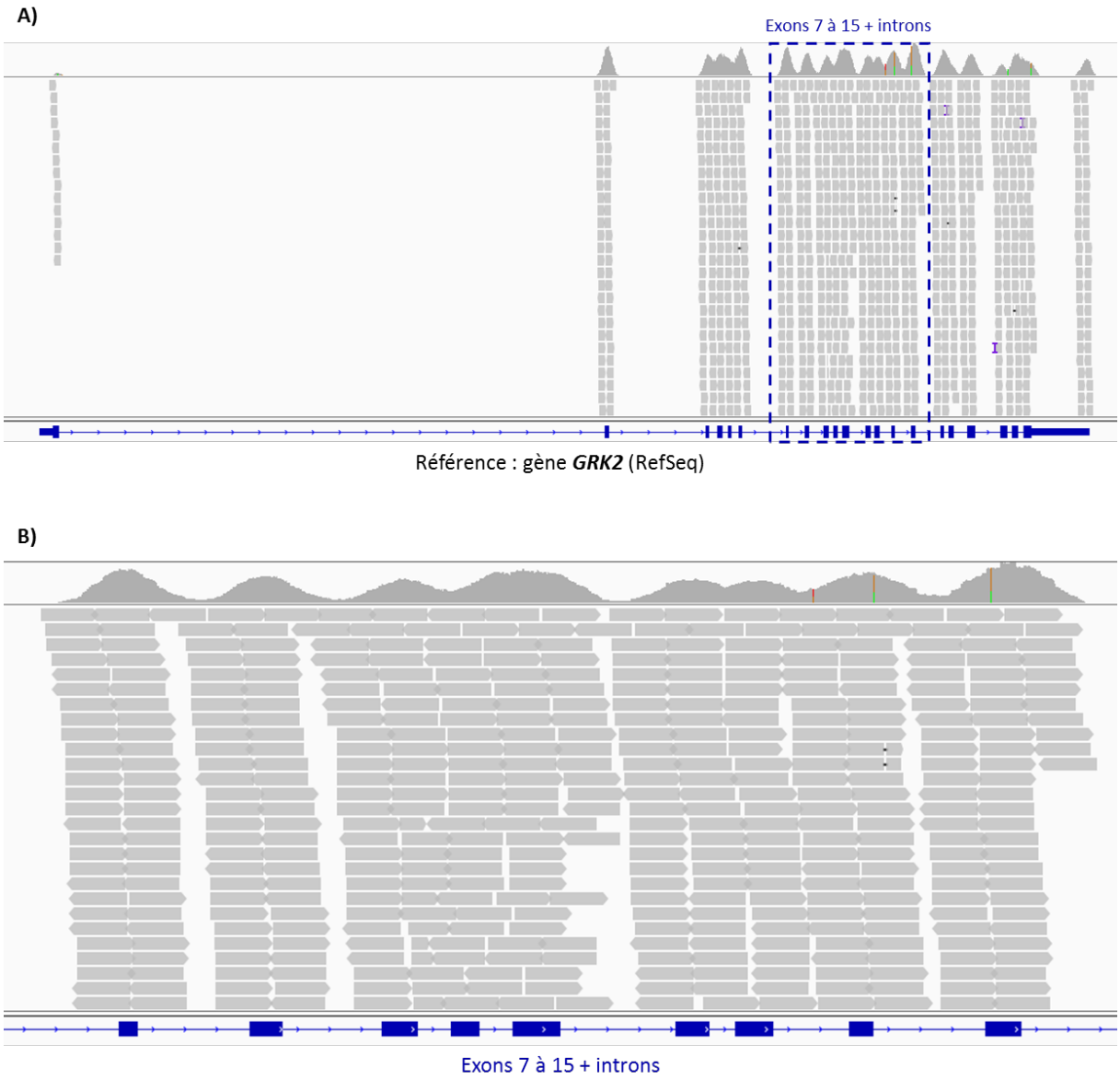


Figure 61 : Exemple de séquençage du gène *GRK2* visualisé sur le logiciel IGV

A) Aperçu global du gène, B) Agrandissement de la région englobant les exons 7 à 15 et les introns en alternance. La totalité de cette séquence est couverte par des reads, régions codantes et non codantes.

Enfin, les 480 polymorphismes sont retrouvés dans 43 des 45 gènes d'intérêt. Aucun variant n'a été retrouvé pour aucun patient dans les gènes *HRK* et *BCL2L1*. Parmi les 480 polymorphismes, 52 (dont 29 indels) ne sont pas répertoriés dans la base de données dbSNP. Il peut s'agir de nouveaux variants ou bien d'une erreur d'annotation des variants. En effet, la

position des polymorphismes (et notamment les indels) peut varier entre le logiciel d'annotation et la base de données dbSNP. De plus, malgré le contrôle qualité, ces variants pourraient également être des artefacts de lecture. C'est pourquoi un séquençage Sanger complémentaire est nécessaire pour tous les polymorphismes candidats identifiés.

Au final, le séquençage de 45 gènes associés à la régulation de l'apoptose chez 24 patients atteints de LMC a permis de mettre en évidence 480 polymorphismes génétiques. Un ou plusieurs de ces variants pourraient être à l'origine du manque de sensibilité des cellules à l'imatinib et donc impliqués dans la résistance au traitement.

4. La population générale : un groupe de référence

Afin d'identifier les polymorphismes génétiques les plus susceptibles d'être associés à la résistance à l'imatinib, la répartition des variants observée pour 24 patients atteints de LMC a été comparée à la répartition estimée de ces mêmes polymorphismes dans la population générale. Pour cela, les MAF des 480 polymorphismes candidats ont été recherchées dans la base de données 1000G.

Avant de poursuivre l'analyse des données, une correction des données a dû être réalisée. En effet, parmi les 480 polymorphismes, 77 possédaient une MAF supérieure à 0,5. Or, par définition, la fréquence de l'allèle variant doit être inférieure à 50%. Pour ces polymorphismes, les fréquences alléliques ont été artificiellement modifiées (MAF modifiée = 1-MAF de la base 1000G) et les génotypes identifiés ont été inversés pour chaque patient. Un patient homozygote de référence est devenu homozygote variant et inversement. Suite à ces ajustements, sept polymorphismes ne présentaient finalement aucun allèle variant chez aucun patient. Ces artefacts, dus au génome de référence, ont donc été exclus de l'étude (Figure 62).

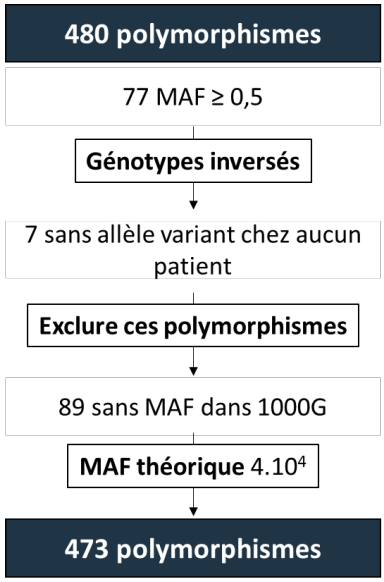


Figure 62 : Ajustement des génotypes observés par NGS en fonction de la MAF décrite dans la base de données 1000G

Un deuxième élément a été pris en compte pour la suite de l'étude. Parmi les 480 polymorphismes, 89 n'étaient pas inclus dans la base de données 1000G. Cela signifie qu'ils n'ont jamais été identifiés chez les 2504 individus inclus dans leur étude. Pour réaliser des analyses statistiques, la valeur nulle ne pouvait pas être conservée. Une MAF arbitraire a donc été définie. La MAF théorique a été choisie selon l'estimation suivante : un polymorphisme sans MAF n'a pas été séquencé dans le projet 1000G, il est donc présent pour moins d'un individu sur 2504 ($1/2504=4.10^{-4}$). A partir de ces nouvelles données, la répartition théorique dans la population générale (groupe artificiel de 12 individus) des 473 polymorphismes restants a été estimée en appliquant le principe d'Hardy-Weinberg (Figure 63).

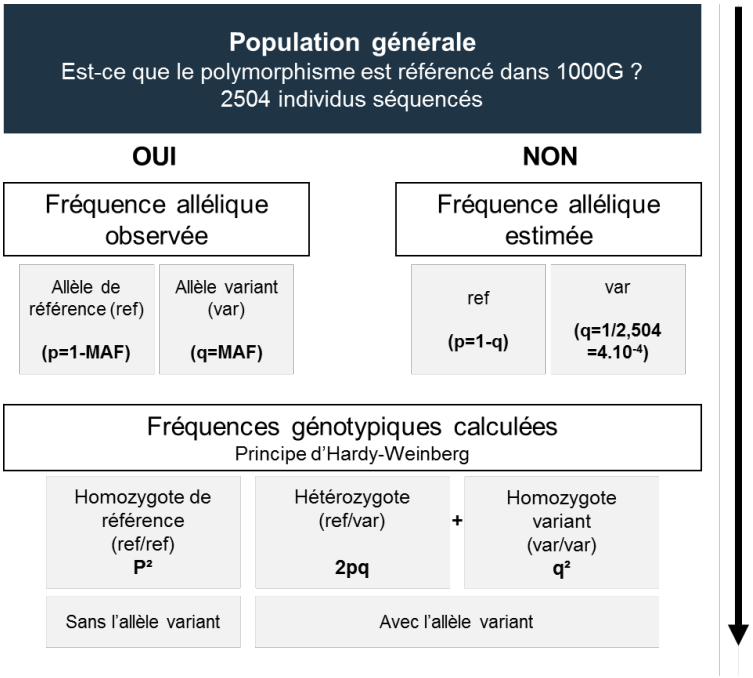


Figure 63 : Simulation d'un groupe de référence à partir des fréquences alléliques référencées dans 1000G ou estimées

Cette approche de séquençage ciblé a permis de mettre en évidence 473 variants présents chez au moins un des 24 patients atteints de LMC, répondeur ou résistant à l'imatinib. Grâce à la base de données 1000G, un troisième groupe d'individus représentant la population générale a été simulé. La distribution de ces polymorphismes est comparée entre les trois groupes dans le but de mettre en évidence les variants les plus susceptibles d'être associés à la résistance au traitement à l'imatinib.

Partie II : Mise en évidence de polymorphismes candidats par de nouvelles approches statistiques

1. Analyser des données volumineuses de NGS obtenues à partir d'un nombre réduit d'échantillons

Afin de mettre en évidence des polymorphismes candidats, la distribution des variants entre les patients répondeurs, les patients résistants et les individus dans la population générale est étudiée par des approches statistiques. Le nombre d'échantillons séquencés étant limité (2 x 12), les statistiques inférentielles classiques ne peuvent pas être appliquées pour analyser les résultats. Cependant, la quantité de données générées étant élevée (473 polymorphismes annotés), une automatisation de l'analyse des résultats était nécessaire. Des approches statistiques alternatives ont donc été développées. Cette nouvelle méthodologie a été mise au point à partir d'un panel de 48 gènes impliqués dans différents processus cellulaires pouvant être impliqués dans la sensibilité à l'imatinib. Les exons codants, les jonctions intron-exon et les régions promotrices de ces gènes ont été séquencés par NGS à partir des mêmes échantillons de patients atteints de LMC. Les résultats de cette analyse font l'objet d'une publication soumise à Human Genomics (Annexe 1).

2. Visualisation des données sur un graphique en deux dimensions

Dans un premier temps, la distribution des 473 polymorphismes est étudiée selon les fréquences de survenue des variants observées dans les trois groupes (patients résistants, patients répondeurs, population générale) déterminées à partir des génotypes observés ou simulés. Pour cette étude, les hétérozygotes et les homozygotes variants ont été classés dans une même catégorie : les individus porteurs de l'allèle variant.

Les données sont visualisées sur un graphique en deux dimensions en appliquant la technique graphique multivariée d'analyse factorielle des correspondances (AFC). A la différence des analyses multivariées classiques, il s'agit d'une méthode sans *a priori* et aucune hypothèse préalable n'est posée. Cette approche est généralement utilisée dans les études épidémiologiques afin de mettre en avant une corrélation entre deux éléments (ex : âge et motricité [228]). Sur le graphique, chaque marqueur représente un polymorphisme et selon sa position sur le graphique, le variant est plus ou moins fréquent dans l'une ou l'autre des populations considérées. Une classification hiérarchique sur composantes principales (HCPC) a permis de générer quatre groupes de polymorphismes présentant les caractéristiques les plus

similaires (Figure 64) (la liste des 473 polymorphismes, la fréquence de survenue du variant chez les patients ou dans la population générale ainsi que le numéro de cluster associé sont indiqués en Annexe 3).

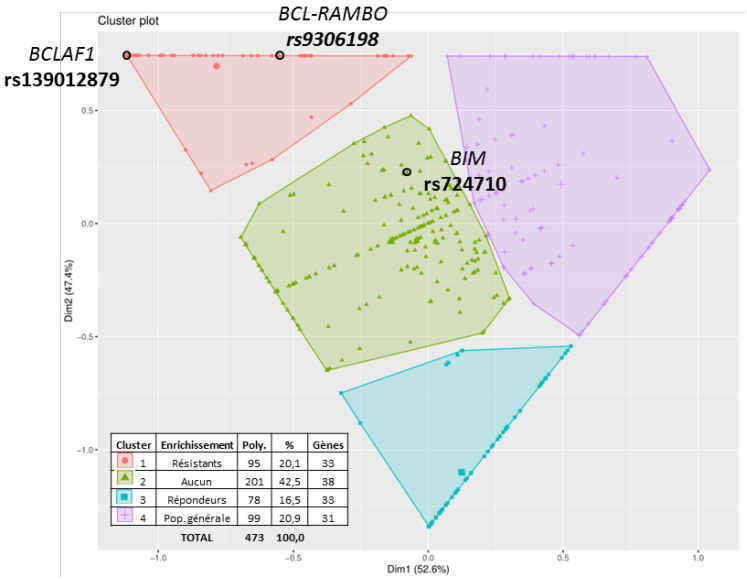


Figure 64 : Représentation des AFC et HCPC réalisées sur les données de NGS

Quatre clusters ont été établis selon la fréquence de survenue du variant. Le variant est plus fréquent parmi 1) les patients résistants, 3) les patients répondeurs, 4) les individus de la population générale. Le cluster 2) englobe des polymorphismes sans différence notable d'enrichissement entre les trois groupes d'individus.

Les variants rs9306198 (*BCL-RAMBO*) et rs139012879 (*BCLAF1*) ont fait l'objet d'une étude plus approfondie (voir ci-dessous).

On retrouve par exemple dans le cluster 2, le polymorphisme rs724710 de *BIM* associé dans une précédente étude à la résistance à l'imatinib dans la LMC [186]. L'allèle variant a été trouvé chez 9 patients résistants (8 hétérozygotes, 1 homozygote variant) (75%) et 5 patients répondeurs au traitement à l'imatinib (hétérozygotes) (42%). Selon la base de données 1000G, la MAF de ce polymorphisme est de 0,69. Il s'agit donc d'un exemple de MAF>0,5. La fréquence de l'allèle minoritaire (T) dans la population générale doit donc être d'environ 31%. D'après ces résultats, il est peu probable que ce variant soit associé à la résistance à l'imatinib dans la LMC.

Il est intéressant de noter également que parmi les 473 polymorphismes identifiés, 105 n'ont été retrouvés chez aucun répondeur (et au maximum chez 5 résistants) et à l'inverse, 98 variants n'ont été retrouvés chez aucun résistant (et au maximum chez répondeurs) (Annexe 3).

L'AFC permet de visualiser rapidement les corrélations entre plusieurs variables. Par cette approche, une liste de 95 polymorphismes candidats (parmi 33 gènes) a été établie. La fréquence de survenue du variant pour ces polymorphismes est supérieure pour les patients atteints de LMC, résistants au traitement, comparée aux patients répondeurs ou aux individus dans la population générale. Cette méthodologie a donc permis de réduire la liste de 473 à 95 polymorphismes susceptibles d'être associés à la résistance au traitement à l'imatinib chez des patients atteints de LMC. Aucune hypothèse préalable ou calcul de valeur statistique n'est nécessaire pour appliquer cette approche de statistique descriptive.

3. Identification de gènes candidats par l'approche des rank products

Une deuxième approche a été développée afin de réduire le nombre de gènes candidats. Cette fois-ci les polymorphismes ne sont pas considérés individuellement mais en combinaison au sein d'un même gène. L'approche des rank products (RP) permet de mettre en évidence les gènes pour lesquels la moyenne des fréquences de survenue du variant de tous ces polymorphismes est supérieure pour les patients résistants comparés aux patients répondeurs et à la population générale. Cette approche a été développée à l'origine pour l'analyse des données de puces à ADN afin de classer les gènes selon les différences d'expression entre deux conditions [223,224]. Pour cette analyse, une sélection des polymorphismes a été réalisée en fonction de l'impact des variants sur la séquence protéique. Les polymorphismes conservés sont non-synonymes, aux sites de jonction intron-exon ou à l'origine d'un décalage du cadre de lecture. A l'issu de ce tri, 42 polymorphismes parmi 23 gènes ont été sélectionnés (liste en Annexe). Les gènes ont été classés selon leur enrichissement en allèles variants parmi les patients atteints de LMC. La taille du gène n'influence pas les résultats et cela a été vérifié par deux approches. Pour un même gène, nous avons réalisé soit la moyenne de la fréquence de survenue de tous ses variants, soit la somme des fréquences divisée par la taille du gène. Dans les deux cas, les gènes les plus enrichis les patients résistants sont identiques. Les cinq premiers gènes, avec une p-value <0,1, sont *BIM*, *GRK2*, *BLK*, *BCL-RAMBO* et *BCLAF1* (Tableau 29).

Tableau 29 : Liste des cinq premiers gènes contenant le plus de variants altérant la protéine parmi les patients résistants selon la méthode RP

Gène	RP/Rsum[a]	FC:(class1/class2)[b]	pfp[c]	P-value
BIM	1,189	7,143000e-38	003607	1,568e-05
GRK2	1,682	2,966000e-37	024680	2,146e-04
BLK	3,224	9,464000e-02	0,0849800	1,108e-02
BCL-RAMBO	4,229	2,387000e-01	0,2246000	3,906e-02
BCLAF1	5,233	6,308000e-01	0,4127000	8,972e-02

[a]RP/Rsum : probabilité pour chaque gène d'être classé premier dans tous les échantillons (des deux conditions). Plus cette valeur est faible, plus la différence entre les deux conditions est importante

[b]FC:(class1/class2) : « fold-change » calculé selon la moyenne des « niveaux d'expression » dans chaque condition

[c]pfp : pourcentage de fausse prédiction.

Le gène le plus enrichi en allèles variants chez les patients résistants comparés aux patients répondeurs et à la population générale est *BIM*. BIM est un facteur majeur dans l'induction de l'apoptose et son inhibition, par des variants génétiques par exemple, peut conduire au développement oncogénique et à la résistance au traitement [229]. BIM est d'ailleurs une cible thérapeutique privilégiée pour favoriser l'élimination des cellules cancéreuses. Le deuxième gène classé par l'approche RP est *GRK2* et son rôle dans la régulation de l'apoptose est plus énigmatique. Une détection du niveau d'apoptose de cellules du myocarde dans des modèles murins n'exprimant plus GRK2 a été réalisée en 2013 [168]. Les résultats suggèrent que GRK2 aurait une fonction pro-apoptotique en régulant les voies PI3K/AKT et MAPK ou en inhibant directement les facteurs anti-apoptotiques tels que BCL2 ou BCLX. Le troisième gène, *BLK*, est un oncogène codant une kinase de la famille SRC. Une surexpression de protéines kinases SRC telles que HCK ou LYN a déjà été associée à une résistance à l'imatinib dans la LMC [230]. BLK pourrait donc être impliqué dans la régulation de l'apoptose *via* BCR-ABL1. Le quatrième gène est *BCL-RAMBO,* de la famille *BCL2*, codant une protéine dont le rôle exact dans la régulation de l'apoptose n'est pas connu. Ce gène a été sélectionné pour la suite de l'étude et une revue détaillée de ce gène et sa fonction probable est proposée ci-dessous. Enfin, le gène *BCLAF1*, a été identifié par l'approche RP comme pouvant être associé à la résistance à l'imatinib dans la LMC. Ce gène code un facteur de transcription réprimant l'expression de gènes anti-apoptotiques de la famille *BCL2* [231].

L'analyse des RP permet de classer les gènes en comparant leur enrichissement en variants génétiques entre plusieurs groupes. Cette méthode a permis d'identifier une nouvelle liste, de 5 gènes susceptibles d'être associés à la résistance à l'imatinib dans la LMC car enrichis en variants chez les patients résistants par rapport aux patients répondeurs ou à la population générale.

Une bonne corrélation entre les deux méthodes statistiques a été identifiée. En effet, la majorité des variants exoniques non-synonymes identifiés parmi les 5 gènes les mieux classés par RP (8/10 variants inclus dans l'analyse) est retrouvée dans le cluster « des patients résistants » (AFC, cluster 1, Figure 64). Afin d'utiliser ces approches innovantes pour l'analyse des données issues de NGS, un outil informatique a été créé (description de l'interface en Annexe 2). Les polymorphismes identifiés peuvent être groupés et représentés sur un graphique en deux dimensions (AFC et HCPC). Les gènes séquencés peuvent être classés selon leur enrichissement en allèle variant chez les patients répondeurs ou résistants au traitement (simulation et RP).

4. Gènes et polymorphismes candidats

A l'aide des résultats obtenus par ces approches statistiques et après une revue de la littérature, le travail s'est concentré sur trois gènes différents : *BCL-RAMBO,* et le variant rs9306198 plus fréquent chez les patients résistants, *BCLAF1,* et le variant rs139012879 identifié uniquement chez un patient résistant, et *BCL2A1* avec le variant rs3826007, susceptible d'altérer la fonction de la protéine et donc la régulation de l'apoptose.

4.1. *BCL-RAMBO*

Le gène *BCL2L13*, ou *BCL-RAMBO*, est un membre de la famille *BCL2.* Ce gène, localisé sur le bras court du chromosome 22 (22q11.21), est divisé en 11 exons et code 13 transcrits différents. La protéine BCL-RAMBO présente une structure globale similaire aux autres facteurs de la famille BCL2 et est ancrée, dans les cellules mammifères, dans la membrane externe mitochondriale. L'isoforme majoritaire comporte 485 acides aminés avec les 4 domaines BH, ainsi qu'un domaine C-terminal d'ancrage à la membrane mitochondriale. Un cinquième domaine protéique a été identifié. Il s'agit d'une séquence de 250 acides aminés présente entre le domaine BH2 et le domaine C-terminal. Cette région, nommée BHNo (*no domain BH*) contient deux séquences d'acides aminés courtes répétées en tandem (RTA et RTB). Ce domaine ne présente aucune homologie apparente avec d'autres domaines connus de

protéines humaines ou d'autres organismes. La partie C-terminale du domaine BH2, le domaine BHNo ainsi que le domaine C-terminal sont tous codés par le dernier exon (n°11) conservé entre tous les transcrits (Figure 65).

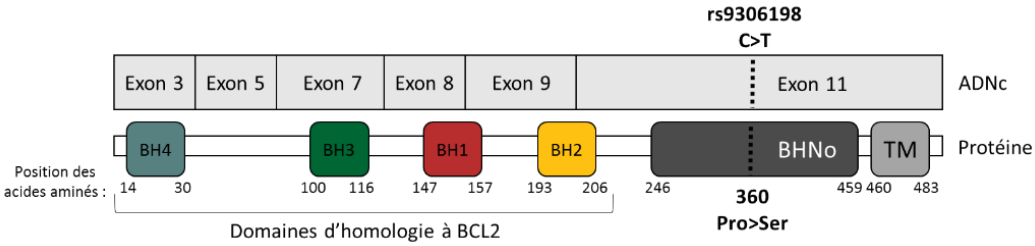


Figure 65 : Schéma de la séquence codante de *BCL-RAMBO* (ADNc) et de la protéine associée

BHNo : no domain BH, TM : domaine transmembranaire, Pro : proline, Ser : sérine

La protéine est exprimée dans de nombreux tissus (UniProt : Q9BXK5). De plus, en se basant sur des données publiées de puces à ADN, *BCL-RAMBO* est fortement exprimé dans la LMC par rapport à d'autres maladies hématologiques (Figure 66). Son rôle exact dans la régulation de l'apoptose est controversé. Peu de publications concernant ce facteur sont disponibles et, selon les modèles étudiés, une fonction pro- ou anti-apoptotique est suggérée.

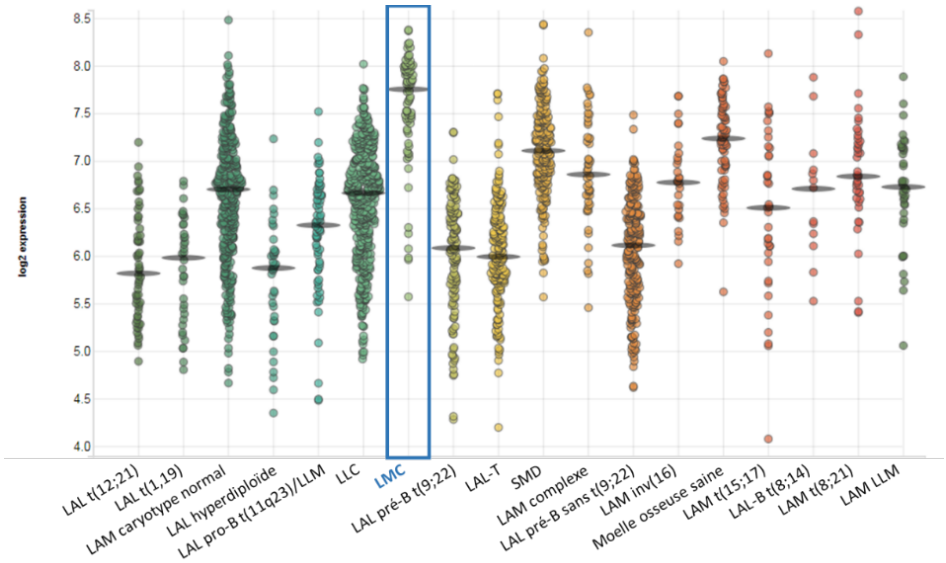


Figure 66 : Expression de l'ARNm de *BCL-RAMBO* dans diverses maladies hématologiques

LAL : leucémie aiguë lymphoblastique, LAM : leucémie aiguë myéloïde, LLM : leucémie de lignée mixte, LLC : leucémie lymphoïde chronique, LMC : leucémie myéloïde chronique, SMD : syndrome myélodisplasique

BloodSpot, Bagger *et al.*, Nucleic Acids Research, 2019 [232]

Trente-cinq polymorphismes dans ce gène ont été mis en évidence chez les 24 patients atteints de LMC (Figure 67). Le variant non-synonyme rs9306198 (C>T), situé au milieu de la séquence codante du domaine BHNo, est retrouvé chez quatre patients résistants (33%), aucun patient répondeur et à une fréquence de 5% dans la population générale selon 1000G (Tableau 30).

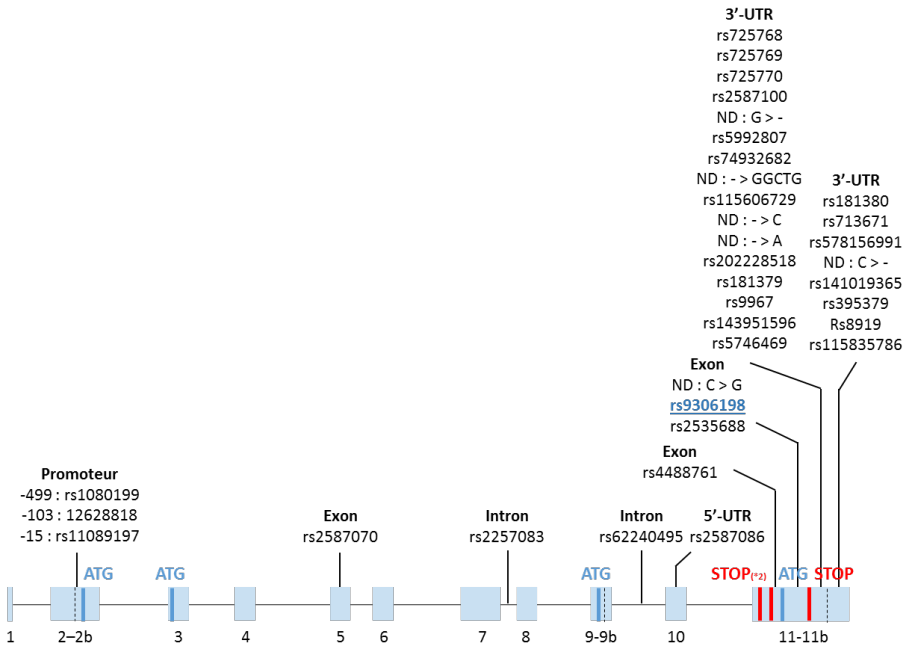


Figure 67 : Schéma du gène *BCL-RAMBO* et des polymorphismes identifiés par NGS

Le gène de 102 kb contient 11 exons représentés par des rectangles bleus.

Tableau 30 : Caractéristiques du polymorphisme rs9306198 de *BCL-RAMBO*

ID dbSNP	Position hg19	Position gène	Variant	Acide aminé	MAF 1000G	Répondeurs	Résistants
rs9306198	22:18209920	Exonique	C>T	Pro>Ser	0,0513	0	4

Ce premier gène a donc été sélectionné pour la suite de l'étude car le variant rs9306198 a été retrouvé enrichi chez les patients résistants et pourrait donc être associé à une dérégulation de l'apoptose et donc intervenir dans la sensibilité au traitement.

4.2. *BCLAF1*

Le deuxième gène sélectionné, *BCLAF1* ou *BTF*, est un facteur de transcription pro-apoptotique agissant notamment en inhibant l'expression des gènes anti-apoptotiques tels que *BCL2* et *BCLX*. Ce facteur est lui-même régulé par certaines protéines anti-apoptotiques de la famille BCL2 en étant séquestré dans le cytoplasme [170]. Ce gène, divisé en 16 exons, est

localisé sur le bras long du chromosome 6 (6q23.3). De manière notable, cette région est souvent délétée dans différentes formes de cancers, en particulier les lymphomes et les leucémies [233]. Ce gène pourrait donc agir comme un suppresseur de tumeur et participer à la résistance à l'imatinib dans la LMC.

Cinq polymorphismes ont été détectés dans ce gène par NGS (Figure 68). Le variant rs139012879 (C>T) est retrouvé chez deux patients résistants (17%) et aucun patient répondeur. Ce variant n'est pas référencé dans la base de données 1000G et sa fréquence dans la base de données ExAC est très faible (T=0475) (Tableau 31).

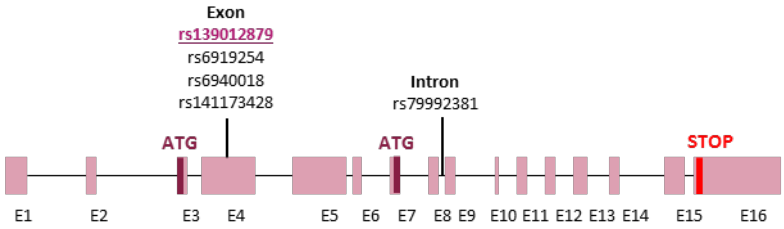


Figure 68 : Schéma du gène *BCLAF1* et des polymorphismes identifiés par NGS

Le gène de 33 kb contient 16 exons représentés par des rectangles roses.

Tableau 31 : Caractéristiques du polymorphisme rs139012879 de *BCLAF1*

ID dbSNP	Position hg19	Position gène	Variant	Acide aminé	MAF 1000G	Répondeurs	Résistants
rs139012879	6:136599585	Exonique	C>T	Arg>His	-	0	2

Ce variant non-synonyme rare est donc retrouvé enrichi chez les patients atteints de LMC, résistants au traitement à l'imatinib. Les bases de données Polyphen-2, Mutation Assessor et Mutation Taster, prédisent une altération de la protéine en présence de ce variant. Le gène *BCLAF1* a donc été sélectionné pour la suite de l'étude.

4.3. *BCL2A1*

A l'inverse, les polymorphismes candidats, susceptibles d'expliquer la sensibilité au traitement, peuvent également être sélectionnés parmi les variants enrichis chez les patients répondeurs. Par exemple, un variant perte de fonction d'un gène anti-apoptotique pourrait être associé à un niveau d'apoptose plus élevée et donc une meilleure réponse au traitement.

Le troisième gène sélectionné, *BCL2A1*, a été sélectionné selon ces critères, avant le développement des approches statistiques. Plusieurs arguments suggéraient une implication de ce facteur dans la réponse des patients à l'imatinib, il a donc été inclus dans les analyses.

Le gène *BCL2A1* est un facteur anti-apoptotique appartenant à la famille *BCL2* souvent désigné *A1*. Ce gène composé de 3 exons est situé sur le chromosome 15 (15q25.1). Deux transcrits de ce gène sont connus. Le plus fréquent est composé des exons 1 et 3 et code une protéine de 175 acides aminés. Le deuxième transcrit identifié contient les 3 exons (codon stop prématuré dans l'exon 3) et code une protéine de 163 acides aminés. Quatre domaines protéiques BH sont identifiés mais aucun segment C-terminal transmembranaire n'est défini. BCL2A1 interagit avec la protéine pro-apoptotique BIM et inhibe son activité *via* son domaine BH1. De plus, la protéine est majoritairement exprimée dans les cellules de la lignée hématopoïétique et est surexprimée dans plusieurs cancers [156].

Huit polymorphismes ont été détectés dans ce gène par NGS (Figure 69). Six variants en déséquilibre de liaison (il existe une association préférentielle entre ces allèles) ont été séquencés et détectés comme enrichis chez les patients répondeurs comparés aux patients résistants (7 contre 2 individus porteurs de l'allèle variant) (Tableau 32).

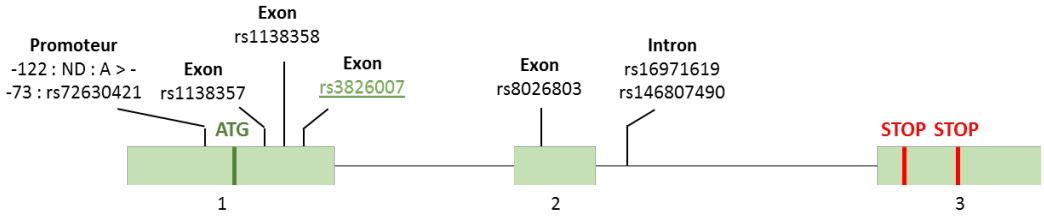


Figure 69 : Schéma du gène *BCL2A1* et des polymorphismes identifiés par NGS.

Le gène de 10 kb contient 3 exons représentés par des rectangles verts.

Tableau 32 : Caractéristiques de six polymorphismes en déséquilibre de liaison dans le gène *BCL2A1*

ID dbSNP	Position hg19	Position gène	Variant	Acide aminé	MAF 1000G	Répondeurs	Résistants
rs1138357	15:80263717	Exonique	G>A	Cys>Tyr	0,3520	7	2
rs1138358	15:80263406	Exonique	A>C	Asn>Lys	0,4589	7	2
rs3826007	15:80263217	Exonique	C>T	Gly>Asp	0,2035	7	2
rs8026803	15:80260014	Exonique	T>C	Thr	0,3894	7	2
rs16971619	15:80259964	Intronique	C>T	-	0,3544	7	2
rs146807490	15:80259956-80259959	Intronique	GTAA>-	-	-	7	2

Le variant rs3826007 situé dans l'exon 1 entraîne un changement d'acide aminé qui serait à l'origine d'une mauvaise conformation du domaine BH1 essentiel pour son activité inhibitrice. Les bases de données Polyphen-2 et SIFT prédisent une altération de la protéine en présence de ce variant.

Les gènes *BCL-RAMBO*, *BCLAF1* et *BCL2A1* ont été sélectionnés pour la suite de l'étude. En effet, plusieurs arguments suggèrent que des polymorphismes non-synonymes, identifiés parmi ces trois gènes (rs9306198, rs139012879, rs3826007), seraient impliqués dans la résistance à l'imatinib. Ces variants pourraient entraîner une altération des protéines codées, fortement exprimées dans les cellules hématopoïétiques, et par conséquent perturber le mécanisme d'apoptose nécessaire à l'élimination des cellules leucémiques.

Partie III : Etude des gènes candidats dans une cohorte de validation

1. Séquençage dans une cohorte de validation

Dans le but de valider les différences de répartition du variant entre les patients répondeurs et résistants, les polymorphismes candidats des gènes *BCL-RAMBO, BCLAF1* et *BCL2A1* ont été étudiés à partir d'échantillons obtenus dans une cohorte de validation. Tout d'abord, le polymorphisme rs9306198 du gène *BCL-RAMBO*, le polymorphisme rs139012879 de *BCLAF1* et les six polymorphismes candidats du gène *BCL2A1* ont été séquencés à partir d'ADNg de leucocytes, prélevés au diagnostic, de 103 patients atteints de LMC, 78 répondeurs et 25 résistants, par une approche de séquençage Sanger.

Une différence significative a été observée entre les patients résistants et répondeurs pour le variant du gène *BCL-RAMBO* (Figure 70). L'allèle T minoritaire est plus fréquent chez les résistants (44%) que chez les patients répondeurs (23%). De manière intéressante, la fréquence du variant est également supérieure pour les patients atteints de LMC (28/103 soit 27%) que dans la population générale d'après la base de données 1000G estimée à environ 5% (MAF=0,0513) ou d'après la base de données ExAC estimée à 9% (MAF=0,0938).

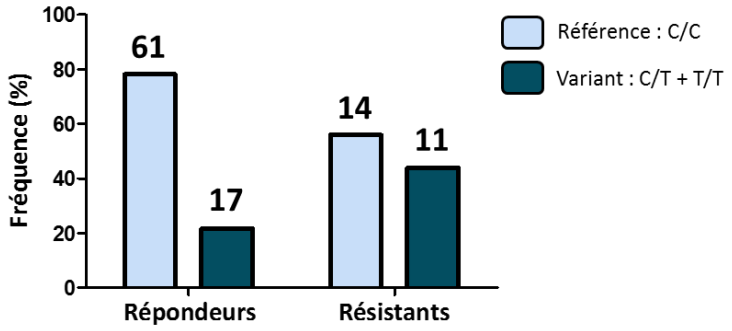


Figure 70 : Répartition des individus porteurs ou non de l'allèle variant du polymorphisme rs9306198 du gène *BCL-RAMBO* après séquençage de 103 patients atteints de LMC, 78 répondeurs et 25 résistants au traitement à l'imatinib

La différence entre les deux groupes est évaluée par le test de Chi², * p = 0,03.

Le variant rare de *BCLAF1* a été retrouvé chez deux autres patients atteints de LMC résistants au traitement à l'imatinib (16%) cependant il est également identifié chez un patient répondeur, il est donc moins susceptible d'être associé à la réponse à l'imatinib (Figure 71). Un séquençage à partir d'une plus grande cohorte de patients serait nécessaire pour obtenir une différence significative.

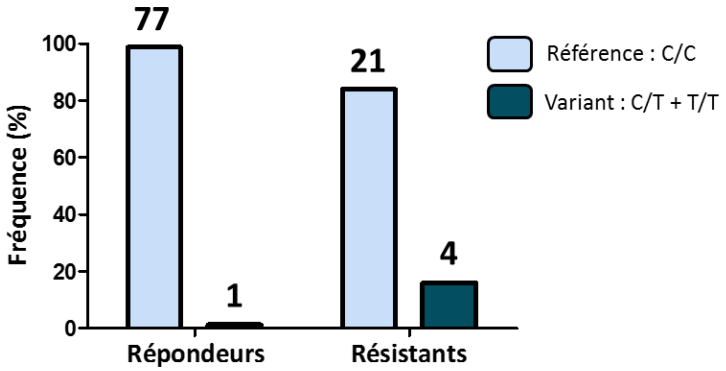


Figure 71 : Répartition des patients porteurs ou non de l'allèle variant du polymorphisme rs139012879 du gène *BCLAF1* chez les patients atteints de LMC répondeurs et résistants au traitement à l'imatinib

La différence entre les deux groupes ne peut pas être évaluée par le test de Chi² car l'effectif est insuffisant.

Les six polymorphismes candidats de *BCL2A1* ont également été séquencés dans la même cohorte. Les polymorphismes sont en quasi-total déséquilibre de liaison mais, à la différence du variant de *BCL-RAMBO*, aucune différence significative n'a été mise en évidence pour ces variants (exemple du polymorphisme rs3826007 en Figure 72).

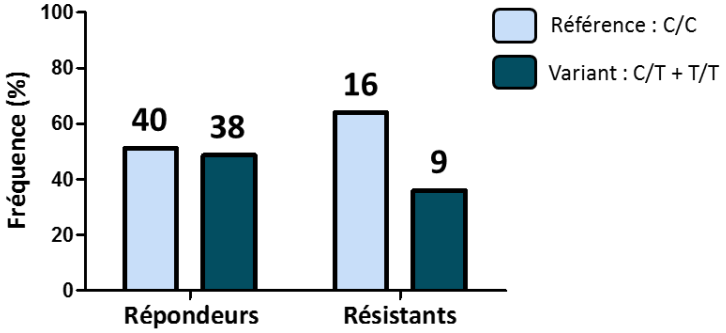


Figure 72 : Répartition des patients porteurs ou non de l'allèle variant du polymorphisme rs3826007 du gène *BCL2A1* chez les patients atteints de LMC répondeurs et résistants au traitement à l'imatinib

La différence entre les deux groupes est évaluée par le test de Chi², ns, p=0,27.

A l'aide du logiciel IGV nous avons pu constater que la séquence promotrice de *BCL2A1* n'avait pas été séquencée. Or, ce gène est hautement régulé par le facteur de transcription NF-κB [156]. Globalement, les séquences promotrices des 45 gènes ciblés ont été peu couvertes par les sondes de capture. Une variation dans la séquence de fixation de ce régulateur pourrait donc modifier l'expression du gène *BCL2A1* et ainsi la régulation du mécanisme d'apoptose. Des amorces de part et d'autre de la séquence d'intérêt ont donc été conçues et ont permis l'amplification par PCR d'une région de près de 1000 pb, non capturée dans les librairies de

séquençage. Ces amplicons ont été séquencés par Sanger à partir d'échantillons de 36 patients atteints de LMC (20 répondeurs, 16 résistants) et trois polymorphismes ont été mis en évidence (Tableau 33).

Tableau 33 : Caractéristiques des trois polymorphismes mis en évidence dans la région promotrice du gène *BCL2A1*

				Nombre (fréquence) de patients porteurs du variant	
ID dbSNP	**Variant**	**Position hg19**	**MAF**	**Répondeurs**	**Résistants**
rs137943308	A>G	15:80264540	G = 07	1 (0,05)	1 (0,06)
rs62025493	G>A	15:80264443	A = 0,202	12 (0,60)	4 (0,25)
rs12912717	G>A	15:80264326	A = 0,1240	4 (0,20)	6 (0,38)

Aucune différence significative n'est identifiée entre les patients répondeurs et résistants. Cependant, il est intéressant de noter que les fréquences observées dans notre cohorte de patients atteints de LMC sont toutes supérieures aux fréquences référencées dans la base de données 1000G.

