

Année : 2021

Thèse N° : 23/19

CENTRE DES ETUDES DOCTORALES DES SCIENCES  
DE LA VIE ET DE LA SANTE  
Formation Doctorale: Biologie Médicale, Pathologie Humaine  
et Expérimentale et Environnement

THESE DE DOCTORAT

**Etude moléculaire des maladies génétiques avec risque de  
cancer : La Xéroderma Pigmentosum et l'anémie de  
Fanconi comme exemples**

Présentée et soutenue publiquement le 07/10/2021

par

Yassamine DOUBAJ

Membres du jury :

**Pr. Taghzouti Mohammed Khalid**

*Faculté des sciences de Rabat ; Université Mohamed V- Rabat*

**Président**

**Pr. Sefiani Abdelaziz**

*Faculté de Médecine et de Pharmacie, Université Mohamed V - Rabat*

**Directeur de thèse**

**Pr. Ouldin Karim**

*Faculté de Médecine et de Pharmacie, Université Hassan II – Fès*

**Pr. Ratbi Ilham**

*Faculté de Médecine et de Pharmacie, Université Mohamed V- Rabat*

**Pr. Lyahyai Jaber**

*Faculté de Médecine et de Pharmacie, Université Mohamed V - Rabat*

**Pr El Kababri Maria**

*Faculté de Médecine et Pharmacie, Université Mohamed V- Rabat*

**Juges**

# *DEDICACES*

*Je remercie, de prime abord, Dieu Tout Puissant de m'avoir aidée et donné la force pour surmonter toutes les contraintes rencontrées tout au long de ce travail.*

*Et c'est avec une énorme joie que je dédie ce travail à toutes celles et tous ceux qui, par leur soutien et leur compréhension, ont permis sa réalisation.*

*A mes très chers parents, qui ont toujours été à mes côtés pour me soutenir et m'encourager, qui ont cru en moi et m'ont épaulée tout au long de mes études, et qui sans eux, je ne serais pas là où j'en suis aujourd'hui...*

*A mon cher mari Adil, qui m'a soutenue et encouragée durant ce travail, je te remercie pour ton aide et ta patience.*

*A mon cher fils Amir et mes chères filles Malak et Nourane qui m'ont aidée à leur façon, même s'ils ne comprennent pas pourquoi je continue à étudier. Je vous remercie pour votre soutien.*

*A mon cher et unique frère Yassine, je prie Dieu pour qu'il t'aide dans ta vie professionnelle et personnelle.*

*A mes chères amies Siham, Ilham et Imane, avec qui nous avons parcouru ensemble des moments agréables et en même temps très difficiles, je vous remercie pour votre soutien et vos encouragements.*

*A mes collègues et amis Aziza, AbdelHafid, Yahya, Latifa, Fatima, Fatima-Zahra, Amina, Hanae et à toute l'équipe du Département de Génétique Médicale - INH à Rabat, qui m'ont soutenue et encouragée, je vous remercie pour votre aide et votre soutien. Permettez-moi de vous exprimer mes remerciements les plus sincères.*

*A toute personne qui a contribué de loin ou de près à la réalisation de ce travail.*

*Aux familles des patients et aux malades. Rien n'aurait été possible sans leur contribution. Qu'ils en soient tous particulièrement remerciés. Merci d'avoir fait avancer la science.*

# REMERCIEMENTS

*J'exprime toute ma gratitude aux membres du jury qui ont bien voulu accepter de juger ce travail.*

***Au Professeur Mohammed Khalid TAGHZOUTI,***

*Vous me faites l'honneur de présider cette soutenance de Thèse. Votre disponibilité, votre gentillesse et votre soutien, m'ont été d'une grande aide. Soyez assuré de ma profonde gratitude et de ma plus grande estime.*

***Au Professeur Abdelaziz SEFIANI,***

*Vous avez accepté d'être le directeur de cette thèse et je vous en remercie. Je vous remercie pour vos conseils, pour votre modestie, votre disponibilité. Je ne saurai par les mots à exprimer mon admiration pour vos qualités humaines, votre compétence et votre rigueur scientifique.*

*J'espère que ce travail sera à la hauteur de vos espérances, et de la confiance que vous m'avez accordée, et dont je suis très honorée.*

***Au Professeur Karim OULDIM,***

*Je vous remercie de m'avoir fait l'honneur de participer à ce jury, afin d'y apporter votre connaissance et votre expérience. Merci pour votre disponibilité, votre soutien et votre confiance.*

*Veillez accepter, l'expression de ma sincère gratitude et ma haute considération.*

***Au Professeur Ilham RATBI,***

*Je vous remercie d'avoir accepté de participer à ce jury. C'est un grand honneur pour moi de vous avoir dans le jury. Veuillez accepter mes remerciements pour votre soutien, vos encouragements et votre disponibilité.*

***Au Professeur Jaber LYAHYAI,***

*Je vous remercie de m'avoir fait l'honneur de participer à ce jury. Et d'avoir accepté d'être rapporteur de cette thèse. Je vous prie de bien vouloir trouver dans ce travail l'assurance de mon profond respect et le témoignage de ma sincère reconnaissance.*

***Au Professeur Maria EL KABABRI***

*Toute ma gratitude pour votre soutien, votre efficacité et votre compétence. Veuillez trouver, dans ce travail, l'expression de mon respect et de mon admiration pour votre modestie, votre gentillesse, et vos grandes qualités humaines.*

# Résumé

**Introduction :** De nombreuses maladies génétiques, notamment celles dues à des anomalies de réparation de l'ADN, sont associées à un risque accru de cancer. L'étude de ces syndromes a considérablement évolué et a conduit à l'identification de gènes impliqués dans la carcinogenèse. Dans ce travail, nous nous sommes intéressés à deux pathologies: La Xérodérma Pigmentosum (XP) et l'Anémie de Fanconi (AF).

**Matériels et méthodes :** Pour la famille avec XP, nous avons appliqué le Séquençage de Sanger pour la recherche de la mutation récurrente c.1643\_1644delTG au niveau de l'exon 9 du gène *XPC*. Concernant les familles atteintes de l'AF, nous avons opté pour le Séquençage haut débit pour analyser les trois gènes les plus fréquemment mutés *FANCA*, *FANCC*, et *FANCG* et qui représentent 84% de tous les gènes impliqués dans l'AF.

**Résultats :** Le séquençage ciblé de l'exon 9 du gène *XPC* a révélé la présence d'une nouvelle mutation à l'état homozygote de type insertion de 4 bases chez le propositus. Les parents ont été testés et se sont révélés porteurs de cette mutation à l'état hétérozygote. L'analyse moléculaire effectuée chez neuf patients atteints d'AF et présentant tous une grande instabilité chromosomique sous mitomycine, a confirmé le diagnostic dans 88.8% des cas avec trois nouvelles mutations prédites comme pathogènes. Des variations du gène *FANCA* et du gène *FANCG* ont été retrouvées respectivement dans 62.5% et 37.5% des cas. .

**Conclusion :** Grâce aux techniques de séquençage, le diagnostic moléculaire est actuellement possible même pour les maladies génétiques avec une grande hétérogénéité génétique. Le choix du test moléculaire dépend du profil génétique de la pathologie. L'identification de la mutation causale permet d'adapter la prise en charge des malades et de prodiguer aux familles un conseil génétique adéquat.

# Abstract

**Introduction:** Many genetic diseases, especially those caused by abnormalities in DNA repair, are associated with an increased risk of cancer. The study of these syndromes predisposing to cancer has evolved considerably and has led to the identification of genes involved in carcinogenesis. In this work, we were interested in two pathologies which are due to a DNA repair defect: Xeroderma Pigmentosum (XP) and Fanconi anemia (FA).

**Materials and methods:** We applied Sanger sequencing for the family with XP to search for the recurrent mutation c.1643\_1644delTG in the exon 9 of the *XPC* gene. Concerning the FA families, we opted for the Next Generation Sequencing (Ion PGM) to analyse the three most frequently mutated genes, *FANCA*, *FANCC*, and *FANCG*, which represent 84% of all genes involved in FA.

**Results:** The targeted sequencing of the exon 9 of the *XPC* gene revealed the presence of a new homozygous insertion of 4 bases in the proband. The parents have been tested and found to be heterozygous carrier of the same mutation. Molecular analysis of nine patients with FA, sharing all a high chromosomal instability to mitomycin, confirmed the diagnosis in 88.8% of cases with three new mutations predicted as pathogenic. Mutations in the *FANCA* and the *FANCG* genes were found in 62.5% and 37.5% of cases respectively.

**Conclusion:** Thanks to sequencing technologies, molecular diagnosis is currently possible even for genetic diseases with genetic heterogeneity. The choice of molecular test depends on the pathology's genetic profile. The identification of the causal mutation allows to adapt the management of patients and to provide an adequate genetic counselling to families.

## ملخص

مقدمة: ترتبط العديد من الأمراض الوراثية، خاصة تلك الناتجة عن خلل في إصلاح الحمض النووي، بزيادة خطر الإصابة بالسرطان. تطورت دراسة هذه المتلازمات المؤدية للسرطان بشكل كبير وأدت إلى تحديد الجينات المشاركة في التسرطن.

في هذا العمل، قمنا بدراسة مرضين ناتجين عن عيب في إصلاح الحمض النووي: جفاف الجلد المصطبغ (Xeroderma Pigmentosum (XP) وفقر الدم فانكوني (Fanconi (AF).

المواد والطرق: طبقنا تسلسل Sanger للعائلة المصابة بجفاف الجلد المصطبغ XP للبحث عن طفرة c.1643\_1644delTG المتكررة على مستوى exon 9 من مورثة XPC. فيما يتعلق بالعائلات المصابة بفقر الدم فانكوني AF، اخترنا التسلسل عالي الإنتاجية (Ion PGM) لتحليل المورثات الثلاثة الأكثر شيوعاً: FANCA و FANCC و FANCG، والتي تمثل 84% من جميع الجينات المسؤولة عن مرض AF.

النتائج: كشف التسلسل المستهدف لـ exon 9 من المورثة XPC عن وجود طفرة جديدة متماثلة الحلائل من نوع الإدراج من 4 قواعد. تم اختبار الوالدين وتبين أنهما حاملان مختلفا الحلائل لنفس الطفرة. تم إجراء التحليل الجزيئي على تسعة مرضى مصابين بفقر الدم، يظهر عليهم عدم استقرار كروموزومي مرتفع في وجود الميتومييسين، الذي أكد التشخيص في 88.8% من الحالات مع وجود ثلاث طفرات جديدة متوقعة كمسببات للمرض. تم العثور على طفرات في مورثة FANCA و في مورثة FANCG في 62.5% و 37.5% من الحالات على التوالي.