Le variant rs9306198 de *BCL-RAMBO* est retrouvé enrichi de manière significative chez les patients réistants au traitement. En revanche, aucune différence significative n'est retrouvée pour les gènes *BCLAF1* et *BCL2A1*. Par la suite, l'expression de *BCL-RAMBO* et *BCL2A1* en fonction des génotypes des polymorphismes identifiés a été évaluée. Le variant du gène *BCLAF1* étant trop rare, il n'a pas été conservé pour la suite des analyses

2. Expression des gènes *BCL-RAMBO* et *BCL2A1* selon le génotype des patients

L'expression des gènes candidats a été évaluée par PCRq en temps réel à partir d'échantillons prélevés au diagnostic pour 78 patients atteints de LMC (60 répondeurs et 18 résistants au traitement à l'imatinib). Le niveau d'expression de ces gènes a été normalisé par l'expression du gène de ménage *ACTB*. Pour le gène *BCL-RAMBO*, aucune différence significative n'a été observée que ce soit en considérant le statut de réponse des patients ou le génotype du polymorphisme rs9306198 (Figure 73).

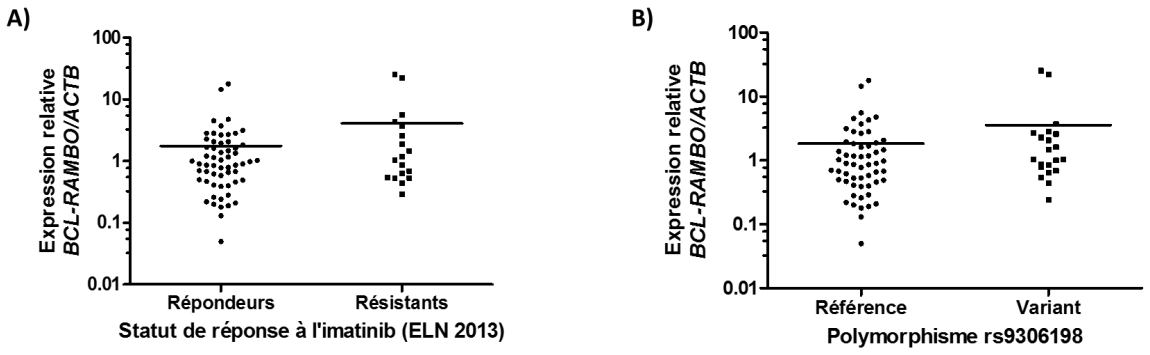


Figure 73 : Expression relative du gène *BCL-RAMBO* selon A) le statut de réponse des patients à l'imatinib, B) le génotype du polymorphisme rs9306198

La présence de l'allèle variant ne semble pas influencer l'expression du gène *BCL-RAMBO.* De plus, aucune différence d'expression n'est détectée entre les patients répondeurs ou résistants ce qui laisse penser qu'une variation de l'expression du gène n'est pas à l'origine d'une différence de sensibilité au traitement à l'imatinib. Pour le gène *BCL2A1*, aucune différence significative n'est identifiée pour aucun des deux transcrits (Figure 74).

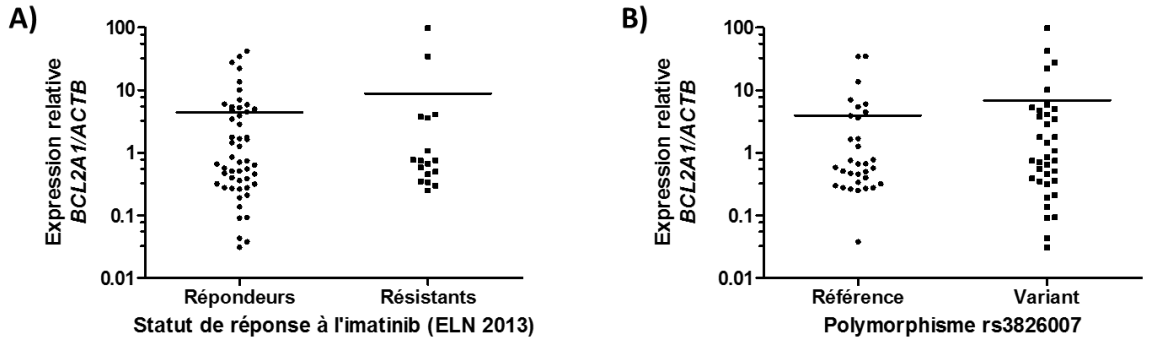


Figure 74 : Expression relative du gène *BCL2A1* selon A) le statut de réponse des patients, B) le génotype du polymorphisme rs3826007

Que ce soit pour le gène *BCL-RAMBO* ou *BCL2A1* aucune différence d'expression n'est mise en évidence lorsque les patients répondeurs et résistants sont comparés. La résistance à l'imatinib pour ces patients atteints de LMC ne semble donc pas corrélée à une différence du niveau d'expression de ces deux gènes. Lorsque l'expression est étudiée en fonction du statut du polymorphisme d'intérêt, de nouveau aucune différence n'est détectée pour les gènes *BCL-RAMBO* et *BCL2A1*. Les polymorphismes rs9306198 et rs3826007 ne semblent pas réguler l'expression des gènes modifiés.

3. Courbes de survie

Enfin, l'association entre la présence des variants et le délai de survenue de la RCH est évaluée à l'aide de courbes de survie dans la cohorte de validation. La date de la RCH n'est pas connue pour deux patients répondeurs qui ont par conséquent été exclus de l'analyse (101 patients atteints de LMC restants). Le délai de survenue de la RCH est représenté en fonction du statut du polymorphisme rs9306198 de *BCL-RAMBO* (Figure 75). Au total, 74 patients sont homozygotes de référence et 27 patients sont porteurs de l'allèle variant (hétérozygotes et homozygotes variants). Il est intéressant de noter que le pourcentage de patients atteints de LMC atteignant la RCH est inférieur pour les patients porteurs de l'allèle variant. Le délai d'obtention de la RCH est également augmenté de manière significative pour ces patients. En revanche, aucune corrélation n'est observée pour le polymorphisme rs3826007 de *BCL2A1* (55 homozygotes de référence, 46 hétérozygotes+homozygotes variants) (Figure 76).

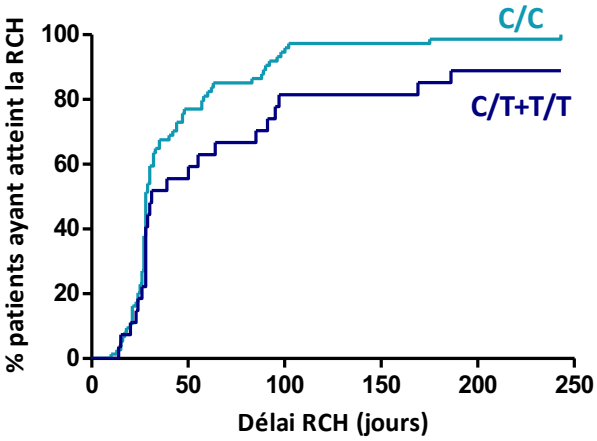


Figure 75 : Evolution du pourcentage de patients atteignant la RCH au cours du temps en fonction du statut du polymorphisme rs9306198 du gène *BCL-RAMBO* pour 101 patients atteints de LMC

C/C : patients homozygotes de référence, C/T+T/T : patients porteurs de l'allèle variant

Test du log rank * p-value = 0,015

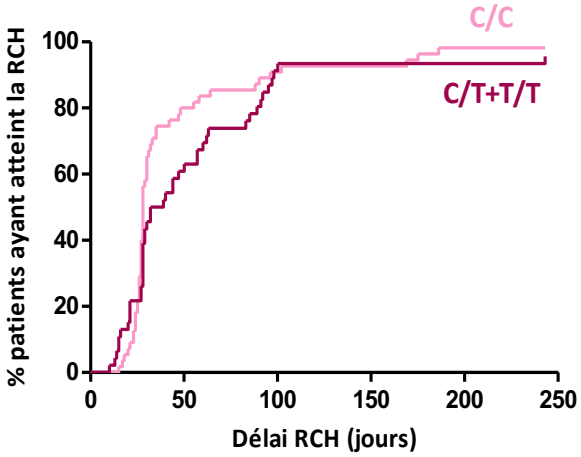


Figure 76 : Evolution du pourcentage de patients atteignant la RCH au cours du temps en fonction du statut du polymorphisme rs3826007 du gène *BCL2A1* pour 101 patients atteints de LMC

C/C : patients homozygotes de référence, C/T+T/T : patients porteurs de l'allèle variant

Test du log rank, ns, p-value = 0,156

L'analyse des gènes candidats dans une cohorte de validation de 103 patients atteints de LMC, 78 répondeurs et 25 résistants, a démontré que le variant du polymorphisme rs9306198 du gène *BCL-RAMBO* était retrouvé plus fréquemment chez les patients en échec de traitement que chez les patients répondeurs. De plus, la présence du variant est corrélée à un délai d'obtention de la RCH plus long. Aucune différence d'expression significative du gène n'a été détectée suggérant que le variant influe en aval de la transcription. En revanche, aucune corrélation entre les variants de *BCL2A1* et la réponse au traitement n'a été mise en évidence dans la cohorte de validation.

Les analyses ont donc été poursuivies pour le gène *BCL-RAMBO.* Des modèles cellulaires ont notamment été développés afin de caractériser le rôle du polymorphisme rs9306198 dans la régulation de l'apoptose en réponse à l'imatinib.

Partie IV : BCL-RAMBO régulation de l'apoptose et sensibilité à l'imatinib

1. Création de lignées cellulaires modifiées par CRISPR-Cas9

L'objectif était dans un premier temps de définir le rôle de la protéine BCL-RAMBO dans la régulation de l'apoptose induite par l'imatinib dans la LMC. Pour cela des lignées cellulaires n'exprimant plus, ou au contraire sur-exprimant le gène *BCL-RAMBO,* ont été créées à partir de la lignée modèle de LMC K-562. Dans un deuxième temps, le but était de caractériser l'impact du variant rs9306198 sur la survie des cellules. Pour cela, des lignées cellulaires exprimant la protéine variante ont été générées grâce à la technologie CRISPR-Cas9 et une nouvelle stratégie de remplacement de l'ADNg (ce travail a fait l'objet d'une publication soumise à Scientific Reports, Annexe).

1.1. Lignées n'exprimant plus BCL-RAMBO (knock-out, KO)

La protéine BCL-RAMBO de la famille BCL2 semble participer à la régulation de l'apoptose dans différents modèles sans que sa fonction exacte ne soit définie. Une lignée cellulaire n'exprimant plus la protéine BCL-RAMBO a tout d'abord été générée grâce à la technologie CRISPR-Cas9 afin de définir le rôle de la protéine dans la régulation de l'apoptose induite par l'imatinib dans la LMC. Afin d'éteindre l'expression du gène *BCL-RAMBO*, deux cassures doubles brins ont été générées soit dans l'exon 2, soit dans l'exon 3 du gène dans la lignée cellulaire modèle K-562. En effet, le transcrit majoritaire de *BCL-RAMBO* n'est pas précisément défini et le codon Start pourrait être codé par l'un ou l'autre des exons (Tableau 34). Des clones cellulaires n'exprimant plus la protéine BCL-RAMBO ont été obtenus après avoir ciblé l'exon 3 uniquement. Les résultats obtenus pour l'exon 2 ne sont donc pas présentés ci-dessous.

Tableau 34 : Transcrits et isoformes associés codés par le gène *BCL-RAMBO*

Les exons transcrits sont représentés en gris foncé, les isoformes associés sont représentés en gris clair, les codons Start (Met) sont indiqués par un « M » bleu et les codons Stop sont indiqués par un « * » rouge.

Les références RefSeq (NM, NP, NR) sont précisées pour chaque transcrit et isoforme.

	E1	E2	E2B	E3	E4	E5	E6	E7	E8	E9	E9B	E10	E11	E11B		
Taille en pb	52	785	303	171	115	108	98	157	70	144	186	168	2854	4179		
Transcrit 1															5132 pb	NM_015367.3
Isoforme a				M										*	485 aa	NP_056182.2
Transcrit 2															4961 pb	NM_001270726.1
Isoforme b			M											*	509 aa	NP_001257655.1
Transcrit 3															4817 pb	NM_001270727.1
Isoforme c			M											*	461 aa	NP_001257656.1
Transcrit 4															4797 pb	NM_001270728.1
Isoforme d				M										*	92 aa	NP_001257657.1
Transcrit 5															4975 pb	NM_001270729.1
Transcrit 6															4804 pb	NM_001270730.1
Transcrit 7															5062 pb	NM_001270731.1
Isoforme e										M				*	323 aa	NP_001257660.1,
Transcrit 8															4905 pb	NM_001270732.1
Isoforme f				M										*	128 aa	NP_001257661.1
Transcrit 9															4347 pb	NM_001270733.1
Isoforme g														S/*	205 aa	NP_001257662.1
Transcrit 10															4918 pb	NM_001270734.1
Isoforme h				M										*	171 aa	NP_001257663.1
Transcrit 11															4761 pb	NM_001270735.1
Isoforme i				M										*	80 aa	NP_001257664.1
Transcrit 12															4730 pb	NR_073068.1
Transcrit 13															4713 pb	NR_073069.1

Si le complexe CRISPR-Cas9 a induit deux coupures doubles brins, une délétion entre les deux sites de coupure doit être générée. L'efficacité du système peut donc être évaluée par PCR et électrophorèse sur gel d'agarose. Comme attendu deux signaux sont observés : une bande à 279 pb correspondant à l'amplification d'un brin non modifié, et une bande à 215 pb correspondant à l'amplification d'un fragment avec une délétion de 64 pb (Figure 77.A). Cette délétion prouve que les deux ARNg ont coupé efficacement l'ADN. Un mélange de cellules non modifiées et modifiées a donc été obtenu. Les différentes cellules sont ensuite individualisées afin d'isoler les clones cellulaires ayant perdu l'expression du gène *BCL-RAMBO*. Les modifications génomiques dans ces différents clones sont de nouveau évaluées par PCR (Figure 77.B).

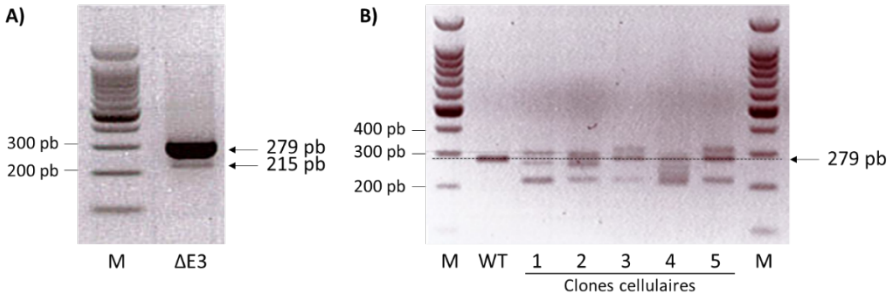


Figure 77 : Résultats d'une éléctrophorèse sur gel d'agarose des produits PCR obtenus à partir de l'amplification de l'ADNg (exon 3 de *BCL-RAMBO*) des cellules modifiées par CRISPR-Cas9

A) PCR à partir d'un mélange de cellules, B) PCR à partir de clones cellulaires non modifiés (WT) et sous-clonées (clones cellulaires 1 à 5), (M) marqueur de taille 100 pb

Des profils d'amplification très hétérogènes sont observés entre les clones cellulaires et aucune lignée avec délétion homozygote n'est obtenue. Ceci suggère que les courtes insertions et délétions aux sites de coupure ont été favorisées au détriment de la délétion du fragment de 64 pb entre les sites ciblés par les ARNg. A plusieurs reprises, plus de deux signaux sont observés pour une même lignée cellulaire ce qui suggère que le gène *BCL-RAMBO* est présent en plus de deux copies dans la lignée cellulaire K-562. En effet, la lignée K-562 présente un caryotype complexe, hypotriploïde (Figure 78). Une analyse des remaniements chromosomiques survenus dans les cellules K-562 a été réalisée par hybridation génomique comparative (CGH) par Naumann *et al.* et un gain du locus 22q11.2, incluant le gène *BCL-RAMBO*, a été observé dans cette lignée ce qui peut expliquer le nombre élevé de signaux obtenus pour chaque clone cellulaire [234]. Le gène *BCL-RAMBO* est donc probablement présent en plus de deux copies dans la lignée modèle de LMC K-562.

Figure 78 : Caryotype hypotriploïde de la lignée cellulaire K-562

Naumann *et al.*, Leukemia Research, 2001 [234]

L'expression protéique de BCL-RAMBO a été évaluée à partir de prélèvements de 12 sous-clones ayant un profil PCR différent des cellules non modifiées (WT) ce qui a permis d'identifier deux clones avec une extinction complète de l'expression de BCL-RAMBO détectée par western blot (Figure 79). De manière intéressante, cette perte d'expression est maintenue au cours du temps. L'expression de la protéine BCL-RAMBO est diminuée par rapport aux cellules WT pour deux clones. Pour les autres clones cellulaires, l'expression est similaire aux cellules WT.

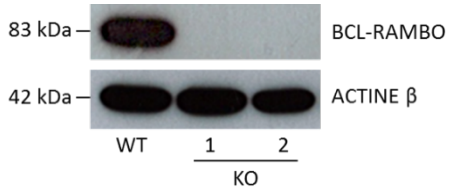


Figure 79 : Expression protéique de la protéine BCL-RAMBO et de l'ACTINE β évaluée par western blot dans trois lignées cellulaires K-562

WT) Lignée sauvage, 1 et 2) Clones modifiés par CRISPR-Cas9 (coupures dans le premier exon codant)

L'anticorps utilisé reconnaît le domaine BHNo de la protéine et ce domaine serait présent dans tous les isoformes de BCL-RAMBO (Tableau 34). Or, un seul signal est détecté à 83 kDa. D'après ces résultats, le transcrit 1 serait le transcrit majoritaire et le seul ARNm traduit en une protéine de 485 aa.

Au final, grâce à l'outil CRISPR-Cas9, deux lignées clonales, issues de la lignée modèle de LMC K-562, présentant une extinction complète du gène *BCL-RAMBO* ont été obtenues.

1.2. Lignées n'exprimant plus les domaines BHNo et TM

De manière similaire, une nouvelle expérience a été conçue afin d'éliminer spécifiquement le domaine BHNo de la protéine BCL-RAMBO. Deux ARNg, ciblant des séquences en amont et en aval du domaine BHNo, ont été conçus afin d'éliminer spécifiquement ce domaine tout en conservant le cadre de lecture et donc le domaine TM. Après sous-clonage, une seule lignée cellulaire clonale modifiée, sur 24 testées, a été obtenue (taille des produits PCR, WT = 1020 pb, avec délétion de 717 pb = 303 pb) (Figure 80).

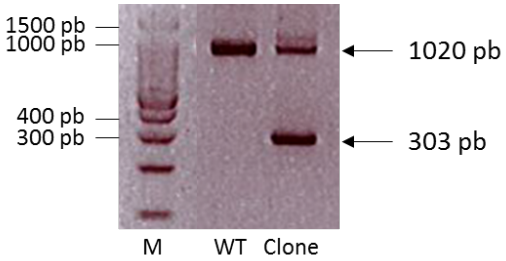


Figure 80 : Résultats d'une éléctrophorèse sur gel d'agarose des produits PCR obtenus à partir de l'amplification de l'ADNg du clone cellulaire modifié par CRISPR-Cas9

M) marqueur de taille 100 pb, WT : PCR à partir d'ADNg extrait de cellules sauvages non modifiées par CRISPR-Cas9, Clone) PCR à partir d'ADNg extrait du clone cellulaire issu du mélange de cellules modifiées par CRISPR-Cas9

L'expression de la protéine est détectée à l'aide d'un anticorps reconnaissant le domaine BHNo de BCL-RAMBO (acides aminés 224 à 459). Les protéines tronquées de BCL-RAMBO ne sont donc pas mises en évidence par western blot. A la différence des lignées KO, toutes les copies du gène n'ont pas été modifiées et la forme WT de la protéine est diminuée mais toujours exprimée dans la lignée clonale modifiée par CRISPR-Cas9 (Figure 81).

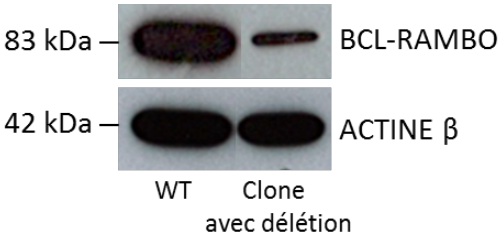


Figure 81 : Expression protéique de la protéine BCL-RAMBO et de l'ACTINE β dans deux lignées cellulaires K-562

WT) Lignée sauvage, Clone avec délétion) Clone modifié par CRISPR-Cas9 (coupures dans l'exon 11)

Après séquençage du produit PCR par Sanger, une délétion de 718 pb a été mise en évidence au lieu de la délétion attendue de 717 pb qui aurait conservé le cadre de lecture et donc le domaine TM. En théorie, le complexe CRISPR-Cas9 génère des cassures doubles brins à 3 pb de la séquence PAM [225]. Cependant, des variations du site de coupure exact ont été observées à plusieurs reprises. Pour ce clone cellulaire, la coupure du deuxième ARNg, en aval du domaine BHNo, est survenue 4 pb avant le PAM.

Les domaines C-terminaux BHNo et TM de BCL-RAMBO ont donc tous les deux été délétés et une lignée cellulaire exprimant la protéine BCL-RAMBO tronquée (mais également la protéine WT) a été obtenue (Figure 82).

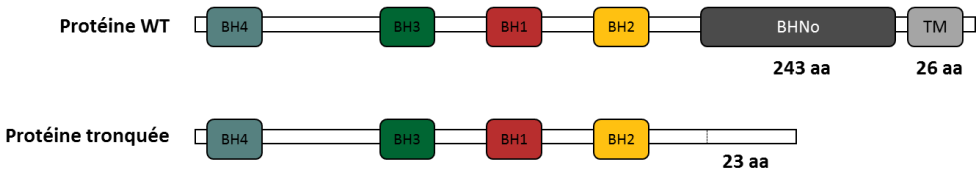


Figure 82 : Schéma de la protéine BCL-RAMBO sous les formes WT et tronquée

Le domaine BHNo de 243 acides aminés (aa) et le domaine transmembranaire (TM) de 26 aa sont éliminés et remplacés par une nouvelle séquence de 23 aa.

Au final, en utilisant la technologie CRISPR-Cas9, le gène *BCL-RAMBO* a été modifié dans plusieurs lignées cellulaires. Deux lignées clonales n'exprimant plus la protéine BCL-RAMBO ont été obtenues. De plus, une lignée clonale exprimant une forme tronquée de BCL-RAMBO (ainsi que la forme WT) a été générée.

1.3. Lignées surexprimant la protéine BCL-RAMBO

Par la suite, afin de comprendre et caractériser le rôle de BCL-RAMBO dans la régulation de l'apoptose et la réponse à l'imatinib, des lignées surexprimant le gène *BCL-RAMBO* ont été développées. La séquence codante de *BCL-RAMBO*, avec ou sans l'allèle variant du polymorphisme rs9306198, a été clonée dans un vecteur de surexpression et transduit dans les cellules K-562 WT. Dans un premier temps, l'intégration de la séquence dans l'ADNg a été confirmée par amplification PCR. Comme attendu, un signal à 2029 pb est détecté pour les lignées transduites avec la séquence codante de *BCL-RAMBO* et un signal à 265 pb est obtenu pour la lignée transduite avec le vecteur de surexpression vide (Figure 83). L'expression du gène *BCL-RAMBO* a également été évaluée dans ces lignées par PCRq. L'expression du gène *BCL-RAMBO* est près de 20 fois supérieure dans les cellules transduites avec la séquence codante que dans les cellules transduites avec le vecteur vide (Figure 84).

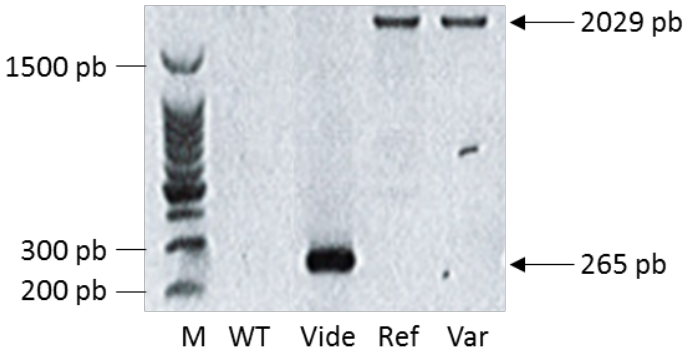


Figure 83 : Résultats d'une électrophorèse sur gel d'agarose des produits PCR obtenus à partir de l'amplification du vecteur de surexpression transduit dans les cellules K-562

Cellules K-562, WT) Sauvages, non transduites, Vide) Avec le vecteur vide, Ref) Avec le vecteur et la séquence codante contenant l'allèle de référence rs9306198, Var) Avec la séquence codante contenant l'allèle variant

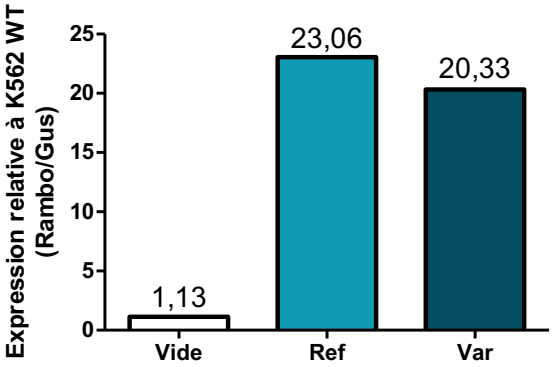


Figure 84 : Expression relative du gène *BCL-RAMBO* normalisée sur l'expression du gène *GUS1* évaluée par PCRq dans trois lignées K-562 transduites avec le vecteur p412 : vide, portant la séquence codante de BCL-RAMBO de référence (Ref) ou variante (Var) (n=1)

Enfin, après avoir confirmé la surexpression du gène *BCL*-RAMBO au niveau de l'ARNm, le niveau d'expression protéique dans le mélange cellulaire a été évalué par western blot (Figure 85). Une séquence codant l'étiquette HA avait été ajoutée par amplification PCR en amont de la séquence codante de BCL-RAMBO (en N-terminal de la protéine). Cependant, aucun signal n'est détecté en utilisant un anticorps anti-HA. De plus, le niveau de surexpression de la protéine détectée par un anticorps anti-BCL-RAMBO est relativement faible comparé à la surexpression détectée en PCRq. La protéine est deux fois plus exprimée dans la lignée de référence, et quatre fois dans la lignée variante comparée aux cellules transduites avec le vecteur vide.

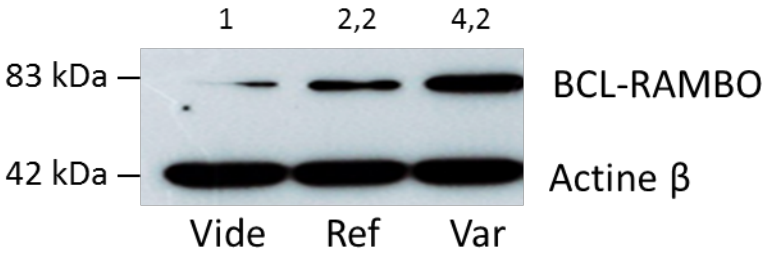


Figure 85 : Expression protéique de la protéine BCL-RAMBO et de l'ACTINE β dans trois lignées cellulaires K-562 transduites avec le vecteur p412 : vide, portant la séquence codante de BCL-RAMBO de référence (Ref) ou variante (Var)

Au final deux lignées avec une surexpression modeste de la protéine BCL-RAMBO, sous sa forme variante ou non, ont été obtenues.

1.4. Lignées exprimant la protéine variante

Afin de caractériser le rôle du variant rs9306198 de *BCL-RAMBO* dans l'activité de la protéine BCL-RAMBO et la sensibilité des cellules à l'imatinib, des lignées cellulaires, ne différant que par la présence ou non de l'allèle variant, ont été créées. La première approche utilisée était la substitution du nucléotide de référence C en nucléotide variant T par l'outil CRISPR-Cas9 et la voie de réparation de la recombinaison homologue.

1.4.1. Recombinaison homologue

Pour cette approche, une matrice de recombinaison homologue double brin de 999 pb a été concue. Dans un premier temps, quatre ARNg ciblant des séquences à proximité du nucléotide d'intérêt ont été testés (Figure 86).

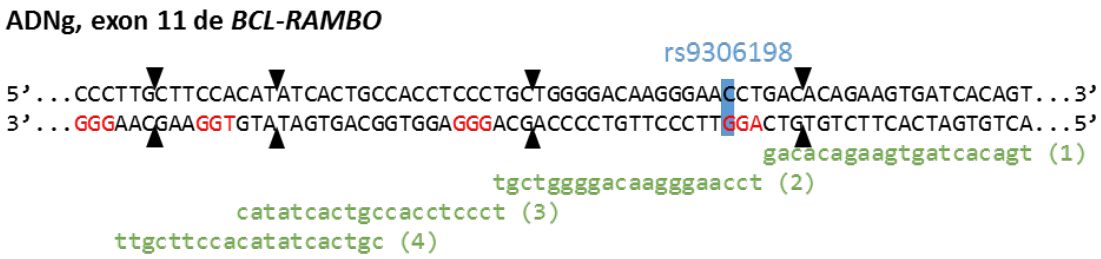


Figure 86 : Séquences ciblées par les 4 ARNg à proximité du polymorphisme rs9306198

La position du polymorphisme rs9306198 est indiquée en bleu, les séquences de 20 nt des ARNg sont représentées en vert, les motifs PAM sont notés en rouge. Les sites de coupures attendus sont indiqués par des flèches.

Un test de clivage a été réalisé. Les résultats de ce test peuvent être observés par électrophorèse sur gel d'agarose (Figure 87). Une bande de taille inférieure est observé pour les quatre échantillons issus de cellules modifiées par CRISPR-Cas9, ce qui prouvent que les 4 ARNg ont coupé efficacement l'ADN, ils sont donc sélectionnés pour la suite des expériences.

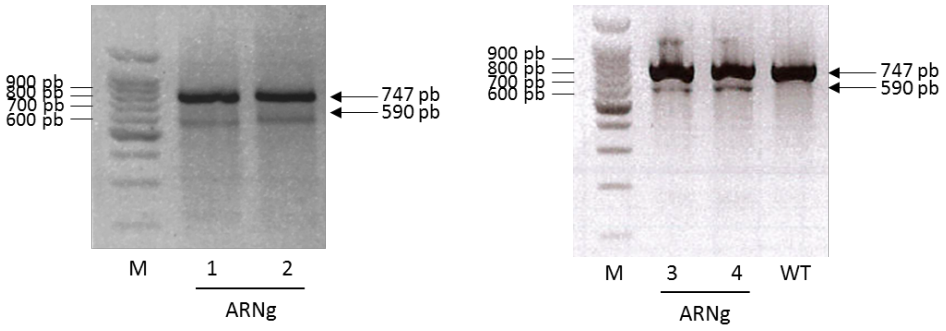


Figure 87 : Résultats d'une éléctrophorèse sur gel d'agarose des produits PCR obtenus suite au test clivage réalisé à partir de l'ADNg de cellules WT ou ciblés par les ARNg 1, 2, 3 et 4

M) marqueur de taille 100 pb, ARNg 1 à 4) PCR et test de clivage à partir d'ADNg extrait des cellules modifiées par CRISPR-Cas9, WT) PCR et test de clivage à partir d'ADNg extrait de cellules sauvages non modifiées par CRISPR-Cas9

Les ARNg étant validés, la réparation par recombinaison homologue a été testée par transfection des ARNg et de la matrice de recombinaison dans les lignées cellulaires 293T et K-562. Si la recombinaison homologue a été efficace, un nouveau site de restriction reconnu par l'enzyme *Sma*I doit être généré à proximité du variant. Une amplification par PCR de la région modifiée à partir d'ADNg extrait des cellules modifiées, suivie d'une digestion des amplicons par l'enzyme *Sma*I, permet d'évaluer l'efficacité de la modification (résultats pour la lignée 293T, Figure 88).

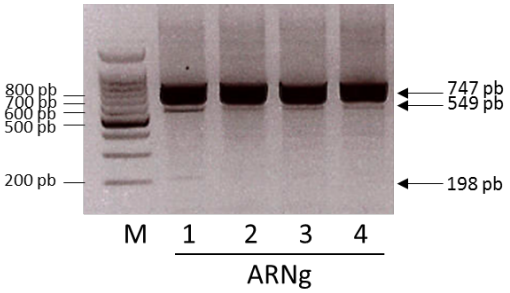


Figure 88 : Résultats d'une éléctrophorèse sur gel d'agarose des produits PCR, digérés par l'enzyme *Sma*I, obtenus à partir de l'ADNg de cellules WT ou transfectés avec quatre ARNg différents ainsi que la matrice de recombinaison de 999 pb

M) marqueur de taille 100 pb, 1 à 4) échantillons extraits après transfection avec les ARNg 1 à 4

Une digestion des produits PCR est observée pour les quatre ARNg dans la lignée cellulaire 293T avec un signal plus intense pour l'ARNg 1. Il s'agit de la séquence ayant généré la coupure la plus proche du nucléotide à modifier (5 pb de distance) (Figure 86). Le site de restriction enzymatique *Sma*I ayant été généré, il est probable que l'allèle variant ait également été intégré à l'ADNg. Les cellules transfectées avec l'ARNg 1 ont été sous clonées. Après séquençage par Sanger, 6 clones sur 33 testés ont effectivement intégré le variant (exemple en Figure 89). Cependant, des indels au site de coupure par l'ARNg sont également observées ce qui signifie que tous les allèles n'ont pas été efficacement modifiés. Ces résultats ont été obtenus pour la lignée contrôle 293T, en revanche aucun signal n'est mis en évidence pour la lignée modèle de LMC K-562.

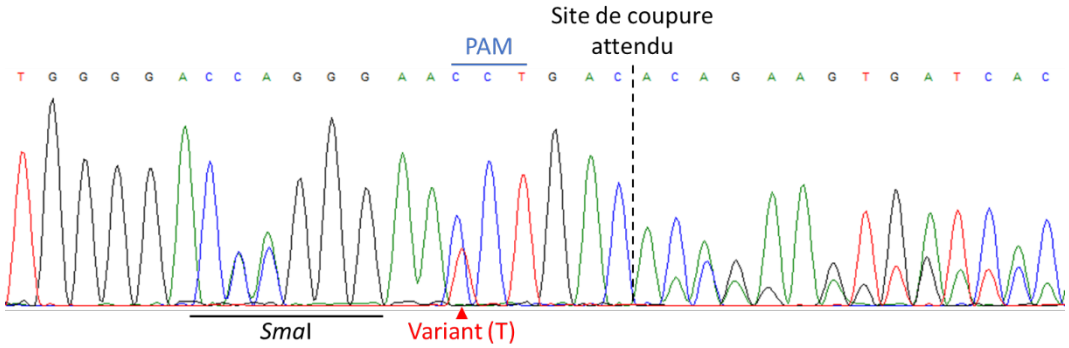


Figure 89 : Chromatogramme obtenu après le séquençage Sanger de la région modifiée de *BCL-RAMBO* d'un clone cellulaire 293T

Grâce à la technologie CRISPR-Cas9 et l'approche de la recombinaison homologue, des lignées 293T porteuses de l'allèle variant pour le polymorphisme rs9306198 de *BCL-RAMBO* ont été obtenues. En revanche, cette méthode n'a pas permis d'obtenir des lignées K-562 modifiées. Une nouvelle stratégie basée sur CRISPR-Cas9 et la réparation par NHEJ a donc été développée.

1.4.2. HIPER

Le mécanisme de recombinaison homologue semble trop minoritaire dans les cellules K-562. Une autre approche de remplacement de l'ADNg basée sur le mécanisme de réparation NHEJ, fréquent dans les cellules K-562, a donc été développée. Ce travail fait l'objet d'une publication soumise à Scientific Reports (Annexe). Brièvement, l'exon 11 WT est remplacé par un exon 11 modifié cloné dans un vecteur donneur (Figure 90). Pour cela quatre cassures doubles brins simultanées sont induites, deux de part et d'autre de l'exon 11 endogène et deux

autres de part et d'autre de la cassette de remplacement. Les séquences cibles des ARNg sont identiques dans l'ADNg WT et le vecteur donneur mais dans le sens opposé. Ainsi, lorsque le fragment de remplacement est inséré à la place de l'exon WT, le système CRISPR-Cas9 ne peut plus couper l'ADN.

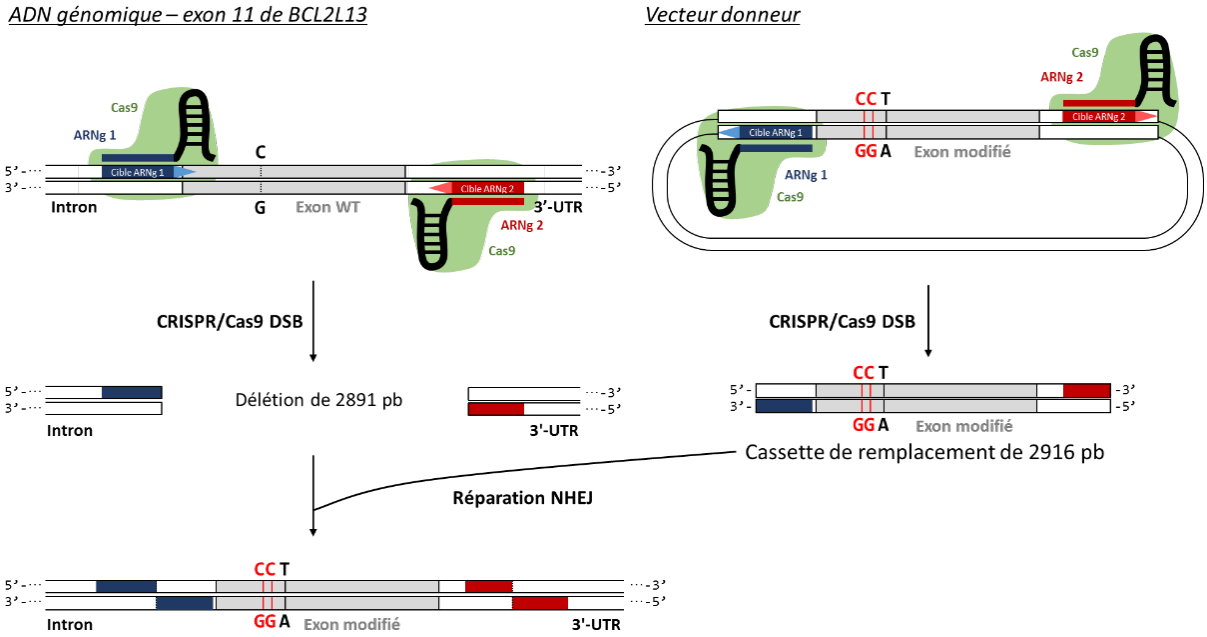


Figure 90 : Schématisation de la stratégie HIPER (homology-independent precise and efficient replacement) permettant l'intégration de l'allèle variant du polymorphisme rs9306198 du gène *BCL-RAMBO* dans la lignée cellulaire K-562

Au final, cette approche a permis d'obtenir deux lignées clonales sur 60 testées exprimant la forme variante de la protéine BCL-RAMBO. Le pourcentage d'allèles modifiés a été estimé par séquençage des clones en NGS (% de modifications estimé à 41% pour la lignée n°1 et 13% pour la lignée n°2) (Figure 91).

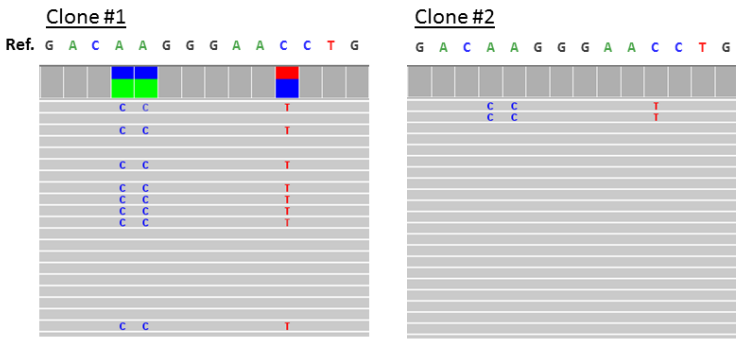


Figure 91 : Modification de la séquence de l'exon 11 de *BCL-RAMBO* séquencée par NGS et visualisée sur le logiciel IGV

Deux substitutions synonymes de A en C et une substitution non-synonyme de C en T (polymorphisme rs9306198) sont observées. 41% et 13% des reads sont modifiés pour les clones 1 et 2 respectivement

2. Induction de l'apoptose par l'imatinib

2.1. Détectée par cytométrie en flux

Grâce à la technologie CRSPR-Cas9, des lignées clonales n'exprimant plus la protéine BCL-RAMBO, exprimant une protéine modifiée ou surexprimant la protéine, ont été obtenues. La caractérisation de ces lignées permet d'étudier le rôle de la protéine BCL-RAMBO dans la régulation de l'apoptose en réponse à l'imatinib dans une lignée modèle de LMC. Les différentes lignées, WT et modifiées, ont été traitées pendant plusieurs jours avec différentes concentrations d'imatinib et le niveau d'apoptose des cellules a été évalué par cytométrie en flux suite à deux marquages différents.

2.1.1. Lignées KO *BCL-RAMBO*

Dans un premier temps, le niveau d'apoptose induite par imatinib a été évalué dans les lignées n'exprimant plus la protéine BCL-RAMBO. Les cellules en apoptose sont tout d'abord mises en évidence par un marquage au $DiOC_6(3)$ détecté par cytométrie en flux. Après 3 jours de traitement à 0,5 µM d'imatinib, une perte de la fluorescence est observée dans les cellules des lignées clonales KO, n'exprimant plus la protéine BCL-RAMBO (exemple en Figure 92). Une augmentation du pourcentage d'apoptose induite est observée dans les deux clones n'exprimant plus la protéine BCL-RAMBO. La différence est plus importante et significative dans la deuxième lignée (Figure 93). Des résultats similaires, mais sans différence significative ont été obtenus à l'aide d'un double marquage à l'annexine V-APC et DAPI (Figure 94).

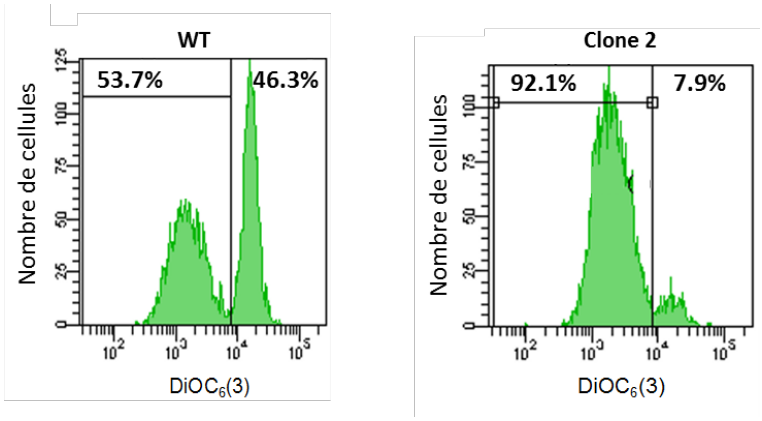


Figure 92 : Détection des cellules en apoptose par marquage DiOC$_6$(3) en cytométrie en flux après 72h de traitement à 0,5 µM d'imatinib

Le pourcentage de cellules en apoptose (DiOC$_6$(3)) est augmenté lorsque la protéine BCL-RAMBO n'est plus exprimée (clone KO n°2 : 92,1% vs WT : 53,7%).

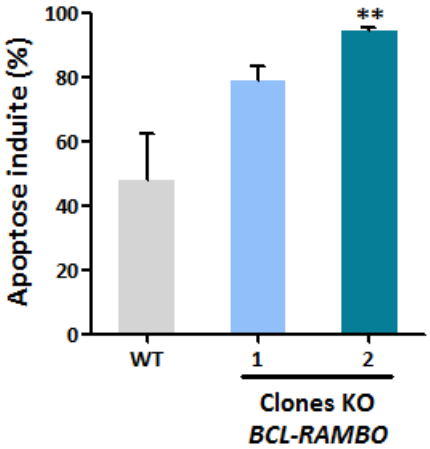


Figure 93 : Pourcentage d'apoptose induite après 72h de traitement à 0,5 µM d'imatinib évalué par marquage DiOC$_6$(3) pour la lignée WT et les deux lignées KO (n=4)

Test de Mann-Whitney ** $p < 0,01$

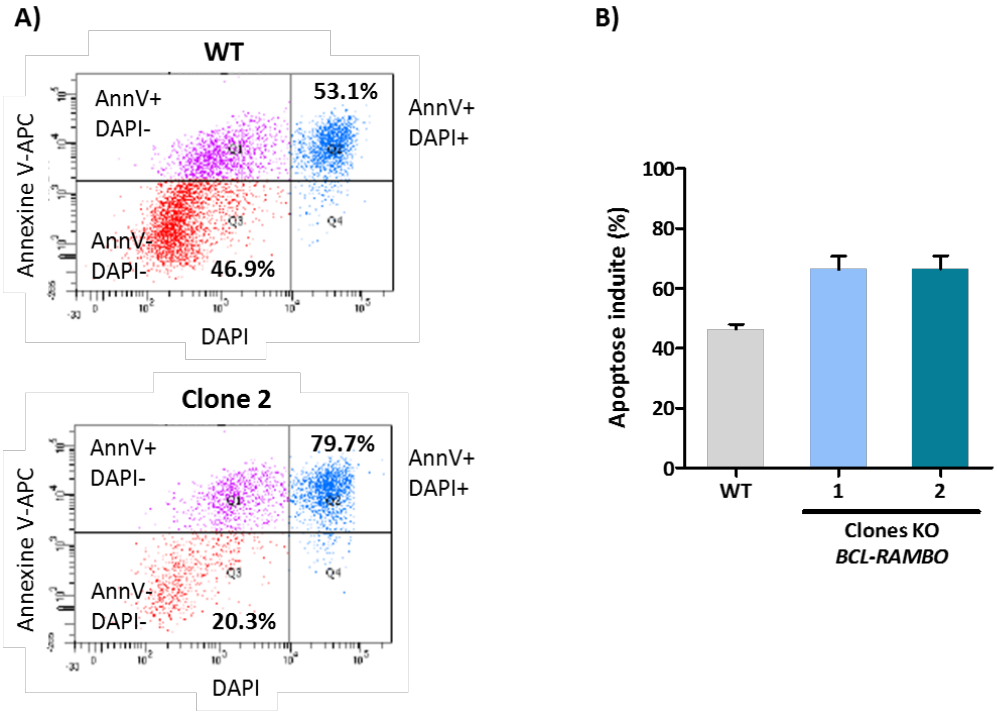


Figure 94 : A) Détection des cellules en apoptose par double marquage annexine V-APC/DAPI et cytométrie en flux après 72h de traitement à 0,5 µM d'imatinib, B) Pourcentage d'apoptose induite après 72h de traitement à 0,5 µM d'imatinib évalué par double-marquage annexine V-APC/DAPI pour la lignée WT et les deux lignées KO (n=3)

L'absence du facteur BCL-RAMBO semble sensibiliser les cellules à l'imatinib. BCL-RAMBO serait donc un facteur anti-apoptotique dans les cellules de la lignée modèle de LMC. Les membres de la famille BCL2 agissent généralement en interagissant entre eux, des mécanismes de compensation peuvent donc intervenir lorsqu'une protéine pro- ou anti-apoptotique est altérée. Or, dans notre modèle, l'absence de la protéine BCL-RAMBO modifie grandement le phénotype avec une augmentation de près de 30% de l'apoptose induite. Ces observations suggèrent que BCL-RAMBO présente un rôle majeur dans la régulation de l'apoptose dans la lignée cellulaire K-562.

2.1.2. Délétion des domaines BHNo et TM

Par la suite, le niveau d'apoptose a été évalué par marquage $DiOC_6(3)$ et annexine V-APC/DAPI pour la lignée exprimant la protéine BCL-RAMBO tronquée, sans les domaines C-terminaux BHNo et TM. De nouveau une augmentation du niveau d'apoptose induite, plus modeste que pour les lignées KO, est détectée dans la lignée modifiée (Figure 95).

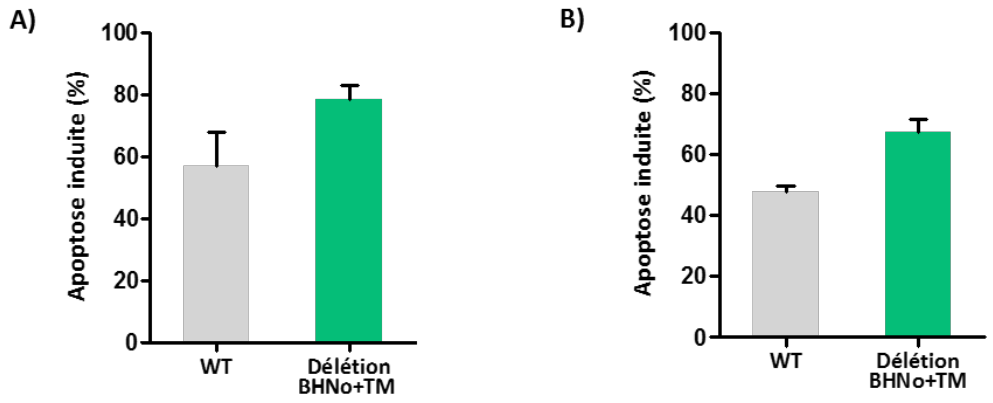


Figure 95 : Pourcentage d'apoptose induite après 72h de traitement à 0,5 µM d'imatinib évalué par marquage A) $DiOC_6(3)$ ou B) Annexine V-APC/DAPI pour la lignée WT et la lignée exprimant la protéine BCL-RAMBO tronquée (sans BHNo et TM) (n=3)

Ces résultats suggèrent que les domaines BHNo et TM sont importants pour la fonction anti-apoptotique de BCL-RAMBO. Le domaine en double hélice hydrophobe permet l'ancrage de la protéine dans la membrane externe mitochondriale. Son rôle dans la régulation de l'apoptose serait donc médié par sa localisation cellulaire. Dans cette lignée la protéine BCL-RAMBO WT est toujours exprimée ce qui pourrait expliquer le niveau d'induction de l'apoptose plus faible comparé aux lignées KO.

Il faut noter que l'augmentation de l'apoptose observée dans ces modèles pourrait également être indépendante de BCL-RAMBO et résulter des cassures doubles brins induites dans l'ADNg. Cependant, cette hypothèse est peu probable pour deux raisons. Tout d'abord, le gène *TP53* est muté dans cette lignée immortelle, il n'y a donc pas d'apoptose induite en réponse aux dommages à l'ADN [235]. Ensuite, des expériences similaires de KO de gènes pro-apoptotiques tels que *BIM* et *BAX* ont montré un phénotype inverse avec une augmentation de la survie comparée aux cellules sauvages (données non publiées, Lewis *et al.*). Pour vérifier cette hypothèse, un nouveau contrôle pourrait être choisi avec un ARNg ciblant une région d'ADN non codante.

2.1.1. Surexpression de la protéine BCL-RAMBO

Le pourcentage d'apoptose induite est évalué par marquage au $DiOC_6(3)$ dans les lignées surexprimant BCL-RAMBO après 3 jours de traitement à 0,5 µM d'imatinib (Figure 96). Aucune différence n'est observée comparée à la lignée transduite avec le vecteur vide.

L'augmentation modeste du niveau protéique de BCL-RAMBO ne module pas l'induction de l'apoptose.

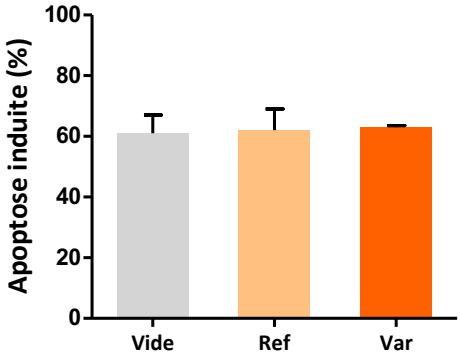


Figure 96 : Pourcentage d'apoptose induite après 72h de traitement à 0,5 µM d'imatinib évalué par marquage $DiOC_6(3)$ dans les lignées K-562 transduites avec le vecteur de surexpression vide (Vide), incluant la séquence codante de BCL-RAMBO de référence (Ref) ou variante (Var) vis-à-vis du polymorphisme rs9306198 (n=2)

Avec ce modèle de surexpression de la protéine, aucun effet n'a été observé sur l'apoptose induite par l'imatinib. Il serait nécessaire de concevoir un nouveau vecteur de surexpression afin d'obtenir une surexpression plus importante de la protéine et potentiellement un phénotype différent comparé à la lignée sauvage.

2.1.1. Changement d'acide aminé causé par le variant rs9306198

Grâce à une nouvelle stratégie de remplacement de fragments d'ADNg (article soumis à Scientific Reports, Annexe 5), l'allèle T (variant) du polymorphisme rs9306198 dans l'exon 11 de *BCL-RAMBO* a été introduit dans la lignée cellulaire K-562, à l'origine porteuse de l'allèle de référence (C). Seule une fraction des copies du gène *BCL-RAMBO* a été modifiée dans deux lignées clonales. Le niveau d'apoptose dans la lignée non modifiée et dans les deux lignées porteuses du variant après 24h de traitement (0,25 ; 0,5 et 1 µM) a été évalué par un marquage annexine V-APC/IP (Figure 97). Le pourcentage de cellules WT mortes après traitement à l'imatinib est dose dépendant. En revanche, la mortalité des cellules exprimant la forme variante de la protéine BCL-RAMBO est quasiment constante quelle que soit la dose d'imatinib utilisée.

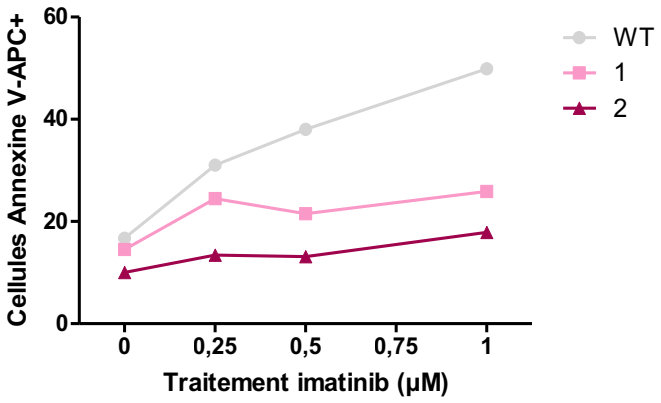


Figure 97 : Pourcentage de cellules annexine V-APC$^+$ à 0,25 ; 0,5 ou 1 µM d'imatinib évalué par marquage DiOC$_6$(3) pour la lignée WT et les deux lignées exprimant la protéine BCL-RAMBO sous la forme variante après 24h de traitement (n=1)

Après 48h, cet effet protecteur est toujours observé dans les deux lignées modifiées après traitement à 0,25 ; 0,5 ou 1 µM d'imatinib (Figure 98).

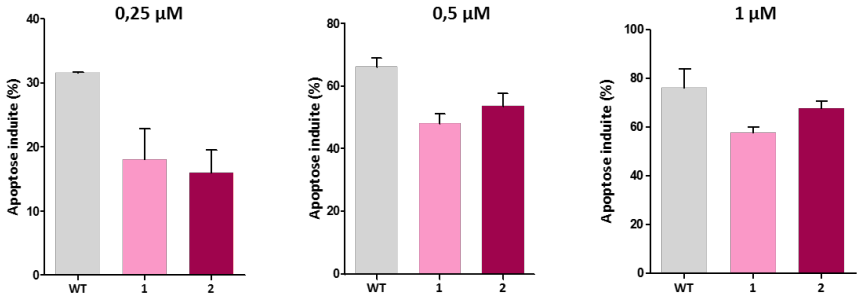


Figure 98 : Pourcentage d'apoptose induite après 48h de traitement à 0,25 ; 0,5 ou 1 µM d'imatinib évalué par marquage DiOC$_6$(3) pour la lignée WT et les deux lignées exprimant la protéine BCL-RAMBO sous la forme variante (n=2)

L'expression de la protéine BCL-RAMBO sous la forme variante dans la lignée K-562 est associée à une diminution de l'apoptose induite. Si la protéine BCL-RAMBO présente un rôle anti-apoptotique dans cette lignée modèle, il semble que la présence du variant soit à l'origine d'un gain de fonction de la protéine.

2.2. Détectée par isolement des mitochondries et western blot

L'induction de l'apoptose peut également être évaluée au niveau protéique par détection de l'expression, de l'activation ou de la localisation cellulaire de différentes protéines. Dans une cellule normale, sans traitement à l'imatinib, le cytochrome c est localisé dans la

mitochondrie. La protéine pro-apoptotique BIM est dégradée dans le cytoplasme et la protéine pro-apoptotique BAX est sous forme de monomère à la membrane mitochondriale externe. Les protéines caspase-3 et PARP1 sont inactives dans le cytoplasme. De plus, il est décrit que BCL-RAMBO est ancré dans la membrane mitochondriale externe (Figure 99).

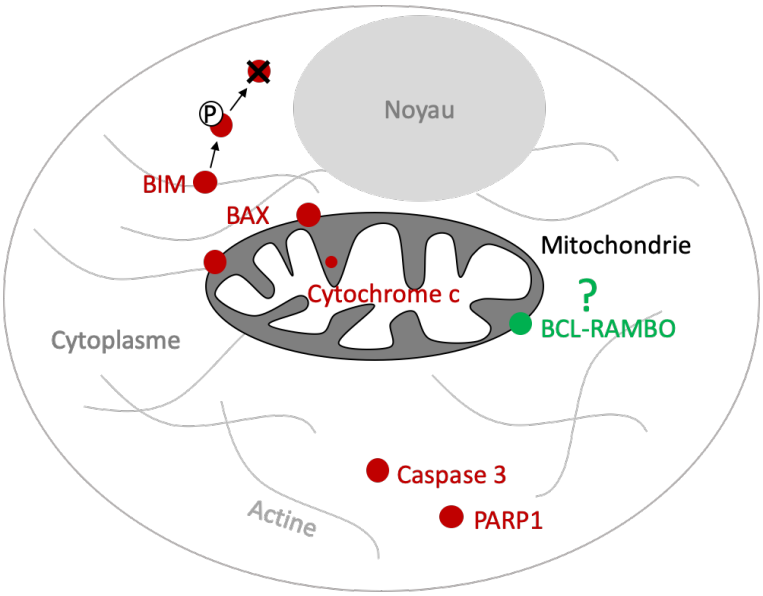


Figure 99 : Représentation d'une cellule WT, en absence de traitement

En présence d'imatinib, la voie de l'apoptose intrinsèque est activée. La protéine BIM est relocalisée dans la membrane mitochondriale et induit la formation de dimères de BAX. Le cytochrome c est alors libéré dans le cytoplasme ce qui entraîne le clivage de la caspase-3 qui entraîne à son tour le clivage de PARP1 à l'origine du signal d'induction de la mort cellulaire (Figure 100).

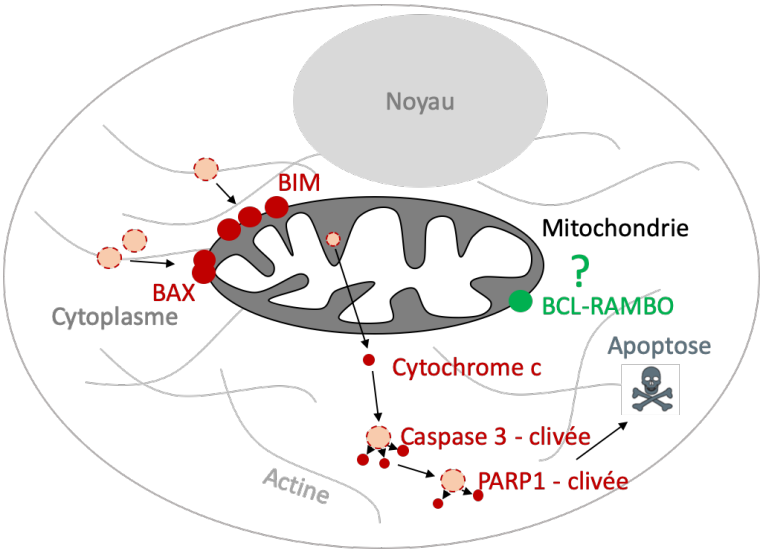


Figure 100 : Représentation d'une cellule WT lorsque l'apoptose est induite notamment par le traitement à l'imatinib

Dans un premier temps, le clivage des protéines caspase-3 et PARP1 a été évalué par western blot dans la lignée K-562 WT, le clone KO n°2 et le clone ΔC-term (=ΔBHNo+TM) pour le gène *BCL-RAMBO* (Figure 101). Une augmentation du pourcentage de ces deux protéines clivées est détectée après 24h de traitement à 0,5 µM d'imatinib. Cette première observation coïncide avec l'augmentation de l'apoptose induite détectée par cytométrie en flux. Le clivage de la caspase-3 est plus important pour la lignée KO que la lignée exprimant à la fois la protéine WT et la protéine tronquée. Le clivage de PARP1 semble équivalent dans les deux lignées. Ces résultats suggèrent que la fonction de BCL-RAMBO dans la régulation de l'apoptose serait en amont du clivage des protéines caspase-3 et PARP1.

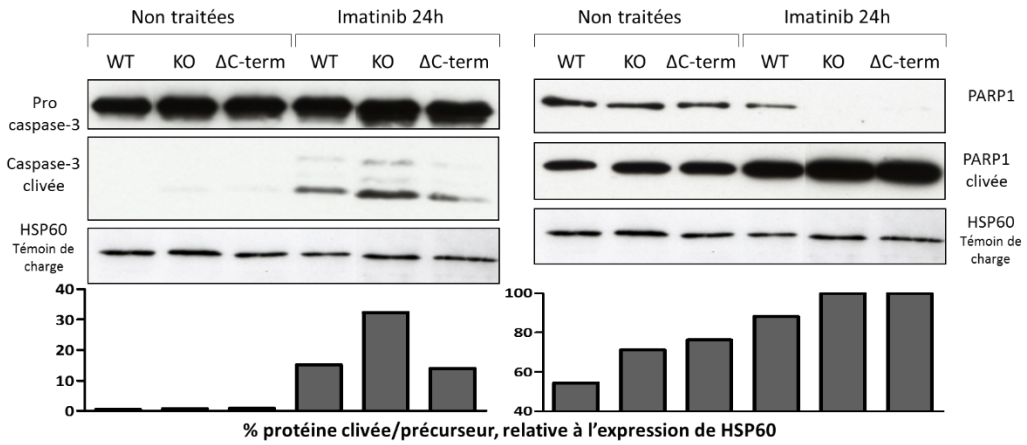


Figure 101 : Expression des protéines caspase-3 et PARP1 (précurseur et protéine clivée) en fonction de l'expression de la protéine HSP60 dans les lignées K-562 WT, KO n°2 et ΔC-term après 24h de traitement à l'imatinib 0,5 µM (n=2)

La quantité de protéines cytochrome c, BIM et BCL-RAMBO dans les mitochondries est également évaluée par western blot (Figure 102). Premièrement, une sortie du cytochrome c de la mitochondrie est détectée pour les lignées WT, KO et tronquée sans différence notable entre les lignées. Deuxièmement, à plusieurs reprises, une disparition de BIM de la mitochondrie a été détectée dans les trois lignées. Cette observation est à l'inverse de ce qui a été décrit précédemment dans la littérature [236]. De manière intéressante, la protéine WT quitte la mitochondrie après induction de l'apoptose. Il est donc possible que la protéine agisse en inhibant directement ou indirectement les protéines BAX ou BAK1 dans la cellule à l'état sauvage avant d'être séquestrée par les protéines pro-apoptotiques initiatrices après induction de l'apoptose (Figure 103).

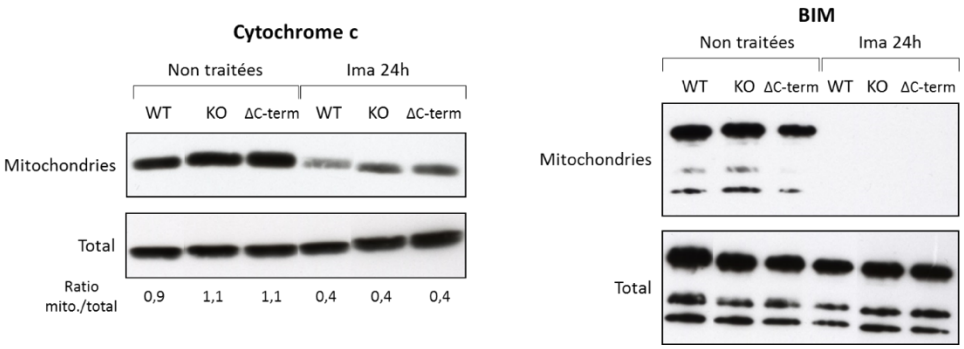


Figure 102 : Expression des protéines cytochrome c et BIM dans la fraction mitochondriale et dans l'extrait total des lignées K-562 WT, KO n°2 et ΔC-term après 24h de traitement à l'imatinib 0,5 µM

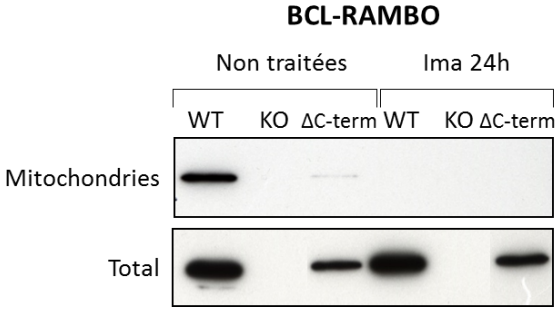


Figure 103 : Expression de la protéine BCL-RAMBO dans la fraction mitochondriale et dans l'extrait total des lignées K-562 WT, KO n°2 et ΔC-term (ΔBHNo+TM) après 24h de traitement à l'imatinib 0,5 µM

La protéine BCL-RAMBO semble donc avoir une fonction anti-apoptotique dans la lignée cellulaire modèle de LMC K-562 en réponse à l'imatinib. Dans les lignées KO ou exprimant la protéine BCL-RAMBO tronquée, l'apoptose induite par l'imatinib est augmentée jusqu'à 30% comparée aux cellules sauvages ce qui suggère un rôle prédominant de BCL-RAMBO dans la régulation de l'apoptose. A l'inverse, le variant, retrouvé enrichi chez les patients en échec de réponse, confère un gain de fonction à la protéine et entraîne une forme de résistance des cellules à l'imatinib.

Discussion

Les premières thérapies ciblées ont fait leur apparition dans les années 2000 et ces nouveaux traitements ont grandement amélioré le pronostic des patients atteints de cancer. Malgré leur efficacité, la résistance au traitement est un problème récurrent. Dans le cas de la LMC, l'imatinib est inefficace pour environ 20% des patients. Au cours de ce travail nous avons recherché de nouveaux mécanismes pouvant être impliqués dans cette résistance au traitement. La variabilité génétique existante entre les individus peut expliquer cette hétérogénéité dans la réponse au traitement.

1. Identification de polymorphismes génétiques par séquençage nouvelle génération

Pour cette étude, une approche de NGS a été choisie afin de rechercher de nouveaux polymorphismes génétiques pouvant être associés à la résistance à l'imatinib dans la LMC. Le séquençage de panel de gènes a été privilégié par rapport au séquençage d'exomes ou génomes entiers, plus coûteux et complexe à analyser. Pour cette étude, le nombre d'échantillons disponibles était limité à 24 (prélevés au diagnostic chez 12 patients répondeurs et 12 patients résistants au traitement à l'imatinib). L'étude de petites cohortes est un problème récurrent dans de nombreuses études pharmacogénétiques et cela réduit les chances d'identifier des marqueurs significativement associés à la réponse au traitement [237]. Le gène *ABCB1* (ou *MDR1*) est un parfait exemple de gènes souvent associés à la résistance à l'imatinib mais jamais inclus dans le diagnostic clinique car des résultats contradictoires sont obtenus d'une étude à l'autre [238]. Le variant 1236C (rs1128503) par exemple a été associé à la résistance à l'imatinib par Dulucq *et al.* (90 patients en phase chronique ou accélérée) [239] et Au *et al.* (215 patients, 107 répondeurs et 103 résistants) [240] et à l'inverse ce même variant a été associé à une meilleure réponse au traitement par Ni *et al.* (52 patients, 26 répondeurs et 25 résistants) [241] ou par Deenik *et al.* (46 patients) [242]. Les critères de sélection des patients varient grandement d'une étude à l'autre (phases de la maladie, traitements utilisés, critères de réponse ...) ce qui augmente d'autant plus la variabilité des résultats. Pour ce travail de thèse une sélection stricte des patients a été réalisée pour la création des cohortes initiale (24 patients) et de validation (103 patients). Les patients étaient tous en phase chronique au diagnostic, traités à l'imatinib 400 mg *de novo*, en échec de réponse ou bon répondeur selon les critères définis par l'ELN 2013. Ces éléments

ont réduit le nombre de patients inclus dans l'étude mais renforcent la puissance des tests statistiques.

Dans notre étude, en plus de l'échantillonnage limité, un panel important de 95 gènes (groupe test de 48 gènes, groupe apoptose de 45 gènes) et par conséquent un grand nombre de polymorphismes a été identifié par séquençage (684 pour le panel test, 473 pour le panel apoptose). Nous avions donc à analyser un grand nombre de polymorphismes pour un petit nombre d'échantillons. Cette situation est fréquente dans les expériences de NGS. Par exemple, en 2018, Shokeen *et al.* ont séquencé les régions exoniques de 4800 gènes à partir de l'ADNg de 18 patients atteints de LMC traités à l'imatinib (10 répondeurs, 8 résistants) [213]. Cette large étude n'a permis de mettre en évidence qu'un seul variant dans le gène *MFGE8* comme marqueur de la résistance à l'imatinib (présents chez 4 patients résistants et aucun répondeur). Dans cette étude, la petite taille de l'échantillon ne permettait pas d'utiliser les approches de statistiques classiques. Les polymorphismes ont donc été triés selon des critères arbitraires mais rigoureux (variant présent pour plus de 50% des patients d'un groupe, absent de l'autre groupe et rare dans la population générale) et un grand nombre de polymorphismes susceptibles d'être associés à la résistance ont par conséquent été exclus de l'analyse. Un séquençage dans une cohorte de validation serait indispensable pour valider le rôle de ce variant dans la résistance à l'imatinib. Pour notre étude, des approches statistiques alternatives ont été développées pour répondre à cette problématique sans passer par l'analyse « à la main » des polymorphismes identifiés. Ce travail a fait l'objet d'une publication soumise à Human Genomics (Annexe 1). Dans un premier temps, afin de visualiser les données sur un graphique, l'approche de l'AFC a été choisie. Cette méthode, souvent utilisée en épidémiologie, a permis de trier rapidement les polymorphismes candidats susceptibles d'être associés à la résistance. Par la suite, la méthode des rank products, développée à l'origine afin de classer des gènes selon leurs différences d'expression entre deux conditions (puces à ADN), a été adaptée pour classer les gènes selon leur différence d'enrichissement en allèles variants entre les patients répondeurs et résistants. De nouveau, cette méthode a permis de mettre en avant facilement une liste réduite de gènes candidats.

La méthode que nous avons développée peut être appliquée à toutes les analyses de données NGS faites à partir d'un petit nombre de patients (<30 patients par groupe). C'est pourquoi, nous avons développé une interface graphique permettant d'appliquer ces nouvelles analyses à n'importe quelle expérience de NGS à partir d'un panel de gènes. Il s'agit de lignes de commandes à utiliser *via* le logiciel de statistiques R et seules quelques notions de base sont nécessaires pour l'utiliser (Annexe 2). Il suffit de créer une matrice d'occurrence regroupant

l'ensemble des génotypes observés pour tous les polymorphismes identifiés. Ce tableau de données doit être importé dans l'interface R pour obtenir en quelques clics un graphique représentant les polymorphismes groupés en clusters et une liste des gènes les plus susceptibles d'être associés à la résistance.

D'autres approches pourraient être utilisées afin d'identifier les meilleurs candidats. L'AFC permet d'analyser les polymorphismes individuellement et la méthode des RP permet d'étudier les polymorphismes en combinaison au sein d'un même gène. En complément, le rôle de plusieurs gènes en combinaison pourrait être également étudié. L'apoptose étant un mécanisme finement régulé grâce à l'interaction de plusieurs facteurs de structures similaires, il est très probable que l'altération simultanée de plusieurs gènes soit la cause de la résistance au traitement.

Au final, plusieurs polymorphismes candidats ont été mis en évidence grâce à cette étude pilote. Le gène *BCL2A1* a été identifié « à la main », avant le développement des outils statistiques. Le variant rs3268007 était présent chez 7 répondeurs et 2 résistants. Après séquençage du polymorphisme candidat rs3268007 dans une cohorte de validation de 103 patients, aucune différence de répartition significative n'a été mise en évidence entre les patients répondeurs et résistants. Ce variant est finalement trop fréquent dans la population générale (environ 20%) et c'est pourquoi il n'est pas apparu dans les analyses statistiques. Il est peu probable qu'un variant si fréquent soit impliqué seul dans la résistance. Un variant rare (rs139012879) du gène *BCLAF1* a également été identifié par l'approche NGS (2 résistants, aucun répondeur porteur de l'allèle variant). Dans la cohorte de validation, un patient répondeur (8%) et quatre patients résistants (33%) étaient porteurs de l'allèle variant alors que la MAF dans la population générale est estimée à 0,5% (ExAC). Ce variant est donc susceptible d'être associé à la résistance mais une cohorte de validation bien plus grande serait nécessaire pour obtenir une différence significative entre les patients répondeurs et résistants au traitement. Enfin, un polymorphisme non-synonyme (rs9306198) du gène *BCL-RAMBO* a été identifié. L'allèle variant T est retrouvé plus fréquemment chez les patients résistants que chez les patients répondeurs et dans la population générale. La différence de répartition de ce variant est significative. De plus, les individus porteurs de l'allèle variant atteignent plus tardivement ou n'atteignent pas la RCH. Ce variant, relativement peu fréquent dans la population générale (environ 5%), est donc susceptible d'être associé à une moins bonne réponse à l'imatinib dans la LMC. De manière intéressante, ce gène de la famille *BCL2* a été très peu étudié (cité dans 42 publications sur Pubmed depuis sa découverte en 2001) et son rôle dans la régulation de

l'apoptose est controversé. C'est pourquoi, la suite de l'étude s'est focalisée sur la caractérisation du gène *BCL-RAMBO* exprimant ou non le variant rs9306198.

2. Création de lignées cellules modifiées grâce à la technologie CRISPR-Cas9

Afin de comprendre le rôle du variant dans la sensibilité des cellules à l'imatinib nous devions obtenir deux lignées cellulaires similaires ne différant que par la présence ou non de l'allèle variant dans le gène *BCL-RAMBO*. L'approche CRISPR-Cas9 a été choisie pour modifier la lignée cellulaire K-562, homozygote de référence pour le polymorphisme rs9306198 de *BCL-RAMBO*.

L'approche couramment utilisée pour générer une substitution spécifique dans l'ADNg nécessite l'activation de la voie de réparation par recombinaison homologue. Or, ce système n'a pas été efficace dans notre modèle et n'a pas permis d'obtenir une lignée K-562 modifiée porteuse de l'allèle variant. Une approche alternative a donc été développée afin d'insérer l'allèle variant dans l'ADNg *via* l'autre voie de réparation, plus fréquente dans les lignées cellulaires humaines : la NHEJ. Cette méthode nommée HIPER (homology-independent precise and efficient remplacement) consiste à supprimer un fragment d'ADNg (l'exon 11) et le remplacer par une séquence exogène sans aucune région d'homologie. Cette méthode a fait l'objet d'une publication soumise à Scientific Reports (Annexe 5). Avec ce système, les cassures doubles brins sont induites dans des régions non codantes pouvant être éloignées de la séquence à modifier, il est donc très facile de concevoir des ARNg efficaces. Cette méthode permet le remplacement de fragments de plusieurs kb. Grâce à cette nouvelle approche de modification du génome, nous avons obtenu deux lignées cellulaires clonales exprimant partiellement la protéine BCL-RAMBO sous la forme variante.

Malgré les promesses de cette approche, un inconvénient majeur existe. L'insertion de la séquence de remplacement n'est pas guidée par des séquences d'homologies, par conséquent, le fragment peut s'insérer dans le sens opposé voire même, plusieurs cassettes peuvent être intégrées dans l'ADNg. Des méthodes de cribles efficaces doivent encore être développées pour valider précisément les modifications induites dans l'ADNg.

Récemment une nouvelle technologie de mutagenèse dirigée par CRISPR-Cas9 a été développée et évolue rapidement depuis 2016 [243–245]. Il s'agit de la technologie de *base editing*. La protéine Cas9, sous forme inactive ou nickase (cassure simple brin), est associée à une enzyme permettant la modification d'une base azotée. D'après les premiers résultats, cette

approche est ciblée et efficace. Cependant, cette méthode présente plusieurs inconvénients. Par exemple, avec l'enzyme xBE4 associée à une cytidine désaminase (substitution d'un C en T), la séquence à modifier doit être entre 13 et 17 pb avant le motif PAM, NG. Le nombre d'options est donc limité et cette stratégie n'est pas applicable à tous les loci. Cependant, pour la création du variant rs9306198, cette méthode aurait été applicable. De plus, aucune autre cytosine n'est présente dans la fenêtre d'action de 5 pb de l'enzyme Cas9. Cette stratégie pourrait donc être testée pour générer de nouvelles lignées cellulaires exprimant le variant d'intérêt dans le gène *BCL-RAMBO*.

Grâce à la technologie CRISPR-Cas9, deux lignées clonales n'exprimant plus la protéine BCL-RAMBO ainsi qu'une lignée exprimant la protéine BCL-RAMBO tronquée (sans les domaines BHNo et TM) ont également été obtenues. Ces différentes lignées ont permis d'étudier le rôle de BCL-RAMBO dans l'apoptose.

3. La régulation de l'apoptose par BCL-RAMBO impliquée dans la réponse à l'imatinib

Dans un premier temps, nous avons mis en évidence que l'absence de BCL-RAMBO dans les cellules K-562 entrainait une augmentation significative de la mort cellulaire programmée. Cette protéine serait donc anti-apoptotique et nécessaire à la régulation de l'apoptose en réponse à l'imatinib. La présence des quatre domaines BH, retrouvés uniquement dans la sous-famille des facteurs comme BCL2, renforce cette hypothèse.

Une fonction pro-apoptotique de BCL-RAMBO avait été décrite dans des modèles *in vitro* dans deux études très similaires [246,247]. Des lignées cellulaires (293T, HeLa, PC-3) avait été transfectées par la séquence codante du gène cloné dans un plasmide. Ces deux équipes décrivent une augmentation du clivage de la caspase-3 lorsque la protéine BCL-RAMBO est surexprimée. Cependant, un rôle anti-apoptotique a également été proposé dans plusieurs publications suite à l'étude de l'expression du gène dans des cellules primaires de patients. Une surexpression de *BCL-RAMBO* a été notamment associée à un mauvais pronostic et une résistance au traitement dans les LAL par deux équipes de recherche différentes (échantillons de 190 et 70 patients) [248,249]. De plus, Jensen *et al.* décrivent une expression élevée du transcrit *BCL-RAMBO* dans de nombreux cancers après une analyse *in silico* (lymphomes analplasiques, LAM, glioblastomes ...) [250]. Ils confirment la surexpression de *BCL-RAMBO* à partir d'échantillons primaires de 31 patients atteints de glioblastomes. Enfin, Tahir *et al.* montrent que la diminution de l'expression du gène est associé à la sensibilité au traitement de

lignées cellulaires de leucémies et lymphomes [251]. D'après ces résultats, la protéine BCL-RAMBO pourrait bien être associée à la sensibilité au traitement voire à la progression tumorale.

Dans un second temps, nous avons observé que le pourcentage de cellules en apoptose était diminué pour les deux lignées exprimant le variant par rapport aux cellules sauvages et ce dès 24h de traitement à l'imatinib. Les premiers résultats suggèrent donc que le changement d'acide aminé (Pro>Ser) induit par le variant dans le domaine BHNo de la protéine BCL-RAMBO serait associé à un gain de fonction de la protéine. Plusieurs hypothèses sont envisagées. Tout d'abord, le variant pourrait moduler l'interaction de BCL-RAMBO avec d'autres protéines. Aucun partenaire de BCL-RAMBO n'a encore été identifié. Une expérience de co-immunoprécipitation à partir des lignées sauvages ou exprimant le variant doit donc être réalisée afin de mieux caractériser les mécanismes d'action de BCL-RAMBO et ses partenaires. Ensuite, le changement d'acide aminé pourrait modifier la localisation de la protéine. A l'heure actuelle, nous n'avons à disposition qu'un anticorps reconnaissant le domaine BHNo de la protéine et aucun anticorps efficace pour marquer l'extrémité N-terminale de la protéine. Il serait intéressant d'intégrer une séquence étiquette fluorescente, en N-terminal de la protéine, afin de suivre la localisation de BCL-RAMBO, sauvage, tronquée ou variante, avec ou sans traitement à l'imatinib. Enfin, le variant pourrait réduire la dégradation de la protéine, par ubiquitinylation notamment, il serait donc intéressant d'évaluer la demi-vie de la protéine avec ou sans changement d'acide aminé en position 360 (Pro>Ser).

Cette étude pilote de NGS nous a donc permis, grâce à de nouvelles méthodes statistiques, de mettre en évidence un variant (rs9306198) enrichi chez les patients résistants dans le gène *BCL-RAMBO* jusqu'alors très peu étudié. Dans notre modèle, la protéine BCL-RAMBO semble avoir un rôle central dans l'inhibition de l'apoptose et par conséquent dans la réponse à l'imatinib. Une stratégie innovante de modification du génome a permis de recréer le variant dans la lignée cellulaire modèle de LMC K-562. Le variant, retrouvé plus fréquemment chez les patients atteints de LMC résistants à l'imatinib, est associé à une perte de la sensibilité des cellules à l'imatinib. Cette approche de séquençage nous a donc permis de détecter un variant gain de fonction dans un gène peu étudié mais finalement important dans la régulation de l'apoptose. Un tel variant gain de fonction n'aurait pas pu être mis en évidence en utilisant des approches de criblage de shRNA ou CRISPR-Cas9. A terme, BCL-RAMBO pourrait donc constituer une cible thérapeutique pour les patients résistants au traitement par altération du mécanisme d'apoptose. Inhiber la fonction de BCL-RAMBO pourrait influencer

la balance entre pro- et anti-apoptotique de la famille BCL2 et déclencher la mort des cellules cancéreuses.

Bibliographie

1. Meier B, Burton JH. Myeloproliferative Disorders. Hematology/Oncology Clinics of North America. 2017;31:1029–44.

2. Bennett JH. Case of hypertrophy of the spleen and liver, which death took place from suppuration of the blood. The Edinburgh medical and surgical journal. 1845;413–23.

3. Virchow R. Weisses Blut. Frorieps Notizen. 1845;151–6.

4. American Cancer Society, editor. Cancer Facts & Figures.

5. National Cancer Institute. Cancer Stat Facts: Leukemia - Chronic Myeloid Leukemia (CML).

6. Huang X, Cortes J, Kantarjian H. Estimations of the increasing prevalence and plateau prevalence of chronic myeloid leukemia in the era of tyrosine kinase inhibitor therapy. Cancer. 2012;118:3123–7.

7. Druker BJ. Translation of the Philadelphia chromosome into therapy for CML. Blood. 2008;112:4808–17.

8. Calabretta B, Perrotti D. The biology of CML blast crisis. Blood. 2004;103:4010–22.

9. Schoch C, Haferlach T, Kern W, Schnittger S, Berger U, Hehlmann R, et al. Occurrence of additional chromosome aberrations in chronic myeloid leukemia patients treated with imatinib mesylate. Leukemia. 2003;17:461–3.

10. Muvarak N, Nagaria P, Rassool FV. Genomic Instability in Chronic Myeloid Leukemia: Targets for Therapy? Current Hematologic Malignancy Reports. 2012;7:94–102.

11. Bernstein R. Cytogenetics of chronic myelogenous leukemia. Semin Hematol. 1988;25:20–34.

12. Feinstein E, Cimino G, Gale RP, Alimena G, Berthier R, Kishi K, et al. p53 in chronic myelogenous leukemia in acute phase. Proc Natl Acad Sci USA. 1991;88:6293–7.

13. Towatari M, Adachi K, Kato H, Saito H. Absence of the human retinoblastoma gene product in the megakaryoblastic crisis of chronic myelogenous leukemia. Blood. 1991;78:2178–81.

14. Sill H, Goldrnan JM. Homozygous Deletions of the p16 Tumor-Suppressor Gene Are Associated With Lymphoid Transformation of Chronic Myeloid Leukemia. :5.

15. Nowell PC, Hungerford D. A minute chromosome in human chronic granulocytic leukemia.

16. Rowley JD. A New Consistent Chromosomal Abnormality in Chronic Myelogenous Leukaemia identified by Quinacrine Fluorescence and Giemsa Staining. Nature. 1973;243:290–3.

17. Apperley JF. Chronic myeloid leukaemia. The Lancet. 2015;385:1447–59.

18. Fialkow PJ, Gartler SM, Yoshida A. Clonal origin of chronic myelocytic leukemia in man. Proc Natl Acad Sci USA. 1967;58:1468–71.

19. Abelson HT, Rabstein LS. Lymphosarcoma: virus-induced thymic-independent disease in mice. Cancer Res. 1970;30:2213–22.

20. Heisterkamp N, Stephenson JR, Groffen J, Hansen PF, de Klein A, Bartram CR, et al. Localization of the c-abl oncogene adjacent to a translocation break point in chronic myelocytic leukaemia. Nature. 1983;306:239–42.

21. Melo JV. The diversity of BCR-ABL fusion proteins and their relationship to leukemia phenotype. Blood. 1996;88:2375–84.