الخلاصة: بفضل تقنيات التسلسل الجزيئي، أصبح التشخيص الجزيئي حالياً ممكناً حتى بالنسبة للأمراض الوراثية ذات التباين الجيني الكبير. يعتمد اختيار الاختبار الجزيئي على المظهر الوراثي للمرض. إن تحديد الطفرة المسببة يسهل عملية التكيف بالمرضى وتزويد العائلات بالاستشارة الوراثية المناسبة.



# SOMMAIRE

<i>Liste des abréviations</i> .....	12
<i>Liste des figures</i> .....	14
<i>Liste des tableaux</i> .....	15
<b>I. INTRODUCTION</b> .....	16
<b>II. MATERIELS METHODES</b> .....	21
1. <i>Patients</i> : .....	22
1.1. <i>Famille avec Xéroderma Pigmentosum</i> :.....	22
1.2. <i>Familles avec suspicion d'Anémie de Fanconi</i> :.....	24
1.2.1. <i>Données cliniques et paracliniques</i> : .....	24
1.2.2. <i>Etude cytogénétique : Caryotype sous mitomycine</i> .....	30
2. <i>Méthodes</i> :.....	34
2.1. <i>Recueil des échantillons pour analyse moléculaire</i> :.....	34
2.2. <i>Extraction par Kit des acides nucléiques à partir du sang total</i> :.....	34
2.3. <i>Dosage et contrôle de la qualité de l'ADN extrait</i> : .....	36
2.3.1. <i>Dosage spectrophotométrique au Nanodrop</i> : .....	36
2.3.2. <i>Dosage fluorométrique par Qubit</i> :.....	37
2.4. <i>Technique de la PCR</i> :.....	37
2.4.1. <i>Détermination de la séquence à amplifier</i> :.....	37
2.4.2. <i>Design des amorces</i> :.....	38
2.4.3. <i>La réaction PCR et le contrôle sur gel</i> : .....	39
2.5. <i>Séquençage par la méthode Sanger pour la famille avec XP</i> : .....	42
2.5.1. <i>Purification du produit de PCR</i> : .....	42
2.5.2. <i>Réaction de Séquence</i> : .....	42
2.5.3. <i>Purification de la réaction de séquence par Sephadex G50</i> : .....	44
2.5.4. <i>Séquençage Sanger</i> :.....	44
2.6. <i>Le séquençage à haut débit par panel de gènes pour les familles avec AF</i> :.....	47
2.6.1. <i>Principe de la technologie PGM Ion Torrent</i> :.....	48
2.6.2. <i>Le protocole de séquençage Ion PGM</i> : .....	48
2.6.2.1. <i>Design de Panel</i> :.....	48

2.6.2.2.	<i>Séquençage PGM proprement dit : Le séquençage Haut Débit sur la plateforme Ion Torrent (ThermoFisherScientific).....</i>	<i>52</i>
2.6.2.3.	<i>Analyse Bio-informatique : .....</i>	<i>62</i>
<b>III.</b>	<b>RESULTATS .....</b>	<b>67</b>
1.	<i>Résultat de Famille avec XP :.....</i>	<i>68</i>
2.	<i>Résultats des familles avec AF : .....</i>	<i>70</i>
<b>IV.</b>	<b>DISCUSSION .....</b>	<b>74</b>
1.	<i>Techniques de séquençage de l'ADN :.....</i>	<i>75</i>
1.1.	<i>Histoire du séquençage de l'ADN : .....</i>	<i>75</i>
1.2.	<i>Séquençage 1ère génération : Séquençage Sanger .....</i>	<i>78</i>
1.3.	<i>Séquençage deuxième génération (Next Generation Sequencing) : .....</i>	<i>80</i>
2.	<i>Applications du séquençage en Génétique médicale : .....</i>	<i>94</i>
2.1.	<i>Les différents niveaux de séquençage : .....</i>	<i>94</i>
2.1.1.	<i>Le séquençage Sanger.....</i>	<i>95</i>
2.1.2.	<i>Panel de gènes d'intérêt : .....</i>	<i>95</i>
2.1.3.	<i>Séquençage de l'Exome clinique (large panel) : .....</i>	<i>96</i>
2.1.4.	<i>Le séquençage de l'exome entier (WES/Whole Exome Sequencing): .....</i>	<i>96</i>
2.1.5.	<i>Le séquençage du génome complet (WGS/Whole Genome Sequencing) :.....</i>	<i>97</i>
2.2.	<i>Choix du test de séquençage : .....</i>	<i>97</i>
2.3.	<i>Bioinformatique et interprétation des variants dans les plateformes NGS : .....</i>	<i>99</i>
2.4.	<i>Limites et difficultés du NGS : .....</i>	<i>103</i>
2.5.	<i>Considérations éthiques :.....</i>	<i>104</i>
3.	<i>Exemples de maladies génétiques avec risque de cancer : .....</i>	<i>107</i>
3.1.	<i>La Xéoderma Pigmentosum : .....</i>	<i>107</i>
3.1.1.	<i>Diagnostic clinique et paraclinique : .....</i>	<i>107</i>
3.1.2.	<i>La génétique : .....</i>	<i>115</i>
3.1.2.1.	<i>Les gènes de la XP.....</i>	<i>117</i>
3.1.2.2.	<i>Diagnostic moléculaire : .....</i>	<i>120</i>
3.1.2.3.	<i>Corrélation génotype-phénotype : .....</i>	<i>124</i>
3.1.2.4.	<i>Pathologies alléliques: .....</i>	<i>124</i>
3.1.2.5.	<i>Gènes XP et la voie de réparation NER : .....</i>	<i>126</i>
3.1.3.	<i>Prise en charge .....</i>	<i>129</i>

<b>3.2. L'Anémie de Fanconi :</b> .....	<b>129</b>
<b>3.2.1. Diagnostic clinique:</b> .....	<b>130</b>
<b>3.2.2. Etude cytogénétique :</b> .....	<b>133</b>
<b>3.2.3. La Génétique de l'AF</b> .....	<b>134</b>
<b>3.2.3.1. Les gènes de l'AF</b> .....	<b>134</b>
<b>3.2.3.2. Corrélation génotype-phénotype</b> .....	<b>140</b>
<b>3.2.3.3. Gènes FANC et la voie FA-BRCA</b> .....	<b>141</b>
<b>3.2.3.4. Prise en charge :</b> .....	<b>146</b>
<b>3.2.3.5. Conseil génétique :</b> .....	<b>146</b>
<b>Conclusion et perspectives</b> .....	<b>147</b>
<b>Références</b> .....	<b>147</b>

# Liste des abréviations

<b>ABI</b>	: Applied BioSystems
<b>AF</b>	: Anémie de Fanconi
<b>ADN</b>	: Acide DésoxyriboNucléique
<b>ARN</b>	: Acide Ribonucléique
<b>BET</b>	: Bromure d'éthidium
<b>BOM</b>	: Biopsie ostéo-médullaire
<b>CES</b>	: Clinical Exome
<b>ddNTP</b>	: Didésoxyribonucléotide tri-phosphate
<b>dNTP</b>	: Désoxyribonucléotide tri-phosphate
<b>DDR</b>	: DNA Damage Response
<b>DS</b>	: Deviation standard
<b>GNOMAD</b>	: genome aggregation database
<b>HGMD</b>	: Human Gene Mutation Database
<b>HR</b>	: Homologous Recombination
<b>ICL</b>	: Interstrand Crosslink
<b>Kb</b>	: Kilobase
<b>LCH</b>	: Luxation congénitale de la hanche
<b>MAR</b>	: Malformation ano-rectale
<b>MMR</b>	: Mismatch Repair

<b>NHEJ</b>	: Non Homologous End Joining
<b>NER</b>	: Nucleotide Excision Repair
<b>NGS</b>	: Next Generation Sequencing
<b>OMIM</b>	: Online Mendelian Inheritance in Man
<b>pb</b>	: Paires de bases
<b>PCR</b>	: Polymerase Chain Reaction
<b>RSP</b>	: Retard staturo-pondérale
<b>SAP</b>	: Shrimp phosphatase alcaline
<b>SBH</b>	: Sequencing based Hybridation
<b>SBS</b>	: Sequencing based synthesis
<b>SHD</b>	: Séquençage à Haut Débit
<b>TE</b>	: Tampon d'élution
<b>TTD</b>	: Trichothiodystrophie
<b>UDS</b>	: Unscheduled DNA synthesis
<b>VCF</b>	: Variant Call File
<b>WES</b>	: Whole Exome Sequencing
<b>WGS</b>	: Whole Genome Sequencing
<b>XP</b>	: Xeroderma Pigmentosum

# Liste des figures

Figure 1: Arbre généalogique de la famille.....	23
Figure 2: Arbres généalogiques des neuf familles avec AF.....	28
Figure 3: Images de chromosomes métaphasiques montrant des cassures et des images triradiales ....	33
Figure 4: Mélange réactionnel et différents cycles de la PCR .....	41
Figure 5: Principe de la réaction de séquence .....	43
Figure 6: Principe de séquençage automatisé d'ADN par la méthode de Sanger .....	46
Figure 7: Panel conçu par l'équipe de DGM-INH .....	51
Figure 8: Spécificité des primers Ion Torrent.....	51
Figure 9: Etapes de la préparation de la librairie (R. BLERVAQUE et al. 2013).....	56
Figure 10: Etapes de la PCR à émulsion (R. BLERVAQUE et al. 2013) .....	58
Figure 11: Etape de l'enrichissement (R. BLERVAQUE et al. 2013) .....	60
Figure 12: Le Séquençage (R. BLERVAQUE et al. 2013).....	61
Figure 13: Analyse des données brutes .....	64
Figure 14: Electrophorégramme de l'exon 9 montrant une insertion CATG à l'état homozygote chez le propositus et à l'état hétérozygote chez le père.....	69
Figure 15: Evolution des capacités de séquençage.....	77
Figure 16: Evolution du coût du séquençage [15].....	77
Figure 17: Principe du séquençage Sanger [19] .....	79
Figure 18: Principe général du séquençage par hybridation (SBH) [19].....	81
Figure 19: Principe du pyroséquençage [19].....	85
Figure 20: Principe du séquençage par Illumina/Solexa (1).....	88
Figure 21 : Principe du séquençage à semi-conducteurs [38].....	91
Figure 22: Photo d'un enfant atteint de XP montrant les lésions cutanées ( <a href="https://dermatomaroc.com/">https://dermatomaroc.com/</a> ) .....	109
Figure 23: Diagnostic différentiel de la XP. A: patiente avec TTD âgée de 3 ans présentant des cheveux cassants [86]. B: patiente atteinte de syndrome de Cockayne avec un aspect cachectique, des yeux enfoncés et des lésions cutanées. C: enfant avec dysmorphie fac .....	114
Figure 24: Les mutations décrites au niveau de l'exon 9 .....	123
Figure 25: Pathologies dues à un défaut de réparation de l'ADN .....	125
Figure 26: Voie de réparation Nucleotide excision repair (NER) .....	128
Figure 27: Distribution géographique des mutations communes du gène FANCA en Afrique du Nord (Modifié à partir de Ben Haj Ali A. et al ;2021). .....	138
Figure 28: Principaux processus cellulaires affectés dans l'anémie de Fanconi. ....	142
Figure 29: Voie de signalisation FA. Les protéines FANC.....	145

# Liste des tableaux

Tableau 1: Caractéristiques cliniques et biologiques des patients avec AF .....	29
Tableau 2: Résultats cytogénétiques avec le nombre des cassures chromosomiques et des formes radiales des patients et des témoins normaux.....	32
Tableau 3: Mutations des gènes FANC identifiées par NGS chez huit patients.....	73
Tableau 4: Bases de données publiques utilisées dans l'interprétation des données de NGS	102
Tableau 5: Les groupes de complémentation XP et les gènes de la XP .....	116
Tableau 6: Groupes de complémentation et gènes FANC responsables de l'AF .....	135

# **I.INTRODUCTION**



Depuis plus de 40 ans, le séquençage de l'ADN est une technique indispensable et commune à tous les laboratoires de biologie moléculaire que ce soit en médecine ou dans d'autres disciplines des sciences de la vie. La première méthode utilisée pour déterminer l'ordre d'enchaînement des bases au niveau des acides nucléiques fut décrite par Frederick Sanger en 1977 [1]. Il s'agit d'une méthode basée sur la synthèse enzymatique dite par « terminaison de chaîne » qui a permis le séquençage de divers génomes dont celui de l'être humain. Toutefois, cette méthode, couteuse en termes de temps et d'argent, n'était pas adaptée au séquençage de grands génomes.

Au fil des années, les innovations dans les outils de biologie moléculaire et l'automatisation des technologies de séquençage ont augmenté les capacités du séquençage tout en diminuant le coût, passant de la lecture de plusieurs centaines de paires de bases d'ADN à une production massive de gigabases de données en un seul run. De nouvelles technologies appelées « séquençage haut débit » ou « de nouvelle génération » ont été développées, révolutionnant ainsi la génomique par le séquençage du génome humain en quelques jours pour moins de 1000 dollars.

La vitesse d'évolution des plateformes, la diminution du coût du séquençage, ainsi qu'une bonne compréhension de l'impact clinique des informations émergentes sur la prise en charge des patients, a conduit à une explosion de plusieurs applications en matière de soins de santé. L'évolution des connaissances sur les gènes et les maladies a changé la pratique actuelle de la médecine et de la santé publique. Une nouvelle entité dite la « médecine génomique » a vu le jour, dans laquelle l'analyse globale et approfondie des informations génomiques sera utilisée à la fois pour prédire, prévenir, diagnostiquer et traiter de nombreuses maladies chroniques et rares courantes.

Actuellement, on peut distinguer plusieurs applications aux nouvelles technologies que ce soit en génétique clinique, en oncologie ou en pharmacogénétique. En plus du séquençage complet d'un panel de gènes candidats, de l'exome entier, ou même du génome entier, on peut analyser, avec ces mêmes technologies, le profil d'expression génique ainsi que les mécanismes épigénétiques.

En conséquence, le mécanisme physiopathologique des syndromes génétiques et des cancers est aujourd'hui mieux élucidé. Longtemps considérées comme des maladies distinctes, il est actuellement admis qu'elles peuvent partager des altérations génétiques causales dans un groupe de troubles du développement associés à des malformations congénitales multiples et un risque plus élevé de cancer. L'étude de ces syndromes avec prédisposition au cancer a conduit à l'identification de gènes impliqués dans la carcinogenèse, dans la régulation de la prolifération et la différenciation des cellules normales. Néanmoins, l'action de ces gènes sur les voies de signalisation et la conséquence de leurs défauts sur le plan phénotypique (anomalies congénitales) ainsi que les altérations sous-jacentes au développement du processus néoplasique ne sont pas, à ce jour, totalement bien clarifiées.

Plusieurs variables peuvent influencer la survenue de tumeurs dans les syndromes génétiques du développement notamment la voie moléculaire impliquée (différenciation, prolifération, mort cellulaire, etc.), le gène affecté (oncogène, suppresseur de tumeur ou gène de réparation de l'ADN), le type d'altération (génétique, génomique, ou épigénétique), et l'origine cellulaire de ces changements (germinale ou somatique). Ces facteurs expliqueraient également le fait que la majorité des anomalies du développement avec risque de cancer peuvent être réparties en groupes avec des manifestations cliniques chevauchantes, des voies moléculaires communes et des mécanismes pathogènes similaires/identiques [2].

La progression d'une cellule d'un état normal à un état tumoral se produit par une série d'étapes de modification du gène qui vont échapper aux mécanismes de maintien de la stabilité génomique et vont aboutir à une instabilité génomique. En effet, l'instabilité génomique est l'une des caractéristiques du cancer, reflétant la voie d'où les cellules cancéreuses sont issues [3]. Dans les cellules normales, l'intégrité génomique est protégée par la réponse aux dommages à l'ADN (DNA damage response : DDR), qui décrit l'ensemble des voies capables de détecter et de réparer différents types des dommages à l'ADN (environnemental, erreurs de réplication de l'ADN, etc.). Les défauts du DDR et de ses voies de réparation, acquises ou héréditaires, peuvent accélérer l'accumulation de mutations et aider à augmenter la perte de stabilité qui entraîne finalement la malignité.

La cellule possède un système complexe de voies de réparation de l'ADN interconnectées, qui lui permet de répondre aux différents types de dommages et de les réparer. Il existe au moins 5 voies de réparation qui sont actives à différentes étapes du cycle cellulaire : la réparation par excision de base (base excision repair (BER)), la réparation par excision de nucléotide (nucleotide excision repair (NER)), la réparation de mésappariement (mismatch repair : MMR), la recombinaison homologue (homologous recombination : HR) et la jonction des extrémités non homologues (non-homologous end joining : NHEJ). Quelques lésions spécifiques peuvent également être éliminées par inversion chimique directe et par la réparation du pontage inter-brin (Interstrand crosslink : ICL). Ces processus de réparation sont essentiels au maintien de la stabilité génétique des cellules.

Nombreux sont les troubles du développement liés à des gènes impliqués dans les voies de réparation notamment l'ataxie télangiectasie, le syndrome de Bloom, le syndrome de Nijmegen, le syndrome de Werner, etc.

Le but de ce travail est d'établir une stratégie de diagnostic moléculaire de certains syndromes génétiques avec risque accru de cancer. Nous nous sommes intéressés à deux maladies dues à un défaut de réparation de l'ADN et caractérisées par une grande hétérogénéité génétique avec plusieurs groupes de complémentations et plusieurs gènes. Nous avons utilisé deux approches diagnostiques différentes pour la confirmation moléculaire de ces deux pathologies.

Pour la famille avec Xéoderma Pigmentosum (XP), nous avons réalisé un séquençage ciblé par la technique Sanger à la recherche de la mutation récurrente avec effet fondateur qui est responsable de 74% de tous les XP. Il s'agit de la mutation c.1643\_1644delTG (p.Val548AlafsX25) au niveau de l'exon 9 du gène *XPC* qui est le principal gène causant la maladie en Afrique du Nord.

Concernant les 9 familles avec suspicion de l'Anémie de Fanconi (AF), nous avons opté pour le séquençage haut débit en utilisant un panel qui comprend les 3 gènes les plus fréquemment mutés dans l'AF et qui représentent environ 84% des cas (*FANCA*, *FANCC* et *FANCG*).

A travers ces observations, nous voulons montrer l'intérêt, l'indication et les avantages de chaque approche de diagnostic moléculaire pour les maladies génétiques avec risque de cancer.

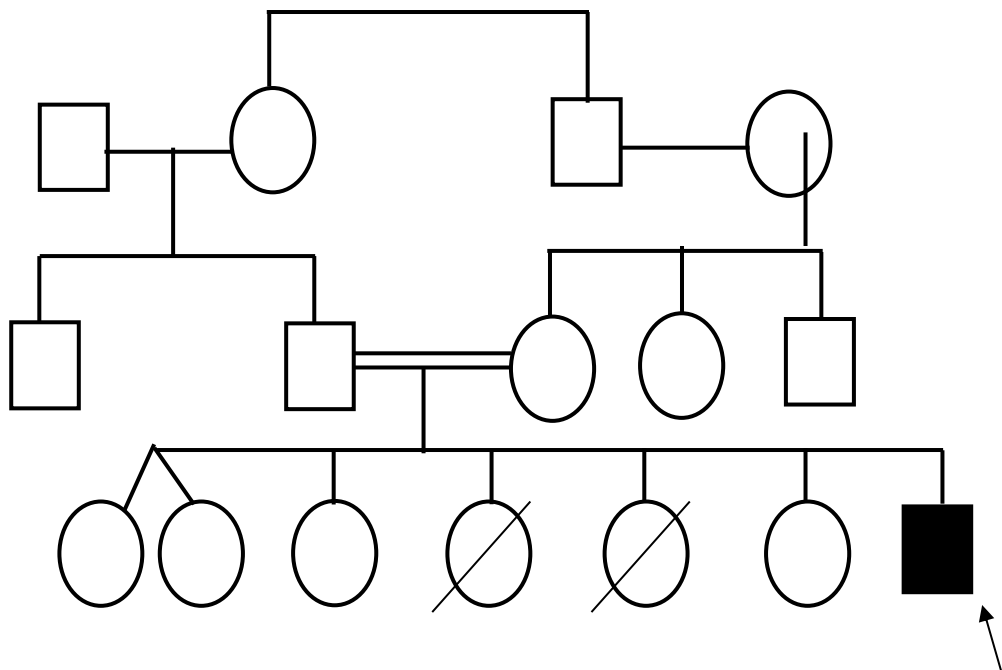
## **II. MATERIELS METHODES**

## **1. Patients :**

### **1.1.Famille avec Xéroderma Pigmentosum :**

Le probant est un jeune homme de 21 ans, le dernier d'une fratrie de 7, issu d'un mariage consanguin sans antécédents familiaux. Il a été adressé par une association au Département de Génétique Médicale à l'Institut National d'Hygiène pour prise en charge. La grossesse et l'accouchement étaient normaux. Il a eu un développement moteur et mental normal. Le début de la symptomatologie remontait à l'âge de 9 mois par l'apparition de troubles de pigmentation dans les zones photo-exposées, en particulier sur le visage. L'évolution a été marquée par la survenue de lésions oculaires bilatérales pour lesquelles le patient fut opéré à deux reprises à l'âge de 15 ans (pas de compte rendu opératoire ni anatomopathologique).