22. Groffen J, Stephenson JR, Heisterkamp N, de Klein A, Bartram CR, Grosveld G. Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22. Cell. 1984;36:93–9.

23. Deininger MW, Goldman JM, Melo JV. The molecular biology of chronic myeloid leukemia. Blood. 2000;96:3343–56.

24. Heisterkamp N, Knoppel E, Groffen J. The first BCR gene intron contains breakpoints in Philadelphia chromosome positive leukemia. Nucleic Acids Res. 1988;16:10069–81.

25. Ravandi F, Cortes J, Albitar M, Arlinghaus R, Qiang Guo J, Talpaz M, et al. Chronic myelogenous leukaemia with p185(BCR/ABL) expression: characteristics and clinical significance. Br J Haematol. 1999;107:581–6.

26. Tanaka K, Takechi M, Hong J, Shigeta C, Oguma N, Kamada N, et al. 9;22 translocation and bcr rearrangements in chronic myelocytic leukemia patients among atomic bomb survivors. J Radiat Res. 1989;30:352–8.

27. Corso A, Lazzarino M, Morra E, Merante S, Astori C, Bernasconi P, et al. Chronic myelogenous leukemia and exposure to ionizing radiation--a retrospective study of 443 patients. Ann Hematol. 1995;70:79–82.

28. Deininger MW, Bose S, Gora-Tybor J, Yan XH, Goldman JM, Melo JV. Selective induction of leukemia-associated fusion genes by high-dose ionizing radiation. Cancer Res. 1998;58:421–5.

29. Kozubek S, Lukásová E, Rýznar L, Kozubek M, Lisková A, Govorun RD, et al. Distribution of ABL and BCR genes in cell nuclei of normal and irradiated lymphocytes. Blood. 1997;89:4537–45.

30. Neves H, Ramos C, da Silva MG, Parreira A, Parreira L. The nuclear topography of ABL, BCR, PML, and RARalpha genes: evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation. Blood. 1999;93:1197–207.

31. Saglio G, Pane F, Martinelli G, Guerrasio A. BCR/ABL rearrangement and leukemia phenotype. Leukemia. 1999;13:S96–S96.

32. Melo JV, Gordon DE, Cross NC, Goldman JM. The ABL-BCR fusion gene is expressed in chronic myeloid leukemia. Blood. 1993;81:158–65.

33. Melo JV, Hochhaus A, Yan XH, Goldman JM. Lack of correlation between ABL-BCR expression and response to interferon-alpha in chronic myeloid leukaemia. Br J Haematol. 1996;92:684–6.

34. de la Fuente J, Merx K, Steer EJ, Müller M, Szydlo RM, Maywald O, et al. ABL-BCR expression does not correlate with deletions on the derivative chromosome 9 or survival in chronic myeloid leukemia. Blood. 2001;98:2879–80.

35. Lugo TG, Pendergast AM, Muller AJ, Witte ON. Tyrosine kinase activity and transformation potency of bcr-abl oncogene products. Science. 1990;247:1079–82.

36. Laneuville P. Abl tyrosine protein kinase. Seminars in Immunology. 1995;7:255–66.

37. Kipreos ET, Wang JY. Cell cycle-regulated binding of c-Abl tyrosine kinase to DNA. Science. 1992;256:382–5.

38. Sawyers CL, McLaughlin J, Goga A, Havlik M, Witte O. The nuclear tyrosine kinase c-Abl negatively regulates cell growth. Cell. 1994;77:121–31.

39. Wang JY. Regulation of cell death by the Abl tyrosine kinase. Oncogene. 2000;19:5643–50.

40. Yuan ZM, Shioya H, Ishiko T, Sun X, Gu J, Huang YY, et al. p73 is regulated by tyrosine kinase c-Abl in the apoptotic response to DNA damage. Nature. 1999;399:814–7.

41. Lewis JM, Schwartz MA. Integrins regulate the association and phosphorylation of paxillin by c-Abl. J Biol Chem. 1998;273:14225–30.

42. Cohen GB, Ren R, Baltimore D. Modular binding domains in signal transduction proteins. Cell. 1995;80:237–48.

43. Van Etten RA, Jackson P, Baltimore D. The mouse type IV c-abl gene product is a nuclear protein, and activation of transforming ability is associated with cytoplasmic localization. Cell. 1989;58:669–78.

44. McWhirter JR, Wang JY. An actin-binding function contributes to transformation by the Bcr-Abl oncoprotein of Philadelphia chromosome-positive human leukemias. EMBO J. 1993;12:1533–46.

45. Reddy EP, Aggarwal AK. The Ins and Outs of Bcr-Abl Inhibition. Genes & Cancer. 2012;3:447–54.

46. Woodring PJ. Regulation of F-actin-dependent processes by the Abl family of tyrosine kinases. Journal of Cell Science. 2003;116:2613–26.

47. Reuther GW, Fu H, Cripe LD, Collier RJ, Pendergast AM. Association of the protein kinases c-Bcr and Bcr-Abl with proteins of the 14-3-3 family. Science. 1994;266:129–33.

48. McWhirter JR, Galasso DL, Wang JY. A coiled-coil oligomerization domain of Bcr is essential for the transforming function of Bcr-Abl oncoproteins. Mol Cell Biol. 1993;13:7587–95.

49. Denhardt DT. Signal-transducing protein phosphorylation cascades mediated by Ras/Rho proteins in the mammalian cell: the potential for multiplex signalling. Biochem J. 1996;318 (Pt 3):729–47.

50. Diekmann D, Brill S, Garrett MD, Totty N, Hsuan J, Monfries C, et al. Bcr encodes a GTPase-activating protein for p21rac. Nature. 1991;351:400–2.

51. Pendergast AM, Muller AJ, Havlik MH, Maru Y, Witte ON. BCR sequences essential for transformation by the BCR-ABL oncogene bind to the ABL SH2 regulatory domain in a non-phosphotyrosine-dependent manner. Cell. 1991;66:161–71.

52. Pendergast AM, Quilliam LA, Cripe LD, Bassing CH, Dai Z, Li N, et al. BCR-ABL-induced oncogenesis is mediated by direct interaction with the SH2 domain of the GRB-2 adaptor protein. Cell. 1993;75:175–85.

53. Ma G, Lu D, Wu Y, Liu J, Arlinghaus RB. Bcr phosphorylated on tyrosine 177 binds Grb2. Oncogene. 1997;14:2367–72.

54. Wu Y, Liu J, Arlinghaus RB. Requirement of two specific tyrosine residues for the catalytic activity of Bcr serine/threonine kinase. Oncogene. 1998;16:141–6.

55. Frazer R, Irvine AE, McMullin MF. Chronic Myeloid Leukaemia in The 21st Century. Ulster Med J. 2007;76:8–17.

56. Sattler M, Griffin JD. Molecular mechanisms of transformation by the BCR-ABL oncogene. Semin Hematol. 2003;40:4–10.

57. Pagliarini R, Shao W, Sellers WR. Oncogene addiction: pathways of therapeutic response, resistance, and road maps toward a cure. EMBO Rep. 2015;16:280–96.

58. Bennour A, Bellâaj H, Ben Youssef Y, Elloumi M, Khelif A, Saad A, et al. Molecular cytogenetic characterization of Philadelphia-negative rearrangements in chronic myeloid leukemia patients. J Cancer Res Clin Oncol. 2011;137:1329–36.

59. Walker LM, Hardy JD. Philadelphia chromosome in acute leukemia: case report. Cancer. 1976;38:1619–24.

60. Hehlmann R, Hochhaus A, Baccarani M, European LeukemiaNet. Chronic myeloid leukaemia. Lancet. 2007;370:342–50.

61. Silver RT, Woolf SH, Hehlmann R, Appelbaum FR, Anderson J, Bennett C, et al. An evidence-based analysis of the effect of busulfan, hydroxyurea, interferon, and allogeneic bone marrow transplantation in treating the chronic phase of chronic myeloid leukemia: developed for the American Society of Hematology. Blood. 1999;94:1517–36.

62. Geary CG. The story of chronic myeloid leukaemia. Historical review. British Journal of Haematology. 2000;110:2–11.

63. Baccarani M, Saglio G, Goldman J, Hochhaus A, Simonsson B, Appelbaum F, et al. Evolving concepts in the management of chronic myeloid leukemia: recommendations from an expert panel on behalf of the European LeukemiaNet. Blood. 2006;108:1809–20.

64. Shenoy S, Smith FO. Hematopoietic stem cell transplantation for childhood malignancies of myeloid origin. Bone Marrow Transplant. 2008;41:141–8.

65. Buchdunger E, Zimmermann J, Mett H, Meyer T, Müller M, Regenass U, et al. Selective inhibition of the platelet-derived growth factor signal transduction pathway by a protein-tyrosine kinase inhibitor of the 2-phenylaminopyrimidine class. Proc Natl Acad Sci USA. 1995;92:2558–62.

66. Druker BJ, Tamura S, Buchdunger E, Ohno S, Segal GM, Fanning S, et al. Effects of a selective inhibitor of the Abl tyrosine kinase on the growth of Bcr-Abl positive cells. Nat Med. 1996;2:561–6.

67. Deininger M, Buchdunger E, Druker BJ. The development of imatinib as a therapeutic agent for chronic myeloid leukemia. Blood. 2005;105:2640–53.

68. Druker BJ, Talpaz M, Resta DJ, Peng B, Buchdunger E, Ford JM, et al. Efficacy and Safety of a Specific Inhibitor of the BCR-ABL Tyrosine Kinase in Chronic Myeloid Leukemia. New England Journal of Medicine. 2001;344:1031–7.

69. Hochhaus A, Druker B, Sawyers C, Guilhot F, Schiffer CA, Cortes J, et al. Favorable long-term follow-up results over 6 years for response, survival, and safety with imatinib mesylate therapy in chronic-phase chronic myeloid leukemia after failure of interferon- treatment. Blood. 2007;111:1039–43.

70. Deininger M, O'Brien SG, Guilhot F, Goldman JM, Hochhaus A, Hughes TP, et al. International Randomized Study of Interferon Vs STI571 (IRIS) 8-Year Follow up: Sustained Survival and Low Risk for Progression or Events in Patients with Newly Diagnosed Chronic Myeloid Leukemia in Chronic Phase (CML-CP) Treated with Imatinib. Blood; 2009.

71. Kantarjian H, Cortes JE. Complete cytogenetic response, not deep molecular response, is associated with survival in chronic myeloid leukemia. J Clin Oncol. 2014;32:3077.

72. Mughal TI, Radich JP, Deininger MW, Apperley JF, Hughes TP, Harrison CJ, et al. Chronic myeloid leukemia: reminiscences and dreams. Haematologica. 2016;101:541–58.

73. Weisberg E, Manley PW, Breitenstein W, Brüggen J, Cowan-Jacob SW, Ray A, et al. Characterization of AMN107, a selective inhibitor of native and mutant Bcr-Abl. Cancer Cell. 2005;7:129–41.

74. Lombardo LJ, Lee FY, Chen P, Norris D, Barrish JC, Behnia K, et al. Discovery of *N* -(2-Chloro-6-methyl- phenyl)-2-(6-(4-(2-hydroxyethyl)- piperazin-1-yl)-2-methylpyrimidin-4-ylamino)thiazole-5-carboxamide (BMS-354825), a Dual Src/Abl Kinase Inhibitor with Potent Antitumor Activity in Preclinical Assays. Journal of Medicinal Chemistry. 2004;47:6658–61.

75. Shah NP, Tran C, Lee FY, Chen P, Norris D, Sawyers CL. Overriding imatinib resistance with a novel ABL kinase inhibitor. Science. 2004;305:399–401.

76. Golas JM, Arndt K, Etienne C, Lucas J, Nardin D, Gibbons J, et al. SKI-606, a 4-Anilino-3-quinolinecarbonitrile Dual Inhibitor of Src and Abl Kinases, Is a Potent Antiproliferative Agent against Chronic Myelogenous Leukemia Cells in Culture and Causes Regression of K562 Xenografts in Nude Mice. :8.

77. Falchi L, Kantarjian HM, Wang X, Verma D, Quintás-Cardama A, O'Brien S, et al. Significance of deeper molecular responses in patients with chronic myeloid leukemia in early chronic phase treated with tyrosine kinase inhibitors. Am J Hematol. 2013;88:1024–9.

78. Koller PB, Kantarjian HM, Nogueras-Gonzalez GM, Jabbour E, Verstovsek S, Borthakur G, et al. Chronic myeloid leukemia among patients with a history of prior malignancies: A tale of dual survivorship. Cancer. 2017;123:609–16.

79. O'Hare T, Shakespeare WC, Zhu X, Eide CA, Rivera VM, Wang F, et al. AP24534, a pan-BCR-ABL inhibitor for chronic myeloid leukemia, potently inhibits the T315I mutant and overcomes mutation-based resistance. Cancer Cell. 2009;16:401–12.

80. Sokal JE, Cox EB, Baccarani M, Tura S, Gomez GA, Robertson JE, et al. Prognostic discrimination in "good-risk" chronic granulocytic leukemia. Blood. 1984;63:789–99.

81. Hasford J, Pfirrmann M, Hehlmann R, Allan NC, Baccarani M, Kluin-Nelemans JC, et al. A new prognostic score for survival of patients with chronic myeloid leukemia treated with interferon alfa. Writing Committee for the Collaborative CML Prognostic Factors Project Group. J Natl Cancer Inst. 1998;90:850–8.

82. Hasford J, Baccarani M, Hoffmann V, Guilhot J, Saussele S, Rosti G, et al. Predicting complete cytogenetic response and subsequent progression-free survival in 2060 patients with CML on imatinib treatment: the EUTOS score. Blood. 2011;118:686–92.

83. Marin D, Ibrahim AR, Goldman JM. European Treatment and Outcome Study (EUTOS) score for chronic myeloid leukemia still requires more confirmation. J Clin Oncol. 2011;29:3944–5.

84. Wang YL, Bagg A, Pear W, Nowell PC, Hess JL. Chronic myelogenous leukemia: laboratory diagnosis and monitoring. Genes Chromosomes Cancer. 2001;32:97–111.

85. Mughal T, Cortes J, Cross NCP, Donato N, Hantschel O, Jabbour E, et al. Chronic myeloid leukemia--some topical issues. Leukemia. 2007;21:1347–52.

86. Baccarani M, Cortes J, Pane F, Niederwieser D, Saglio G, Apperley J, et al. Chronic Myeloid Leukemia: An Update of Concepts and Management Recommendations of European LeukemiaNet. Journal of Clinical Oncology. 2009;27:6041–51.

87. Baccarani M, Deininger MW, Rosti G, Hochhaus A, Soverini S, Apperley JF, et al. European LeukemiaNet recommendations for the management of chronic myeloid leukemia: 2013. Blood. 2013;122:872–84.

88. Mauro MJ. Defining and Managing Imatinib Resistance. Hematology. 2006;2006:219–25.

89. O'Brien SG, Guilhot F, Larson RA, Gathmann I, Baccarani M, Cervantes F, et al. Imatinib Compared with Interferon and Low-Dose Cytarabine for Newly Diagnosed Chronic-Phase Chronic Myeloid Leukemia. New England Journal of Medicine. 2003;348:994–1004.

90. Gorre ME, Mohammed M, Ellwood K, Hsu N, Paquette R, Rao PN, et al. Clinical resistance to STI-571 cancer therapy caused by BCR-ABL gene mutation or amplification. Science. 2001;293:876–80.

91. Deininger MW, Hodgson JG, Shah NP, Cortes JE, Kim D-W, Nicolini FE, et al. Compound mutations in BCR-ABL1 are not major drivers of primary or secondary resistance to ponatinib in CP-CML patients. Blood. 2016;127:703–12.

92. Wang D, Pan H, Wang Y. T315L: a novel mutation within BCR-ABL kinase domain confers resistance against ponatinib. Leukemia & Lymphoma. 2017;58:1733–5.

93. Soverini S, de Benedittis C, Mancini M, Martinelli G. Mutations in the BCR-ABL1 Kinase Domain and Elsewhere in Chronic Myeloid Leukemia. Clin Lymphoma Myeloma Leuk. 2015;15 Suppl:S120-128.

94. Corbin AS, Buchdunger E, Pascal F, Druker BJ. Analysis of the Structural Basis of Specificity of Inhibition of the Abl Kinase by STI571. Journal of Biological Chemistry. 2002;277:32214–9.

95. Shah NP, Nicoll JM, Nagar B, Gorre ME, Paquette RL, Kuriyan J, et al. Multiple BCR-ABL kinase domain mutations confer polyclonal resistance to the tyrosine kinase inhibitor imatinib (STI571) in chronic phase and blast crisis chronic myeloid leukemia. Cancer Cell. 2002;2:117–25.

96. le Coutre P, Tassi E, Varella-Garcia M, Barni R, Mologni L, Cabrita G, et al. Induction of resistance to the Abelson inhibitor STI571 in human leukemic cells through gene amplification. Blood. 2000;95:1758–66.

97. Mahon FX, Deininger MW, Schultheis B, Chabrol J, Reiffers J, Goldman JM, et al. Selection and characterization of BCR-ABL positive cell lines with differential sensitivity to the tyrosine kinase inhibitor STI571: diverse mechanisms of resistance. Blood. 2000;96:1070–9.

98. Picard S, Titier K, Etienne G, Teilhet E, Ducint D, Bernard M-A, et al. Trough imatinib plasma levels are associated with both cytogenetic and molecular responses to standard-dose imatinib in chronic myeloid leukemia. Blood. 2007;109:3496–9.

99. le Coutre P, Kreuzer K-A, Pursche S, Bonin M v, Leopold T, Baskaynak G, et al. Pharmacokinetics and cellular uptake of imatinib and its main metabolite CGP74588. Cancer Chemother Pharmacol. 2004;53:313–23.

100. Peng B, Lloyd P, Schran H. Clinical pharmacokinetics of imatinib. Clin Pharmacokinet. 2005;44:879–94.

101. Chen S, Sutiman N, Chowbay B. Pharmacogenetics of drug transporters in modulating imatinib disposition and treatment outcomes in chronic myeloid leukemia & gastrointestinal stromal tumor patients. Pharmacogenomics. 2016;

102. Floyd MD, Gervasini G, Masica AL, Mayo G, George AL, Bhat K, et al. Genotype-phenotype associations for common CYP3A4 and CYP3A5 variants in the basal and induced metabolism of midazolam in European- and African-American men and women. Pharmacogenetics. 2003;13:595–606.

103. Ankathil R, Azlan H, Dzarr AA, Baba AA. Pharmacogenetics and the treatment of chronic myeloid leukemia: how relevant clinically? An update. Pharmacogenomics [Internet]. 2018

[cited 2018 Apr 3]; Available from: https://www.futuremedicine.com/doi/10.2217/pgs-2017-0193

104. Smith KD, Paterson S. Binding of alpha-1-acid glycoprotein to imatinib following increased dosage of drug. Haematologica. 2005;90 Suppl:ELT01.

105. Thomas J, Wang L, Clark RE, Pirmohamed M. Active transport of imatinib into and out of cells: implications for drug resistance. Blood. 2004;104:3739–45.

106. Crossman LC, Druker BJ, Deininger MWN, Pirmohamed M, Wang L, Clark RE. hOCT 1 and resistance to imatinib. Blood. 2005;106:1133–4; author reply 1134.

107. Angelini S, Soverini S, Ravegnini G, Barnett M, Turrini E, Thornquist M, et al. Association between imatinib transporters and metabolizing enzymes genotype and response in newly diagnosed chronic myeloid leukemia patients receiving imatinib therapy. Haematologica. 2013;98:193–200.

108. Ughachukwu P, Unekwe P. Efflux pump-mediated resistance in chemotherapy. Ann Med Health Sci Res. 2012;2:191–8.

109. Mahon F-X, Belloc F, Lagarde V, Chollet C, Moreau-Gaudry F, Reiffers J, et al. MDR1 gene overexpression confers resistance to imatinib mesylate in leukemia cell line models. Blood. 2003;101:2368–73.

110. Mukai M, Che X-F, Furukawa T, Sumizawa T, Aoki S, Ren X-Q, et al. Reversal of the resistance to STI571 in human chronic myelogenous leukemia K562 cells. Cancer Sci. 2003;94:557–63.

111. Hughes TP, Kaeda J, Branford S, Rudzki Z, Hochhaus A, Hensley ML, et al. Frequency of Major Molecular Responses to Imatinib or Interferon Alfa plus Cytarabine in Newly Diagnosed Chronic Myeloid Leukemia. New England Journal of Medicine. 2003;349:1423–32.

112. Gallipoli P, Abraham SA, Holyoake TL. Hurdles toward a cure for CML: the CML stem cell. Hematol Oncol Clin North Am. 2011;25:951–66, v.

113. Holtz MS, Forman SJ, Bhatia R. Nonproliferating CML CD34+ progenitors are resistant to apoptosis induced by a wide range of proapoptotic stimuli. Leukemia. 2005;19:1034–41.

114. Corbin AS, Agarwal A, Loriaux M, Cortes J, Deininger MW, Druker BJ. Human chronic myeloid leukemia stem cells are insensitive to imatinib despite inhibition of BCR-ABL activity. J Clin Invest. 2011;121:396–409.

115. Schepers K, Campbell TB, Passegué E. Normal and leukemic stem cell niches: insights and therapeutic opportunities. Cell Stem Cell. 2015;16:254–67.

116. O'Dwyer ME, Mauro MJ, Kurilik G, Mori M, Balleisen S, Olson S, et al. The impact of clonal evolution on response to imatinib mesylate (STI571) in accelerated phase CML. Blood. 2002;100:1628–33.

117. O'Dwyer ME, Mauro MJ, Blasdel C, Farnsworth M, Kurilik G, Hsieh Y-C, et al. Clonal evolution and lack of cytogenetic response are adverse prognostic factors for hematologic

relapse of chronic phase CML patients treated with imatinib mesylate. Blood. 2004;103:451–5.

118. Cortes JE, Talpaz M, Giles F, O'Brien S, Rios MB, Shan J, et al. Prognostic significance of cytogenetic clonal evolution in patients with chronic myelogenous leukemia on imatinib mesylate therapy. Blood. 2003;101:3794–800.

119. Lahaye T, Riehm B, Berger U, Paschka P, Müller MC, Kreil S, et al. Response and resistance in 300 patients with BCR-ABL-positive leukemias treated with imatinib in a single center: a 4.5-year follow-up. Cancer. 2005;103:1659–69.

120. Jabbour E, Kantarjian H, Jones D, Talpaz M, Bekele N, O'Brien S, et al. Frequency and clinical significance of BCR-ABL mutations in patients with chronic myeloid leukemia treated with imatinib mesylate. Leukemia. 2006;20:1767–73.

121. Donato NJ, Wu JY, Stapley J, Gallick G, Lin H, Arlinghaus R, et al. BCR-ABL independence and LYN kinase overexpression in chronic myelogenous leukemia cells selected for resistance to STI571. Blood. 2003;101:690–8.

122. Schindler T, Bornmann W, Pellicena P, Miller WT, Clarkson B, Kuriyan J. Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. Science. 2000;289:1938–42.

123. Deininger MW, Goldman JM, Lydon N, Melo JV. The tyrosine kinase inhibitor CGP57148B selectively inhibits the growth of BCR-ABL-positive cells. Blood. 1997;90:3691–8.

124. Wood W, Turmaine M, Weber R, Camp V, Maki RA, McKercher SR, et al. Mesenchymal cells engulf and clear apoptotic footplate cells in macrophageless PU.1 null mouse embryos. Development. 2000;127:5245–52.

125. Kerr JF, Wyllie AH, Currie AR. Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer. 1972;26:239–57.

126. Horvitz HR. Genetic control of programmed cell death in the nematode Caenorhabditis elegans. Cancer Res. 1999;59:1701s–6s.

127. Jeanteur P, Galas S. [2002 Nobel Prize in Physiology and Medecine: from nematodes to programmed cell death]. Bull Cancer. 2002;89:838.

128. Abou-Ghali M, Stiban J. Regulation of ceramide channel formation and disassembly: Insights on the initiation of apoptosis. Saudi Journal of Biological Sciences. 2015;22:760–72.

129. Dikic I, Elazar Z. Mechanism and medical implications of mammalian autophagy. Nature Reviews Molecular Cell Biology. 2018;19:349–64.

130. Mariño G, Niso-Santano M, Baehrecke EH, Kroemer G. Self-consumption: the interplay of autophagy and apoptosis. Nature Reviews Molecular Cell Biology. 2014;15:81–94.

131. Ichim G, Tait SWG. A fate worse than death: apoptosis as an oncogenic process. Nat Rev Cancer. 2016;16:539–48.

132. Pop C, Salvesen GS. Human Caspases: Activation, Specificity, and Regulation. Journal of Biological Chemistry. 2009;284:21777–81.

133. Locksley RM, Killeen N, Lenardo MJ. The TNF and TNF receptor superfamilies: integrating mammalian biology. Cell. 2001;104:487–501.

134. Ashkenazi A, Dixit VM. Death receptors: signaling and modulation. Science. 1998;281:1305–8.

135. Chicheportiche Y, Bourdon PR, Xu H, Hsu Y-M, Scott H, Hession C, et al. TWEAK, a New Secreted Ligand in the Tumor Necrosis Factor Family That Weakly Induces Apoptosis. Journal of Biological Chemistry. 1997;272:32401–10.

136. Suliman A, Lam A, Datta R, Srivastava RK. Intracellular mechanisms of TRAIL: apoptosis through mitochondrial-dependent and -independent pathways. Oncogene. 2001;20:2122–33.

137. Wajant H. The Fas signaling pathway: more than a paradigm. Science. 2002;296:1635–6.

138. Kischkel FC, Hellbardt S, Behrmann I, Germer M, Pawlita M, Krammer PH, et al. Cytotoxicity-dependent APO-1 (Fas/CD95)-associated proteins form a death-inducing signaling complex (DISC) with the receptor. EMBO J. 1995;14:5579–88.

139. Adams JM, Cory S. The Bcl-2 protein family: arbiters of cell survival. Science. 1998;281:1322–6.

140. García-Sáez AJ, Mingarro I, Pérez-Payá E, Salgado J. Membrane-Insertion Fragments of Bcl-x_L, Bax, and Bid †. Biochemistry. 2004;43:10930–43.

141. Cory S, Adams JM. The Bcl2 family: regulators of the cellular life-or-death switch. Nature Reviews Cancer. 2002;2:647–56.

142. Elmore S. Apoptosis: a review of programmed cell death. Toxicol Pathol. 2007;35:495–516.

143. Li H, Zhu H, Xu C, Yuan J. Cleavage of BID by Caspase 8 Mediates the Mitochondrial Damage in the Fas Pathway of Apoptosis. Cell. 1998;94:491–501.

144. Luo X, Budihardjo I, Zou H, Slaughter C, Wang X. Bid, a Bcl2 Interacting Protein, Mediates Cytochrome c Release from Mitochondria in Response to Activation of Cell Surface Death Receptors. Cell. 1998;94:481–90.

145. Jiang P, Du W, Wu M. p53 and Bad: remote strangers become close friends. Cell Research. 2007;17:283–5.

146. Pawlowski J, Kraft AS. Bax-induced apoptotic cell death. Proceedings of the National Academy of Sciences. 2000;97:529–31.

147. Fu Z, Tindall DJ. FOXOs, cancer and regulation of apoptosis. Oncogene. 2008;27:2312–9.

148. de Groot RP, Raaijmakers JAM, Lammers J-WJ, Koenderman L. STAT5-Dependent CyclinD1 and Bcl-xL Expression in Bcr-Abl-Transformed Cells. Molecular Cell Biology Research Communications. 2000;3:299–305.

149. Gottlieb TM, Leal JFM, Seger R, Taya Y, Oren M. Cross-talk between Akt, p53 and Mdm2: possible implications for the regulation of apoptosis. Oncogene. 2002;21:1299–303.

150. Cheong R, Hoffmann A, Levchenko A. Understanding NF-kappaB signaling via mathematical modeling. Mol Syst Biol. 2008;4:192.

151. Riley T, Sontag E, Chen P, Levine A. Transcriptional control of human p53-regulated genes. Nat Rev Mol Cell Biol. 2008;9:402–12.

152. Bai D, Ueno L, Vogt PK. Akt-mediated regulation of NFkappaB and the essentialness of NFkappaB for the oncogenicity of PI3K and Akt. Int J Cancer. 2009;125:2863–70.

153. Hay N. Interplay between FOXO, TOR, and Akt. Biochim Biophys Acta. 2011;1813:1965–70.

154. Zhang X, Tang N, Hadden TJ, Rishi AK. Akt, FoxO and regulation of apoptosis. Biochim Biophys Acta. 2011;1813:1978–86.

155. Tekedereli I, Alpay SN, Tavares CDJ, Cobanoglu ZE, Kaoud TS, Sahin I, et al. Targeted Silencing of Elongation Factor 2 Kinase Suppresses Growth and Sensitizes Tumors to Doxorubicin in an Orthotopic Model of Breast Cancer. Ahmad A, editor. PLoS ONE. 2012;7:e41171.

156. Vogler M. BCL2A1: the underdog in the BCL2 family. Cell Death Differ. 2012;19:67–74.

157. Waibel M, Solomon VS, Knight DA, Ralli RA, Kim S-K, Banks K-M, et al. Combined targeting of JAK2 and Bcl-2/Bcl-xL to cure mutant JAK2-driven malignancies and overcome acquired resistance to JAK2 inhibitors. Cell Rep. 2013;5:1047–59.

158. Bahrami-B F, Ataie-Kachoie P, Pourgholami MH, Morris DL. p70 Ribosomal protein S6 kinase (Rps6kb1): an update. J Clin Pathol. 2014;67:1019–25.

159. Shoemaker MD RC, Ryan PhD JC. A Gene Primer for Health Care Providers: the Genomics of Cirs and Associated Molecular Pathways. [Internet]. Cork: BookBaby; 2018 [cited 2019 Apr 1]. Available from: http://public.eblib.com/choice/publicfullrecord.aspx?p=5335794

160. Ley R, Balmanno K, Hadfield K, Weston C, Cook SJ. Activation of the ERK1/2 signaling pathway promotes phosphorylation and proteasome-dependent degradation of the BH3-only protein, Bim. J Biol Chem. 2003;278:18811–6.

161. Kandoth C, McLellan MD, Vandin F, Ye K, Niu B, Lu C, et al. Mutational landscape and significance across 12 major cancer types. Nature. 2013;502:333–9.

162. Leblanc V, May P. Activation et modifications post-traductionnelles de p53 après dommage de l'ADN. médecine/sciences. 2002;18:577–84.

163. Bode AM, Dong Z. Post-translational modification of p53 in tumorigenesis. Nat Rev Cancer. 2004;4:793–805.

164. Subramanian M, Shaha C. Up-regulation of Bcl-2 through ERK phosphorylation is associated with human macrophage survival in an estrogen microenvironment. J Immunol. 2007;179:2330–8.

165. Penela P, Murga C, Ribas C, Salcedo A, Jurado-Pueyo M, Rivas V, et al. G protein-coupled receptor kinase 2 (GRK2) in migration and inflammation. Arch Physiol Biochem. 2008;114:195–200.

166. O'Reilly LA, Kruse EA, Puthalakath H, Kelly PN, Kaufmann T, Huang DCS, et al. MEK/ERK-mediated phosphorylation of Bim is required to ensure survival of T and B lymphocytes during mitogenic stimulation. J Immunol. 2009;183:261–9.

167. Roux PP. RPS6KA1 (ribosomal protein S6 kinase, 90kDa, polypeptide 1). Atlas of Genetics and Cytogenetics in Oncology and Haematology [Internet]. 2011 [cited 2019 Apr 1]; Available from: http://hdl.handle.net/2042/44385

168. Fan Q, Chen M, Zuo L, Shang X, Huang MZ, Ciccarelli M, et al. Myocardial Ablation of G Protein-Coupled Receptor Kinase 2 (GRK2) Decreases Ischemia/Reperfusion Injury through an Anti-Intrinsic Apoptotic Pathway. PLoS ONE. 2013;8:e66234.

169. Lizcano JM, Göransson O, Toth R, Deak M, Morrice NA, Boudeau J, et al. LKB1 is a master kinase that activates 13 kinases of the AMPK subfamily, including MARK/PAR-1. EMBO J. 2004;23:833–43.

170. Sarras H, Alizadeh Azami S, McPherson JP. In Search of a Function for BCLAF1. The Scientific World JOURNAL. 2010;10:1450–61.

171. Zhou W, Zhang J, Marcus AI. LKB1 Tumor Suppressor: Therapeutic Opportunities Knock when LKB1 Is Inactivated. Genes Dis. 2014;1:64–74.

172. Moody SE, Schinzel AC, Singh S, Izzo F, Strickland MR, Luo L, et al. PRKACA mediates resistance to HER2-targeted therapy in breast cancer cells and restores anti-apoptotic signaling. Oncogene. 2015;34:2061–71.

173. Sood R, Kamikubo Y, Liu P. Role of RUNX1 in hematological malignancies. Blood. 2017;129:2070–82.

174. Ferrao PT. Overexpression of P-glycoprotein in K562 cells does not confer resistance to the growth inhibitory effects of imatinib (STI571) in vitro. Blood. 2003;102:4499–503.

175. Dai Y, Rahmani M, Corey SJ, Dent P, Grant S. A Bcr/Abl-independent, Lyn-dependent Form of Imatinib Mesylate (STI-571) Resistance Is Associated with Altered Expression of Bcl-2. Journal of Biological Chemistry. 2004;279:34227–39.

176. Frenzel A, Grespi F, Chmelewskij W, Villunger A. Bcl2 family proteins in carcinogenesis and the treatment of cancer. Apoptosis. 2009;14:584–96.

177. Thomas S, Quinn BA, Das SK, Dash R, Emdad L, Dasgupta S, et al. Targeting the Bcl-2 family for cancer therapy. Expert Opinion on Therapeutic Targets. 2013;17:61–75.

178. Greenberg EF, Lavik AR, Distelhorst CW. Bcl-2 regulation of the inositol 1,4,5-trisphosphate receptor and calcium signaling in normal and malignant lymphocytes: potential new target for cancer treatment. Biochim Biophys Acta. 2014;1843:2205–10.

179. Tsujimoto Y, Finger LR, Yunis J, Nowell PC, Croce CM. Cloning of the chromosome breakpoint of neoplastic B cells with the t(14;18) chromosome translocation. Science. 1984;226:1097–9.

180. Bakhshi A, Jensen JP, Goldman P, Wright JJ, McBride OW, Epstein AL, et al. Cloning the chromosomal breakpoint of t(14;18) human lymphomas: clustering around JH on chromosome 14 and near a transcriptional unit on 18. Cell. 1985;41:899–906.

181. Kuribara R, Honda H, Matsui H, Shinjyo T, Inukai T, Sugita K, et al. Roles of Bim in Apoptosis of Normal and Bcr-Abl-Expressing Hematopoietic Progenitors. Molecular and Cellular Biology. 2004;24:6172–83.

182. Ng KP, Hillmer AM, Chuah CTH, Juan WC, Ko TK, Teo ASM, et al. A common BIM deletion polymorphism mediates intrinsic resistance and inferior responses to tyrosine kinase inhibitors in cancer. Nature Medicine. 2012;18:521–8.

183. Shinohara Y, Takahashi N, Nishiwaki K, Hino M, Kashimura M, Wakita H, et al. A multicenter clinical study evaluating the confirmed complete molecular response rate in imatinib-treated patients with chronic phase chronic myeloid leukemia by using the international scale of real-time quantitative polymerase chain reaction. Haematologica. 2013;98:1407–13.

184. Chen X, Liu H, Xing H, Sun H, Zhu P. The BIM deletion polymorphism cannot account for intrinsic TKI resistance of Chinese individuals with chronic myeloid leukemia. Nature Medicine. 2014;20:1090–1090.

185. Xu J, Gu J, Zhao Y, Meng H, Du L, Zhang R, et al. The BCL2L11 deletion polymorphism is not associated with imatinib resistance in chronic myeloid leukemia patients: meta-analysis. Oncotarget. 2017;8:99041–8.

186. Augis V, Airiau K, Josselin M, Turcq B, Mahon F-X, Belloc F. A single nucleotide polymorphism in cBIM is associated with a slower achievement of major molecular response in chronic myeloid leukaemia treated with imatinib. PLoS ONE. 2013;8:e78582.

187. Gagné V, Rousseau J, Labuda M, Sharif-Askari B, Brukner I, Laverdière C, et al. Bim polymorphisms: influence on function and response to treatment in children with acute lymphoblastic leukemia. Clin Cancer Res. 2013;19:5240–9.

188. Maxam AM, Gilbert W. A new method for sequencing DNA. Proceedings of the National Academy of Sciences. 1977;74:560–4.

189. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences. 1977;74:5463–7.

190. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, et al. DNA sequencing at 40: past, present and future. Nature. 2017;550:345–53.

191. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes JC, et al. Nucleotide sequence of bacteriophage φX174 DNA. Nature. 1977;265:687–95.

192. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature. 2005;437:376–80.

193. Valouev A, Ichikawa J, Tonthat T, Stuart J, Ranade S, Peckham H, et al. A high-resolution, nucleosome position map of C. elegans reveals a lack of universal sequence-dictated positioning. Genome Res. 2008;18:1051–63.

194. Fox EJ, Reid-Bayliss KS, Emond MJ, Loeb LA. Accuracy of Next Generation Sequencing Platforms. Next Gener Seq Appl. 2014;1.

195. Glenn TC. Field guide to next-generation DNA sequencers: FIELD GUIDE TO NEXT-GEN SEQUENCERS. Molecular Ecology Resources. 2011;11:759–69.

196. McCarthy A. Third Generation DNA Sequencing: Pacific Biosciences' Single Molecule Real Time Technology. Chemistry & Biology. 2010;17:675–6.

197. Roberts RJ, Carneiro MO, Schatz MC. The advantages of SMRT sequencing. Genome Biology [Internet]. 2013 [cited 2019 Mar 13];14. Available from: http://genomebiology.biomedcentral.com/articles/10.1186/gb-2013-14-7-405

198. Metzker ML. Sequencing technologies — the next generation. Nature Reviews Genetics. 2010;11:31–46.

199. Loman NJ, Quick J, Simpson JT. A complete bacterial genome assembled de novo using only nanopore sequencing data. Nat Methods. 2015;12:733–5.

200. Perkel J. TechBlog: The nanopore toolbox [Internet]. NatureJobs blog. 2017 [cited 2019 Mar 13]. Available from: http://blogs.nature.com/naturejobs/2017/10/16/techblog-the-nanopore-toolbox/

201. International Human Genome Sequencing Consortium. Initial sequencing and analysis of the human genome. Nature. 2001;409:860–921.

202. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. Nature. 2004;431:931–45.

203. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The Sequence of the Human Genome. Science. 2001;291:1304–51.

204. Levy S, Sutton G, Ng PC, Feuk L, Halpern AL, Walenz BP, et al. The diploid genome sequence of an individual human. PLoS Biol. 2007;5:e254.

205. Schloss JA. How to get genomes at one ten-thousandth the cost. Nature Biotechnology. 2008;26:1113–5.

206. 1000 Genomes Project Consortium, Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.

207. Sudmant PH, Rausch T, Gardner EJ, Handsaker RE, Abyzov A, Huddleston J, et al. An integrated map of structural variation in 2,504 human genomes. Nature. 2015;526:75–81.

208. Exome Aggregation Consortium, Lek M, Karczewski KJ, Minikel EV, Samocha KE, Banks E, et al. Analysis of protein-coding genetic variation in 60,706 humans. Nature. 2016;536:285–91.

209. Koko M, Abdallah MOE, Amin M, Ibrahim M. Challenges imposed by minor reference alleles on the identification and reporting of clinical variants from exome data. BMC Genomics. 2018;19:46.

210. Lavrov AV, Chelysheva EY, Smirnikhina SA, Shukhov OA, Turkina AG, Adilgereeva EP, et al. Frequent variations in cancer-related genes may play prognostic role in treatment of patients with chronic myeloid leukemia. BMC Genetics [Internet]. 2016 [cited 2019 Mar 17];17. Available from: http://bmcgenet.biomedcentral.com/articles/10.1186/s12863-015-0308-7

211. Lavrov AV, Chelysheva EYu, Adilgereeva EP, Shukhov OA, Smirnikhina SA, Kochergin-Nikitsky KS, et al. Exome, transcriptome and miRNA analysis don't reveal any molecular markers of TKI efficacy in primary CML patients. BMC Medical Genomics [Internet]. 2019 [cited 2019 Mar 17];12. Available from: https://bmcmedgenomics.biomedcentral.com/articles/10.1186/s12920-019-0481-z

212. Kim T, Tyndel MS, Zhang Z, Ahn J, Choi S, Szardenings M, et al. Exome sequencing reveals DNMT3A and ASXL1 variants associate with progression of chronic myeloid leukemia after tyrosine kinase inhibitor therapy. Leukemia Research. 2017;59:142–8.

213. Shokeen Y, Sharma NR, Vats A, Taneja V, Minhas S, Jauhri M, et al. Identification of prognostic and susceptibility markers in chronic myeloid leukemia using next generation sequencing. Ethiopian Journal of Health Sciences. 2018;28:135.

214. Zheng Q, Cao J, Hamad N, Kim H-J, Moon JH, Sohn SK, et al. Single nucleotide polymorphisms in apoptosis pathway are associated with response to imatinib therapy in chronic myeloid leukemia. Journal of Translational Medicine [Internet]. 2016 [cited 2019 Mar 17];14. Available from: http://translational-medicine.biomedcentral.com/articles/10.1186/s12967-016-0837-5

215. Yates B, Braschi B, Gray KA, Seal RL, Tweedie S, Bruford EA. Genenames.org: the HGNC and VGNC resources in 2017. Nucleic Acids Res. 2017;45:D619–25.

216. Afgan E, Baker D, Batut B, van den Beek M, Bouvier D, Čech M, et al. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2018 update. Nucleic Acids Research. 2018;46:W537–44.

217. Vuong H, Che A, Ravichandran S, Luke BT, Collins JR, Mudunuri US. AVIA v2.0: annotation, visualization and impact analysis of genomic variants and genes: Fig. 1. Bioinformatics. 2015;31:2748–50.

218. Patel RK, Jain M. NGS QC Toolkit: a toolkit for quality control of next generation sequencing data. PLoS ONE. 2012;7:e30619.

219. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.

220. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nat Biotechnol. 2011;29:24–6.

221. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 2010;20:1297–303.

222. Benzécri J-P. Correspondence analysis handbook. New York: Marcel Dekker; 1992.

223. Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett. 2004;573:83–92.

224. Koziol JA. The rank product method with two samples. FEBS Letters. 2010;584:4481–4.

225. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science. 2012;337:816–21.

226. Labun K, Montague TG, Gagnon JA, Thyme SB, Valen E. CHOPCHOP v2: a web tool for the next generation of CRISPR genome engineering. Nucleic Acids Research. 2016;44:W272–6.

227. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nature Protocols. 2013;8:2281–308.

228. Sourial N, Wolfson C, Zhu B, Quail J, Fletcher J, Karunananthan S, et al. Correspondence analysis is a useful tool to uncover the relationships among categorical variables. J Clin Epidemiol. 2010;63:638–46.

229. Sionov RV, Vlahopoulos SA, Granot Z. Regulation of Bim in Health and Disease. Oncotarget. 2015;6:23058–134.

230. Pene-Dumitrescu T, Smithgall TE. Expression of a Src family kinase in chronic myelogenous leukemia cells induces resistance to imatinib in a kinase-dependent manner. J Biol Chem. 2010;285:21446–57.

231. Kasof GM, Goyal L, White E. Btf, a novel death-promoting transcriptional repressor that interacts with Bcl-2-related proteins. Mol Cell Biol. 1999;19:4390–404.

232. Bagger FO, Kinalis S, Rapin N. BloodSpot: a database of healthy and malignant haematopoiesis updated with purified and single cell mRNA sequencing profiles. Nucleic Acids Research. 2019;47:D881–5.

233. Mohamed AN. del(6q). Atlas of Genetics and Cytogenetics in Oncology and Haematology [Internet]. 2018 [cited 2019 Mar 27]; Available from: http://hdl.handle.net/2042/69001

234. Naumann S, Reutzel D, Speicher M, Decker HJ. Complete karyotype characterization of the K562 cell line by combined application of G-banding, multiplex-fluorescence in situ

hybridization, fluorescence in situ hybridization, and comparative genomic hybridization. Leuk Res. 2001;25:313–22.

235. Law JC, Ritke MK, Yalowich JC, Leder GH, Ferrell RE. Mutational inactivation of the p53 gene in the human erythroid leukemic K562 cell line. Leuk Res. 1993;17:1045–50.

236. Kale J, Osterlund EJ, Andrews DW. BCL-2 family proteins: changing partners in the dance towards death. Cell Death Differ. 2018;25:65–80.

237. Ross S, Anand SS, Joseph P, Paré G. Promises and challenges of pharmacogenetics: an overview of study design, methodological and statistical issues. JRSM Cardiovascular Disease. 2012;1:1–13.

238. Maia R, Vasconcelos F, Souza P, Rumjanek V. Towards Comprehension of the ABCB1/P-Glycoprotein Role in Chronic Myeloid Leukemia. Molecules. 2018;23:119.

239. Dulucq S, Bouchet S, Turcq B, Lippert E, Etienne G, Reiffers J, et al. Multidrug resistance gene (MDR1) polymorphisms are associated with major molecular responses to standard-dose imatinib in chronic myeloid leukemia. Blood. 2008;112:2024–7.

240. Au A, Aziz Baba A, Goh AS, Wahid Fadilah SA, Teh A, Rosline H, et al. Association of genotypes and haplotypes of multi-drug transporter genes ABCB1 and ABCG2 with clinical response to imatinib mesylate in chronic myeloid leukemia patients. Biomedicine & Pharmacotherapy. 2014;68:343–9.

241. Ni L-N, Li J-Y, Miao K-R, Qiao C, Zhang S-J, Qiu H-R, et al. Multidrug resistance gene (MDR1) polymorphisms correlate with imatinib response in chronic myeloid leukemia. Medical Oncology. 2011;28:265–9.

242. Deenik W, van der Holt B, Janssen JJWM, Chu IWT, Valk PJM, Ossenkoppele GJ, et al. Polymorphisms in the multidrug resistance gene MDR1 (ABCB1) predict for molecular resistance in patients with newly diagnosed chronic myeloid leukemia receiving high-dose imatinib. Blood. 2010;116:6144–5; author reply 6145-6146.

243. Komor AC, Kim YB, Packer MS, Zuris JA, Liu DR. Programmable editing of a target base in genomic DNA without double-stranded DNA cleavage. Nature. 2016;533:420–4.

244. Marx V. Base editing a CRISPR way. Nature Methods. 2018;15:767–70.

245. Eid A, Alshareef S, Mahfouz MM. CRISPR base editors: genome editing without double-stranded breaks. Biochem J. 2018;475:1955–64.

246. Kataoka T, Holler N, Micheau O, Martinon F, Tinel A, Hofmann K, et al. Bcl-rambo, a Novel Bcl-2 Homologue That Induces Apoptosis via Its Unique C-terminal Extension. Journal of Biological Chemistry. 2001;276:19548–54.

247. Kim J-Y, So K-J, Lee S, Park J-H. Bcl-rambo induces apoptosis via interaction with the adenine nucleotide translocator. FEBS Letters. 2012;586:3142–9.

248. Holleman A. The expression of 70 apoptosis genes in relation to lineage, genetic subtype, cellular drug resistance, and outcome in childhood acute lymphoblastic leukemia. Blood. 2006;107:769–76.

249. Yang Y-L, Lin S-R, Chen J-S, Lin S-W, Yu S-L, Chen H-Y, et al. Expression and prognostic significance of the apoptotic genes BCL2L13, Livin, and CASP8AP2 in childhood acute lymphoblastic leukemia. Leukemia Research. 2010;34:18–23.

250. Jensen SA, Calvert AE, Volpert G, Kouri FM, Hurley LA, Luciano JP, et al. Bcl2L13 is a ceramide synthase inhibitor in glioblastoma. Proceedings of the National Academy of Sciences. 2014;111:5682–7.

251. Tahir SK, Wass J, Joseph MK, Devanarayan V, Hessler P, Zhang H, et al. Identification of Expression Signatures Predictive of Sensitivity to the Bcl-2 Family Member Inhibitor ABT-263 in Small Cell Lung Carcinoma and Leukemia/Lymphoma Cell Lines. Molecular Cancer Therapeutics. 2010;9:545–57.

Annexes

Partie I : Annexe 1

Article soumis à Human Genomics

Novel analytical methods to interpret large sequencing data from small sample sizes

Florence Lichou[1], Sébastien Orazio[2], Stéphanie Dulucq[1], Gabriel Etienne[1], Michel Longy[1], Christophe Hubert[3], Alexis Groppi[4], Alain Monnereau[2], François-Xavier Mahon[1], Béatrice Turcq[1]

[1]Laboratory of Mammary and Leukaemic Oncogenesis, Inserm U1218 ACTION, University of Bordeaux, Bergonié Cancer Institute, Bordeaux, France.

[2]Team EPICENE, Inserm U1219 BPH, University of Bordeaux, Bergonié Cancer Institute, Bordeaux, France.

[3]Inserm U1211 MRGM, University of Bordeaux, Bordeaux, France

[4]The Bordeaux Bioinformatics Center (CBiB), University of Bordeaux, Bordeaux, France

Corresponding author: Béatrice Turcq. Laboratory of Mammary and Leukaemic Oncogenesis, Inserm U1218 ACTION, University of Bordeaux. 146 rue Léo Saignat, bâtiment TP 4ème étage, case 50, 33076 Bordeaux, France. Tel: (+33) 5 57 57 10 51. E-mail address: beatrice.turcq@u-bordeaux.fr

Abstract

Background: Targeted therapies have greatly improved cancer patients prognosis. For instance, chronic myeloid leukaemia is now well treated with imatinib, a tyrosine kinase inhibitor. Around 80% of the patients reach complete remission. However, despite its great efficiency, some patients are resistant to the drug. This heterogeneity in the response might be associated with pharmacokinetic parameters, varying between individuals because of genetic variants. To assess this issue, next-generation sequencing of large panels of genes can be performed from patient samples. However, the common problem in pharmacogenetic studies is the availability of samples, often limited. In the end, large sequencing data are obtained from small sample sizes, therefore, classical statistical analyses can't be applied to identify interesting targets. To overcome this concern, here we described original and underused statistical methods to analyze large sequencing data from a restricted number of samples.

Results: To evaluate the relevance of our method, 48 genes involved in pharmacokinetics were sequenced by next-generation sequencing from 24 chronic myeloid leukaemia patients, either sensitive or resistant to imatinib treatment. Using a graphical representation, from 708 identified polymorphisms a reduced list of 108 candidates was obtained. Then, by analyzing each gene and the distribution of variant alleles, several candidates were highlighted such as *UGT1A9*, *PTPN22,* and *SLC22A1*. These genes were already associated with the transport, the metabolism and even the sensitivity to imatinib.

Conclusions: These relevant tests are great alternatives to inferential statistics not applicable to next-generation sequencing experiments performed on small sample sizes. These approaches permit to reduce the number of targets and find good candidates for further treatment sensitivity studies.

Keywords: chronic myeloid leukaemia, next-generation sequencing, pharmacogenetics, small sample size, statistics, factorial correspondence analysis, hierarchical clustering on principal components, rank products

Background

Pharmacokinetics refers to drug transport, absorption or metabolism affecting treatment efficacy. Inter-individual variability in drug response has been described and may be associated with genetic variants (pharmacogenetics) [1]. Identification of these variants, differentially enriched in patients, could help to predict their response to the treatment. However, only a few published studies have reported genetic predictors of efficacy that met the criteria with statistical significance [2,3]. Targeted approaches were often performed, using single nucleotide polymorphism (SNP) arrays or Sanger sequencing of selected genes. In this way, only common and known variants were investigated. The emergence of next-generation sequencing (NGS) since the 2000s opened up new perspectives and allows to identify rare and non-described variants [4]. Even though, generating highly confident results is still a concern [5,6]. Small sample sizes are a recurrent issue in pharmacogenetic studies mainly due to a lack of sample availability and high sequencing costs. As the number of sequenced patients is low, conventional statistical approaches are not applicable to highlight the best polymorphisms, differentially expressed or mutated between a control and a treated group. In this study, we propose novel analytical approaches to identify and filter the best variants, more likely to be associated with drug sensitivity, from a massive amount of NGS data using a limited sample size. To assess and to test these approaches, a targeted sequencing was performed on chronic myeloid leukaemia (CML) patients. CML is a clonal myeloproliferative disorder characterized by the aberrant Philadelphia chromosome, arising from a reciprocal translocation t(9;22) (q34;q11) [7]. This event results in the *BCR-ABL1* (*breakpoint cluster region-abelson 1*) fusion gene encoding a constitutively active tyrosine kinase that upregulates many kinase pathways. These alterations lead to leukemogenesis: a cell proliferation increase, apoptosis inhibition and persistence of hematopoietic stem cells. This BCR-ABL1 chimeric protein is specifically targeted and inhibited by tyrosine kinase inhibitors (TKIs), such as imatinib mesylate (IM, Gleevec®), commonly used as first-line therapy for CML patients. This treatment shows impressive results with a 10-year event-free survival of 83% in 2013 [8]. Despite these convincing results, resistance to treatment is a persistent clinical problem. Mutations in *ABL1* kinase domain or BCR-ABL1 overexpression are known mechanisms of resistance for about 50% of patients. For other cases, the resistance is still unexplained and surely involves more complex and heterogeneous mechanisms

[9]. Notably, alterations of proteins implicated in pharmacokinetics could be participating [10]. To identify genetic variants associated with IM resistance, samples at diagnosis from CML patients, either sensitive or resistant to the IM treatment, were sequenced. A small sample size was available: 12 sensitive patients and 12 resistant patients. 48 genes, selected from previous pharmacogenetic studies, were analyzed by a custom approach using NGS. In this way, all polymorphisms in splicing sites, promoting and coding regions, already described in public databases or new ones, have been identified. They were then filtered and classified according to the variant allele frequency. Novel approaches using descriptive statistics, simulation studies and non-parametric statistics were performed to investigate the results generated from this NGS study using a small cohort of patients.

Results

Selection of 48 genes involved in pharmacokinetics

48 genes were selected and sequenced by NGS (Table 1). They encode proteins involved in several pathways potentially linked to IM resistance by directly regulating TKIs or different cell processes in the leukemic cells. They were classified into 6 groups. The first group includes 10 genes encoding plasma proteins, membrane transporters and regulators, involved in the transport and the diffusion of IM through the cell membrane. Genes involved in this process were largely studied in the field of IM resistance. In particular, three exonic polymorphisms (rs1045642 3435C>T, rs1128503 1236C>T, rs2032582 2677G>T/A) in *ATP binding cassette subfamily B member 1* (*ABCB1* also known as *multidrug resistance protein 1, MDR1*), encoding a major IM membrane efflux transporter, have already been identified and associated with lower IM efficiency in several studies, although there are some conflicting results [11,12]. *ATP binding cassette subfamily G member 2* (*ABCG2*) and *solute carrier family 22 member 1* (*SLC22A1* also known as *human organic cation transporter type 1*, *hOCT1 or OCT1*), both encoding major IM membrane transporters, are also widely studied and several non-synonymous polymorphisms in these two genes were associated with a lower IM response [12]. The second group includes 12 genes encoding metabolic enzymes and regulators. IM is mainly metabolized by the cytochrome P450 (CYP) isoenzymes. Polymorphisms in *CYP3A4* and *CYP3A5*, both encoding dominant enzymes involved in IM metabolism, have been described in

several studies as correlated with IM resistance in CML patients [12]. The third and fourth groups include genes encoding proteins involved in cell cycle and proliferation regulation (n=5) and proteins involved in DNA repair in response to damages (n=10). Alterations of these cell processes might impair the IM efficiency by enhancing the capacity of cells to proliferate. The next groups include factors involved in cytokine pathways (n=6) and kinases and phosphatases regulating BCR-ABL1 (n=5): both have been suggested to be involved in IM sensitivity.

NGS quality control and genetic polymorphism characteristics

After sequencing of the 48 genes from 24 CML patient samples, around 9 million reads were obtained. 95% of the reads passed the quality filter (Phred score over or equal to 30). A high read depth was obtained with more than 90% of targeted regions covered with more than 35 short sequences (35X). SNPs and small insertions and deletions (INDELs) were detected and filtered according to the sequencing quality and the depth (30X threshold). After quality filtering, 708 polymorphisms were identified: 41 deletions (the largest of 9bp), 27 insertions (the largest of 6bp) and 640 SNPs (Tables 2 and 3, Table S1. Description of ANNOVAR annotations).

Table 2. ANNOVAR annotations of all sequenced polymorphisms

ANNOVAR annotation	Deletion	Insertion	SNP	TOTAL	% of total
Upstream to the promoter	0	0	6	6	0.8
Downstream to the promoter	0	0	3	3	0.4
UTR5	1	2	53	56	7.9
UTR3	1	2	29	32	4.5
Exonic	5	1	164	170	24.0
Exonic. splicing	0	0	5	5	0.7
Splicing	0	0	1	1	0.2
Intergenic	0	0	3	3	0.4
Intronic	34	22	375	431	60.9
ncRNA_exonic	0	0	1	1	0.2
TOTAL	41	27	640	708	100.0
% of total	5.8	3.8	90.4	100.0	

Table 3. ANNOVAR annotations of exonic sequenced polymorphisms

Polymorphisms	Exonic and exonic splicing	% of all exonic polymorphisms	% of all polymorphisms
Deletion	5	2.9	0.7
Frameshift	2	1.1	0.3
Nonframeshift	2	1.1	0.3
Stopgain	1	0.6	0.1
Insertion	1	0.6	0.1
Frameshift	1	0.6	0.1
SNP	169	96.6	23.9
Nonsynonymous	104	59.4	14.7
Synonymous	64	36.6	9.0
Stoploss	1	0.6	0.1
TOTAL	175	100.0	24.7

Generating a third group representing non-CML individuals: the general population

The observed genotypes of the 708 polymorphisms identified in 24 CML patients were reported in a matrix. The frequency of patients carrying the variant allele (heterozygous and variant homozygous) for each polymorphism was calculated for both the sensitive and the resistant groups. To highlight the polymorphisms more likely to be involved in IM resistance, these results were compared to the frequencies of individuals carrying the variant allele in the general population (non-CML individuals). This information was estimated using the minor allele frequency (MAF), the frequency of the variant allele, less frequent in the general population than the reference allele. MAF are reported for over 88 million of variants thanks to one of the large sequencing initiatives called 1000G (phase 3 released in 2014, http://www.internationalgenome.org/) [13] (Table S2. Initial co-occurrence matrix). Among the 708 identified polymorphisms, 118 did not have a MAF reported on the 1000G database (Table 4).

Table 4. Repartition of the sequenced polymorphisms in 1000G database

	Total polymorphisms	Polymorphisms with MAF in 1000G	% with MAF
Upstream to the promoter	6	3	50.0
Downstream to the promoter	3	2	66.7
UTR5	56	52	92.9
UTR3	32	28	87.5
Exonic	170	144	84.7
Exonic splicing	5	3	60.0
Splicing	1	1	100.0
Intergenic	3	3	100.0
Intronic	431	356	82.6
ncRNA_exonic	1	1	100.0
TOTAL	708	593	83.7

For the polymorphisms without any MAF (16.3% of all polymorphisms), an arbitrary MAF was defined. The alternative allele was estimated as a rare variant, present in less than 1 individual in the whole sequenced cohort. As 2504 individuals were included in 1000G, the theoretical MAF was 1/2504, about 4.10^{-4} [13]. The same analysis could also be performed without setting a theoretical MAF by removing from the dataset the polymorphisms with no MAF in the 1000G database (data not shown). The other element to note is that 130 MAFs (21.9% of all reported MAFs) were superior to 0.5. By definition, the MAF must be the frequency of the less common allele and so, inferior to 0.5. These errors are due to the reference genome, initially obtained from a whole genome sequencing and so, containing both major and minor alleles for different polymorphisms. For these 130 polymorphisms, the reference and the variant alleles were inverted and hence the genotypes. After this adjustment 24 polymorphisms showed no variant allele in the 24 CML patients and they were removed from the analysis. All these steps are resumed in Figure 1.

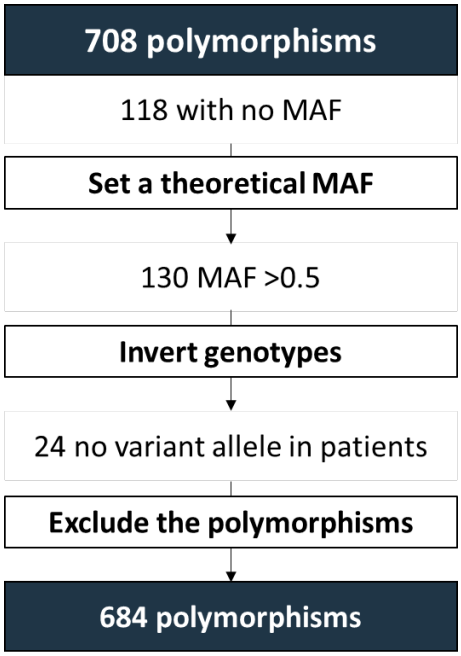


Fig. 1. Workflow to adjust the co-occurrence matrix according to the MAF

First, for polymorphisms with no MAF a theoretical MAF is added. Second, genotypes for polymorphisms with a MAF >0.5 are inverted. Third, polymorphisms with no variant allele identified by NGS are excluded.

Finally, genotypic frequencies of the 684 remaining polymorphisms were estimated in the general population applying the Hardy-Weinberg principle (Fig. 2). At the end, a matrix containing the frequencies of individuals carrying the variant allele for 684 polymorphisms in three different groups (sensitive CML patients, resistant CML patients and the general population) was obtained.

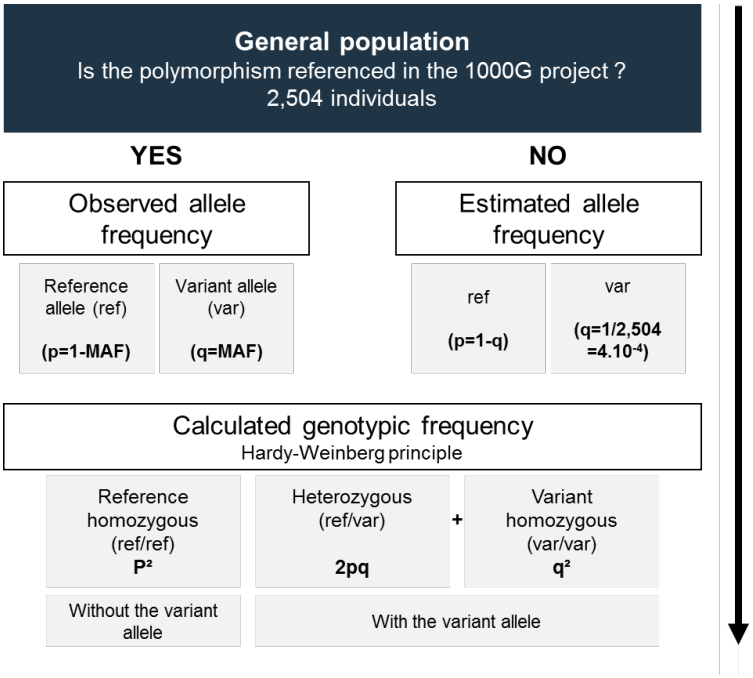


Fig. 2. Workflow to generate a simulated distribution of the 684 identified polymorphisms in a group representing the general population. Calculations are performed for each polymorphism. First, genotypic frequencies are calculated. If the polymorphism is referenced in the 1000 genome database (1000G) the observed minor allele frequency (MAF) is used, if not the estimated MAF is chosen (4.10^{-4}). Second, the frequency of individuals carrying or not the variant allele is calculated.

Observing individually variant allele distribution using factorial correspondence analysis and hierarchical clustering on principal components

Because of the limited number of sequenced patients, the conventional statistical analysis could not be applied. Indeed, in our experiment, common testing assumptions are violated. To overcome this issue, descriptive statistics were preferentially used and factorial correspondence analysis (FCA) was performed. This approach permitted to display the 684 identified polymorphisms on a two-dimensional graph according to the number of individuals carrying the variant allele in each group (sensitive CML patients, resistant CML patients, general population). After hierarchical clustering on principal components (HCPC), 5 distinct groups were identified (Fig. 3, Table S3. Contingency table and list of the clusters).

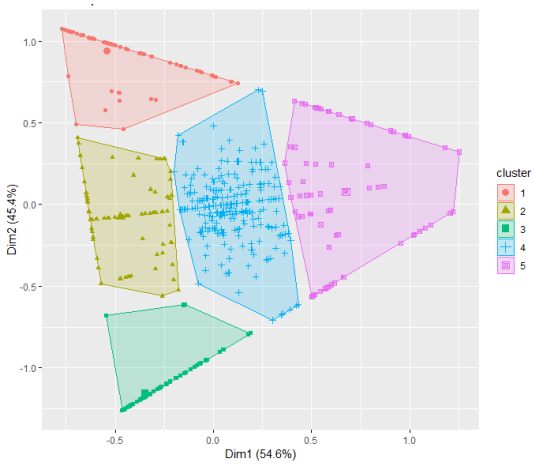


Fig. 3. Factorial correspondence analysis (FCA) map of the distribution of the variant allele for each identified polymorphism (684) among the three groups: CML sensitive patients, CML resistant patients and the general population

Five clusters were obtained after hierarchical clustering on principal components (HCPC). 1: highest variant frequency in CML sensitive patients, 2: highest variant frequency in both resistant and sensitive CML patients, 3: highest variant frequency in CML resistant patients, 4: no difference between populations, 5: highest variant frequency in the simulated general population group.

The cluster 3 showed polymorphisms with variant alleles more frequent in CML resistant patients than CML sensitive patients or individuals in the general population (simulated group). The cluster 1 contained variants enriched in CML sensitive patients. The cluster 2 contained variants enriched in CML patients, either sensitive or resistant to the treatment. In contrast, cluster 5 showed variants more frequent in the general population than in our sequenced CML patients. In cluster 4, there was no visible difference in the distribution between the three sub-populations. 108 polymorphisms were enriched in resistant patients (cluster 3). These polymorphisms were more likely to be associated with IM resistance because the variant allele was found to be more frequent in these patients than sensitive patients or non-CML patients. For example, a non-synonymous variant (rs2476601, chr1:114377568, 1858G>A, R620W) in the gene encoding the protein tyrosine phosphatase, non-receptor type 22 (PTPN22) was highlighted. According to the MAF, the variant allele was found in less than 3% in the general population. Around 42% of resistant patients (5 out 12) were carrying this variant allele, whereas all sensitive patients were carrying the wild-type allele. This polymorphism was also highlighted in 2011 by Guillem *et al.* and linked to a bad prognosis for CML patients [14]. CML patients with the G/A genotype were seven times more likely to experience primary failure to IM treatment than the G/G (reference allele)

carriers. The FCA and the HCPC displayed globally the variant distribution between the three groups and permitted the discrimination of the different polymorphisms according to the variant allele distribution in CML patients and the general population. Using this approach, from 684 polymorphisms, a reduced list of 108 polymorphisms more likely to be associated with IM resistance was identified.

Classifying the genes using the rank product method

Secondly, instead of analyzing individual polymorphisms, the variants were grouped together according to the gene they belong to. This approach permitted to identify the genes more likely to be associated with IM resistance because they were carrying multiple genetic variants. The rank products (RP) method was applied. The genes were ranked according to the genotype relative frequencies of their variants. Two approaches were tested. Either the sum of frequencies divided by the gene size was performed or the mean of variant frequencies. In both approaches, one frequency was obtained for each gene. The results were similar for both calculations (Table S4. Rank product results): the use of variant allele frequency means seems not to generate a bias in the analysis. It was chosen for the following steps. Several conditions were tested: all polymorphisms, polymorphisms with MAF in 1000G or exonic variants causing protein alterations. Results were comparable (Table S4. Rank product results), however, the last condition was the most biologically relevant. 105 polymorphisms in 30 genes were included in the RP test. The three top genes (pfp<0.05) are reported in Table 5.

Table 5. Top genes identified by the rank product method

Gene	RP/Rsum[a]	FC:(class1/class2)[b]	pfp[c]	P-value
UGT1A9	1.000	0.001031	4.555e-05	1.627e-06
PTPN22	2.000	0.045410	4.550e-03	3.250e-04
SLC22A1	3.000	0.423800	3.452e-02	3.699e-03

[a]RP/Rsum (rank products statistics): the probability that the gene would be classified first in all samples (from both conditions). The lower it is, more the difference between control and treated conditions is important

[b]FC:(class1/class2): computed fold change of the average “expression levels” under two conditions

[c]pfp: percentage of false prediction

The first ranked gene, most significantly variant in CML resistant patients (p-value <0.05), was u*ridine diphosphate-glucuronosyltransferase family 1 member A9*

(*UGT1A9*). The UGT1 enzymes catalyze the transformation of small lipophilic molecules into active metabolites [15]. Interestingly, IM is not a UGTs substrate but it can inhibit their activity. Indeed, it has been shown that, when IM (same or weaker effect with other TKIs: sorafenib, dasatinib, nilotinib …) was co-administrated with acetaminophen (paracetamol), it decreased its efficacy by inhibiting UGT activity (UGT1A9 and UGT2B15) and so paracetamol glucuronidation [16]. Furthermore, in IM resistant patients treated by the second-generation TKI, nilotinib, some *UGT1A9* variants affected nilotinib efficacy and were associated with adverse events (hyperbilirubinemia) [17]. This enzyme seems to interact with TKIs and variants in *UGT1A9* might affect their efficacy and provoke resistance to the treatment. The second gene was, once again, *PTPN22*. SLC22A1, the third factor, is an organic cation transporter [18]. It has been reported as the main IM influx transporter in CML cells. Several studies associated *SLC22A1* expression or genetic variants to IM response heterogeneity with conflicting results [12,19]. Using this method, the genes were classified according to their enrichment with nonsynonymous variants. A reduced list of three factors more likely to be associated with IM resistance was obtained.

Discussion

With the NGS emergence, recent pharmacogenetic studies generate massive amounts of data that are difficult to interpret. Despite this, the small sample size is still a recurrent issue. In this way, conventional statistical approaches are not applicable and it is necessary to develop novel analytical approaches to highlight interesting polymorphisms. Here we proposed several methods to get around inferential statistics limitations using descriptive statistics, a simulation method, and non-parametric statistics. These approaches give new paths to follow pharmacokinetic studies. FCA and HCPC are very useful but underused tools to visualize quickly the data and estimate their distribution. In our study, in a simple way, the candidate variant list was reduced from 684 to 108 polymorphisms likely to be associated with IM resistance in CML patients. As it is descriptive statistics, no assumptions are made and there is no statistical limitation in contrary to inferential statistics, today commonly used. The RP method, designed for microarray analysis, seems a powerful tool to classify the candidate genes and again, gives clues for further studies. In our study, it was associated with a Monte Carlo simulation to generate simulated experiments and give

more weight to the test. It is another way to bypass small sample size limitations. Using this technic the genes carrying exonic variants causing protein alteration were ranked according to their enrichment in the resistant CML patient group compared to sensitive CML patient group but also the general population. According to the needs, the list of genes to be ranked can be modulated. In our study, results were comparable while using all the variants or a restricting list of variants more likely to be associated with protein loss of function. Several intriguing genes were highlighted using this method. *UGT1A9* was never directly associated with IM resistance but correlated with other TKI resistance [17]. *PTPN22*, an interesting gene, already linked to IM resistance. A non-synonymous variant (rs2476601, 1858G>A, R620W), already associated with IM resistance, was found in resistant patients in our study [14]. It was included in the cluster enriched in resistant CML patients in the FCA analysis. SLC22A1 is the main IM influx transporter in CML cells and it has been associated several times to IM sensitivity although there are some contradictory results [19].

Conclusions

In this study, we proposed and tested underused and uncommon statistical strategies. From large amount of data generated by NGS approaches and few samples, we can highlight interesting targets for future studies. An informatics tool was developed to perform all these analyses in a simple manner and transpose this approach to other NGS experiments performed from small sample sizes.

Methods

Patient characteristics

24 Philadelphia chromosome positive CML patients newly diagnosed were included in the study either in optimal response (n=12) or in failure response (n=12) according to the European leukaemia net (ELN) criteria released in 2013 [20]. None of them carried *BCR-ABL1* alterations at diagnosis. All patients enrolled in this study provided informed consent according to the Declaration of Helsinki. They were daily treated with IM 400mg. Sex ratio, the median age at diagnosis and Sokal risk group are reported in Table 6.

Table 6. Patients' characteristics

Characteristics	All patients	Optimal response	Failure response
No. of patients (%)	24 (100)	12 (50)	12 (50)
Gender			
Male	15	8	7
Female	9	4	5
Median age at diagnosis (range)	59 (19-86)	61 (19-86)	57 (20-77)
Sokal score (Low/Intermediate/High)	8/7/9	6/2/4	2/5/5

Overall, there were 9 females (37.5%) and 15 males (62.5%). The median age of patients at diagnosis was 59 years (range 19 to 86 years old). There was no significant difference in ages between males and females or sensitive patients and resistant patients.

Genomic DNA extraction

Genomic DNA was extracted from leucocyte dry pellets obtained at diagnosis using "DNA extraction kit" (Agilent Technologies) according to the manufacturer's instructions. The quantity and extraction quality was assessed using a "Nanodrop 2000" spectrophotometer (Thermo scientific). The genomic DNA integrity (high molecular weight) was verified on a 0.5% agarose gel electrophoresis.

Targeted sequence capture and next-generation sequencing

The genotyping was performed by NGS. A custom DNA library was prepared using the "SureSelect^{QXT} Target Enrichment for Illumina Multiplexed Sequencing kit" (Agilent technologies). Probes of 120 nucleotides long (Table S5. Probe sequences) were designed using the software "SureDesign" (Agilent technologies) to specifically capture exons, intron-exon junctions and promoter regions of the 48 genes of interest (reported in Table 1). A paired-end sequencing (2x150bp) has been performed on a MiSeq device (Illumina) using two "MiSeq Reagent Micro kit v2" (Illumina) according to the manufacturer's instructions. Obtained sequences were mapped to the human reference assembly GRCh37/*hg19* using the Burrows-Wheeler Aligner (BWA) software (http://sourceforge.net/projects/bio-bwa/) [21]. SNPs and small INDELs were identified using the Genome Analysis Toolkit (GATK, http://www.broadinstitute.org/gatk/) [22]. Variant annotation was performed using the

Annotation, Visualization and Impact Analysis (AVIA) online resource (http://avia-abcc.ncifcrf.gov/) [23].

Contingency table with the observed genotype relative frequencies

To identify polymorphisms correlated with IM resistance, the number of individuals carrying the variant allele of each polymorphism was analyzed. To compare our results to the repartition in non-CML individuals, the theoretical genotypes frequencies in the general population for each polymorphism were estimated using the Hardy-Weinberg principle and the MAF. Before applying calculating the genotypic frequencies some adjustments were performed as the MAF shows some limitations. First, some polymorphisms have no MAF reported: a theoretical MAF was added for these variants. Second, some MAFs were superior to 0.5: the observed genotypes were inverted for the samples corresponding to these polymorphisms. Third, after these modifications some polymorphisms had no variant allele in all the samples: these polymorphisms were excluded from the analysis (Fig. 1). After these adjustments, the frequency of individuals carrying the variant allele in the general population was used to calculate the theoretical genotype frequencies in a non-CML individuals group (Fig. 2). This distribution was compared to the distribution observed in CML patients using a contingency table. It displayed the frequencies of patients carrying each variant allele (either homozygous or heterozygous) with the polymorphisms reported in rows and the three groups indicated in columns.

Factorial correspondence analysis and hierarchical clustering on principal components

An FCA was performed to display the distribution of individuals carrying variant allele in the three groups: sensitive patients, resistant patients, and the general population [24]. This multivariate graphical technique is used to highlight relationships among categorical variables from a contingency table. Unlike common multivariate analyses, this approach makes no distributional assumptions and preserves the categorical nature of the variables. It can be used, without any restriction, to small cohorts. A two-dimensional “map” was obtained with each dot corresponding to one identified polymorphism. An HCPC was then performed [25]. The data were separated into different groups (clusters) according to the closeness of the different points on the graph correlated to the repartition of the variant allele in the three groups. The number

of clusters was automatically defined. An unsupervised classification was realized and clusters were generated. Then, the k-means method was applied. The centroid of each cluster was moved according to the average of all the points in the cluster. A new classification was then performed using the new centroid value (Table S3. Contingency table and list of the clusters).

Monte Carlo simulation and rank products method

The RP method is a non-parametric statistical method, developed to analyze microarray experiments data, to detect genes differentially expressed between two conditions [26,27]. This approach, based on ranks of fold changes, could also be applied to our experiment. This second method was used to classify the identified genes according to the number and the frequencies of variants found in each one of them. Two approaches were tested. First, the sum of all the variant genotype frequencies was performed. The number of polymorphisms in one gene can be increased with gene size, therefore, the sum of frequencies was divided by the gene size (values defined in probe design, Table S6. Size of the captured genes). Second, the mean of all the variant genotype frequencies for one gene was calculated independently of the gene size. The RP method requires experiment replicates. A Monte Carlo simulation was performed to generate biological replicates from the observed results [28]. A naïve bootstrap was chosen as it is a simulation approach applied in a predicted non-Gaussian subset [29]. Eight simulated experiments were generated (4 from sensitive patients observed data, 4 from resistant patients observed data) with k=999 bootstrap-type resamples for 12 patients each. From these replicates, variant frequencies for each gene (either sum or mean of all the variants per gene) were determined in the two patient groups. For each gene, variant frequencies in CML patients groups were then compared to variant frequencies in the general population by calculating the risk of variant occurrence (ODDS ratio) for each simulated experiments (8 values).

$$ODDS_i = \frac{\sum_{k=1}^{n} \frac{q_{ki} \times (1 - p_{ki})}{p_{ki} \times (1 - q_{ki})}}{999}$$

$\Sigma ODDS_i$ is the plug-in estimation of the $ODDS_i$ in the CML patient group for k=999 bootstrap resample, q is the variant frequency for the gene i in the CML patient group, p is the variant frequency for the gene i estimated in the general population.

The RP method was applied to these calculated values. First, the ODDS ratios for each gene were compared between sensitive and resistant CML patients (fold changes). Genes were then ranked according to the ODDS ratio differences between sensitive and resistant patient groups (two-class analysis).

R software

All these analyses were performed using RStudio software (version 1.1.383) and several dedicated packages. A graphic interface was developed to facilitate the use of these novel analytical methods. A detailed manual and the annotated script are given in File S7.

Abbreviations: 1000G, 1000 genome project; ABCB1, ATP binding cassette subfamily B member 1; ABCG2, ATP binding cassette subfamily G member 2; BCR-ABL1, breakpoint cluster region-abelson 1; CML, chronic myeloid leukaemia; CYP, cytochrome 450; FCA, factorial correspondence analysis; HCPC, hierarchical clustering on principal components; hOCT1, human organic cation transporter type 1; IM, imatinib mesylate; INDEL, insertion and deletion; MAF, minor allele frequency; MDR1, multidrug resistance protein 1; NGS, next-generation sequencing; PTPN22, protein tyrosine phosphatase, non-receptor type 22; RP, rank products; SNP, single nucleotide polymorphism; SLC22A1, solute carrier family 22 member 1; TKI, tyrosine kinase inhibitor; UGT1A9, uridine diphosphate-glucuronosyltransferase family 1 member A9

Declarations

Consent for publication: Not applicable

Availability of data and materials: Not applicable

Competing interests: The authors declare that they have no competing interests

Funding: This work was supported by la Fondation ARC. We thank La region Nouvelle-Aquitaine and SIRIC BRIO for salary support for FL.

Author Contributions: BT, FL, CH, SD and GE conceived, designed and performed the NGS experiment. BT, FL, ML and AG performed the NGS data analyses. BT, FL and SO conceived and tested the statistical methods. All coauthors contributed to writing the paper. FL, SO and BT prepared the manuscript.

Acknowledgments: Not applicable

References

1. Vogel F. Moderne Probleme der Humangenetik. In: Heilmeyer L, Schoen R, Rudder B, editors. Ergebnisse der Inneren Medizin und Kinderheilkunde [Internet]. Berlin, Heidelberg: Springer Berlin Heidelberg; 1959 [cited 2018 Jan 23]. p. 52–125. Available from: http://www.springerlink.com/index/10.1007/978-3-642-94744-5_2

2. Nelson MR, Johnson T, Warren L, Hughes AR, Chissoe SL, Xu C-F, et al. The genetics of drug efficacy: opportunities and challenges. Nature Reviews Genetics. 2016;17:197–206.

3. Relling MV, Evans WE. Pharmacogenomics in the clinic. Nature. 2015;526:343–50.

4. Hong H. Next-Generation Sequencing and Its Impact on Pharmacogenetics. Journal of Pharmacogenomics & Pharmacoproteomics [Internet]. 2012 [cited 2018 Jan 23];03. Available from: https://www.omicsonline.org/next-generation-sequencing-and-its-impact-on-pharmacogenetics-2153-0645.1000e119.php?aid=6088

5. Ross S, Anand SS, Joseph P, Paré G. Promises and challenges of pharmacogenetics: an overview of study design, methodological and statistical issues. JRSM Cardiovascular Disease. 2012;1:1–13.

6. Bacanu SA, Whittaker JC, Nelson MR. How informative is a negative finding in a small pharmacogenetic study? The pharmacogenomics journal. 2012;12:93–95.

7. Sawyers CL. Chronic Myeloid Leukaemia. New England Journal of Medicine. 1999;340:1330–40.

8. Mughal TI, Radich JP, Deininger MW, Apperley JF, Hughes TP, Harrison CJ, et al. Chronic myeloid leukaemia: reminiscences and dreams. Haematologica. 2016;101:541–58.

9. Milojkovic D, Apperley J. Mechanisms of Resistance to Imatinib and Second-Generation Tyrosine Inhibitors in Chronic Myeloid Leukaemia. Clinical Cancer Research. 2009;15:7519–27.

10. Dulucq S, Krajinovic M. The pharmacogenetics of imatinib. Genome medicine. 2010;2:85.

11. Maia R, Vasconcelos F, Souza P, Rumjanek V. Towards Comprehension of the ABCB1/P-Glycoprotein Role in Chronic Myeloid Leukaemia. Molecules. 2018;23:119.

12. Ankathil R, Azlan H, Dzarr AA, Baba AA. Pharmacogenetics and the treatment of chronic myeloid leukaemia: how relevant clinically? An update. Pharmacogenomics [Internet]. 2018 [cited 2018 Apr 3]; Available from: https://www.futuremedicine.com/doi/10.2217/pgs-2017-0193

13. Auton A, Abecasis GR, Altshuler DM, Durbin RM, Abecasis GR, Bentley DR, et al. A global reference for human genetic variation. Nature. 2015;526:68–74.

14. Guillem V, Amat P, Cervantes F, Alvarez-Larrán A, Cervera J, Maffioli M, et al. Functional polymorphisms in SOCS1 and PTPN22 genes correlate with the response to imatinib treatment in newly diagnosed chronic-phase chronic myeloid leukaemia. Leuk Res. 2012;36:174–81.

15. Meech R, Mackenzie PI. Structure and function of uridine diphosphate glucuronosyltransferases. Clin Exp Pharmacol Physiol. 1997;24:907–15.

16. Liu Y, Ramírez J, Ratain MJ. Inhibition of paracetamol glucuronidation by tyrosine kinase inhibitors: Inhibition of paracetamol glucuronidation by TKIs. British Journal of Clinical Pharmacology. 2011;71:917–20.

17. Takahashi N, Miura M, Kuroki J, Mitani K, Kitabayashi A, Sasaki O, et al. Multicenter phase II clinical trial of nilotinib for patients with imatinib-resistant or-intolerant chronic myeloid leukaemia from the East Japan CML study group evaluation of molecular response and the efficacy and safety of nilotinib. Biomarker research. 2014;2:6.

18. Koehler MR, Wissinger B, Gorboulev V, Koepsell H, Schmid M. The two human organic cation transporter genes SLC22A1 and SLC22A2 are located on chromosome 6q26. Cytogenet Cell Genet. 1997;79:198–200.

19. Watkins DB, Hughes TP, White DL. OCT1 and imatinib transport in CML: is it clinically relevant? Leukaemia. 2015;29:1960–9.

20. Baccarani M, Deininger MW, Rosti G, Hochhaus A, Soverini S, Apperley JF, et al. European LeukaemiaNet recommendations for the management of chronic myeloid leukaemia: 2013. Blood. 2013;122:872–84.

21. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.

22. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research. 2010;20:1297–303.

23. Vuong H, Che A, Ravichandran S, Luke BT, Collins JR, Mudunuri US. AVIA v2.0: annotation, visualization and impact analysis of genomic variants and genes: Fig. 1. Bioinformatics. 2015;31:2748–50.

24. Benzécri J-P. Correspondence analysis handbook. New York: Marcel Dekker; 1992.

25. Josse J, Husson F. missMDA: A Package for Handling Missing Values in Multivariate Data Analysis. Journal of Statistical Software [Internet]. 2016 [cited 2018 Jan 23];70. Available from: http://www.jstatsoft.org/v70/i01/

26. Breitling R, Armengaud P, Amtmann A, Herzyk P. Rank products: a simple, yet powerful, new method to detect differentially regulated genes in replicated microarray experiments. FEBS Lett. 2004;573:83–92.

27. Koziol JA. The rank product method with two samples. FEBS Letters. 2010;584:4481–4.

28. Sheppard CW. Computer simulation of stochastic processes through model-sampling (Monte Carlo) techniques. FEBS letters. 1969;2:S14–S21.