À l'examen clinique, il présentait une photophobie et des taches de rousseur au niveau du visage, du cou et des membres. L'examen ophtalmologique avait objectivé une atteinte oculaire bilatérale en rapport avec des lésions inflammatoires conjonctivales de la cornée. Par ailleurs, le patient ne présentait pas de lésions précancéreuses ni de signes neurologiques.



*Figure 1: Arbre généalogique de la famille*

## **1.2. Familles avec suspicion d'Anémie de Fanconi :**

En collaboration avec l'équipe du service CHOP à l'Hôpital d'Enfants de Rabat, nous avons fait une étude génétique chez les enfants atteints d'une anémie de Fanconi. Il s'agit d'une étude prospective s'étalant sur la période de 2017 à 2019. L'étude regroupe 9 patients issus de 9 familles différentes ayant un tableau évocateur de l'anémie de Fanconi.

### **1.2.1. Données cliniques et paracliniques :**

#### **1<sup>er</sup> patient :**

Fille âgée de 5 ans, consanguine. Le développement psychomoteur était normal. Le début de la symptomatologie remonte à l'âge de 3 ans par une stagnation staturo-pondérale.

A l'examen, elle présente un retard staturo-pondéral entre -3 et -4DS. Le reste de l'examen est sans particularités en dehors d'un pertuis au niveau de la région sacrée. Sur le plan biologique, on note une pancytopenie avec présence d'une aplasie médullaire à la biopsie ostéo-médullaire (BOM). L'échographie abdomino-pelvienne a objectivé une agénésie rénale.

#### **2<sup>ème</sup> Patient :**

Fille âgée de 6 ans, non consanguine mais les parents sont originaires de la même région. Elle est suivie pour malformation ano-rectale (MAR) de type anus vulvaire et une luxation congénitale bilatérale de la hanche (LCH).

A l'examen, elle présente un visage triangulaire, un retard staturo-pondéral entre -3 et -4DS et des anomalies squelettiques avec un pouce gauche bas implanté et une hypoplasie de l'hémithorax gauche. La numération formule sanguine a révélé une bicytopenie (Taux de plaquettes et d'hémoglobine diminués). La patiente a un rein gauche ectopique à l'échographie abdomino-pelvienne.



**3<sup>ème</sup> patient :**

Fille consanguine âgée de 5 ans, ayant eu un développement psychomoteur normal. Selon la maman, elle a présenté une cassure de la courbe pondérale et un bilan a été fait à l'âge de 5 ans qui a objectivé une thrombopénie.

A l'examen, on note un retard staturo-pondéral à -2DS, une dysmorphie faciale avec un visage triangulaire et une micrognathie associée à des anomalies cutanées (deux taches café au lait au niveau du tronc et de la région fessière). Le bilan biologique a montré une anémie et une thrombopénie. Une biopsie ostéo-médullaire a été faite montrant une moelle hypoplasique sans présence de lésion tumorale.

**4<sup>ème</sup> patient :**

Fille non consanguine âgée de 11 ans, issue d'une grossesse suivie avec un accouchement prématuré par césarienne à 8 mois et un poids de naissance de 1kg900. Une béance du cardia a été diagnostiquée à l'âge de 1 an avec notion d'otites à répétition. A l'âge de 11 ans, découverte d'une neutropénie avec une anémie lors d'une hospitalisation pour angine pultacée.

A l'examen, on trouve un retard staturo-pondéral à -2DS, des cheveux bas implantés au niveau du cou et du front, une polydactylie préaxiale gauche au niveau du membre supérieur avec un défaut de supination à droite. L'examen cutanéomuqueux révèle une tache café au lait au niveau du pli de l'aîne. Le bilan radiologique a objectivé un retard de l'âge osseux estimé à 8 ans et 10 mois pour un âge chronologique de 11 ans et une synostose radio-cubitale.

**5<sup>ème</sup> patient :**

Fille consanguine âgée de 5 ans avec un bon développement psychomoteur. L'enfant fut hospitalisé pour un épistaxis à l'âge de 5 ans, le bilan biologique réalisé a objectivé une bicytopenie (une anémie associée à une neutropénie).

A l'examen, elle ne présente pas de dysmorphie faciale avec un poids à 16 kg (-1,-2DS) et une taille à 100cm (-1,-2DS). On note également une polydactylie préaxiale gauche et une tache café au lait au niveau du pli de l'aîne.

**6<sup>ème</sup> patient :**

Garçon non consanguin, âgé de 8 ans ayant un bon développement psychomoteur. Le début de la symptomatologie remonte à l'âge de 5 ans par l'apparition de gingivorragies et d'hématomes multiples au niveau des 2 jambes pour lesquels l'enfant fut hospitalisé. Le bilan biologique a objectivé une pancytopenie. L'enfant fut hospitalisé 1 an après pour des hématomes au niveau du cerveau.

A l'examen, il a un poids à 26 kg (-1DS), une taille à 123cm (-1DS), il ne présente pas de dysmorphie faciale ni d'anomalies des membres. L'examen cutanéomuqueux retrouve une hémorragie conjonctivale bilatérale et des taches café au lait au niveau membre supérieur droit et une tache achromique au niveau du dos (2cm de diamètre). L'examen génital trouve un hypospadias.

La BOM a objectivé une hypoplasie médullaire.

**7<sup>ème</sup> patient :**

Fille consanguine âgée de 4 ans, issue d'une grossesse suivie avec un AVB médicalisé. Notion d'un retard de croissance intra-utérin (PN: 1kg200) ; elle fut mise dans une couveuse

pendant 10j. Découverte d'une bicytopenie avec thrombopenie (Plq à 45000/mm<sup>3</sup>) et une anémie (Hb à 10,7 g/dl).

A l'examen, elle présente un retard staturo-pondéral -3DS, un faciès triangulaire et un pouce droit hypoplasique bas implanté avec absence de signes cutanés.

**8<sup>ème</sup> patient :**

Garçon consanguin âgé de 7 ans. A l'âge de 4 mois, il a présenté une anémie nécessitant des transfusions à répétition. A l'examen, il présente un visage triangulaire avec un retard staturo-pondéral entre -2 et -3DS. La numération formule sanguine a révélé une anémie et une thrombopenie.

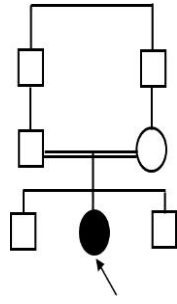
**9<sup>ème</sup> patient :**

Garçon consanguin âgé de 4 ans et demi. Le développement psychomoteur fut normal. Découverte d'une anémie à l'âge de 1 an. Adressé au Département pour hypoplasie médullaire. A l'examen, l'enfant n'a pas de retard staturo-pondéral, il présente une dysmorphie faciale avec un visage triangulaire et des oreilles légèrement décollées, un pouce droit hypoplasique. L'examen cutanéomuqueux révèle la présence de quatre taches hyperchromiques au niveau du tronc dont la plus grande est de 2cm de diamètre et de deux taches hypochromiques.

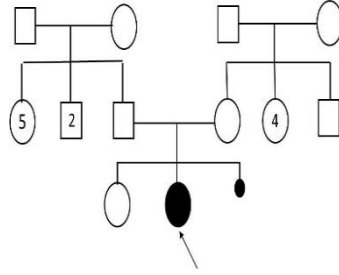
La BOM a été réalisé et a révélé une hypoplasie médullaire.

La figure 2 montre les arbres généalogiques des neuf familles avec AF.

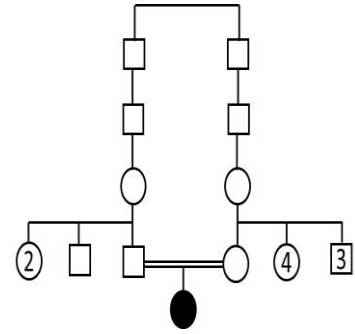
Le tableau 1 résume l'ensemble des données cliniques et paracliniques des patients atteints d'AF.



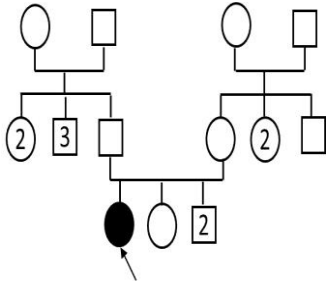
**Famille 1**



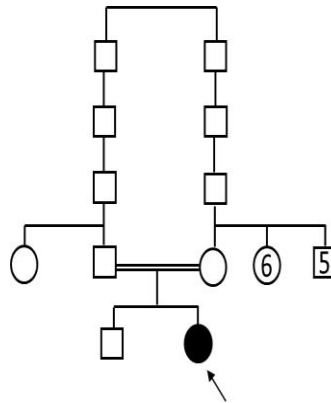
**Famille 2**



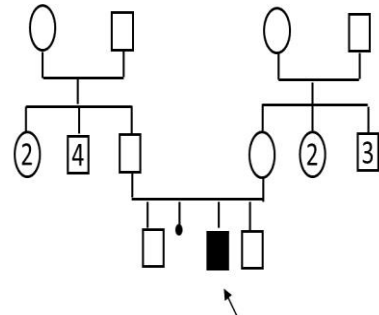
**Famille 3**



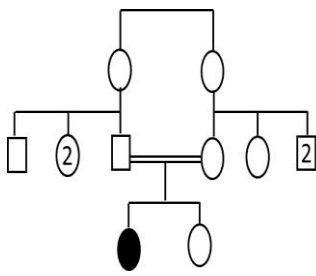
**Famille 4**



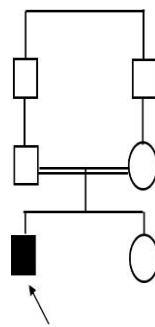
**Famille 5**



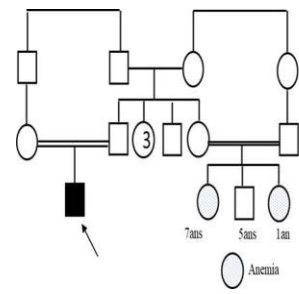
**Famille 6**



**Famille 7**



**Famille 8**



**Famille 9**

**Figure 2: Arbres généalogiques des neuf familles avec AF**

Patients	Age (ans)	Sexe	Consanguinité parentale	RSP	DF	AS	SC	Autres	Hémogramme	Biopsie médullaire
P1	5	F	+	+	-	-	-	Agénésie rénale	Pancytopénie	Aplasie médullaire
P2	6	F	-	+	+	Pouce hypoplasique, scoliose	-	Malformation ano-rectal, rein gauche ectopique	Anémie+ thrombocytopénie	ND
P3	5	F	+	+	+	-	+	-	Anémie+ thrombocytopénie	Hypoplasie médullaire
P4	11	F	-	+	-	Polydactylie ; synostose radio-cubitale	+	-	Anémie+ neutropénie	ND
P5	5	F	+	+	-	Polydactylie	-	-	Anémie+ neutropénie	Hypoplasie médullaire
P6	8	M	-	-	-	-	+	Hypospadias	Pancytopénie	Hypoplasie médullaire
P7	4	F	+	+	+	Pouce bas implanté	-	-	Anémie+ thrombocytopénie	ND
P8	7	M	+	+	+	-	-	-	Anémie+ thrombocytopénie	ND
P9	4,5	M	+	-	+	Pouce droit hypoplasique	+	-	Pancytopénie	Hypoplasie médullaire

**Tableau 1** *Caractéristiques cliniques et biologiques des patients avec AF (RSP : Retard staturo-pondéral, DF : Dysmorphie Faciale, AS : Anomalies Squelettiques, SP : Signes cutanés, ND : Non Disponible)*

### **1.2.2. Etude cytogénétique : Caryotype sous mitomycine**

Chez tous ces patients atteints d'AF, nous avons réalisé une étude cytogénétique avant d'effectuer l'étude moléculaire.

Lors d'une consultation de génétique, les informations médicales sont recueillies dans un dossier cytogénétique. Un prélèvement de 2-3ml est effectué sur tube hépariné, doté d'un numéro cytogénétique selon la procédure de culture cellulaire et gardé à +4°C avant la culture cellulaire (Annexe 1).

Le caryotype sous mitomycine se fait en plusieurs étapes :

➤ ***Culture cellulaire :***

Dans deux tubes de 8ml de milieu de culture cellulaire prêt à l'emploi (Gibco, milieu de caryotypage PB-MAX), on réalise une culture simultanée du sang prélevé et d'un sang de contrôle apparié à l'âge et au sexe. Les tubes sont ensuite incubés dans une étuve à 37°C pendant 72 heures.

➤ ***Ajout de la mitomycine C :***

Il s'agit d'un agent alkylant qui entraîne l'apparition de lésions au niveau de l'ADN génomique ; les cellules AF ayant un défaut de réparation sont incapables de corriger ces endommagements. La mitomycine C à une concentration de 36ng/ml est ajoutée 24 heures avant la récolte des métaphases.

➤ ***Obtention des chromosomes métaphasiques :***

Après 72h de culture, on ajoute la colchicine (solution de karyomaxcolcémide) pour arrêter le cycle cellulaire, un traitement hypotonique au KCl (0,075 M) est ensuite réalisé pour entraîner un éclatement de la membrane cellulaire, suivi d'une fixation par le Carnoy, solution

préparé au préalable à base de méthanol (trois volumes) et d'acide acétique (un volume) [4]. Les préparations chromosomiques obtenues sont étalées sur des lames sur plaque chauffante puis colorées avec le colorant Giemsa.

Une dénaturation thermique dans un tampon phosphate dihydrogène à 84 est réalisé pour chaque patient afin de déterminer la formule chromosomique (Annexe 2).

➤ ***Analyse chromosomique au cytoscan :***

L'analyse est réalisée à l'aide d'une station semi-automatique liée à un microscope. Au moins 50 métaphases sont observées chez le patient et le témoin à la recherche de cassures chromosomiques (cassures chromatidiennes, chromosomes dicentriques et acentriques), d'images en double minute, d'images tri et tétra radiales.

➤ ***Interprétation et rendu de résultat :***

Le résultat final est rédigé pour être validé lors d'un staff cytogénétique.

Le nombre total des cassures est calculé comme suit :

- Une cassure au niveau d'une chromatide est comptée comme 1 cassure
- Une cassure des 2 brins est comptée comme 2 cassures
- les chromosomes acentriques et doubles minutes sont comptés comme 2 cassures

Le tableau 2 résume les résultats cytogénétiques chez les neuf patients étudiés avec le nombre total des cassures par rapport au témoin et le nombre des images radiales.

Patients	Cassures chromosomiques		Formes radiales
	Patient	Témoin	
P1	36	8	2
P2	64	8	11
P3	34	8	4
P4	21	2	1
P5	39	8	4
P6	83	7	5
P7	92	8	27
P8	128	8	16
P9	170	8	41

*Tableau 2: Résultats cytogénétiques avec le nombre des cassures chromosomiques et des formes radiales des patients et des témoins normaux*





*Figure 3: Images de chromosomes métaphasiques montrant des cassures et des images triradiales*

## **2. Méthodes :**

### **2.1. Recueil des échantillons pour analyse moléculaire :**

Après avoir obtenu un consentement libre et éclairé du patient lui-même ou des parents pour les mineurs, un dossier génétique est établi pour chaque famille au cours d'une consultation de génétique, un prélèvement de 3 à 4ml de sang périphérique du patient sur tube EDTA (Ethylène Diamine Tétracétique acide) est ensuite réalisé. Au niveau des échantillons sanguins, on mentionne le nom complet du patient, son numéro de dossier (spécifique à notre laboratoire) ainsi que le nom de la pathologie pour laquelle il a été prélevé. Les prélèvements sont ensuite enregistrés et conservés à -20°C. Dans ces conditions, la décongélation douce des échantillons sanguins à +4°C ou à température ambiante est souhaitable juste avant l'extraction d'ADN.

### **2.2. Extraction par Kit des acides nucléiques à partir du sang total :**

Les acides nucléiques des patients de la présente étude ont été extraits par des kits commerciaux. Les deux kits utilisés sont : PureLink™ Genomic DNA Mini Kit (Invitrogen) et le Kit QIAamp DSP DNA Blood Mini (Qiagen).

Vu que les deux Kits sont basés sur le même principe et comportent les mêmes étapes, nous allons décrire ci-dessous l'extraction de l'ADN par Kit QIAGEN.

Le kit Qiagen (Qiagen®: Kit QIAamp DSP DNA Blood Mini) ([www.qiagen.com/resources/download.aspx?id=d11105d4-3499](http://www.qiagen.com/resources/download.aspx?id=d11105d4-3499)) est un système qui permet l'isolation et la purification de l'ADN génomique à partir d'échantillons biologiques et qui a recours à la technologie des membranes de silice (QIAamp). A partir de 200 µL de sang total (frais ou congelé) traités à l'EDTA ou au citrate, l'extraction de l'ADN génomique est réalisée

sur mini colonnes et permet d'obtenir de l'ADN purifié prêt à l'emploi. Cette technique s'effectue en deux étapes (Annexe 3):

- ❖ **Etape de lyse** : La lyse cellulaire est effectuée immédiatement après la décongélation du sang à +4°C ou à température ambiante. Dans un tube eppendorf de 1,5 mL, on prend 200 µL du sang total auquel on ajoute 200 µL de tampon de lyse AL et 20 µL de protéinase K (Protéase QIAGEN - QP), le tout est incubé à 56°C pendant 10 minutes.
- ❖ **Etape de purification** : On ajoute 200 µL d'éthanol absolu et on vortexe brièvement. Pour chaque échantillon, on prépare une colonne QIAamp posée sur un tube collecteur, cette étape permet de fixer l'ADN génomique à la membrane de la colonne, on y dépose 620 µL de la solution obtenue et on centrifuge à 1771 G pendant 1 min. La colonne est ensuite placée sur un nouveau tube collecteur et on jette l'ancien tube et l'éluât. Le lysat passe à travers la membrane alors que l'ADN génomique est adsorbé sur la membrane de silice. Un lavage de la membrane par les tampons du kit est réalisé, on dépose 500 µL de tampon AW1 fourni par le kit, on centrifuge à 1771 G pendant 1 min et on replace la colonne sur un nouveau tube collecteur. Ensuite, on dépose 500 µL du tampon AW2 et on centrifuge à 2877 G pendant 3 min. Une fois l'ADN est lavé, on place la colonne QIAamp sur un tube eppendorf de 1,5 mL, on dépose 50 µL de buffer AE et on centrifuge à 1771 G pendant 1 min. A cette étape, l'ADN est élué. La colonne QIAamp est jetée et le tube eppendorf contenant l'ADN extrait et purifié est conservé à +4°C si l'échantillon est utilisé dans les jours qui suivent l'extraction ou à -20°C pour une utilisation ultérieure.

## **2.3. Dosage et contrôle de la qualité de l'ADN extrait :**

### **2.3.1. Dosage spectrophotométrique au Nanodrop :**

Au sein de notre laboratoire, l'appareil NANODROP 2000 (Thermo Scientific Nano Drop TM 2000) est utilisé pour mesurer le dosage des ADN. Il suffit de déposer dans l'appareil 1  $\mu$ l du tampon de l'élution (TE) comme « blanc » puis 1  $\mu$ l d'ADN extrait pour déterminer sa concentration et obtenir la courbe qui nous renseigne sur la pureté et la qualité de l'ADN.

En effet, ce dosage s'effectue par mesure de l'absorption A qui se définit par l'unité de densité optique mesurée à 260 nm en lumière ultraviolette (UV). Une unité de densité optique correspond à l'absorption d'une solution d'ADN double brin à la concentration de 50  $\mu$ g/ml ou à l'absorption d'une solution d'ADN simple brin (ou d'ARN) à la concentration de 25  $\mu$ g/ml.

La mesure spectrophotométrique de la quantité de rayons UV absorbés par les bases est une opération facile et précise en l'absence de quantité significative de contaminants tels que des protéines, du phénol ou autres. En général, la concentration d'acides nucléiques est déterminée par une mesure effectuée à 260 nm contre un échantillon appelé « blanc » (Tampon aqueux à faible concentration ionique, dans notre cas c'est le tampon d'élution TE utilisé lors de l'extraction de l'ADN génomique).

Le calcul du rapport entre l'absorption de l'ADN et celles des protéines permet de reconnaître une interférence par des contaminants. Les protéines absorbant à 280 nm, le ratio A260/A280 est utilisé pour estimer la pureté de l'acide nucléique. L'ADN pur devrait avoir un ratio d'environ 1,8.

Tous nos ADNs extraits dans cette étude représentent un dosage qui varie entre 30 et 150 ng/ $\mu$ l avec des rapports d'absorption variant entre 1,8 et 2, ce qui nous permet de procéder à toutes les techniques de biologie moléculaire.