29. Efron B. The Jackknife, the Bootstrap and Other Resampling Plans [Internet]. Society for Industrial and Applied Mathematics; 1982 [cited 2018 Apr 4]. Available from: http://epubs.siam.org/doi/book/10.1137/1.9781611970319

Tables

Table 1. List and characteristics of the 48 sequenced genes (obtained from GeneCards® database)

Gene symbol	Chromosomal location	Gene name
Plasma proteins, membrane transporters and regulators (n=10)		
ABCB1	7q21.12	ATP binding cassette subfamily B member 1
ABCC2	10q24	ATP binding cassette subfamily C member 2
ABCG2	4q22.1	ATP binding cassette subfamily G member 2 (Junior blood group)
HFE	6p21.3	hemochromatosis
HIF1A	14q23.2	hypoxia inducible factor 1 alpha subunit
ORM1	9q32	orosomucoid 1
SLC22A1	6q25.3	solute carrier family 22 member 1
SLC22A4	5q23.3	solute carrier family 22 member 4
SLCO1A2	12p12	solute carrier organic anion transporter family member 1A2
SLCO1B1	12p12	solute carrier organic anion transporter family member 1B1
Metabolism enzymes and regulators (n=12)		
CYP1A1	15q24.1	cytochrome P450 family 1 subfamily A member 1
CYP1A2	15q24.1	cytochrome P450 family 1 subfamily A member 2
CYP2C19	10q24	cytochrome P450 family 2 subfamily C member 19
CYP2C8	10q24.1	cytochrome P450 family 2 subfamily C member 8
CYP2C9	10q24.1	cytochrome P450 family 2 subfamily C member 9
CYP2D6	22q13.1	cytochrome P450 family 2 subfamily D member 6
CYP3A4	7q21.1	cytochrome P450 family 3 subfamily A member 4
CYP3A5	7q21.1	cytochrome P450 family 3 subfamily A member 5
NR1I2	3q12-q13.3	nuclear receptor subfamily 1 group I member 2
NR1I3	1q23.3	nuclear receptor subfamily 1 group I member 3
UGT1A1	2q37.1	UDP glucuronosyltransferase family 1 member A1
UGT1A9	2q37	UDP glucuronosyltransferase family 1 member A9

Cell cycle and proliferation (n=5)		
CCND1	11q13	cyclin D1
PPP2R2A	8p21.2	protein phosphatase 2 regulatory subunit B. alpha
RPA1	17p13.3	replication protein A1
RPA2	1p35	replication protein A2
RPA3	7p21.3	replication protein A3
DNA repair (n=10)		
ERCC2	19q13.3	excision repair cross-complementation group 2
ERCC3	2q21	excision repair cross-complementation group 3
ERCC4	16p13.3	excision repair cross-complementation group 4
ERCC5	13q22-q34	excision repair cross-complementation group 5
ERCC6	10q11	excision repair cross-complementation group 6
ERCC8	5q12.1	excision repair cross-complementation group 8
LIG1	19q13.33	DNA ligase 1
RAD23B	9p31.2	RAD23 homolog B. nucleotide excision repair protein
XPA	9p22.3	xeroderma pigmentosum. complementation group A
XPC	3p25.1	xeroderma pigmentosum. complementation group C
Cytokine pathways (n=6)		
CXCL8	4q13-q21	C-X-C motif chemokine ligand 8
IFNG	12q14	interferon. gamma
IFNGR1	6q23-q24	interferon gamma receptor 1
IFNGR2	21q22.1	interferon gamma receptor 2 (interferon gamma transducer 1)
SOCS1	16p13.13	suppressor of cytokine signaling 1
SOCS2	12q	suppressor of cytokine signaling 2
Kinases and phosphatases (n=5)		
AKT1	14q32.33	v-akt murine thymoma viral oncogene homolog 1
ULK3	15q24.1	unc-51 like kinase 3
PTPN1	20q12.1-q13.2	protein tyrosine phosphatase. non-receptor type 1
PTPN2	18p11.3-p11.2	protein tyrosine phosphatase. non-receptor type 2
PTPN22	1p13.2	protein tyrosine phosphatase. non-receptor type 22

Supporting information

Table S1. Description of ANNOVAR annotations (.xlsx)

Table S2. Initial co-occurrence matrix (.csv)

Table S3. Contingency table and list of the clusters (.xlsx)

Table S4. Rank product results (.xlsx)

Table S5. Probe sequences (.xslx)

Table S6. Size of the captured genes (.csv)

File S7. Script to perform the analyses and guide on how to use it (.docx)

Partie II : Annexe 2

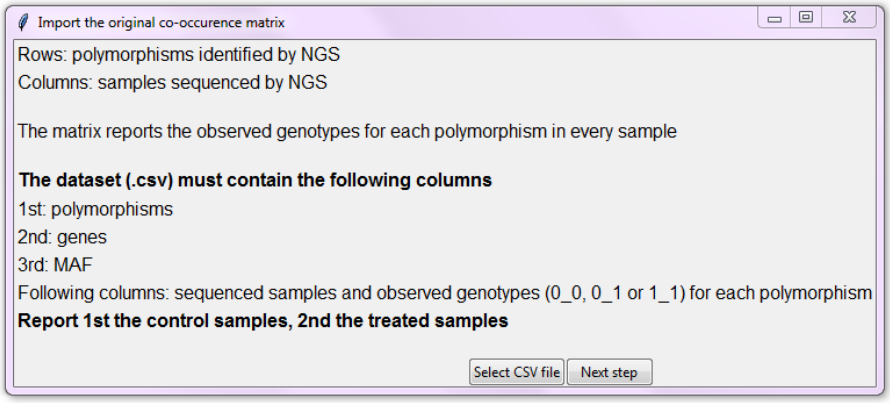
Une interface graphique a été développée pour :

- la correction de la matrice d'occurrence selon la MAF
- la création du graphique associé à l'AFC/HCPC
- la simulation de données et le classement des meilleurs gènes par rank products

Un script, fourni en données supplémentaires de l'article « Novel analytical methods to interpret large sequencing data from small sample sizes » (File S7), doit être chargé sur le logiciel R (version >3.5.). Une série de fenêtres apparaissent et indiquent à l'utilisateur la démarche à suivre.

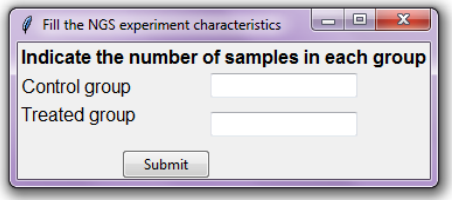
1- Correction de la matrice d'occurrence selon la MAF

Fenêtre 1 :



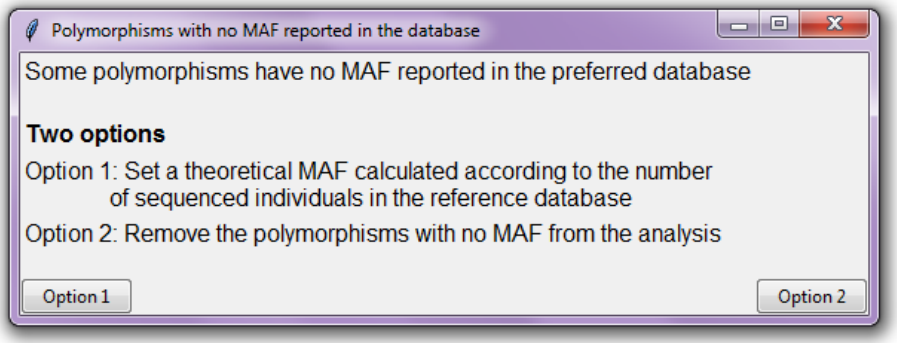
L'utilisateur doit importer la matrice d'occurrence contenant la liste des polymorphismes, les MAF associées et les génotypes identifiés par NGS (groupe contrôle en premier, groupe traité en deuxième) au format .csv.

Fenêtre 2 :



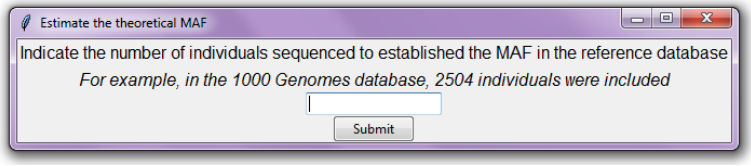
Le nombre d'individus dans chaque groupe doit être renseigné.

Fenêtre 3 :



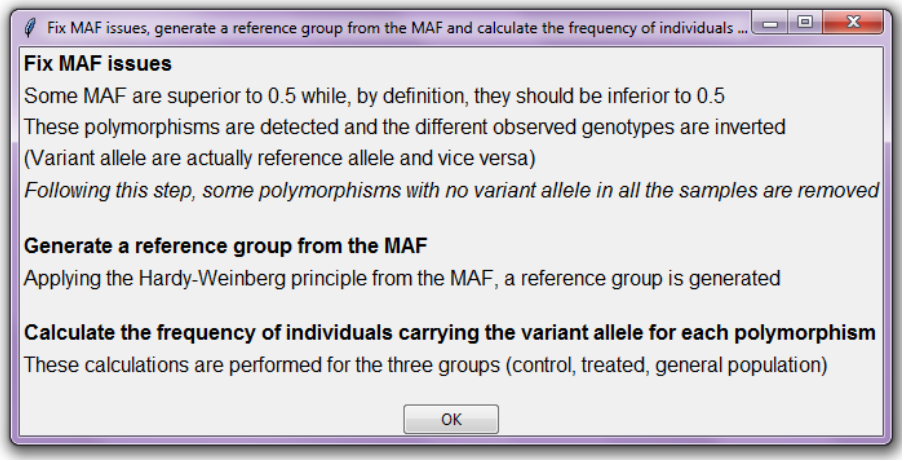
Certains polymorphismes n'ont pas de MAF définie dans la base de données utilisée.
L'utilisateur doit choisir entre deux options :

- Définir une MAF théorique selon le nombre d'individus séquencés dans la base de données (l'information doit être indiquée par la suite)



- Supprimer ces polymorphismes de l'analyse

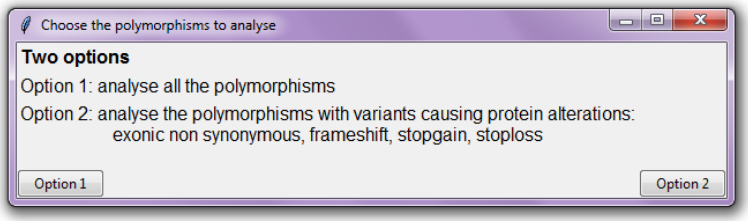
Fenêtre 4 :



Plusieurs calculs se lancent automatiquement :

- Les MAF supérieures à 0,5 sont inversées et les génotypes associés également (si certains polymorphismes n'ont plus de variant chez aucun individu, ils sont éliminés pour la suite des analyses)
- Un troisième groupe de référence est généré en se basant sur la MAF
- La probabilité de survenue du variant est estimée dans les groupes selon les génotypes observés ou estimés

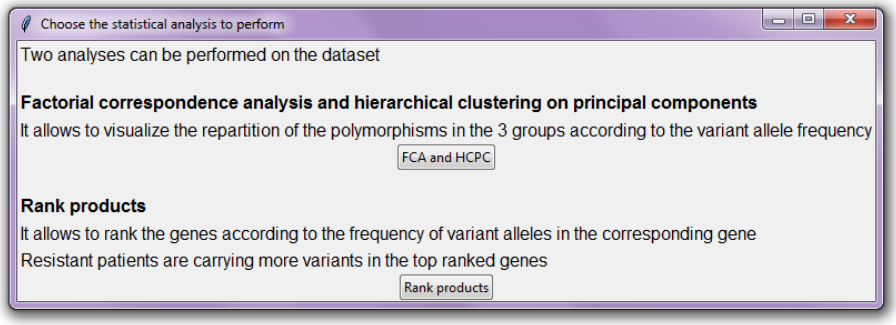
Fenêtre 5 :



Une fois la matrice corrigée et le troisième groupe généré, deux nouvelles options sont proposées à l'utilisateur :

- Etudier tous les polymorphismes
- Ne conserver que les polymorphismes pouvant modifier la conformation protéique (non-synonyme, avec décalage du cadre de lecture, gain de codon STOP, perte de codon STOP)

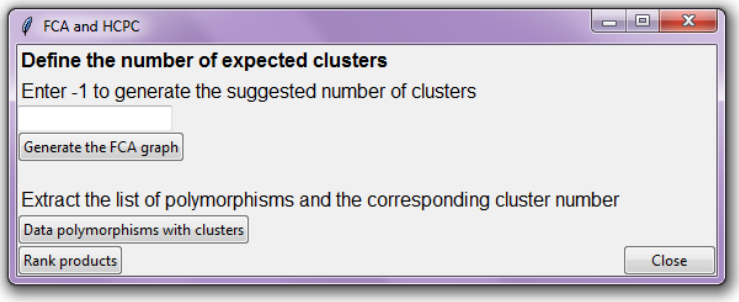
2- Analyses statistiques



L'utilisateur doit choisir quelle analyse statistique, développée dans l'article, réaliser (possibilité de réaliser les deux) :

- AFC et HCPC
- Simulation de données et rank products

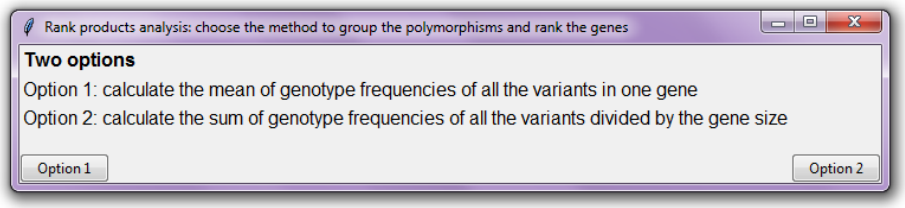
a. AFC et HCPC



L'utilisateur choisit le nombre de clusters à définir (-1 = sélection automatique par l'algorithme). Le graphique et la liste des clusters peuvent être visualisés et enregistrés *via* l'interface du logiciel R. Plusieurs combinaisons peuvent être testées.

b. Simulation de données et rank products

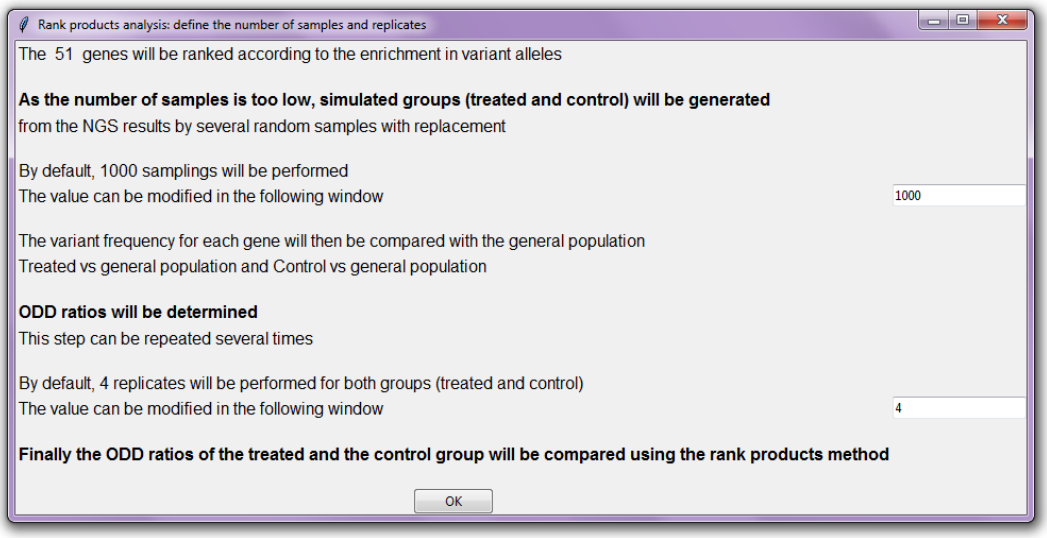
Fenêtre 1 :



De nouveau l'utilisateur a deux possibilités. Rassembler les variants d'un même gène selon :

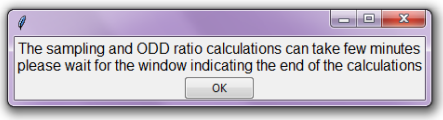
- La moyenne des fréquences de survenue du variant d'un même gène
- La somme des fréquences de survenue du variant d'un même gène

Fenêtre 2 :

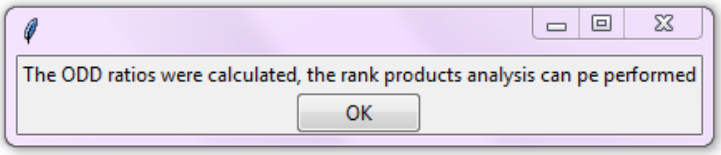


Le nombre de tirages et de réplicas pour la simulation de Monte Carlo sont définis.
Par défaut 4 réplicas de 1000 tirages aléatoires sont réalisés pour chaque groupe d'individus séquencés.

Fenêtre 3 :

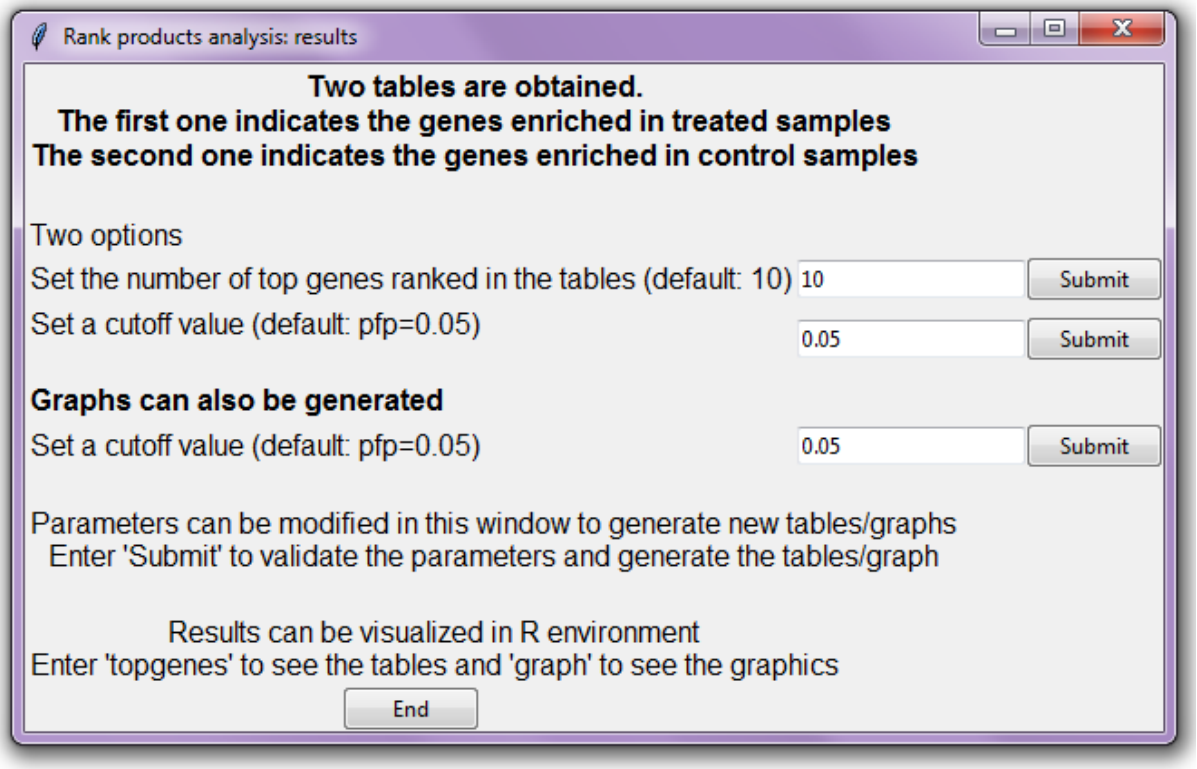


Le prochain algorithme permettant le calcul des rapports de cote (ODDS ratios) démarre lorsque cette fenêtre est validée (bouton « OK » enclenché).



Au bout de plusieurs minutes la fin du calcul des ODDS ratios est notifiée.

Fenêtre 4 :



L'utilisateur choisit le nombre de gènes classés par les rank products. Le classement des gènes et la représentation graphique peuvent être visualisés et enregistrés *via* l'interface du logiciel R. Plusieurs combinaisons peuvent être testées.

Partie III : Annexe 3

Liste des 473 polymorphismes identifiés parmi 43 gènes du panel « apoptose ». Certains polymorphismes sont situés dans des régions intergéniques. Le nom des deux gènes indiqués est donc indiqué (ex : *RPS6KB1,TUBD1°*).

Ref : allèle de référence, Var : allèle variant, Rep : patients répondeurs, Res : patients résistants, Pop : population générale, Clust : numéro de cluster (AFC/HCPC)

					Fréquence d'individus porteurs de l'allèle variant			
Position du variant (hg19)	**Allèle de référence**	**Allèle variant**	**Type de polymorphisme**	**Gène**	**Rep**	**Res**	**Pop**	**Clust**
2:111876941	A	C	SNP	*ACOXL,BCL2L11*	0,92	0,83	0,58	2
2:111877104	C	T	SNP	*ACOXL,BCL2L11*	0,92	0,83	0,61	2
2:111877174	C	G	SNP	*ACOXL,BCL2L11*	0,17	0,08	0,21	4
2:111877175	T	G	SNP	*ACOXL,BCL2L11*	0,17	0,08	0,21	4
11:67050614	G	A	SNP	*ADRBK1*	0	0,08	0	1
11:67052901	G	A	SNP	*ADRBK1*	0	0,08	0	1
11:67049534	C	T	SNP	*ADRBK1*	0	0,08	0,05	1
11:67051082	T	C	SNP	*ADRBK1*	0,08	0	0	3
11:67050123	T	G	SNP	*ADRBK1*	0	0,17	0,26	4
11:67047400	G	C	SNP	*ADRBK1*	0	0,08	0,18	4
11:67047248	T	C	SNP	*ADRBK1*	0	0,08	0,19	4
11:67050292	G	A	SNP	*ADRBK1*	0	0,25	0,58	4
11:67034266	C	A	SNP	*ADRBK1*	0	0,25	0,58	4
11:67052466	G	A	SNP	*ADRBK1*	0	0,17	0,53	4
1:114429340	T	C	SNP	*AP4B1-AS1,BCL2L15*	0,50	0,50	0,61	4
17:7569542	G	A	SNP	*ATP1B2,TP53*	0,25	0,17	0,02	2
17:7569698	A	C	SNP	*ATP1B2,TP53*	0,75	0,67	0,48	2
11:64051823	C	T	SNP	*BAD*	0,08	0,08	0,02	2
11:64044284	C	A	SNP	*BAD*	0,67	0,25	0,31	2
11:64037407	C	T	SNP	*BAD*	0	0,08	0,29	4
6:33541871	G	A	SNP	*BAK1*	0	0,08	0	1
6:33545178	C	T	SNP	*BAK1*	0,17	0,25	0	2
6:33545176	C	A	SNP	*BAK1*	0,08	0,08	0	2
6:33548090	A	AG	insertion	*BAK1*	0,33	0,75	0,38	2
6:33545340	G	A	SNP	*BAK1*	0,50	0,33	0,15	2
6:33541507	A	G	SNP	*BAK1*	0,33	0,67	0,38	2
6:33541719	A	G	SNP	*BAK1*	0,75	0,75	0,57	2
6:33542172	A	T	SNP	*BAK1*	0,08	0	0,06	4
6:33543116	C	T	SNP	*BAK1*	0,08	0	0,08	4
19:49465094	C	T	SNP	*BAX*	0	0,33	0,02	1
19:49464363	C	T	SNP	*BAX*	0,08	0,17	0,03	2
19:49465060	G	A	SNP	*BAX*	0,08	0,08	0,02	2

19:49457938	A	G	SNP	*BAX*	0,33	0,25	0,15	2
19:49459212	A	G	SNP	*BAX*	0,33	0,25	0,16	2
19:49458570	T	C	SNP	*BAX*	0,33	0,33	0,31	2
19:49464971	G	A	SNP	*BAX*	0,33	0,17	0,17	2
19:49459104	A	G	SNP	*BAX*	0,50	0,58	0,61	2
19:47731829	C	T	SNP	*BBC3*	0,17	0,42	0,22	2
19:47735846	A	G	SNP	*BBC3*	0,08	0	0	3
19:47735745	G	A	SNP	*BBC3*	0,08	0	0,01	3
19:47735875	T	C	SNP	*BBC3*	0,08	0	0,03	3
18:60985879	T	C	SNP	*BCL2*	0,58	0,67	0,43	2
18:60986837	G	T	SNP	*BCL2*	0,75	0,67	0,62	2
18:60985600	G	A	SNP	*BCL2*	0,08	0	0,05	3
15:80263765	GA	G	délétion	*BCL2A1*	0	0,08	0	1
15:80259955	TGTAA	T	délétion	*BCL2A1*	0,58	0,17	0	3
15:80263217	C	T	SNP	*BCL2A1*	0,58	0,17	0,37	2
15:80263717	C	T	SNP	*BCL2A1*	0,08	0	0,04	3
15:80263406	C	T	SNP	*BCL2A1*	0,58	0,17	0,58	4
15:80259964	C	T	SNP	*BCL2A1*	0,58	0,17	0,58	4
15:80260014	T	C	SNP	*BCL2A1*	0,58	0,17	0,63	4
15:80263345	A	C	SNP	*BCL2A1*	0,58	0,17	0,71	4
15:52404813	G	A	SNP	*BCL2L10*	0	0,08	0	1
15:52404862	A	C	SNP	*BCL2L10*	0,92	0,92	0,71	2
15:52401951	G	A	SNP	*BCL2L10*	0,25	0,33	0,28	2
15:52404934	C	T	SNP	*BCL2L10*	0,33	0,33	0,39	2
2:111907651	T	G	SNP	*BCL2L11*	0	0,08	0	1
2:111908270	A	G	SNP	*BCL2L11*	0	0,08	0	1
2:111908481	TTGTG	T	délétion	*BCL2L11*	0,83	0,83	0	2
2:111908100	AAAC	A	délétion	*BCL2L11*	0,17	0,17	0	2
2:111881263	A	AT	insertion	*BCL2L11*	0,33	0,25	0	2
2:111902124	G	A	SNP	*BCL2L11*	0,08	0,08	0,01	2
2:111908459	T	TTTTG	insertion	*BCL2L11*	0,17	0,08	0	2
2:111907691	T	C	SNP	*BCL2L11*	0,42	0,75	0,52	2
2:111880600	C	T	SNP	*BCL2L11*	0,92	0,83	0,61	2
2:111908262	G	A	SNP	*BCL2L11*	0,67	0,83	0,70	2
2:111899945	A	T	SNP	*BCL2L11*	0,08	0	0,02	3
2:111899881	G	A	SNP	*BCL2L11*	0,17	0,08	0,18	4
2:111881810	C	T	SNP	*BCL2L11*	0,17	0	0,23	4
2:111908485	G	T	SNP	*BCL2L11*	0,17	0,08	0,68	4
19:50172434	G	A	SNP	*BCL2L12*	0,08	0,08	0,01	2
19:50170197	G	C	SNP	*BCL2L12*	0,17	0,08	0,02	2
19:50168601	G	C	SNP	*BCL2L12*	0,08	0	0	3
19:50172541	C	T	SNP	*BCL2L12*	0,08	0	0	3
19:50170370	C	T	SNP	*BCL2L12*	0,08	0	0	3
19:50168871	G	A	SNP	*BCL2L12*	0,42	0,42	0,61	4
19:50177171	G	A	SNP	*BCL2L12*	0,17	0,08	0,15	4

19:50168940	A	G	SNP	*BCL2L12,IRF3*	0,17	0,08	0,07	2
19:50168927	A	C	SNP	*BCL2L12,IRF3*	0,42	0,42	0,61	4
19:50169020	T	C	SNP	*BCL2L12,IRF3*	0,42	0,42	0,64	4
22:18209866	C	G	SNP	*BCL2L13*	0	0,08	0	1
22:18212631	G	C	SNP	*BCL2L13*	0	0,08	0	1
22:18211333	A	G	SNP	*BCL2L13*	0	0,08	0,01	1
22:18210409	T	G	SNP	*BCL2L13*	0	0,08	0,02	1
22:18210800	TG	T	délétion	*BCL2L13*	0,17	0,17	0	2
22:18209920	C	T	SNP	*BCL2L13*	0	0,33	0,10	1
22:18212776	GC	G	délétion	*BCL2L13*	0,58	0,50	0	2
22:18211020	G	GA	insertion	*BCL2L13*	0,42	0,33	0	2
22:18211021	T	C	SNP	*BCL2L13*	0,42	0,33	0	2
22:18210865	A	AGGCTG	insertion	*BCL2L13*	0,25	0,17	0,05	2
22:18178859	C	T	SNP	*BCL2L13*	0	0,42	0,26	1
22:18210618	G	A	SNP	*BCL2L13*	0,17	0,17	0,10	2
22:18209718	C	T	SNP	*BCL2L13*	0,17	0,17	0,10	2
22:18212923	A	G	SNP	*BCL2L13*	0,17	0,17	0,11	2
22:18210560	T	G	SNP	*BCL2L13*	0,17	0,17	0,12	2
22:18165995	T	G	SNP	*BCL2L13*	0,17	0,17	0,14	2
22:18189469	G	A	SNP	*BCL2L13*	0,17	0,17	0,14	2
22:18211983	T	C	SNP	*BCL2L13*	0,17	0,33	0,28	2
22:18209613	A	G	SNP	*BCL2L13*	0,50	0,58	0,56	2
22:18210828	T	A	SNP	*BCL2L13*	0,08	0	0	3
22:18211205	T	C	SNP	*BCL2L13*	0,67	0,67	0,71	2
22:18121246	T	C	SNP	*BCL2L13*	0,33	0,08	0,11	3
22:18213057	G	A	SNP	*BCL2L13*	0,67	0,67	0,71	2
22:18211185	G	C	SNP	*BCL2L13*	0	0,08	0,08	4
22:18210822	C	G	SNP	*BCL2L13*	0,33	0,08	0,12	3
22:18212530	C	T	SNP	*BCL2L13*	0,33	0,17	0,21	2
22:18210704	G	C	SNP	*BCL2L13*	0,75	0,42	0,56	2
22:18212531	G	A	SNP	*BCL2L13*	0,33	0,08	0,15	2
22:18212855	G	A	SNP	*BCL2L13*	0,08	0	0,02	3
22:18186179	G	A	SNP	*BCL2L13*	0	0,08	0,10	4
22:18121334	G	C	SNP	*BCL2L13*	0,58	0,25	0,44	2
22:18213563	C	T	SNP	*BCL2L13*	0,08	0	0,02	3
22:18120850	C	T	SNP	*BCL2L13*	0,58	0,25	0,47	4
22:18210895	C	T	SNP	*BCL2L13*	0,08	0	0,02	3
22:18210896	G	GC	insertion	*BCL2L13*	0,08	0	0,13	4
12:12247921	AT	A	délétion	*BCL2L14*	0,92	0,83	0	2
12:12224484	G	T	SNP	*BCL2L14*	0,17	0,17	0,06	2
12:12223842	G	A	SNP	*BCL2L14*	0,17	0,17	0,07	2
12:12225302	C	T	SNP	*BCL2L14*	0,08	0,33	0,24	2
12:12247616	T	C	SNP	*BCL2L14*	0,58	0,42	0,29	2
12:12247594	A	G	SNP	*BCL2L14*	0,58	0,42	0,30	2
12:12240199	C	T	SNP	*BCL2L14*	0,92	0,75	0,71	2

12:12251804	G	C	SNP	*BCL2L14*	0,33	0,33	0,34	2
12:12230938	C	G	SNP	*BCL2L14*	0,33	0,42	0,44	2
12:12225207	C	T	SNP	*BCL2L14*	0,75	0,42	0,50	2
12:12223868	T	C	SNP	*BCL2L14*	0,58	0,33	0,42	2
12:12224543	C	T	SNP	*BCL2L14*	0,58	0,33	0,43	2
12:12224718	G	A	SNP	*BCL2L14*	0,58	0,33	0,49	2
12:12224812	C	G	SNP	*BCL2L14*	0,58	0,33	0,49	2
12:12247360	A	G	SNP	*BCL2L14*	0,25	0,33	0,44	4
12:12247793	A	T	SNP	*BCL2L14*	0,17	0	0,11	3
12:12243828	C	A	SNP	*BCL2L14*	0,17	0	0,12	4
1:114423653	C	T	SNP	*BCL2L15*	0,50	0,50	0,64	4
14:23778198	G	A	SNP	*BCL2L2*	0,08	0,25	0,03	1
14:23777099	G	A	SNP	*BCL2L2,BCL2L2-PABPN1*	0,50	0,33	0,44	2
14:23777081	G	A	SNP	*BCL2L2,BCL2L2-PABPN1*	0,08	0,08	0,12	4
6:136599585	C	T	SNP	*BCLAF1*	0	0,17	0	1
6:136599391	C	T	SNP	*BCLAF1*	0	0,08	0	1
6:136599404	T	C	SNP	*BCLAF1*	1,00	1,00	0	2
6:136594187	C	T	SNP	*BCLAF1*	0	0,08	0,03	1
6:136599393	G	C	SNP	*BCLAF1*	1,00	1,00	0,75	2
22:18218422	T	G	SNP	*BID*	0,08	0,17	0,02	2
22:18218464	T	C	SNP	*BID*	0	0,42	0,25	1
22:18218475	C	G	SNP	*BID*	0	0,42	0,28	1
22:18220831	A	G	SNP	*BID*	0,17	0,17	0,12	2
22:18222263	A	G	SNP	*BID*	0,33	0,58	0,45	2
22:18222868	GGCCACGCTCAACT	G	délétion	*BID*	0,17	0	0	3
22:18220746	G	A	SNP	*BID*	0,08	0	0	3
22:18226764	T	C	SNP	*BID*	0,33	0,08	0,09	3
22:18226612	A	G	SNP	*BID*	0,33	0,08	0,10	3
22:18233000	G	A	SNP	*BID*	0,83	0,67	0,72	2
22:18218441	A	C	SNP	*BID*	0,33	0,08	0,16	2
22:43524684	G	A	SNP	*BIK*	0	0,08	0	1
22:43525332	G	T	SNP	*BIK*	0	0,08	0,01	1
22:43524487	CCT	C	délétion	*BIK*	0,92	0,67	0	2
22:43520258	A	C	SNP	*BIK*	0,08	0,50	0,24	1
22:43523689	T	C	SNP	*BIK*	0,08	0	0	3
22:43525234	GTGCTGCTGGCGCTGCTGC	G	délétion	*BIK*	0,08	0	0	3
22:43525262	C	T	SNP	*BIK*	0,08	0	0,01	3
22:43520263	A	G	SNP	*BIK*	0,33	0,67	0,73	4
22:43525176	C	T	SNP	*BIK*	0,08	0,50	0,60	4
8:11351798	C	G	SNP	*BLK*	0	0,08	0,01	1
8:11405576	G	A	SNP	*BLK*	0	0,08	0,02	1
8:11366666	C	T	SNP	*BLK*	0	0,08	0,03	1
8:11367357	T	C	SNP	*BLK*	0	0,08	0,03	1
8:11367042	C	T	SNP	*BLK*	0,08	0,42	0,14	1

8:11367092	T	C	SNP	*BLK*	0,17	0,58	0,35	2
8:11366670	C	T	SNP	*BLK*	0,50	0,58	0,31	2
8:11422045	G	A	SNP	*BLK*	0,50	0,75	0,42	2
8:11422130	T	C	SNP	*BLK*	0,50	0,75	0,42	2
8:11418766	G	A	SNP	*BLK*	0,25	0,25	0,14	2
8:11400944	G	T	SNP	*BLK*	0,42	0,75	0,51	2
8:11400680	G	A	SNP	*BLK*	0,42	0,75	0,51	2
8:11406593	T	C	SNP	*BLK*	0,75	0,92	0,66	2
8:11421793	C	T	SNP	*BLK*	0,83	0,92	0,69	2
8:11415597	C	T	SNP	*BLK*	0,67	0,92	0,70	2
8:11415572	A	G	SNP	*BLK*	0,67	0,92	0,71	2
8:11422170	A	G	SNP	*BLK*	0,83	0,92	0,72	2
8:11418773	C	T	SNP	*BLK*	0,83	0,92	0,75	2
8:11367037	G	C	SNP	*BLK*	0,08	0,17	0,14	2
8:11367016	G	T	SNP	*BLK*	0,08	0	0	3
8:11418564	G	A	SNP	*BLK*	0,08	0	0	3
8:11414237	T	C	SNP	*BLK*	0,50	0,33	0,31	2
8:11367046	C	A	SNP	*BLK*	0,08	0	0,01	3
8:11367038	C	T	SNP	*BLK*	0,58	0,25	0,30	2
8:11418341	C	T	SNP	*BLK*	0,08	0	0,01	3
8:11352056	A	G	SNP	*BLK*	0,50	0,17	0,32	2
8:11351912	C	T	SNP	*BLK*	0,50	0,42	0,67	4
8:11418385	G	A	SNP	*BLK*	0,50	0,33	0,71	4
15:40400420	A	AAC	insertion	*BMF*	0,75	1,00	0	2
15:40398718	A	AAC	insertion	*BMF*	0,17	0,08	0	2
15:40401122	G	GCC	insertion	*BMF*	0,08	0	0	3
15:40401122	G	GC	insertion	*BMF*	0,50	0,42	0,37	2
15:40397936	G	T	SNP	*BMF*	0,83	0,67	0,72	2
15:40384132	G	A	SNP	*BMF*	0,17	0	0,15	4
2:242498150	C	T	SNP	*BOK*	0	0,08	0	1
2:242498984	C	G	SNP	*BOK*	0,08	0,08	0,01	2
1:155233051	C	T	SNP	*CLK2*	0	0,08	0	1
1:155237759	T	C	SNP	*CLK2*	0	0,08	0	1
1:155240609	G	A	SNP	*CLK2*	0	0,08	0	1
1:155236555	C	T	SNP	*CLK2*	0,08	0	0	3
1:155240656	C	G	SNP	*CLK2*	0,08	0	0	3
12:68054629	C	T	SNP	*DYRK2*	0	0,08	0	1
12:68051420	C	A	SNP	*DYRK2*	0	0,08	0,01	1
12:68054877	CT	C	délétion	*DYRK2*	0,42	0,50	0	2
12:68055812	T	C	SNP	*DYRK2*	0,33	0,67	0,44	2
12:68054877	C	CT	insertion	*DYRK2*	0,08	0	0	3
12:68051571	A	G	SNP	*DYRK2*	0,08	0	0	3
12:68054686	T	C	SNP	*DYRK2*	0,75	0,50	0,72	2
12:68055199	G	A	SNP	*DYRK2*	0,75	0,50	0,74	4
12:68052178	G	T	SNP	*DYRK2*	0,75	0,50	0,74	4

16:22268874	C	T	SNP	*EEF2K*	0	0,08	0	1
16:22262095	C	T	SNP	*EEF2K*	0	0,08	0	1
16:22260058	G	A	SNP	*EEF2K*	0,08	0,25	0,03	1
16:22261891	G	A	SNP	*EEF2K*	0	0,08	0,02	1
16:22291442	C	T	SNP	*EEF2K*	0,17	0,33	0,11	2
16:22268526	A	C	SNP	*EEF2K*	0,17	0,42	0,32	2
16:22262514	C	A	SNP	*EEF2K*	0,92	0,92	0,73	2
16:22269946	C	T	SNP	*EEF2K*	0,08	0	0	3
16:22291612	C	T	SNP	*EEF2K*	0,08	0	0	3
16:22285007	G	A	SNP	*EEF2K*	0,08	0	0	3
16:22277965	T	C	SNP	*EEF2K*	0,33	0,33	0,44	4
16:22261900	G	A	SNP	*EEF2K*	0,42	0,42	0,74	4
16:22269122	G	A	SNP	*EEF2K*	0,08	0	0,04	3
16:22255898	C	T	SNP	*EEF2K*	0,17	0,17	0,39	4
16:22237273	C	G	SNP	*EEF2K*	0,08	0,08	0,24	4
16:22237118	A	G	SNP	*EEF2K*	0,17	0,08	0,40	4
12:12207881	C	CAA	insertion	*ETV6,BCL2L14*	0,50	0,67	0	2
12:12207881	C	CA	insertion	*ETV6,BCL2L14*	0,42	0,17	0	3
12:12208613	T	C	SNP	*ETV6,BCL2L14*	0,17	0,42	0,20	2
12:12208614	C	T	SNP	*ETV6,BCL2L14*	0,17	0,42	0,20	2
12:12208578	T	C	SNP	*ETV6,BCL2L14*	0,25	0,50	0,28	2
12:12208009	G	C	SNP	*ETV6,BCL2L14*	0,25	0,50	0,29	2
12:12204745	A	G	SNP	*ETV6,BCL2L14*	0,50	0,42	0,30	2
12:12208651	G	A	SNP	*ETV6,BCL2L14*	0,58	0,75	0,75	2
6:108881021	C	T	SNP	*FOXO3*	0	0,08	0	1
6:108882830	C	T	SNP	*FOXO3*	0,08	0,08	0,05	2
6:108911071	A	C	SNP	*FOXO3*	0,08	0,08	0,05	2
6:108882570	C	T	SNP	*FOXO3*	0,33	0,50	0,33	2
6:108977663	T	C	SNP	*FOXO3*	0,92	0,67	0,75	2
6:108985893	C	T	SNP	*FOXO3*	0,08	0	0,02	3
1:74808631	C	T	SNP	*FPGT-TNNI3K,TNNI3K*	0	0,17	0	1
1:74818986	A	G	SNP	*FPGT-TNNI3K,TNNI3K*	0	0,08	0	1
1:74929048	T	G	SNP	*FPGT-TNNI3K,TNNI3K*	0	0,08	0,01	1
1:74801871	G	C	SNP	*FPGT-TNNI3K,TNNI3K*	0	0,08	0,01	1
1:74716515	CT	C	délétion	*FPGT-TNNI3K,TNNI3K*	0,25	0,33	0	2
1:74901739	CT	C	délétion	*FPGT-TNNI3K,TNNI3K*	0,67	0,83	0	2
1:74737274	C	T	SNP	*FPGT-TNNI3K,TNNI3K*	0,08	0,08	0,01	2
1:74901739	CTT	C	délétion	*FPGT-TNNI3K,TNNI3K*	0,33	0,17	0	2
1:74902244	A	G	SNP	*FPGT-TNNI3K,TNNI3K*	0,17	0,25	0,13	2
1:74954856	C	CT	insertion	*FPGT-TNNI3K,TNNI3K*	0	0,25	0,16	1
1:74901829	A	C	SNP	*FPGT-TNNI3K,TNNI3K*	0,17	0,25	0,15	2
1:75006027	A	G	SNP	*FPGT-TNNI3K,TNNI3K*	0,83	0,83	0,54	2