### **2.3.2. Dosage fluorométrique par Qubit:**

Pour les échantillons qui vont passer par séquençage haut débit, le dosage des ADNs se fait dans un premier temps par l'appareil de NANODROP 2000 puis nous effectuons un dosage par l'appareil Qubit3.0 (Invitrogen™ Qubit™ 3.0 Quantitation Starter). On dilue ces ADNs pour que leurs concentrations finales ne dépassent pas 100ng/ $\mu$ l. On met dans un tube 1,5 ml spécial Qubit, 199  $\mu$ l du tampon Qubit dsDNA HS Buffer et 1  $\mu$ l de Reagent 200x avec 2  $\mu$ l d'ADN. On fait analyser l'échantillon par le Qubit et on dilue notre ADN jusqu'à l'obtention d'une concentration finale entre 10 et 20 ng/  $\mu$ l.

### **2.4. Technique de la PCR :**

En effet, le séquençage par la méthode de Sanger est précédé par une étape cruciale qui est la réaction de polymérisation en chaîne (Polymerase Chain Reaction : PCR) et dont la procédure complète a été décrite et appliquée expérimentalement par Kary Mullis en 1985. C'est une technique d'amplification puissante qui peut générer une quantité suffisante d'un segment spécifique d'ADN (c'est-à-dire un amplicon) à partir d'une petite quantité d'ADN [5].

Nous avons utilisé cette technique pour amplifier les fragments d'ADN chez la famille avec XP qui vont être séquençés par Sanger.

#### **2.4.1. Détermination de la séquence à amplifier :**

Pour pouvoir lancer une PCR, il faut d'abord déterminer le fragment d'ADN qu'on souhaite amplifier. Pour cela, il existe plusieurs bases de données génomiques bioinformatiques disponibles sur internet qui sont utilisées pour déterminer la séquence

d'ADN cible (NCBI, Ensembl, UCSC, GeneAtlas, etc.). Celles-ci permettent également de définir les séquences de référence (RefSeq). Il s'agit d'une référence stable pour l'annotation du génome, l'identification des gènes et l'analyse des mutations et des polymorphismes.

Pour un gène donné, un ou plusieurs RefSeq peuvent lui être annotées selon les transcrits étudiés de ce gène, on lui attribue alors une référence NM (pour « Natural mRNA »).

Le gène *XPC* est localisé sur le chromosome 3, comporte 16 exons et code pour une protéine de 940 acides aminés (NM\_004628) selon l'hg19 (CRCh37.p13).

#### **2.4.2. Design des amorces :**

Pour amplifier sélectivement une séquence d'ADN par PCR, il est indispensable de disposer d'une paire d'amorces synthétisées par voie chimique. Une amorce est faite d'oligonucléotides (20 nucléotides en moyenne). Plusieurs critères doivent être pris en compte pour la conception et l'emploi des amorces :

- Elles doivent être complémentaires avec les deux extrémités de la séquence d'intérêt.
- La séquence des amorces doit comporter la plus grande proportion possible de guanines et de cytosines pour une meilleure stabilité de la liaison au moment de l'hybridation des amorces et une hybridation plus spécifique.
- Les pourcentages en guanines et cytosines de chaque membre du couple d'amorces doivent être aussi proches que possible pour avoir une température d'hybridation ( $T_m$  pour melting temperature) convenable qui permettrait une hybridation équilibrée des amorces.

- Les amorces ne doivent pas présenter de fortes complémentarités de séquence pour éviter le risque d'association entre elles, ceci peut perturber l'amorçage de la réplication.

Actuellement et pour faciliter le design d'amorces, des logiciels spécialisés sont disponibles avec des bases de données nucléotidiques en respectant tous ces critères. Dans notre étude, nous avons utilisé *UCSC* et *Primer3*.

Pour rechercher la mutation récurrente c.1643\_1644delTG au niveau de l'exon 9 du gène *XPC* en utilisant les amorces suivantes:

XPC-9F: GGCATCCTCAAGCTCTTCAA

XPC-9R: GGGCTCTGGTATGGTCTCAA

### **2.4.3. La réaction PCR et le contrôle sur gel :**

La réaction de polymérisation en chaîne (PCR) est effectuée dans un tube de 200 µl sur un volume final de 50 µl. Le mélange réactionnel comprend l'ADN génomique, un mix comprenant une solution tampon comprenant du MgCl<sub>2</sub> et les dNTP, la Taq polymérase et de l'eau distillée. Ce mélange est mis dans un thermocycleur ABI 9700 (AppliedBiosystems) pour l'amplification du gène (Annexe 5).

L'amplification des fragments d'ADN concernés par la méthode de PCR (Réaction de Polymérisation en Chaîne) est basée sur le fonctionnement cyclique d'une ADN polymérase particulière, la Taq polymérase. Les exons ciblés sont multipliés à l'aide des amorces oligonucléotidiques et de l'ADN polymérase thermostable en suivant les trois cycles de la PCR : Dénaturation à 95°C, hybridation dont la température T<sub>m</sub> varie en général entre 52°C et 62°C, et élongation à 72°C (Figure 4). La durée de chaque cycle peut varier de 15 secondes

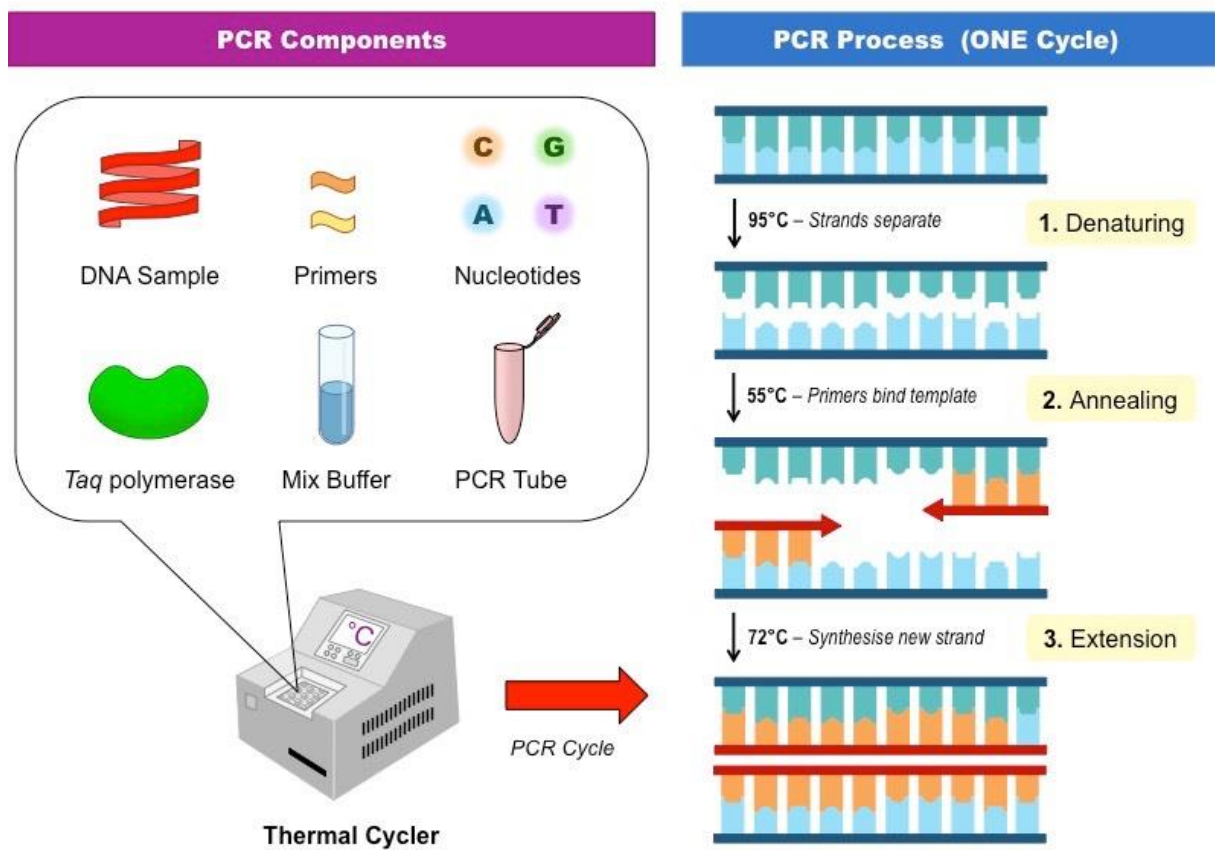
jusqu'à 7 minutes. L'exon à séquencer étant doublé à chaque cycle, cette amplification rend l'exon cible facilement séquençable.

**Le programme utilisé est :**

95°C pendant 1 minute, suivie de 35 cycles de 95°C pendant 15 secondes, 61°C pendant 15 secondes et 72°C pendant 10 secondes.

Le contrôle de ces produits est réalisé par électrophorèse sur gel d'agarose (1%). Les séquences d'ADN sont révélées au moyen d'une coloration au bromure d'éthidium (BET) qui les rend instantanément visibles sous forme d'une bande unique par transillumination aux ultraviolets.





*Figure 4: Mélange réactionnel et différents cycles de la PCR*

## **2.5.Séquençage par la méthode Sanger pour la famille avec XP :**

### **2.5.1. Purification du produit de PCR :**

Après la PCR, une purification du produit est essentielle afin d'éliminer les amorces et les dNTP résiduels, non incorporés dans le mélange réactionnel, autrement, ces résidus risquent d'interférer dans les réactions en aval (séquençage). Dans cette étude, nous avons utilisé la purification enzymatique par ExoSAP-IT™ PCR Product Cleanup Reagent (Annexe 6).

Cette méthode utilise deux enzymes hydrolytiques réunies dans un même mélange avec un tampon conçu pour optimiser leur activité en une seule réaction :

- L'exonucléase qui a pour rôle d'éliminer l'excès d'amorces dans le produit de PCR,
- La Shrimp phosphatase alcaline (SAP) permettant de déphosphoryler les dNTP résiduels pour altérer leur migration et réincorporation aux fragments d'ADN.

L'incubation à 37°C pendant 15 min active ces enzymes et le chauffage à 80°C pendant 15 min les inactivent irréversiblement.

### **2.5.2. Réaction de Séquence :**

Pour la réaction de séquence, nous avons eu recours à un mélange contenant une taq polymérase, des dNTP, des ddNTP marqués par un fluorochrome, du MgCl<sub>2</sub> et une solution tampon, c'est le kit BigDyeterminator v3.1 cycle sequencing kit (Thermo Fisher Scientific). Dans un tube de 200 µl contenant le produit de PCR purifié, on ajoute le BigDye à une amorce sens ou anti-sens et à de l'eau distillée (Figure 5). On soumet ensuite le mélange réactionnel à une amplification de 30 cycles dans le thermocycleur à des conditions spécifiques (Annexe 7).

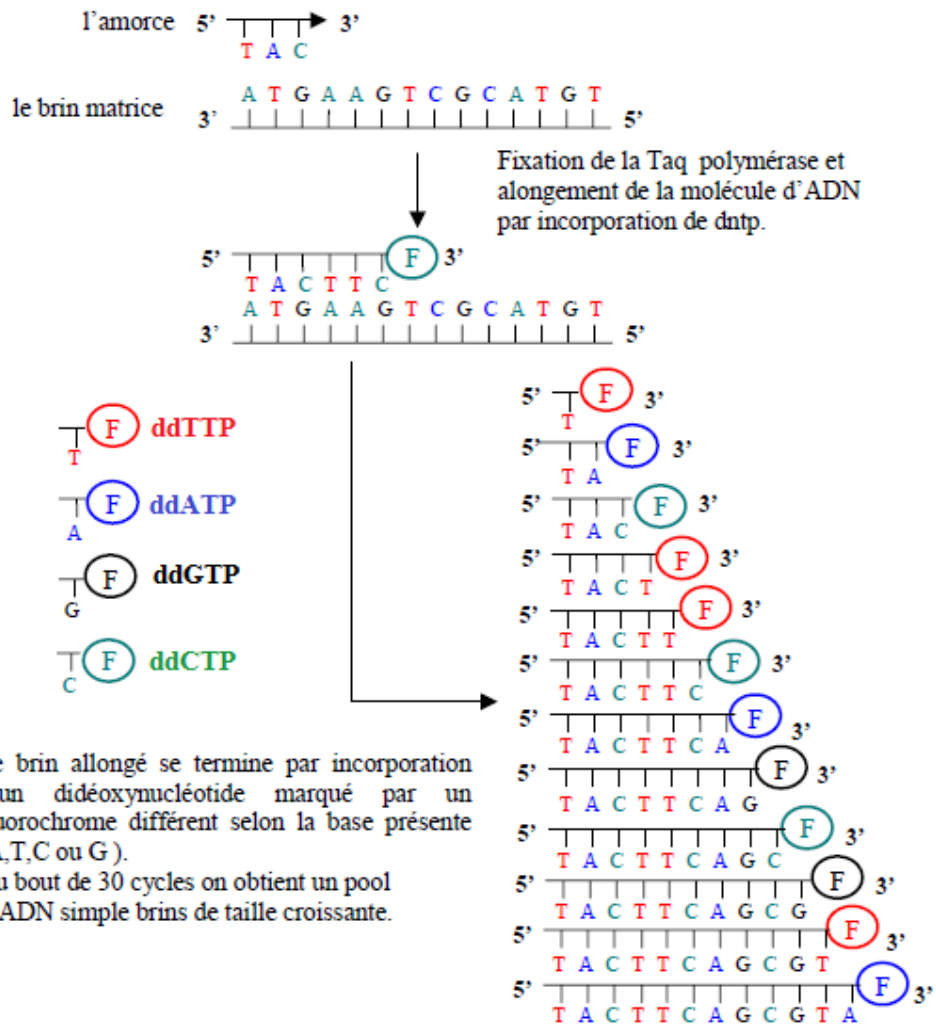


Figure 5: Principe de la réaction de séquence

### **2.5.3. Purification de la réaction de séquence par Sephadex G50 :**

La réaction de séquence est purifiée par ce qu'on appelle la chromatographie d'exclusion, appelée également chromatographie de filtration sur gel ou tamisage moléculaire. Il s'agit d'une technique apparue en 1959, utilisant le Sephadex. C'est un gel de dextrane qui se présente sous forme de billes poreuses très hydrophiles qui gonflent dans l'eau après réticulation. Il existe différents types de Sephadex en fonction de la taille des billes et de leur porosité.

Le but de cette purification du produit de la réaction de séquence est de piéger les sels et les ddNTP libres en excès qui pourraient parasiter les signaux de fluorescence spécifiques lors du séquençage automatique (Annexe 8).

### **2.5.4. Séquençage Sanger:**

Le séquençage de Sanger est un moyen de détermination de l'ordre des quatre nucléotides dans un brin d'ADN, afin de voir la succession de nucléotides constituant un fragment d'ADN donné. Son application dans divers domaines de la recherche de base, de la biotechnologie et de la médecine est devenue indispensable.

En 1977, Frederick Sanger a mis au point une méthode qui lui a valu le prix Nobel de chimie en 1980 et qui permet de lire la séquence de nucléotides. Cette technique est basée sur le principe de surveillance des nucléotides ajoutés un par un lors de la synthèse des brins d'ADN.

Après purification de nos produits de réaction de séquence, nous les avons soumis à l'appareil ABI prism 3130 DNA séquençer à 16 capillaires (life technologies) pour séquençage automatique. Ensuite, nous avons procédé à l'alignement et à l'analyse des

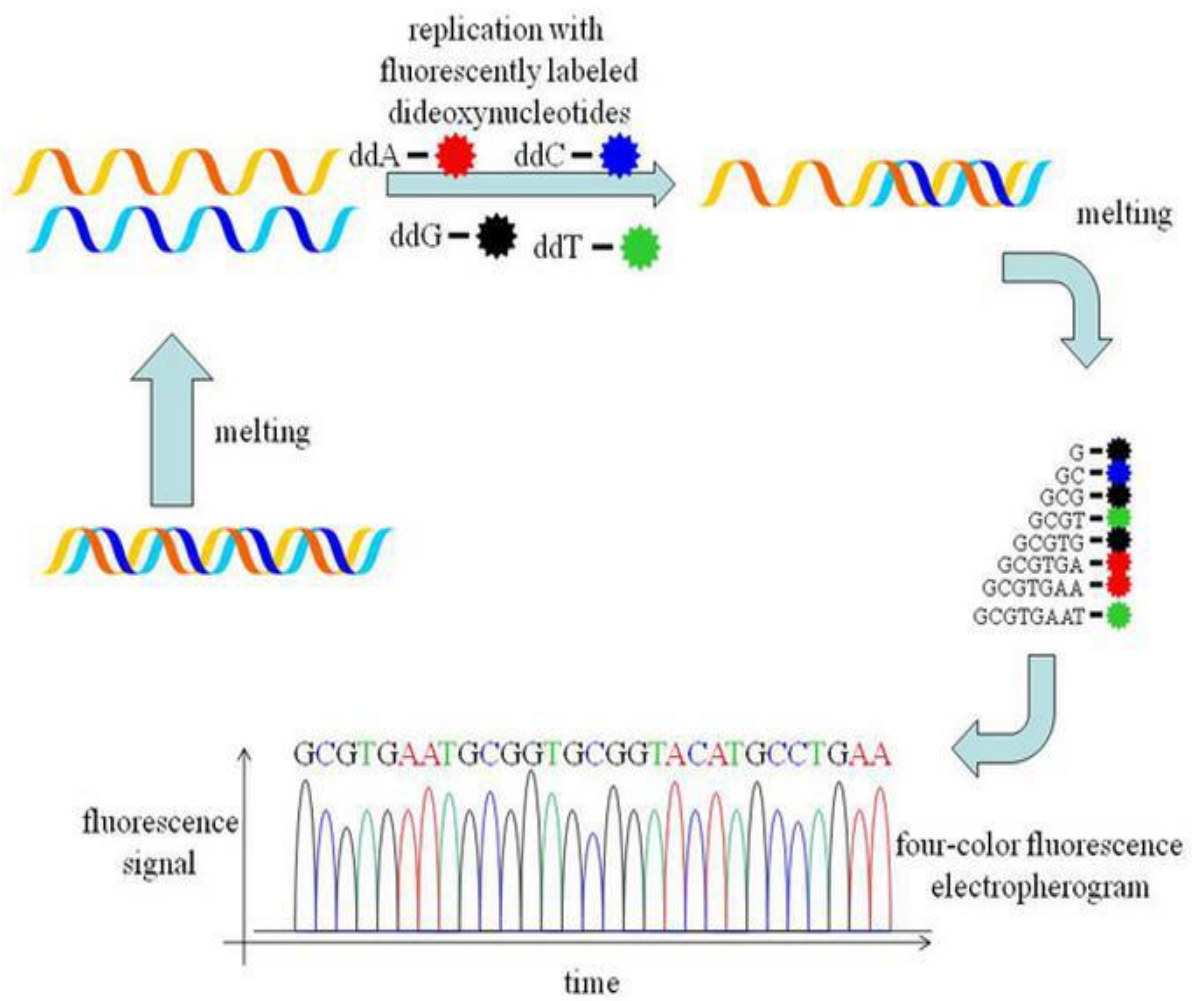
séquences au moyen du logiciel d'analyse des variants d'ADN (Mutation Surveyor® software).

Le séquençage Sanger repose sur la réplication de l'ADN par l'ADN polymérase et des amorces spécifiques de séquence en incorporant les dNTP présents dans le milieu réactionnel. En addition à ces dNTP, les didésoxyribonucléotides interrupteurs de chaîne (ddNTP), qui seront incorporés aléatoirement par l'ADN polymérase lors de la réplication. Les ddNTP diffèrent des dNTP par leur extrémité 3'. L'extrémité 3' hydroxyle (OH) des dNTP est remplacée par un hydrogène (H) dans les ddNTP. Cette modification empêche la formation de la liaison phospho-diester entre le ddNTP incorporé et le nucléotide suivant, ce qui interrompt la synthèse du fragment d'ADN [6].

Dans le milieu réactionnel, les ddNTP sont ajoutés de manière à ce qu'ils soient incorporés statistiquement à toutes les positions possibles de la séquence d'ADN à étudier.

Ainsi, à la fin de la réaction de séquence, nous obtenons un produit qui contient des fragments d'ADN de tailles différentes selon la position d'incorporation du ddNTP (Figure 6).

Au moyen d'un appareil automatique, le produit de cette réaction de séquence sera soumis à l'électrophorèse sur un gel de polyacrylamide qui va séparer tous les fragments présents en fonction de leur taille. Chaque ddNTP (A, C, G et T) est marqué par un fluorochrome avec un spectre d'émission spécifique. Dans le séquenceur automatique et au cours de l'électrophorèse capillaire des fragments d'ADN, les fluorochromes sont excités par un rayon laser permettant ainsi de libérer des signaux fluorescents qui seront enregistrés à des points fixes. Au fur et à mesure que les signaux sont capturés, la séquence est enregistrée puis visualisée sous forme de pics de quatre couleurs représentant la succession des nucléotides qui constituent une séquence d'ADN.



*Figure 6: Principe de séquençage automatisé d'ADN par la méthode de Sanger*

## **2.6. Le séquençage à haut débit par panel de gènes pour les familles avec AF :**

Le séquençage de nouvelle génération (NGS) ou le séquençage massif en parallèle décrivent une technologie de séquençage d'ADN qui a révolutionné la recherche génomique. En utilisant le NGS, un génome humain entier peut être séquençé en quelques jours.