1:74836076	G	C	SNP	*FPGT-TNNI3K,TNNI3K*	0,08	0	0	3
1:74705494	G	T	SNP	*FPGT-TNNI3K,TNNI3K*	0,33	0,17	0,14	2
1:74715099	C	G	SNP	*FPGT-TNNI3K,TNNI3K*	0,08	0	0	3
1:74836154	T	G	SNP	*FPGT-TNNI3K,TNNI3K*	0,17	0,08	0,10	2
1:74819077	T	G	SNP	*FPGT-TNNI3K,TNNI3K*	0,17	0,08	0,12	2
1:74701950	T	A	SNP	*FPGT-TNNI3K,TNNI3K*	0,75	0,42	0,58	2
1:74716310	C	A	SNP	*FPGT-TNNI3K,TNNI3K*	0	0,08	0,13	4
1:74716515	C	CT	insertion	*FPGT-TNNI3K,TNNI3K*	0,50	0,25	0,62	4
1:74836204	C	T	SNP	*FPGT-TNNI3K,TNNI3K*	0,25	0	0,12	3
1:74705684	G	C	SNP	*FPGT-TNNI3K,TNNI3K*	0,17	0	0,10	3
1:74929075	T	C	SNP	*FPGT-TNNI3K,TNNI3K*	0,17	0	0,13	4
1:74801864	A	G	SNP	*FPGT-TNNI3K,TNNI3K*	0	0,08	0,21	4
1:74957696	A	G	SNP	*FPGT-TNNI3K,TNNI3K*	0,17	0	0,25	4
1:75010153	A	G	SNP	*FPGT-TNNI3K,TNNI3K*	0,08	0	0,14	4
1:75007226	A	T	SNP	*FPGT-TNNI3K,TNNI3K*	0,08	0	0,16	4
1:75010117	T	C	SNP	*FPGT-TNNI3K,TNNI3K*	0,08	0	0,16	4
1:75006076	G	A	SNP	*FPGT-TNNI3K,TNNI3K*	0,08	0	0,16	4
22:22142954	C	T	SNP	*MAPK1*	0	0,08	0,01	1
22:22162126	A	G	SNP	*MAPK1*	0	0,17	0,06	1
22:22160384	T	C	SNP	*MAPK1*	0,17	0,25	0,09	2
22:22160095	G	A	SNP	*MAPK1*	0,08	0	0	3
22:22153239	G	A	SNP	*MAPK1*	0,08	0	0,19	4
6:36040642	AT	A	délétion	*MAPK14*	0	0,08	0	1
6:36065255	C	CT	insertion	*MAPK14*	0	0,08	0	1
6:36065687	G	GTT	insertion	*MAPK14*	0	0,08	0	1
6:36067877	G	C	SNP	*MAPK14*	0	0,08	0	1
6:36063722	T	C	SNP	*MAPK14*	0	0,08	0,01	1
6:36040937	C	T	SNP	*MAPK14*	0	0,08	0,01	1
6:36065664	T	C	SNP	*MAPK14*	0,25	0,50	0,10	2
6:36041707	G	T	SNP	*MAPK14*	0,17	0,25	0,15	2
6:36041431	A	G	SNP	*MAPK14*	0,92	0,92	0,62	2
6:36064036	G	A	SNP	*MAPK14*	0,92	0,83	0,62	2
6:36075219	A	C	SNP	*MAPK14*	0,92	0,83	0,64	2
6:36065687	G	GT	insertion	*MAPK14*	0,92	0,75	0,60	2
6:36070338	G	A	SNP	*MAPK14*	0,08	0	0	3
6:36065185	T	C	SNP	*MAPK14*	0,58	0,33	0,31	2
6:35995633	A	G	SNP	*MAPK14*	0,08	0,17	0,22	4
6:36076280	C	T	SNP	*MAPK14*	0,08	0,17	0,25	4
6:36065622	A	C	SNP	*MAPK14*	0,08	0,17	0,25	4
6:36027142	A	G	SNP	*MAPK14*	0,08	0,17	0,31	4
16:30127888	T	C	SNP	*MAPK3*	0,08	0	0,02	3

16:30134312	C	G	SNP	*MAPK3*	0,08	0,08	0,12	4
10:49643507	TAC	T	délétion	*MAPK8*	0	0,17	0	1
10:49618045	T	C	SNP	*MAPK8*	0	0,08	0	1
10:49633878	C	T	SNP	*MAPK8*	0,42	0,67	0	2
10:49633874	A	ACAAGAATAC ATTCT	insertion	*MAPK8*	0,33	0,50	0	2
10:49633876	G	GTAA	insertion	*MAPK8*	0,33	0,50	0	2
10:49633879	T	G	SNP	*MAPK8*	0,50	0,67	0	2
10:49634868	G	A	SNP	*MAPK8*	0	0,17	0,06	1
10:49635072	T	A	SNP	*MAPK8*	0	0,17	0,06	1
10:49638471	A	G	SNP	*MAPK8*	0,17	0,08	0	2
10:49643226	A	G	SNP	*MAPK8*	0,17	0,17	0,08	2
10:49643570	T	G	SNP	*MAPK8*	0,92	0,92	0,73	2
10:49643679	A	G	SNP	*MAPK8*	0,08	0	0	3
10:49609578	C	A	SNP	*MAPK8*	0	0,08	0,08	4
10:49644416	T	C	SNP	*MAPK8*	0,67	0,75	0,73	2
10:49612844	A	C	SNP	*MAPK8*	0,08	0	0,18	4
10:49618017	A	G	SNP	*MAPK8*	0,08	0	0,18	4
10:49635505	G	A	SNP	*MAPK8*	0,08	0	0,18	4
10:49634179	G	A	SNP	*MAPK8*	0,08	0	0,18	4
10:49635277	T	G	SNP	*MAPK8*	0,08	0	0,18	4
11:63669896	T	C	SNP	*MARK2*	0	0,08	0	1
11:63676562	G	A	SNP	*MARK2*	0	0,08	0,02	1
11:63676517	C	T	SNP	*MARK2*	0	0,08	0,03	1
11:63668215	T	C	SNP	*MARK2*	0	0,08	0,03	1
11:63670441	G	A	SNP	*MARK2*	0	0,08	0,03	1
11:63670724	G	A	SNP	*MARK2*	0,17	0,08	0,01	2
11:63663010	T	G	SNP	*MARK2*	0,08	0	0	3
11:63675849	C	T	SNP	*MARK2*	0,08	0	0	3
11:63676313	G	GC	insertion	*MARK2*	0,75	0,75	0,72	2
11:63606638	C	G	SNP	*MARK2*	0	0,08	0,16	4
1:150550924	T	C	SNP	*MCL1*	0	0,08	0,01	1
1:150552202	C	T	SNP	*MCL1*	0,08	0,25	0,05	1
1:150551295	T	C	SNP	*MCL1*	0,08	0,25	0,05	1
1:150552042	G	A	SNP	*MCL1*	0,50	0,33	0,12	2
19:39659472	T	C	SNP	*PAK4*	0	0,08	0,03	1
19:39660277	C	T	SNP	*PAK4*	0	0,08	0,03	1
19:39665934	C	T	SNP	*PAK4*	0,42	0,17	0,10	2
19:39668371	C	T	SNP	*PAK4*	0,08	0	0	3
19:39664466	G	C	SNP	*PAK4*	0,08	0	0	3
19:39669105	G	T	SNP	*PAK4*	0,25	0	0,03	3
19:39616493	G	A	SNP	*PAK4*	0,50	0,25	0,34	2
19:39616521	T	C	SNP	*PAK4*	0,50	0,25	0,34	2
19:39616685	C	T	SNP	*PAK4*	0,42	0,25	0,34	2
19:39659486	G	T	SNP	*PAK4*	0,33	0,42	0,66	4

18:57567451	C	T	SNP	*PMAIP1*	0,08	0,08	0,03	2
18:57567699	G	A	SNP	*PMAIP1*	0,50	0,33	0,26	2
18:57567638	G	C	SNP	*PMAIP1*	0,17	0,50	0,41	2
18:57567284	G	A	SNP	*PMAIP1*	0,08	0	0	3
18:57567228	A	G	SNP	*PMAIP1*	0,08	0	0,02	3
19:14217547	C	G	SNP	*PRKACA*	0	0,08	0,02	1
19:14208760	A	C	SNP	*PRKACA*	0,08	0,08	0,01	2
19:14217672	C	T	SNP	*PRKACA*	0,08	0	0	3
19:14208782	C	T	SNP	*PRKACA*	0,25	0,17	0,19	2
14:24805462	T	TG	insertion	*RIPK3*	0	0,08	0	1
14:24806852	C	T	SNP	*RIPK3*	0	0,17	0,11	1
14:24806902	G	A	SNP	*RIPK3*	0	0,17	0,11	1
14:24807221	G	A	SNP	*RIPK3*	0	0,17	0,11	1
14:24807302	C	T	SNP	*RIPK3*	0	0,17	0,11	1
14:24808842	G	A	SNP	*RIPK3*	0	0,17	0,12	1
14:24808936	C	T	SNP	*RIPK3*	0	0,17	0,12	1
14:24808388	A	G	SNP	*RIPK3*	0	0,17	0,12	1
14:24808168	G	A	SNP	*RIPK3*	0,75	0,67	0,49	2
14:24806229	G	A	SNP	*RIPK3*	0,75	0,67	0,50	2
14:24806800	G	C	SNP	*RIPK3*	0,75	0,67	0,68	2
14:24805463	G	T	SNP	*RIPK3*	0,25	0,17	0,23	2
14:24808924	C	T	SNP	*RIPK3*	0	0,17	0,33	4
14:24808961	T	G	SNP	*RIPK3*	0	0,17	0,33	4
14:24809185	T	C	SNP	*RIPK3*	0	0,17	0,33	4
14:24809191	T	C	SNP	*RIPK3*	0	0,17	0,33	4
1:26881028	G	C	SNP	*RPS6KA1*	0	0,08	0	1
1:26857674	G	T	SNP	*RPS6KA1*	0	0,08	0	1
1:26881565	AG	A	délétion	*RPS6KA1*	0,42	0,42	0	2
1:26883450	G	C	SNP	*RPS6KA1*	0,08	0,08	0	2
1:26901007	G	A	SNP	*RPS6KA1*	0,08	0,08	0,01	2
1:26900805	G	A	SNP	*RPS6KA1*	0,08	0,25	0,12	2
1:26882710	C	T	SNP	*RPS6KA1*	0,08	0,25	0,13	2
1:26900920	A	T	SNP	*RPS6KA1*	0,08	0,17	0,09	2
1:26900708	G	A	SNP	*RPS6KA1*	0,08	0,25	0,15	2
1:26883334	G	A	SNP	*RPS6KA1*	0,08	0	0	3
1:26898142	A	G	SNP	*RPS6KA1*	0,75	0,83	0,71	2
1:26887150	A	G	SNP	*RPS6KA1*	0,75	0,83	0,74	2
1:26887592	T	C	SNP	*RPS6KA1*	0,75	0,83	0,74	2
1:26881806	C	A	SNP	*RPS6KA1*	0,75	0,83	0,74	2
1:26872554	A	G	SNP	*RPS6KA1*	0,75	0,83	0,74	2
1:26880288	C	T	SNP	*RPS6KA1*	0,42	0,75	0,71	2
1:26880989	T	C	SNP	*RPS6KA1*	0,42	0,75	0,71	2
1:26899687	C	A	SNP	*RPS6KA1*	0,67	0,42	0,55	2
1:26879920	T	C	SNP	*RPS6KA1*	0,08	0	0,02	3
1:26869610	G	A	SNP	*RPS6KA1*	0,58	0,25	0,47	4

1:26900735	C	G	SNP	*RPS6KA1*	0,58	0,17	0,38	2
1:26873245	A	G	SNP	*RPS6KA1*	0,58	0,17	0,39	2
1:26899671	C	G	SNP	*RPS6KA1*	0,25	0,17	0,35	4
1:26883511	A	C	SNP	*RPS6KA1*	0,42	0,08	0,30	4
1:26857954	T	A	SNP	*RPS6KA1*	0,17	0	0,12	4
17:58024656	CT	C	délétion	*RPS6KB1*	0,08	0,33	0	1
17:57992145	TATC	T	délétion	*RPS6KB1*	0,25	0,33	0	2
17:57991988	T	A	SNP	*RPS6KB1*	0,08	0,08	0,01	2
17:57971299	C	CT	insertion	*RPS6KB1*	0,50	0,67	0,42	2
17:58013909	A	G	SNP	*RPS6KB1*	0,33	0,25	0,20	2
17:57991947	T	C	SNP	*RPS6KB1*	0,08	0	0	3
17:58008121	AC	A	délétion	*RPS6KB1*	0,08	0	0	3
17:58011498	G	T	SNP	*RPS6KB1*	0,92	0,75	0,66	2
17:58024275	A	G	SNP	*RPS6KB1*	0,50	0,58	0,56	2
17:58024224	C	CAA	insertion	*RPS6KB1*	0,50	0,58	0,56	2
17:58024324	A	G	SNP	*RPS6KB1*	0,25	0,33	0,33	2
17:58013804	T	A	SNP	*RPS6KB1*	0,08	0	0,01	3
17:58018142	G	C	SNP	*RPS6KB1*	0,25	0,33	0,41	4
17:57970371	T	C	SNP	*RPS6KB1,TUBD1*	0	0,08	0	1
21:36252862	C	T	SNP	*RUNX1*	0	0,08	0	1
21:36421692	AT	A	délétion	*RUNX1*	0,17	0,25	0	2
21:36259324	A	G	SNP	*RUNX1*	0	0,08	0,02	1
21:36421412	TA	T	délétion	*RUNX1*	0,67	0,75	0	2
21:36421331	T	C	SNP	*RUNX1*	0,17	0,25	0,18	2
21:36171699	C	CCCAGGTATTGG	insertion	*RUNX1*	0,08	0	0	3
21:36259312	G	A	SNP	*RUNX1*	0,08	0	0	3
21:36165041	T	C	SNP	*RUNX1*	0,67	0,83	0,73	2
21:36164405	G	T	SNP	*RUNX1*	0,83	0,67	0,60	2
21:36259125	T	C	SNP	*RUNX1*	0,08	0	0	3
21:36193871	T	C	SNP	*RUNX1*	0,25	0,17	0,31	4
21:36231613	G	A	SNP	*RUNX1*	0,08	0	0,04	3
21:36260953	G	A	SNP	*RUNX1*	0,08	0	0,04	3
17:40500566	C	T	SNP	*STAT3*	0	0,08	0	1
17:40485635	C	T	SNP	*STAT3*	0	0,08	0	1
17:40478263	T	C	SNP	*STAT3*	0,08	0,25	0	1
17:40481394	AC	A	délétion	*STAT3*	0,08	0,25	0	1
17:40491273	CA	C	délétion	*STAT3*	0,67	1,00	0	2
17:40489939	T	G	SNP	*STAT3*	0,58	0,50	0,33	2
17:40491273	C	CA	insertion	*STAT3*	0,17	0	0	3
17:40481529	G	C	SNP	*STAT3*	0,67	0,67	0,57	2
17:40475650	G	GA	insertion	*STAT3*	0,58	0,33	0,64	4
17:40483466	C	A	SNP	*STAT3*	0,08	0	0,16	4
17:40489741	G	C	SNP	*STAT3*	0,08	0	0,16	4
17:40489919	T	G	SNP	*STAT3*	0,08	0	0,17	4

17:40442227	C	A	SNP	*STAT5A*	0	0,08	0	1
17:40440336	C	T	SNP	*STAT5A*	0	0,08	0	1
17:40440227	G	C	SNP	*STAT5A*	0	0,17	0	1
17:40461352	C	T	SNP	*STAT5A*	0	0,08	0,01	1
17:40453422	C	A	SNP	*STAT5A*	0,08	0	0	3
17:40462531	C	T	SNP	*STAT5A*	0,08	0	0,01	3
17:40460428	G	A	SNP	*STAT5A*	0,08	0	0,02	3
17:40452675	C	A	SNP	*STAT5A*	0,58	0,42	0,69	4
17:40461003	C	T	SNP	*STAT5A*	0,50	0,17	0,42	4
17:40459737	C	T	SNP	*STAT5A*	0,50	0,17	0,42	4
17:40460989	A	G	SNP	*STAT5A*	0,50	0,17	0,44	4
17:40459562	A	G	SNP	*STAT5A*	0,50	0,17	0,45	4
17:40452031	C	T	SNP	*STAT5A*	0	0,17	0,30	4
17:40353658	T	C	SNP	*STAT5B*	0	0,08	0	1
17:40354615	C	A	SNP	*STAT5B*	0	0,08	0	1
17:40359505	GCA	G	délétion	*STAT5B*	0,92	0,92	0	2
17:40375309	A	G	SNP	*STAT5B*	0,08	0,08	0,01	2
17:40359539	ACAC	A	délétion	*STAT5B*	0,08	0	0	3
17:40353744	G	A	SNP	*STAT5B*	0,08	0	0,03	3
17:40370237	G	T	SNP	*STAT5B*	0,08	0	0,04	3
17:40353722	A	G	SNP	*STAT5B*	0,08	0	0,27	4
19:1219274	G	A	SNP	*STK11*	0,92	0,83	0,72	2
19:1206602	C	T	SNP	*STK11*	0,08	0	0,01	3
19:1218523	G	T	SNP	*STK11*	0,25	0,42	0,47	4
19:1220321	T	C	SNP	*STK11*	0,25	0,50	0,59	4
19:1207238	G	T	SNP	*STK11*	0,17	0,50	0,58	4
19:1206483	C	T	SNP	*STK11*	0,08	0	0,02	3
19:1222012	G	C	SNP	*STK11*	0,50	0,17	0,31	2
19:1220757	T	C	SNP	*STK11*	0,08	0	0,05	3
19:1221161	C	T	SNP	*STK11*	0,17	0	0,19	4
1:74701107	C	T	SNP	*TNNI3K*	0,08	0	0,01	3
17:7573897	T	A	SNP	*TP53*	0	0,08	0,01	1
17:7578210	T	C	SNP	*TP53*	0	0,08	0,01	1
17:7577678	CT	C	délétion	*TP53*	0,33	0,83	0	1
17:7578711	CTTT	C	délétion	*TP53*	0,92	1,00	0	2
17:7579643	CCCCCAGCCCTC CAGGT	C	délétion	*TP53*	0,92	0,92	0	2
17:7576501	G	A	SNP	*TP53*	0,17	0,25	0,03	2
17:7578711	CT	C	délétion	*TP53*	0,08	0	0	3
17:7577407	A	C	SNP	*TP53*	0	0,25	0,32	4
17:7577427	G	A	SNP	*TP53*	0	0,25	0,32	4
17:7579619	G	T	SNP	*TP53*	0,08	0,08	0,15	4
17:7578679	A	G	SNP	*TP53*	0,08	0,08	0,16	4
17:7579801	G	C	SNP	*TP53*	0,33	0,33	0,66	4
17:7579472	G	C	SNP	*TP53*	0,33	0,33	0,71	4

17:7578645	C	T	SNP	*TP53*	0,25	0	0,29	4
17:7578115	T	C	SNP	*TP53*	0,25	0	0,31	4

Partie IV : Annexe 4

Liste des 42 polymorphismes non-synonymes, aux sites de jonction intron-exon ou à l'origine d'un décalage du cadre de lecture. Certains polymorphismes sont situés dans des régions intergéniques. Le nom des deux gènes concernés est donc indiqué (ex : FPGT-TNNI3K, TNNI3K).

Ref : allèle de référence, Var : allèle variant, SNP : polymorphisme d'un seul nucléotide, MAF : fréquence de l'allèle mineur, Local : localisation du polymorphisme dans le gène, ID dbSNO : identifiant de la base de données dbSNP.

Position du variant (hg19)	Ref	Var	SNP/ indel	Gène	MAF	Rep	Res	Local	Annotation	ID dbSNP
11:67050614	G	A	SNP	*ADRBK1*	-	0	0,08	exon	non-synonyme	-
11:67052901	G	A	SNP	*ADRBK1*	-	0	0,08	exon	non-synonyme	rs1030574788
19:47735846	A	G	SNP	*BBC3*	0,0016	0,08	0	exon	non-synonyme	rs45477493
15:80263345	A	C	SNP	*BCL2A1*	0,4589	0,58	0,17	exon	non-synonyme	rs1138358
15:80263406	C	T	SNP	*BCL2A1*	0,352	0,58	0,17	exon	non-synonyme	rs1138357
15:80263217	C	T	SNP	*BCL2A1*	0,2035	0,58	0,17	exon	non-synonyme	rs3826007
15:52404862	A	C	SNP	*BCL2L10*	0,5365	0,92	0,92	exon	non-synonyme	rs2231292
2:111907651	T	G	SNP	*BCL2L11*	0,0004	0	0,08	exon	non-synonyme	rs114585494
19:50170370	C	T	SNP	*BCL2L12*	0,0018	0,08	0	exon	non-synonyme	rs149021967
22:18209866	C	G	SNP	*BCL2L13*	-	0	0,08	exon	non-synonyme	rs990839577
22:18165995	T	G	SNP	*BCL2L13*	0,0711	0,17	0,17	exon	non-synonyme	rs2587070
22:18209920	C	T	SNP	*BCL2L13*	0,0513	0	0,33	exon	non-synonyme	rs9306198
12:12247793	A	T	SNP	*BCL2L14*	0,0541	0,17	0	exon	non-synonyme	rs61739220
6:136599585	C	T	SNP	*BCLAF1*	-	0	0,17	exon	non-synonyme	rs139012879
6:136599393	G	C	SNP	*BCLAF1*	0,4982	1	1	exon	non-synonyme	rs6940018
6:136599391	C	T	SNP	*BCLAF1*	0,0008	0	0,08	exon	non-synonyme	rs141173428
22:18226764	T	C	SNP	*BID*	0,0483	0,33	0,08	exon	non-synonyme	rs8190315
22:43525262	C	T	SNP	*BIK*	0,0034	0,08	0	exon	non-synonyme	rs148407250
8:11405576	G	A	SNP	*BLK*	0,0124	0	0,08	exon	non-synonyme	rs55758736
2:242498984	C	G	SNP	*BOK*	0,007	0,08	0,08	exon	non-synonyme	rs142158341
1:155236555	C	T	SNP	*CLK2*	-	0,08	0	exon	non-synonyme	rs769997230
1:155240609	G	A	SNP	*CLK2*	-	0	0,08	exon	non-synonyme	rs200641226
1:155240656	C	G	SNP	*CLK2*	0,0014	0,08	0	exon	non-synonyme	rs141175277
12:68051420	C	A	SNP	*DYRK2*	0,0052	0	0,08	exon	non-synonyme	rs34166200
12:68051571	A	G	SNP	*DYRK2*	0,0014	0,08	0	exon	non-synonyme	rs56293072
16:22237118	A	G	SNP	*EEF2K*	0,2252	0,17	0,08	exon	non-synonyme	rs9935059
16:22237273	C	G	SNP	*EEF2K*	0,1274	0,08	0,08	exon	non-synonyme	rs17841292

16:22285007	G	A	SNP	*EEF2K*	0,0014	0,08	0	exon	non-synonyme	rs55982915
6:108882830	C	T	SNP	*FOXO3*	0,0246	0,08	0,08	exon	non-synonyme	rs111556510
1:74808631	C	T	SNP	*FPGT-TNNI3K , TNNI3K*	0,002	0	0,17	exon	non-synonyme	rs34521608
1:74836076	G	C	SNP	*FPGT-TNNI3K , TNNI3K*	0,0002	0,08	0	exon; site d'épissage	non-synonyme	rs145260115
19:39664466	G	C	SNP	*PAK4*	-	0,08	0	exon	non-synonyme	-
19:14217672	C	T	SNP	*PRKACA*	0,0002	0,08	0	exon	non-synonyme	rs56085217
14:24805463	G	T	SNP	*RIPK3*	0,1202	0,25	0,17	exon	non-synonyme	rs3212254
14:24806902	G	A	SNP	*RIPK3*	0,0547	0	0,17	exon; site d'épissage	non-synonyme	rs34106261
14:24805462	T	TG	insertion	*RIPK3*	0,0024	0	0,08	exon	décalage de l'ORF	-
1:26883511	A	C	SNP	*RPS6KA1*	0,1609	0,42	0,08	exon	non-synonyme	rs2229712
21:36171699	C	CCCAGG TATTGG	insertion	*RUNX1*	-	0,08	0	exon	décalage de l'ORF	-
21:36252862	C	T	SNP	*RUNX1*	-	0	0,08	exon	non-synonyme	-
21:36259312	G	A	SNP	*RUNX1*	-	0,08	0	exon	non-synonyme	rs765314703
21:36259324	A	G	SNP	*RUNX1*	0,0104	0	0,08	exon	non-synonyme	rs111527738
17:7579472	G	C	SNP	*TP53*	0,5429	0,33	0,33	exon	non-synonyme	rs1042522

Partie V : Annexe 5

Article soumis à Scientific Reports

Precise and Efficient CRISPR/Cas9 Mediated Exon Replacement by Homology-Independent DNA Repair

Florence Lichou[1], Valérie Prouzet-Mauléon[1], Elodie Bierge[1], Matthieu Lewis[1], François-Xavier Mahon[1], Béatrice Turcq[1]*

[1]Laboratory of Mammary and Leukemic Oncogenesis, Inserm U1218 ACTION, University of Bordeaux, Bergonié Cancer Institute, Bordeaux, France.

Corresponding author: Béatrice Turcq. Laboratory of Mammary and Leukemic Oncogenesis, Inserm U1218 ACTION, University of Bordeaux, 146 rue Léo Saignat, bâtiment TP 4ème étage, case 50, 33076 Bordeaux, France. Tel: (+33) 5 57 57 10 51. E-mail address: beatrice.turcq@u-bordeaux.fr

ABSTRACT

CRISPR/Cas9 is a powerful tool for genome editing. Gene knock-out is now widely performed in various organisms. However, specific modifications such as point mutations remain challenging. Indeed, the common methods require a DNA repair template with homology arms to induce the homology directed repair (HDR) mechanism. Yet, this approach is limited as the frequency of HDR is relatively low in most mammalian cells.

Here we propose a novel method to achieve DNA replacement, with no homology sequences, via non-homologous end joining (NHEJ) pathway, more frequent in cells than HDR. We wanted to generate a non-synonymous substitution (rs9306198, 1431C>T, Pro>Ser) in the *BCL2L13* gene (*Bcl-Rambo*) in the K-562 cancer cell line. We wanted to assess the role of this amino acid change in anticancer drug resistance. To allow an efficient nucleotide replacement without affecting the coding sequence we removed the whole exon containing the nucleotide of interest. At the same time, we introduced in the cells a replacement cassette with the variant nucleotide, which inserted between free genomic DNA ends.

This new method, the homology-independent precise and efficient replacement (HIPER) strategy, allows to easily replace a large DNA fragment to induce a specific mutation in cell lines with predominant NHEJ repair.

INTRODUCTION

CRISPR/Cas9 technology has extensively been used to perform efficient gene knock-outs in numerous different species [1–3]. This requires a complex composed of a Cas9 endonuclease and a single guide RNA (sgRNA) molecule, specific to the gene coding sequence. Once hybridized to the DNA, the Cas9-sgRNA complex generates targeted double-stranded breaks (DSBs). In the end, these alterations may be repaired by the error-prone non-homologous end joining (NHEJ) pathway, result in frame-shift mutations and thus lead to the extinction of gene expression [4–6]. CRISPR/Cas9 is also a powerful technology to model human diseases caused by the mutation of a single base [6,7]. To generate such point mutations, in addition to the sgRNA-Cas9 complex, a molecule of donor DNA with homology arms, containing the nucleotide variant, must be introduced into model cells. In this case, the homology-directed repair (HDR) will repair the DSB [5,6]. However, HDR is inefficient in most mammalian cells and substitution or insertion of specific DNA sequences remain challenging [4,7,8].

Recently in our lab, the *Bcl-Rambo* gene was found to be involved in treatment resistance in chronic myeloid leukemia (CML), a clonal myeloproliferative syndrome (unpublished data). Our goal is to reveal the role of a specific missense substitution (rs9306198, 1431C>T, Pro>Ser) in the last coding exon of this gene. The K-562 cell line, established from a CML patient, was chosen to assess this question. It carries the reference allele (cytosine, C) for the polymorphism of interest. In these cells, first attempts to introduce the variant using a repair template, homologous to the targeted sequence, failed despite the use of different strategies to improve the rate of HDR [6,9]. The relatively low efficiency of this repair mechanism, compared to NHEJ, might explain these underwhelming results. It is also well known that the percentage of edited cells may greatly vary according to the targeted locus [10,11]. To overcome these limitations, we developed a method to replace one or several nucleotides at a specific locus only via NHEJ machinery. Several teams previously described that DNA fragments up to >10 kb can be excised from the genome if flanked by two DSBs [12–15]. Moreover, the integration of large DNA fragments into the genome after DSB, based on homology-independent repair or with short homology sequences, were already developed [16–24].

We combined both approaches to develop a novel homology-independent precise and efficient replacement strategy (HIPER). Basically, to introduce specific point mutations into the genomic DNA (gDNA), endogenous DNA sequences are

removed and replaced with similar exogenous DNA fragments containing the wanted modifications. To avoid coding sequence alterations, the whole exon, flanked by non-coding sequences, is replaced. We successfully employed this approach to introduce point mutations into two human cancer cell lines: 293T and K-562.

RESULTS

Design of the HIPER strategy and plasmid constructions

Overview of the HIPER strategy

First, the wild-type (WT) exon needs to be excised from the gDNA. To this end, two sgRNAs are designed with target sites (20 nt complementary to the sgRNA) and protospacer adjacent motifs (PAMs) surrounding the coding region. Second, a replacement cassette, highly similar to the deleted fragment except for the nucleotides of interest, needs to be present in the cell together with the exon removal. Therefore, the sequence is cloned into a donor vector flanked by target sites and PAMs recognized by the same sgRNAs. When the two Cas9-sgRNA complexes are expressed in the cell they could simultaneously cleave the donor plasmid DNA and the gDNA. It leads to the deletion of the endogenous *Bcl-Rambo* exon and the release of the replacement cassette from the vector. In the end, the donor replacement sequence is inserted between free gDNA ends (Figure 1).

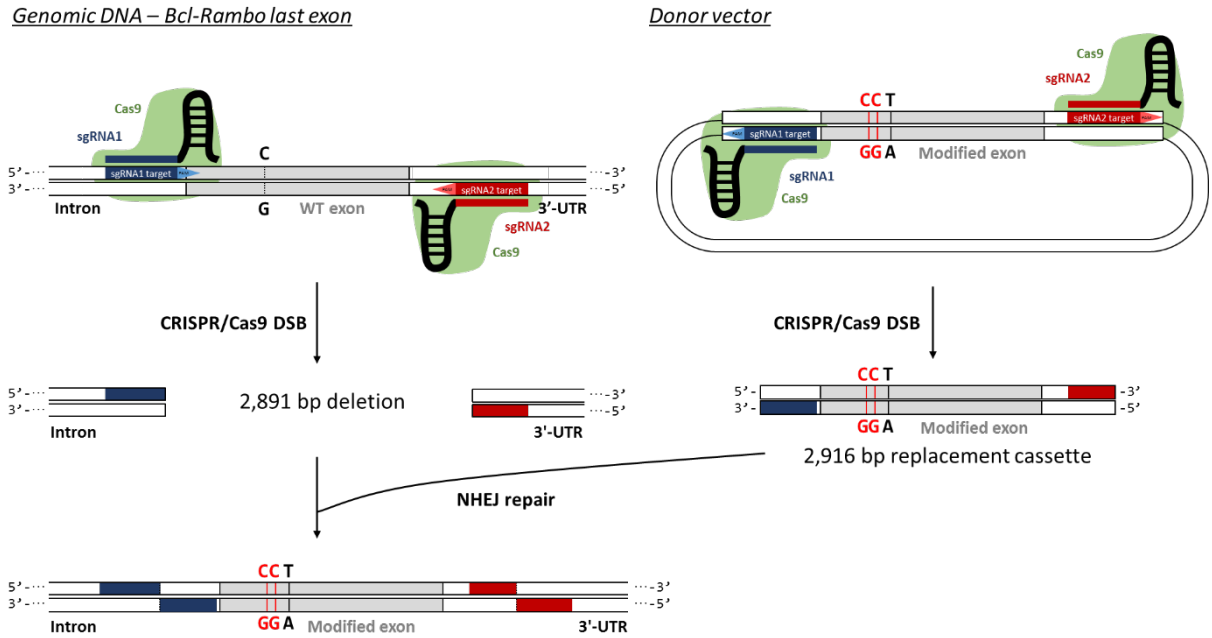


Figure 1: Overview of the HIPER strategy to replace the Bcl-Rambo last exon.
Two single guide RNAs (sgRNAs) were designed to induce double stranded breaks (DSBs) in non-coding sequences on both sides of the Bcl-Rambo last exon. The sgRNA1 and the sgRNA2 were targeting the last intron and the 3'-UTR region, respectively. It allowed the deletion of the last coding exon. The donor vector was constructed with the modified last exon flanked by target sites, complementary to the same sgRNAs, and the protospacer adjacent motifs (PAMs). It permitted the release of the replacement cassette. These sequences were designed in reverse complement and in the opposite direction compared to the endogenous Bcl-Rambo gene. In this way, when the modified exon was introduced in the genomic DNA (gDNA), the PAMs were eliminated and the CRISPR/Cas9 could no longer induces targeted DSBs.

sgRNA-Cas9 vectors

The sgRNA1 targeted the intron upstream the exon of interest. The target site was localized on the DNA top strand and its protospacer adjacent motif (PAM, AGG) was thereby towards the exon. The sgRNA2 targeted the 3'-UTR region downstream the exon. The target site was localized on the DNA bottom strand and hence its PAM (TGG) was also towards the exon. In this way when the two Cas9-sgRNA complexes cleaved the gDNA, the exon and the PAM sequences were removed (Figure 2a). To this end, two plasmids encoding the endonuclease Cas9 with either the sgRNA1 or the sgRNA2 were constructed.

Donor vector containing the DNA replacement cassette

A third plasmid, the donor vector, was generated. This plasmid contained the *Bcl-Rambo* last coding exon sequence with three different nucleotides compared to the WT sequence. First, a T was substituted for a C to recreate the 1431C>T variant and thus, generate the desired amino acid change (Pro>Ser). Second, five nucleotides upstream to the variant, two C were substituted for adenine (A) to create a new restriction enzyme site (*Sma*I, CCCGGG) (Figure 2b). These modifications did not induce an amino acid change (WT: ACA AGG >Thr Arg; variant: ACC CGG >Thr Arg). To allow the release of the replacement cassette, the sgRNA1 and sgRNA2 target sites along with their PAMs were added from either side of the exon cloned into the vector. These sequences were designed in reverse direction on the complementary strand compared to the WT genomic sequence (Figure 2b). Therefore, the PAM sequences were outward facing. In this way, when the two CRISPR complexes cleaved the donor vector, only the DNA fragment containing the modified exon, and its flanking non-coding regions without the PAMs, was released. Consequently, when the exogenous DNA was introduced instead of the endogenous gDNA, the sgRNA target sites were truncated and all the PAMs were definitively removed, preventing future CRISPR/Cas9 complex binding (Figure 2c).

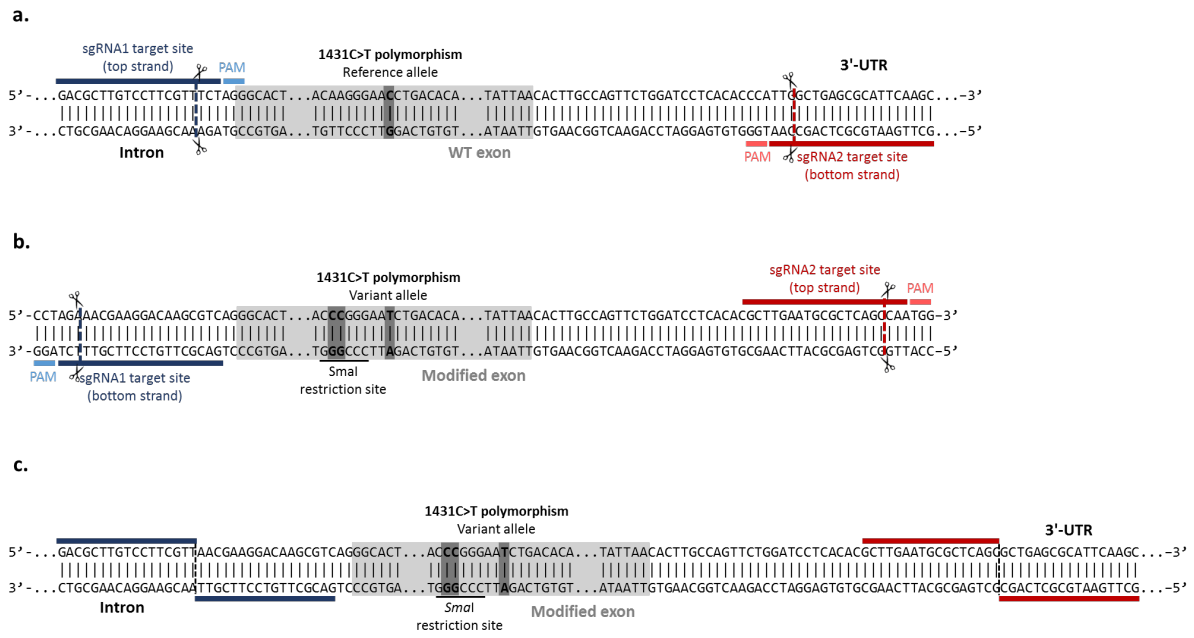


Figure 2: Outline of the tools needed to replace the last Bcl-Rambo exon and generate the 1431C>T variant.

a. Wild-type (WT) genomic DNA (gDNA): focus on the Bcl-Rambo last exon and its flanking non-coding DNA regions. The WT sequence contained the reference allele (C) for the 1431C>T polymorphism. The PAM associated to the sgRNA1 target site was overlapping the intron-exon junction (PAM sequence = AGG, AG was in the intron, the last G of the PAM was in the exon). On the side of the 3'-UTR region, the sgRNA2 target site was further on the bottom strand. The DSB induced by CRISPR/Cas9-sgRNA2 was expected 32 bp after the end of the exon.

b. DNA insert, containing the replacement cassette, cloned into the donor vector. This sequence contained the Bcl-Rambo last exon, with three nucleotides modified compared to the WT sequence. To avoid coding sequence alterations, upstream the exon, two nucleotides AG were also incorporated to rebuild the intron-exon junction. Downstream the exon, the 26 bp from the 3'-UTR region present in the WT gDNA and removed after DSB, were added. Finally, sgRNA1 and sgRNA2 target sites and their PAMs, were included at both extremities of the replacement cassette.

c. Modified gDNA, after WT sequence deletion and insertion of the replacement cassette. This is the expected sequence resulting from CRISPR/Cas9 DSBs occurring 3 bp before the PAM and a correct NHEJ repair, with no random insertion or deletion.

Validation of the sgRNA cleavage efficiency

The first step consisted of assessing if the designed sgRNA sequences and the CRISPR/Cas9 system could efficiently induce DSBs in the gDNA. K-562 cells were co-transfected with both plasmids containing sgRNA and Cas9 coding sequences. Ten days after transfection and antibiotic selection, gDNA from WT and transfected K-562 cells was extracted. A polymerase chain reaction (PCR), targeting both borders outside

the two sgRNA target sites, was performed (P1/P2 PCR, Figure 3a). In both samples a fragment around 3 kb, corresponding to the WT exon amplification, was observed (Figure 3b). Additionally, a fragment around 350 bp was amplified by PCR on the gDNA of transfected cells (Figure 3b). This fragment resulted from the direct ligation of both DSBs to each other, resulting in the removal of the region containing the exon and flanking sequences. Both designed sgRNAs cut the gDNA and were selected for the replacement strategy.

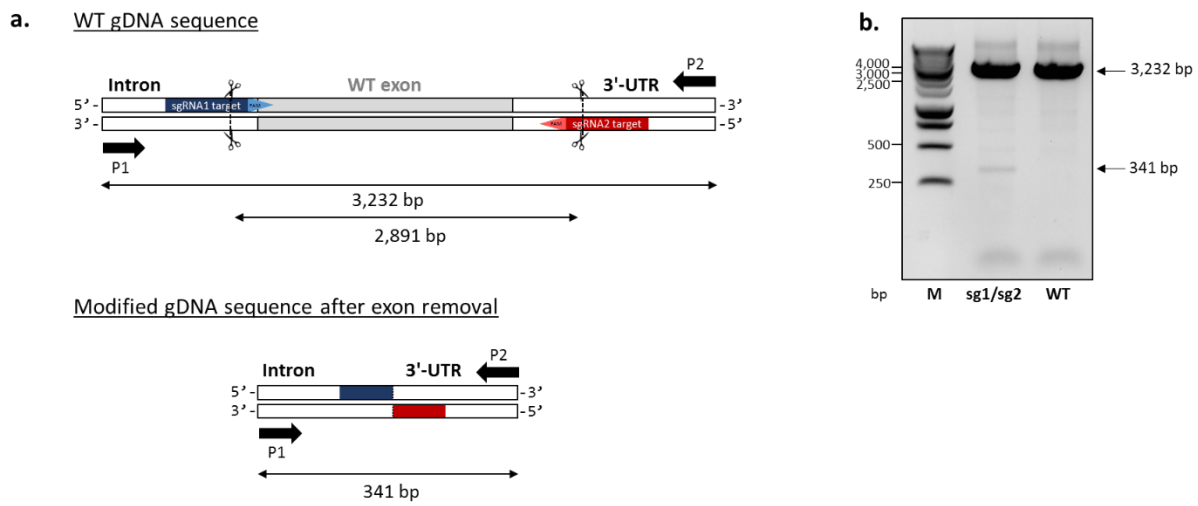


Figure 3: Validation of the sgRNA cleavage efficacy by PCR

a. Design of the P1/P2 PCR and overview of the obtained amplicons.

b. PCR products resolved on a 1% agarose gel electrophoresis. (M) 1kb DNA marker, (sg1/sg2) PCR performed on gDNA extracted from K-562 cells transfected with both vectors carrying sgRNA1, sgRNA2 and Cas9 coding sequences, (WT) PCR performed on gDNA extracted from not transfected WT K-562 cells.

Efficient replacement of the endogenous Bcl-Rambo in a pool of cells

In the second step, the sgRNA-Cas9 vectors were co-transfected with the donor vector in both cell lines (293T and K-562). Ten days after transfection and antibiotic selection, gDNA was extracted from WT and transfected cells. Several strategies were developed to estimate the DNA replacement efficiency.

Insertion of modified nucleotides assessed by enzymatic restriction

To detect the insertion of the modified DNA fragments, a PCR with primers targeting the intron upstream of the sgRNA1 target site and the exon downstream the modified nucleotides (P1/P3 PCR), was achieved and then digested by *Sma*I

(Figure 4a). The *Sma*I restriction site was absent in the WT sequence and should be created after correct exon replacement. A fragment around 1,200 bp was obtained for both cell lines corresponding to the DNA fragment amplified by P1/P3 PCR (Figure 4b). After digestion by the restriction enzyme *Sma*I a relatively weak additional fragment >500 bp was detected for the 293T cells (Figure 4b). It resulted from the PCR product digestion by *Sma*I and thus it indicates that some alleles in these cells were accurately edited. On the other hand, no smaller amplicon was detected for the K-562 cells (Figure 4b). Signals obtained with the digestion by *Sma*I were not sufficient to detect DNA replacement in K-562 cells. The percentage of edited alleles might be too low in these cells to pass the sensitivity threshold or, the DNA replacement was unsuccessful. Therefore, a different screening approach was developed.

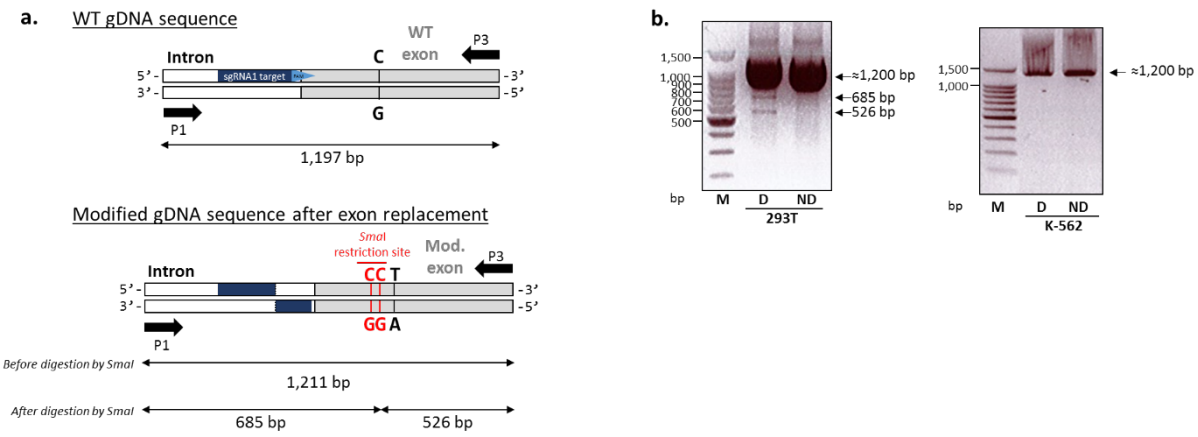


Figure 4: Insertion of modified nucleotides assessed by enzymatic restriction.

a. Design of the P1/P3 PCR and overview of the obtained amplicons, before and after SmaI digestion.

b. PCR performed on gDNA extracted from 293T and K-562 cells transfected with the three vectors. PCR products were resolved on 1.5% agarose gel electrophoresis. (M) DNA marker, (D) PCR products digested by SmaI, (ND) Non-digested PCR products.

Specific amplification of the modified exon

Another PCR amplification was conducted to improve the detection sensitivity of the modified DNA alleles. The forward primer (P1) annealed to the intron, upstream of the sgRNA1 target site. The reverse primer (P4) was conceived to anneal specifically to the modified sequence but not to the WT sequence. To this end, the three expected modified nucleotides were included in the 3' end of the primer sequence generating 3 mismatches with the WT exon preventing the WT fragment amplification (Figure 5a). No signal was detected in WT cell lines whereas an amplicon around 700 bp was

observed in both transfected cell lines (Figure 5b). This amplification resulted from an efficient insertion of the DNA replacement cassette at the expected locus in the correct orientation. With the same amount of transfected vectors, the signal was almost 3 times more intense for the 293T cells compared to the K-562 cells. This difference of signal between both cell lines may be due to differences in NHEJ activation between the two cell lines or, the variation in transfection efficiency. Indeed, transfection was less efficient in K-562 cells and induced higher cell mortality. This technique seems more sensitive than the digestion of PCR products that did not permit to detect gDNA modifications in K-562 cells. These experiments confirmed that, in some cells, the replacement cassette was inserted after the intron in the correct orientation. The percentage of edited alleles in a pool of cells was high enough to be easily detected by classical PCR. However, it was not elucidated yet if the gDNA was accurately edited. Indeed, we confirmed that at least one replacement cassette was added into the gDNA although we didn't evaluate if the WT exon was removed and if only a unique fragment was inserted.

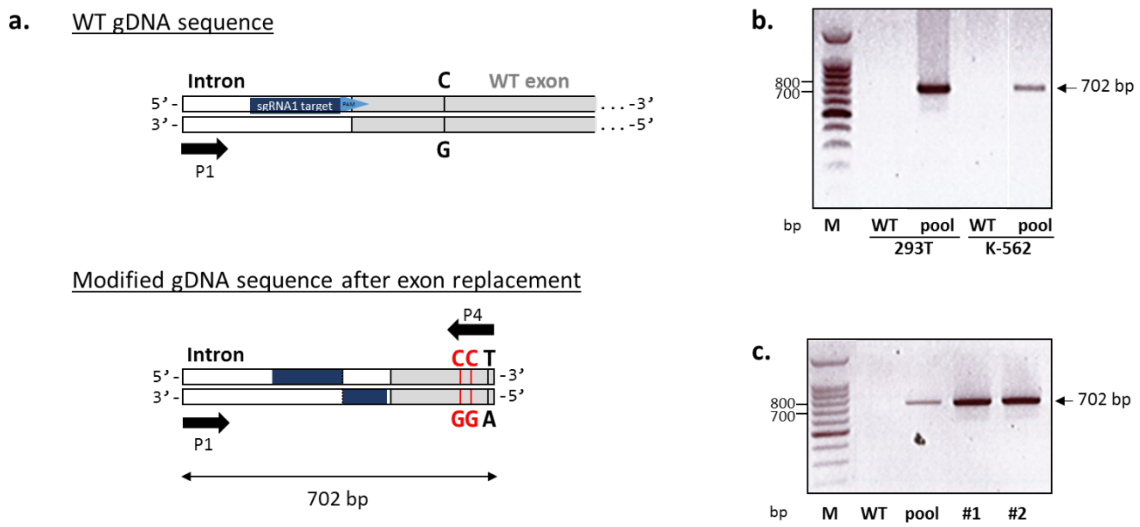


Figure 5: Specific amplification of the modified exon.

a. Design of the P1/P4 PCR and overview of the obtained amplicons. b. and c. PCR performed on gDNA extracted from 293T and K-562 cells. PCR products were resolved on 1.5% agarose gel electrophoresis. (M) 100bp DNA marker, (WT) Non-modified cells, (pool) Mix of modified cells, (#1) and (#2) Modified clonal cell lines.

Characterization of the DNA modifications in clonal cell lines

As the replacement strategy seemed efficient in a pool of cells, sub-cloning was performed to isolate the positive clones in the K-562 cell line. The P1/P4 PCR (primer with mismatches) was used to discriminate the clones carrying the mutations. Two out of 60 clones (3.3 %) were positive in PCR meaning that at least one copy of the gene contained the replacement cassette in these cells.

Characterization of the gDNA modifications by PCR

To obtain an overview of the modifications that occurred in gDNA, five PCRs were performed with different pairs of primers. It allowed evaluating, in the two clonal cell lines, if the WT exon was properly removed, if a unique cassette was integrated and if this cassette was in the correct orientation. The two primers annealing from either side of the target sites (P1 and P2) were used. The reverse primer, annealing specifically to the modified sequence (P4), was also chosen. A forward primer (P5), annealing specifically to the modified sequence along the same principle as the P4 reverse primer, was designed. Another pair of primers (P6 and P7) was also chosen to analyze the junctions created after the DSBs (Figure 6). If the gDNA was accurately edited, a positive signal was presumed for the P1/P4 PCR (702 bp) but no signal was expected for the P1/P5, P4/P2, P5/P4, and P6/P7 PCR. These PCRs were performed on gDNA extracted from K-562 cells: WT cells, the pool of transfected cells and the two clonal cell lines. For both K-562 modified clones, the expected signals were observed with a signal around 700 bp for the P1/P4 PCR (Figure 5c) and no amplification with the other PCRs. These findings suggest that the gDNA was efficiently modified in the two clones with the replacement, in the correct orientation, of the *Bcl-Rambo* last exon. Still, the number of edited copies of the gene remains unknown.

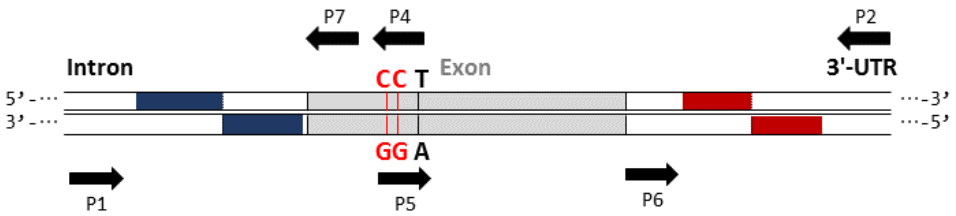


Figure 6: Characterization of the gDNA modifications by PCR.

Design of five PCR using six different primers (P1, P2, P4, P5, P6 and P7) to elucidate the gDNA modifications.

Detection of nucleotide modifications by Sanger and next-generation sequencing

To further characterize the modifications, we analyzed the targeted genomic locus by Sanger sequencing in the two modified clones and the WT cells. The P1/P2 PCR products were sequenced with three different primers. The P1 and P2 primers were chosen to study the junction sequences. One additional primer (P8), annealing to the *Bcl-Rambo* exon upstream the variant of interest, was used to verify the three-base modification (Figure 7a). Near the junctions, with the P1 and P2 primers, several signals were detected for both K-562 clones meaning that the DSBs occurred at both target sites and random insertions or deletions were generated. For each clone, there were too many signals to distinguish the sequences on the chromatogram. Indeed, the number of copies of the *Bcl-Rambo* gene is unknown in the K-562 cells and more than two copies are suspected. For the K-562 clone #1, both reference and modified nucleotides were detected with the P8 primer (Figure 7b). These results suggest that the replacement was efficient in a fraction of the *Bcl-Rambo* gene copies in this clone. On the other hand, no modified nucleotides were detected in the K-562 clone #2 despite the positive signal in P1/P4 PCR (Figure 7b). The limit of detection for Sanger sequencing is around 15-20%. If the percentage of modified alleles was lower in this clonal cell line, it could not be detected by this technique. Therefore, the modified sequence was also analyzed in the two K-562 clones and the WT cells by next-generation sequencing (NGS, limit of detection around 1-5%). A PCR was performed to amplify a 407 bp fragment containing the locus of interest (Figure 7c). Both primers were designed from either side of the modified nucleotides, into the *Bcl-Rambo* last exon. The fraction of modified reads was determined. For the first clone, 41% of the reads (149 aligned reads in total) were correctly edited and for the second clone, 13% of the reads (214 aligned reads in total) were modified (Figure 7d). This result coincides with the fact that no signal was detected by Sanger in clone #2. The initial pool of cells was also sequenced by NGS and 11% of the reads (255 aligned reads in total) contained the expected modifications. The PCR and NGS approaches confirmed that some alleles were accurately edited in both cell lines. The proportion of modified alleles suggests that more than two copies of *Bcl-Rambo* gene are indeed present in the targeted cell lines.

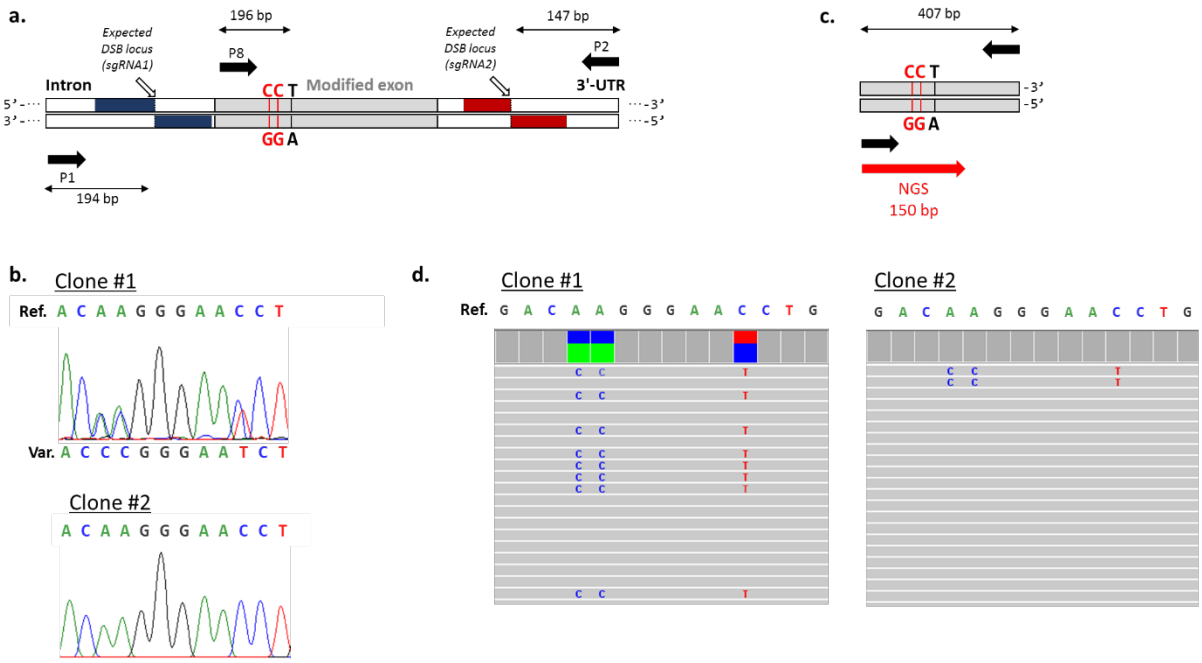


Figure 7: Detection of nucleotide modifications by Sanger and next-generation sequencing (NGS).

a. P1/P2 PCR products were sequenced using three different primers (P1, P2 and P8).

b. Sanger chromatograms with the P8 primer. For the clone #1, two signals were overlapping. The three targeted nucleotides were accurately edited. No modification was detected for the clone #2.

c. Design of a 407 bp PCR to sequence the modified locus by NGS.

d. NGS results for the K-562 clones #1 and #2 visualized on the Integrative Genomics Viewer (IGV) software (Broad Institute). 41% and 13% of the reads, respectively, were containing the modified nucleotides compared to the reference genome.

Efficient transcription of the modified DNA assessed by RT-PCR

Then, to evaluate if the DNA replacement did not prevent *Bcl-Rambo* gene expression, a PCR annealing specifically to the modified exon (P8/P4) was performed on cDNA obtained from total RNA extracted from the different K-562 cell lines.

As expected, an amplicon around 200 bp was detected in the cells modified by CRISPR/Cas9 but no signal appeared for the WT cells (Figure 8a). To confirm the purity of the cDNA (no contamination by gDNA) the same PCR was performed on RNA samples. As a positive control, gDNA from the pool of modified K-562 cells was amplified. A signal around 200 bp was detected from the gDNA sample whereas no amplicon was observed from the RNA samples (Figure 8b). It confirmed that the modified gDNA in the K-562 cell lines was accurately transcribed.

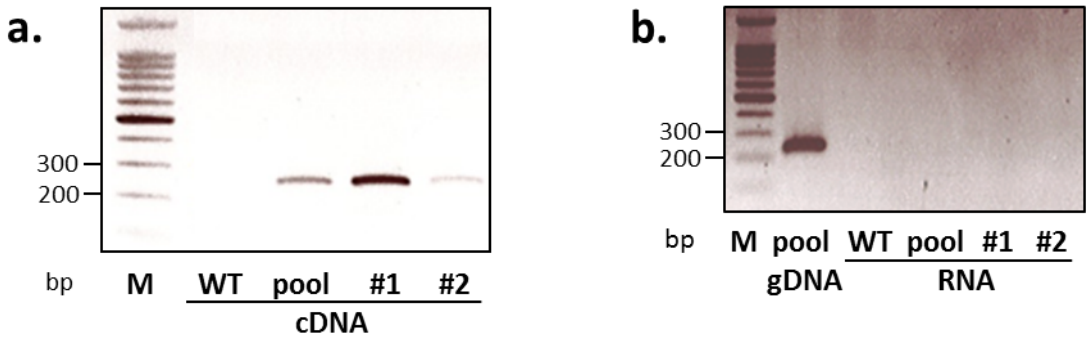


Figure 8. Efficient transcription of the modified DNA assessed by RT-PCR.
Reverse transcription and P8/P4 PCR were performed on RNA extracted from K-562 WT cells, pool of modified cells and the two modified clonal cell lines. As a positive control in the P8/P4 PCR on RNA extracts, gDNA extracted from the pool of modified cells was used. The P8/P4 PCR products were resolved on 1.5% agarose gel electrophoresis. (M) 100 bp DNA marker, (WT) Non-modified cells, (pool) Mix of modified cells, (#1) and (#2) Modified clonal cell lines. Input samples a. cDNA, b. RNA.

DISCUSSION

We described a novel efficient strategy, HIPER, based on the CRISPR/Cas9 technology, to specifically generate a point mutation in two mammalian cell lines (293T and K-562) via NHEJ repair. We were able to remove a 3 kb gDNA fragment and replace it with a novel fragment of similar size, without any homology sequences. This method was conceived on the basis of previous studies. Nakade *et al.* and later Sakuma *et al.* described the PITCh (Precise Integration into Target Chromosome) system [17,22]. This strategy was based on the repair mechanism named microhomology-mediated end joining (MMEJ). A selection cassette was inserted into a vector flanked by short microhomology sequences and sgRNA target sites. The vector was co-transfected in cells expressing the Cas9 and sgRNAs targeting the region of interest in the gDNA and the donor vector to free the cassette. This repair mechanism is more predominant in cells than HDR and more precise than NHEJ but still required homology sequences. Yao *et al.* developed a similar strategy with the homology-mediated end joining (HMEJ) machinery, similar to MMEJ with larger homology arms [23]. Suzuki *et al.* conceived another method via the classical NHEJ repair named homology-independent targeted integration (HITI) [21]. They were able to perform efficient gene knock-in *in vitro* and *in vivo* by inducing one single DSB in gDNA and adding simultaneously a replacement cassette in various forms. HITI was more efficient to integrate the cassette than the PITCh strategy or repair by HDR. When they designed one cut in the gDNA and two cuts from either side of the knock-in cassette they reached

around 25% of modified cells compared to less than 5% via HDR. In addition, several teams observed that two targeted DSBs in DNA could result in large deletions based on NHEJ repair in several eukaryotic cell lines. Zheng *et al.* were able to induce removal of fragments up to 10 kb with relatively high-efficiencies (up to 28% of edited fragments detected by quantitative PCR) [12].

We decided to test our strategy, combining both DNA deletion and insertion, to generate a specific missense substitution in the last coding exon of the *Bcl-Rambo* gene (rs9306198, 1431C>T, Pro>Ser). Several elements were considered to design the technique. First, DSBs repaired by NHEJ can generate random insertions or deletions, that is why non-coding regions were targeted by the CRISPR complexes. Still, it is essential to keep in mind that even if these regions are non-coding they might be involved in gene regulation. Second, it is also interesting to note that the two cell lines used in this study, 293T and K-562, are polyploid and the exact number of copies of the *Bcl-Rambo* gene is unknown. We didn't succeed to edit all the alleles in one clonal cell line. It would be interesting to evaluate the efficiency of this technique to modify a gene with a known number of copies or to test it in haploid cells (e.g. KBM-7) [25]. Nevertheless, the missense substitution that was generated was found mostly in the heterozygous state in CML patients (unpublished data). In this way, the modified cell lines remain a great model to assess the effect of this variant on treatment sensitivity. Third, it is well known that great variability exists in CRISPR/Cas9 editing according to the targeted locus primarily due to epigenetic regulation [10,11]. The efficiency could either be decreased or increased in other loci and must be tested for each project. Finally, the efficiency of gene editing can be easily assessed by PCR. It allows for the evaluation of different scenarios in a simple and robust manner. NGS could also be performed to validate in a more precise manner the results. Nevertheless, the exact modifications in each allele cannot be known with the techniques that we used. To be certain of the sequence of each modified fragment, a long-read sequencing approach will be necessary [26]. Interestingly, other recent approaches were also developed in order to generate point mutations, such as CRISPR base editors [27]. These methods are based on using modified Cas9 proteins which are unable to perform breaks in the DNA. These tools reduce the occurrence of random insertions or deletions and the off-target effects. But the application of these new strategies is limited. It allows only to modify a C to T, A to G and the opposite

substitutions. Moreover, the need for a specific PAM sequence close to the target DNA limits also their utilization.

With the HIPER method, DSBs can be performed in non-coding regions and far from the nucleotide of interest. Therefore, a PAM sequence can be easily found in the intron sequence. No homology is needed and the inserted sequence can be similar or entirely different to the targeted DNA, in size and composition. DSB is repaired by the widely expressed NHEJ pathway and thus high-efficiency rates could be easily obtained, in dividing or post-mitotic cells. Multiple gene editing approaches could be envisioned with this strategy: create point mutations but also insert tag sequences or switch DNA region encoding a specific protein domain.

METHODS

sgRNAs and Cas9 plasmids

sgRNAs were designed using the CHOPCHOP v2 web tool (http://chopchop.cbu.uib.no). sgRNA sequences (sgRNA1, targeting the intron region: GACGCTTGTCCTTCGTTTCT, sgRNA2, targeting the 3'-UTR region: GCTTGAATGCGCTCAGCCAA) were cloned into three individual modified PX462 plasmid (a gift from Feng Zhang, Addgene #48141) as previously described using annealed and phosphorylated oligonucleotides [28]. In this plasmid, the D10A Cas9 nickase was replaced by the enhanced specificity nuclease Cas9 from *Streptococcus pyogenes* (eSPcas9(1.1)). The puromycin resistance was improved by a Puro-lowCpG sequence.

HIPER donor vector with the replacement cassette

A donor vector containing the replacement cassette (modified *Bcl-Rambo* exon) was generated. First, an overhang PCR was performed on K-562 gDNA (carrying the reference allele for the rs9306198 polymorphism, 1431C). The last coding exon of *Bcl-Rambo* was amplified, as well as the acceptor splice site (AG) of the last exon and 16 bp of the 3'-UTR region (corresponding to the sequence until the PAM recognized by the sgRNA2). Primers contained also sgRNA target sites and their associated PAM in reverse complement compared to the gDNA (Forward: 5' CCTAGAAACGAAGGACAAGCGTCagggcactgtgtttagtctt 3'; Reverse: 5' CCATTGGCTGAGCGCATTCAAGCGTgtgtgaggatccagaactgg 3'). The KAPA HiFi PCR Kit (Kapa Biosystems) was used. The 2,930 bp amplicon was inserted into a

pCR™-Blunt II-TOPO® vector using the Zero Blunt® TOPO® PCR Cloning Kit (Thermo Fisher Scientific) following the manufacturer's instructions (with 1 µL of PCR product in a final volume of 6 µL). The variant allele and two point mutations (new *Sma*I restriction site) were added by targeted mutagenesis. The QuikChange® Site-Directed Mutagenesis (Stratagene) kit was used with the following primers (Forward: 5' CTCCCTGCTGGGGACCCGGGAATCTGACACAGAAGTG 3', Reverse: 5' CACTTCTGTGTCAGATTCCCGGGTCCCCAGCAGGGAG 3') using the KOD Hot Start DNA Polymerase (Merck) (PCR program in Supplementary Table S2). Prior mutagenesis, primers were phosphorylated using the T4 polynucleotide kinase (Thermo Fisher Scientific) following the manufacturer's instructions.

Bacterial transformation and plasmid extraction

Plasmids were amplified in competent *Escherichia coli* TOP10 bacteria (Thermo Fisher Scientific). The transformed bacteria were cultured on LB agar containing either 100 µg/mL of ampicillin or kanamycin. Plasmids were extracted by an endotoxin-free midiprep method (NucleoBond® Xtra MidiEF, Macherey Nagel). Plasmid integrity was assessed by enzymatic digestion. The correct sequence and insertion of the sgRNA coding sequences and the replacement cassette into the different plasmids were validated by Sanger sequencing using different primers. For the sgRNAs, a forward primer annealing to the human U6 promoter was used (5' CAAGGCTGTTAGAGAGATAATTGGA 3'). For the replacement cassette, four different primers were chosen. Two primers annealing to the M13 priming sites provided in the Zero Blunt® TOPO® PCR Cloning Kit (Thermo Fisher Scientific) (Forward: 5' GTAAAACGACGGCCAG 3', Reverse: 5' CAGGAAACAGCTATGAC 3'). Two forward primers annealing to the *Bcl-Rambo* last exon (5' GCACGTGGAGAAAGAAGAGG 3' and 5' GACATTGCCAAATGTCCCGT 3'). The same primers were used to sequence the donor vector after mutagenesis and no alteration was detected.

Cell culture

293T cells (ATCC® CRL-3216™) were grown in Dulbecco's Modified Eagle's Medium (DMEM, Thermo Fisher Scientific). K-562 cells (ATCC® CCL-243™) were grown in Roswell Park Memorial Institute (RPMI) 1640 medium (Thermo Fisher Scientific). Both media were supplemented with 10 % fetal calf serum (FCS, Eurobio),

1 mM L-glutamine, 100 units/mL penicillin, 0.1 mg/mL streptomycin (Thermo Fisher Scientific). Cells were maintained in a humidified atmosphere at 37°C with 5 % CO_2. Mycoplasm contamination was monthly checked.

Transfection

The calcium phosphate co-precipitation method was chosen to introduce plasmids into 293T cells. DNA was mixed with $CaCl_2$ 140 mM and HEPES buffered saline (HBS) 1X buffer. The 4D-NucleofectorTM system (Lonza) was used to introduce plasmids into K-562 cells. The SF Cell Line 4D-Nucleofector™ X Kit L (Lonza) was used as transfection reagent according to the manufacturer's instructions. The FF-120 program was applied. For each transfection, one million cells were transfected with 2 µg of each sgRNA-Cas9 plasmid and 15 µg of HIPER donor vector. Transfected cells were selected with 3.5 µg/mL puromycin for 48 h, one day after transfection.

Cell sub-cloning

K-562 cells have the ability to form colonies in a semi-solid methylcellulose-based medium. This property was used to isolate cells and obtain clonal cell lines. Cells were diluted at a ratio 1:5 in a MethoCult™ H4534 classic without EPO medium (Stemcell technologies). Cells were cultured 10 days in humidified atmosphere at 37°C with 5% CO_2. Therefore, colonies were individually picked up and deposited in a 96 wells plate with RPMI 1640 supplemented medium.

Genomic DNA and RNA extraction

293T and K-562 cells were harvested and a phenol-chloroform based extraction was performed to extract gDNA and total RNA. Each cell pellet was resuspended in 1 mL of TRIzol™ Reagent (Thermo Fisher Scientific). Cells were totally lysed using a Disruptor Genie (Scientific Industries). The solutions were mixed with 200 µL Chloroform:isoamyl alcohol 24:1 (Sigma-Alrich), incubated 3 min at room temperature then centrifuged 15 min at 11,300 g. White interphase, containing gDNA molecules, appeared between the TRIzol™ reagent and the upper aqueous phase containing the RNA. gDNA and RNA phases were collected independently. The white interphase with gDNA was mixed with 300 µL of absolute ethanol. After 3 min incubation at room temperature, the filaments of DNA were visible. They were picked up and washed with ethanol 75%. After drying at room temperature, the filaments were resuspended in

UltraPure™ DNase/RNase-Free Distilled Water (Thermo Fisher Scientific). The aqueous phase with RNA was mixed again with 200 µL Chloroform:isoamyl alcohol 24:1 (Sigma-Alrich), incubated 3 min at room temperature then centrifuged 15 min at 11,300 g. The top phase was collected, 500 µL isopropanol (Sigma-Aldrich) was added and the solution was centrifuged 15 min at 11,300 g. The supernatant was discarded and a series of washes, with ethanol 75%, were then performed. After drying at room temperature, the pellet was resuspended in UltraPure™ DNase/RNase-Free Distilled Water (Thermo Fisher Scientific).

DNAse of total RNA and reverse transcription

To eliminate potential gDNA contamination of extracted RNA the DNA-*free*™ DNA Removal Kit (Thermo Fisher Scientific) was used following the manufacturer's instructions. cDNA from 1 µg of total RNA was then obtained using the Transcriptor First Strand cDNA Synthesis Kit (Roche) and random hexamer primers. The manufacturer's instructions were followed.

Validation of accurate gDNA replacement and transcription

PCR

Several PCR amplifications were performed to validate correct DNA modifications (exon removal and replacement with the donor cassette) and efficient transcription. PCRs were performed on each DNA sample using the primer sequences detailed in Supplementary Table S1. The P1/P2, P1/P5, and P5/P4 PCRs were performed using the Q5® Hot Start High-Fidelity DNA Polymerase (New England Biolabs), an accurate enzyme for large amplicons. The P1/P3, P1/P4, P4/P2, P6/P7, and P8/P4 PCRs, amplifying smaller fragments, were performed using the GoTaq® G2 DNA Polymerase (Promega) (PCR programs in Supplementary Table S2). Every PCR was performed in a final volume of 25 µL with 100 ng gDNA. PCR products were resolved either on 1% or 1.5% agarose gel electrophoresis according to the size of expected amplicons. According to Jinek *et al.*, both sgRNAs should cut 3 bp before the PAM [29]. This assumption was made for this project to estimate all the edited DNA sequences sizes.

*Sma*I digestion

P1/P3 PCR products were digested with the FastDigest *Sma*I restriction enzyme (Thermo Fisher Scientific) following the manufacturer's instructions. A

negative control, without enzyme, was realized simultaneously. The negative control and the products of digestion were resolved on 1.5% agarose gel electrophoresis.

Sanger sequencing

P1/P2 PCR products were Sanger sequenced by Eurofins Genomics. Three primers were chosen (P1, P2, and P8). Sequenced were visualized using the FinchTV software (Geospiza).

Next-generation sequencing

The last exon of *Bcl-Rambo* was sequenced in the K-562 cell lines established during this project. Two PCR were performed on gDNA to generate a custom DNA library. The library design was developed by Shalem *et al.* to amplify modified gDNA after transduction of a CRISPR/Cas9 knock-out bank [30]. The first PCR added generic priming sites flanking the targeted sequence using a unique pair of primers (Forward: 5' TCTTGTGGAAAGGACGAAACACCGctgcaagaggcacttcctga 3', Reverse: 5' TCTACTATTCTTTCCCCTGCACTGTggcatgggcttcatctcacc 3', WT expected size: 408 bp). The KAPA HiFi PCR Kit (Kapa Biosystems) was used. PCR products were purified using the Nucleospin kit (Macherey Nagel). The second PCR added the Illumina primers and barcodes specific to each sample (WT F04/R01, Pool F12/R05, Clone #1 F12/R06, Clone #2 F12/R07, primer sequences in Supplementary Table S1). PCR products were purified using 0.65X AMPure XP Beads (Beckman Coulter). Size of PCR products, DNA concentration and purity were assessed by Tape Station electrophoresis (Agilent Technologies). High Sensitivity D1000 tapes were used. The mean sample size was 600 bp (range) and the mean DNA concentration was 85 ng/µL (range). A single-end sequencing (150 bp) has been performed on a MiSeq device (Illumina) using a MiSeq Reagent Micro kit v2 (Illumina) according to the manufacturer's instructions. Data was analyzed using the web-based platform Galaxy (http://usegalaxy.org). The DNA sequences (reads) were mapped to the human reference assembly GRCh37/*hg19* using Burrows-Wheeler Aligner (BWA, http://sourceforge.net/projects/bio-bwa/). SNPs and small INDELs were identified using Unified Genotyper. Variant annotation was performed using the Annotation, Visualization and Impact Analysis (AVIA) online resource (http://avia-abcc.ncifcrf.gov/) [31,32]. The results were visualized on the Integrative Genomics Viewer (IGV) software (Broad Institute).

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study within the paper and its Supplementary Information files are available from the corresponding author upon request.

REFERENCES

1. Mali, P. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science* **339**, 823–826 (2013).

2. Cong, L. *et al.* Multiplex Genome Engineering Using CRISPR/Cas Systems. *Science* **339**, 819–823 (2013).

3. D'Agostino, Y. & D'Aniello, S. Molecular basis, applications and challenges of CRISPR/Cas9: a continuously evolving tool for genome editing. *Briefings in Functional Genomics* elw038 (2017). doi:10.1093/bfgp/elw038

4. Lieber, M. R. The Mechanism of Double-Strand DNA Break Repair by the Nonhomologous DNA End-Joining Pathway. *Annual Review of Biochemistry* **79**, 181–211 (2010).

5. Hsu, P. D., Lander, E. S. & Zhang, F. Development and applications of CRISPR-Cas9 for genome engineering. *Cell* **157**, 1262–1278 (2014).

6. Salsman, J. & Dellaire, G. Precision genome editing in the CRISPR era. *Biochemistry and Cell Biology* **95**, 187–201 (2017).

7. Ran, F. A. *et al.* Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity. *Cell* **154**, 1380–1389 (2013).

8. Rong, Z., Zhu, S., Xu, Y. & Fu, X. Homologous recombination in human embryonic stem cells using CRISPR/Cas9 nickase and a long DNA donor template. *Protein & Cell* **5**, 258–260 (2014).

9. Kan, Y., Ruis, B., Takasugi, T. & Hendrickson, E. A. Mechanisms of precise genome editing using oligonucleotide donors. *Genome Res.* **27**, 1099–1111 (2017).

10. Hsu, P. D. *et al.* DNA targeting specificity of RNA-guided Cas9 nucleases. *Nature Biotechnology* **31**, 827–832 (2013).

11. Jensen, K. T. *et al.* Chromatin accessibility and guide sequence secondary structure affect CRISPR-Cas9 gene editing efficiency. *FEBS Letters* **591**, 1892–1901 (2017).

12. Zheng, Q. *et al.* Precise gene deletion and replacement using the CRISPR/Cas9 system in human cells. *BioTechniques* **57**, (2014).

13. Canver, M. C. *et al.* Characterization of Genomic Deletion Efficiency Mediated by Clustered Regularly Interspaced Palindromic Repeats (CRISPR)/Cas9 Nuclease System in Mammalian Cells. *Journal of Biological Chemistry* **289**, 21312–21324 (2014).

14. Kraft, K. *et al.* Deletions, Inversions, Duplications: Engineering of Structural Variants using CRISPR/Cas in Mice. *Cell Rep* (2015). doi:10.1016/j.celrep.2015.01.016

15. Boroviak, K., Fu, B., Yang, F., Doe, B. & Bradley, A. Revealing hidden complexities of genomic rearrangements generated with Cas9. *Sci Rep* **7**, 12867 (2017).

16. Cristea, S. *et al.* In vivo cleavage of transgene donors promotes nuclease-mediated targeted integration. *Biotechnology and Bioengineering* **110**, 871–880 (2013).

17. Nakade, S. *et al.* Microhomology-mediated end-joining-dependent integration of donor DNA in cells and animals using TALENs and CRISPR/Cas9. *Nature Communications* **5**, (2014).

18. Auer, T. O., Duroure, K., De Cian, A., Concordet, J.-P. & Del Bene, F. Highly efficient CRISPR/Cas9-mediated knock-in in zebrafish by homology-independent DNA repair. *Genome Research* **24**, 142–153 (2014).

19. Hisano, Y. *et al.* Precise in-frame integration of exogenous DNA mediated by CRISPR/Cas9 system in zebrafish. *Scientific Reports* **5**, (2015).

20. Li, J. *et al.* Intron targeting-mediated and endogenous gene integrity-maintaining knockin in zebrafish using the CRISPR/Cas9 system. *Cell Res.* **25**, 634–637 (2015).

21. Suzuki, K. *et al.* In vivo genome editing via CRISPR/Cas9 mediated homology-independent targeted integration. *Nature* **540**, 144–149 (2016).

22. Sakuma, T., Nakade, S., Sakane, Y., Suzuki, K.-I. T. & Yamamoto, T. MMEJ-assisted gene knock-in using TALENs and CRISPR-Cas9 with the PITCh systems. *Nat Protoc* **11**, 118–133 (2016).

23. Yao, X. *et al.* Homology-mediated end joining-based targeted integration using CRISPR/Cas9. *Cell Research* **27**, 801–814 (2017).

24. Suzuki, K. & Izpisua Belmonte, J. C. In vivo genome editing via the HITI method as a tool for gene therapy. *J. Hum. Genet.* **63**, 157–164 (2018).

25. Andersson, B. S. *et al.* Ph-positive chronic myeloid leukemia with near-haploid conversion in vivo and establishment of a continuously growing cell line with similar cytogenetic pattern. *Cancer Genetics and Cytogenetics* **24**, 335–343 (1987).

26. Pollard, M. O., Gurdasani, D., Mentzer, A. J., Porter, T. & Sandhu, M. S. Long reads: their purpose and place. *Human Molecular Genetics* **27**, R234–R241 (2018).

27. Eid, A., Alshareef, S. & Mahfouz, M. M. CRISPR base editors: genome editing without double-stranded breaks. *Biochem. J.* **475**, 1955–1964 (2018).

28. Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols* **8**, 2281–2308 (2013).

29. Jinek, M. *et al.* A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. *Science* **337**, 816–821 (2012).

30. Shalem, O. *et al.* Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science* **343**, 84–87 (2014).

31. Vuong, H., Stephens, R. M. & Volfovsky, N. AVIA: an interactive web-server for annotation, visualization and impact analysis of genomic variations. *Bioinformatics* **30**, 1013–1014 (2014).

32. Vuong, H. *et al.* AVIA v2.0: annotation, visualization and impact analysis of genomic variants and genes: Fig. 1. *Bioinformatics* **31**, 2748–2750 (2015).

ACKNOWLEDGMENTS

This work was supported by la Fondation ARC. We thank La region Nouvelle-Aquitaine and SIRIC BRIO for salary support for FL.

AUTHOR CONTRIBUTIONS

B.T., F.L., V.P-M. and M.L. conceived and designed experiments. F.L., E.B. and V.P-M. performed the experiments. B.T. and F.-X.M. supervised the work. All coauthors contributed to writing the paper.

COMPETING INTERESTS

The authors declare no conflict of interest.

SUPPLEMENTARY DATA

Sequence of the primers are listed in Supplementary Table 1. PCR programs are listed in Supplementary Table 2.

SUPPLEMENTARY INFORMATION

Precise and Efficient CRISPR/Cas9 Mediated Exon Replacement by Homology-Independent DNA Repair

Florence Lichou[1], Valérie Prouzet-Mauléon[1], Elodie Bierge[1], Matthieu Lewis[1], François-Xavier Mahon[1], Béatrice Turcq[1]*

[1]Laboratory of Mammary and Leukemic Oncogenesis, Inserm U1218 ACTION, University of Bordeaux, Bergonié Cancer Institute, Bordeaux, France.

Corresponding author: Béatrice Turcq. Laboratory of Mammary and Leukemic Oncogenesis, Inserm U1218 ACTION, University of Bordeaux, 146 rue Léo Saignat, bâtiment TP 4ème étage, case 50, 33076 Bordeaux, France. Tel: (+33) 5 57 57 10 51. E-mail address: beatrice.turcq@u-bordeaux.fr

Supplementary Table S1 – Primer list

Primer ID	Forward/ Reverse	Primer sequence
Replacement cassette-F	Forward	CCTAGAAACGAAGGACAAGCGTCagggcactgtgtttagtctt
Replacement cassette-R	Reverse	CCATTGGCTGAGCGCATTCAAGCGTgtgtgaggatccagaactgg
Mutagenesis-F	Forward	CTCCCTGCTGGGGACCCGGGAATCTGACACAGAAGTG
Mutagenesis-R	Reverse	CACTTCTGTGTCAGATTCCCGGGTCCCCAGCAGGGAG
U6 promoter	Forward	CAAGGCTGTTAGAGAGATAATTGGA
M13-F	Forward	GTAAAACGACGGCCAG
M13-R	Reverse	CAGGAAACAGCTATGAC
Sanger-Rambo-plasmid-1	Forward	GCACGTGGAGAAAGAAGAGG
Sanger-Rambo-plasmid-2	Forward	GACATTGCCAAATGTCCCGT
P1	Forward	GAGACTGCTCAGAAGACTGTC
P2	Reverse	AAGCAGCAAGTAGCCGCTTC
P3	Reverse	CATGCCACTGCCATGTTGAG
P4	Reverse	TTCTGTGTCAGATTCCCGGG
P5	Forward	GACCCGGGAATCTGACACAG
P6	Forward	AGGTGAAGAGCAAACTGGTG
P7	Reverse	TCTCAGACTGCCAGGAAGTG
P8	Forward	ACGGAGCTGGAGAGAAGAGT
NGS-F1	Forward	TCTTGTGGAAAGGACGAAACACCGctgcaagaggcacttcctga
NGS-R1	Reverse	TCTACTATTCTTTCCCCTGCACTGTggcatgggcttcatctcacc
F04	Forward	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTcgatCATGATCGtcttgtggaaaggacgaaacaccg
F12	Forward	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCTgatAGGTCGCAtcttgtggaaaggacgaaacaccg
R01	Reverse	CAAGCAGAAGACGGCATACGAGATAAGTAGAGGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtTCTACTATTCTTTCCCCTGCACTGT
R05	Reverse	CAAGCAGAAGACGGCATACGAGATCGTTACCAGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTtcgatTCTACTATTCTTTCCCCTGCACTGT
R06	Reverse	CAAGCAGAAGACGGCATACGAGATTCCTTGGTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTatcgatTCTACTATTCTTTCCCCTGCACTGT
R07	Reverse	CAAGCAGAAGACGGCATACGAGATAACGCATTGTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTgatcgatTCTACTATTCTTTCCCCTGCACTGT

Supplementary Table S2 – PCR programs

PCR 1 - Replacement cassette
KAPA HiFi PCR Kit (Kapa Biosystems)
PCR product: 2,930 bp

1X	95°C	3 min
35X	98°C	20 sec
	72°C	15 sec
	72°C	3 min 30
1X	72°C	3 min 30

PCR 2 - P1/P2
Q5® Hot Start High-Fidelity DNA Polymerase (New England BioLabs)
WT product size: 3,232 bp

1X	98°C	30 sec
35X	98°C	10 sec
	68°C	30 sec
	72°C	1 min 40
1X	72°C	2 min

PCR 3 - P1/P3
GoTaq® G2 DNA Polymerase (Promega)
WT product size: 1,197 bp

1X	95°C	2 min
35X	95°C	20 sec
	60°C	20 sec
	73°C	50 sec
1X	73°C	5 min

PCR 4 - P1/P4
GoTaq® G2 DNA Polymerase (Promega)
Modified product size: 702 bp

1X	95°C	2 min
35X	95°C	20 sec
	60°C	20 sec
	73°C	50 sec
1X	73°C	5 min

PCR 5 - P1/P5
Q5® Hot Start High-Fidelity DNA Polymerase (New England BioLabs)
Wrong direction cassette product size: 2,624 bp

1X	98°C	30 sec
35X	98°C	10 sec
	65°C	30 sec
	72°C	1 min 40
1X	72°C	2 min

PCR 6 - P4/P2
GoTaq® G2 DNA Polymerase (Promega)
Wrong direction cassette product size: 655 bp

1X	95°C	2 min
35X	95°C	20 sec
	60°C	20 sec
	73°C	50 sec
1X	73°C	5 min

PCR 8 - P6/P7
GoTaq® G2 DNA Polymerase (Promega)
Several cassettes product size: 283 bp

1X	95°C	2 min
35X	95°C	20 sec
	60°C	20 sec
	73°C	20 sec
1X	73°C	5 min

PCR 9 - P8/P4
GoTaq® G2 DNA Polymerase (Promega)
Modified product size: 215 bp

1X	95°C	2 min
35X	95°C	20 sec
	60°C	20 sec
	73°C	20 sec
1X	73°C	5 min

PCR 10 - NGS-1st PCR
KAPA HiFi PCR Kit (Kapa Biosystems)
WT PCR product size: 408 bp

1X	95°C	3 min
35X	98°C	20 sec
	65°C	15 sec
	72°C	1 min
1X	72°C	1 min

PCR 11 - NGS-2nd PCR
KAPA HiFi PCR Kit (Kapa Biosystems)
PCR product size range: 546-550 bp

1X	95°C	3 min
30X	98°C	20 sec
	55°C	15 sec
	72°C	1 min
1X	72°C	1 min