En effet, il existe plusieurs plates-formes de NGS utilisant des technologies différentes mais elles réalisent toutes un séquençage de millions de petits fragments d'ADN en parallèle. Des analyses bioinformatiques sont utilisées pour reconstituer ces fragments en cartographiant l'individu avec le génome humain de référence. Chacun des trois milliards de bases du génome humain est séquençé plusieurs fois, offrant une grande profondeur à livrer des données précises et un aperçu des variations de l'ADN.

Le NGS peut être utilisé pour séquençer des génomes entiers (Whole Genome Sequencing) ou limité à des domaines d'intérêt spécifiques, y compris les 22000 gènes codants (un exome entier : Whole Exome Sequencing) ou un petit nombre de gènes individuels (Panel de gènes dédiés à une pathologie ou un groupe précis de maladies).

L'anémie de Fanconi est une maladie très hétérogène sur le plan génétique avec plus de 20 gènes. Dans le cadre de mon sujet de thèse, nous avons décidé d'ajouter les 3 gènes les plus fréquemment mutés *FANCA*, *FANCG* et *FANCC* à notre panel destiné pour le diagnostic moléculaire de certaines maladies héréditaires.

### **2.6.1. Principe de la technologie PGM Ion Torrent:**

(<https://www.thermofisher.com/us/en/home/life-science/sequencing/nextgeneration-sequencing.html>)

La technologie Ion Torrent <sup>TM</sup> convertit directement la séquence nucléotidique en informations numériques sur une puce semi-conductrice [7]. Dans une réaction de synthèse d'ADN, quand un nucléotide est incorporé en face de sa base complémentaire dans une chaîne d'ADN en croissance, un ion hydrogène est libéré. Cela modifie le pH de la solution qui peut être enregistré comme un changement de tension par un capteur d'ions, un peu comme un pH-mètre. Si aucun nucléotide n'est incorporé, aucune pointe de tension ne se produit. Quand deux nucléotides adjacents incorporent le même nucléotide, deux hydrogènes sont libérés et la tension est double. Ainsi, les lectures ou « runs » d'un seul nucléotide peuvent également être déterminées. Les réactions de séquençage Ion Torrent se produisent dans des millions de puits qui recouvrent une puce semi-conductrice contenant des millions de pixels qui convertissent les informations chimiques en information de séquences.

### **2.6.2. Le protocole de séquençage Ion PGM :**

Il s'agit d'un processus complexe comprenant plusieurs étapes et nécessitant environ 2 jours de manipulations pour parvenir à l'étape de séquençage.

#### ***2.6.2.1. Design de Panel :***

Une étape préliminaire de design des amorces est requise, le design des panels s'effectue sur Ion AmpliSeq<sup>TM</sup> Designer. Ce dernier est un site en ligne gratuit qui permet de créer et de commander des panels personnalisés, comprenant des gènes spécifiques selon la demande et le besoin du laboratoire. Ion AmpliSeq<sup>TM</sup> Designer propose trois différents types de panels de gènes, et à partir de là on peut choisir la stratégie de design la mieux adaptée aux



besoins du laboratoire. Les différents types de panels proposés sont : On-Demand, Made-to-Order et Ready-to-Use.

- Panels On-Demand : Ion AmpliSeq™ On-Demand panels permet le design de panels customisés à partir de gènes pré-testés pour les petites réactions (jusqu'à 96), ce qui aide à réduire les coûts et les risques. Le design peut se faire en choisissant un panel avec les gènes impliqués dans un groupe de pathologies hétérogènes (ex : les dystrophies musculaires, les maladies cardiovasculaires, etc.), ou en cherchant des gènes d'intérêt dans le catalogue Ion AmpliSeqOnDemand composé de 5000 gènes prétestés, ce qui permet un design de panels personnalisés pour le diagnostic simultané de plusieurs maladies génétiques. Il est à noter que seuls les gènes impliqués dans les pathologies de la lignée germinale sont inclus dans ce panel.

- Panels Made-to-order : Ce type de panel offre un maximum de flexibilité à l'utilisateur. Il permet le design d'un panel customisé des gènes et/ou des régions chromosomiques. Il est destiné pour les larges réactions (3000 réactions). Un Panel 'Made-to-Order' testé et dont l'efficacité a été prouvée peut être partagé avec les utilisateurs de la communauté de Ion AmpliSeq. (ex : panel pharmacogénétique, etc.).

- Panels Ready-to-use : Ce sont des panels prêts à l'emploi, élaborés par Ion AmpliSeq et destinés aux petites réactions de 8. Ces panels constituent un moyen rapide et économique pour commencer les tests de diagnostic et de recherche. Ce sont des panels testés et dont l'efficacité a été prouvée et optimisée par l'équipe de Ion Ampliseq.

Dans le cadre de mon sujet de thèse, nous avons décidé de mettre au point dans mon laboratoire d'accueil un panel de gènes customisé consacré à certaines maladies génétiques héréditaires. Ce panel contient les trois gènes majoritaires de l'anémie de Fanconi associés à 19 autres gènes impliqués dans des maladies mendéliennes vues en consultation de génétique du DGM de l'INH (Figure 7).

- Spécificité des Primers Ion Torrent : Les amorces qui composent le panel désigné par Ion Ampliseq contiennent des séquences spécifiques permettant d'associer les Adaptateurs et le Code barre (barcode) (Figure 8) :

- Adaptateurs : petits oligonucléotides de séquence connue, qui permettent leur amplification simultanément à partir d'amorces universelles. Pour éviter l'association de plusieurs molécules d'adaptateurs entre elles, celles-ci sont déphosphorylées sur leurs extrémités 5'.
- Code à barre = barcode : séquences d'environ 10 nucléotides permettant d'identifier les ADN par la suite quand tous les échantillons sont mélangés pour le séquençage.

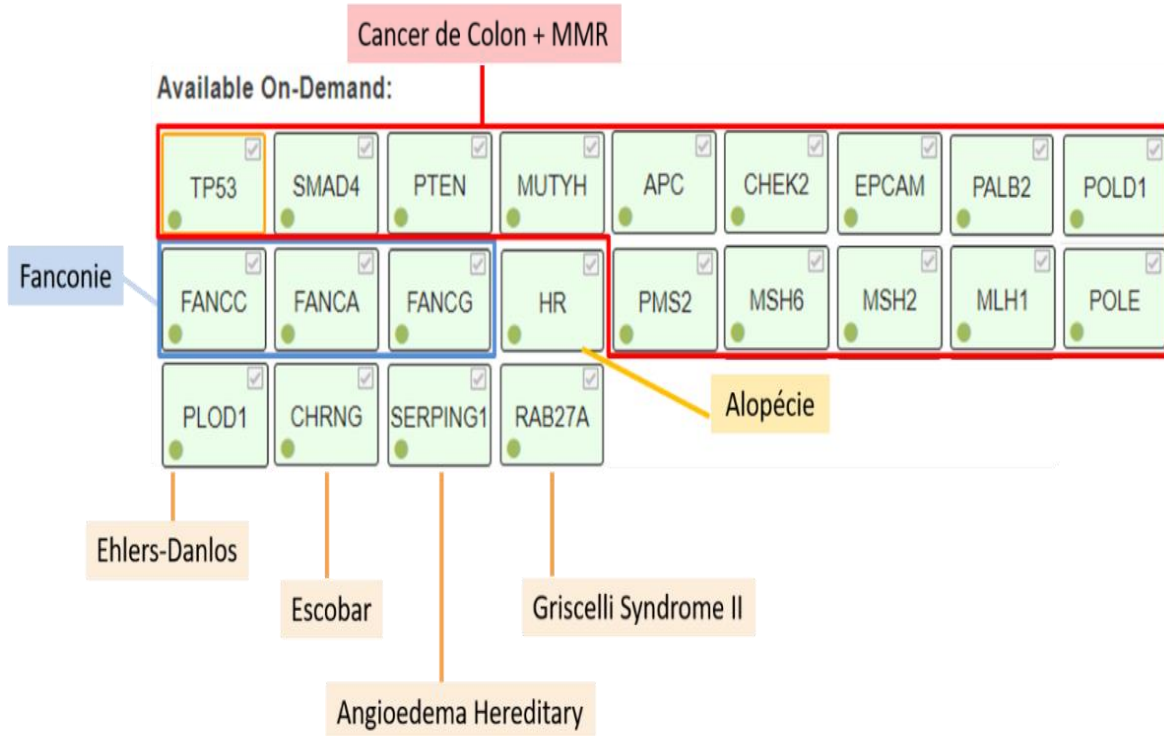


Figure 7: Panel conçu par l'équipe de DGM-INH

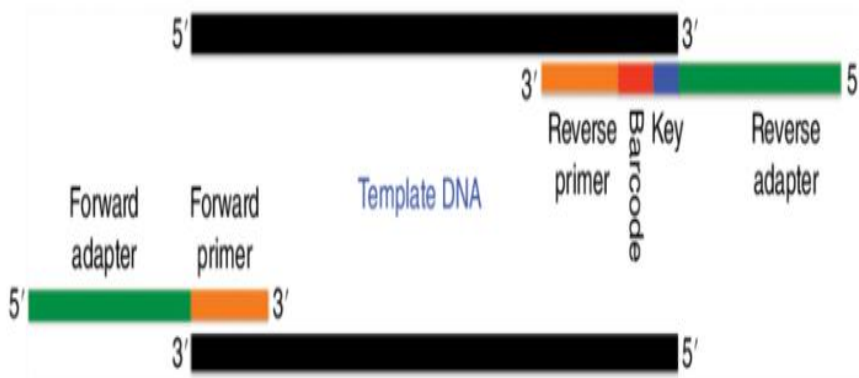



Figure 8: Spécificité des primers Ion Torrent

### 2.6.2.2. Séquençage PGM proprement dit : Le séquençage Haut Débit sur la plateforme Ion Torrent (ThermoFisherScientific)

La partie technique de NGS-PGM Ion Torrent (workflow) demande la réalisation de quatre grandes étapes : la préparation de la librairie, la PCR à émulsion, l'enrichissement et le séquençage. C'est une technique qui nécessite de la part de l'opérateur un investissement en temps (3 jours successifs), en concentration et en rigueur (car on manipule plusieurs patients à la fois). Dans les premières étapes de cette technique (avant de mettre les barcodes), on numérote tous les tubes contenant les ADNs des patients et on mentionne leurs noms et leurs numéros de dossiers au niveau du cahier de paillasse (propre à chaque manipulateur) pour éviter toute erreur.

#### ❖ Préparation de la librairie (Figure 9) :

Dans un tube de 0,2 ml, nous mélangeons 1 µl de 5<sup>x</sup> Ion AmpliSeqHiFi Mix, 2,5 µl de Ion AmpliSeq Primer Pool 2<sup>x</sup>, 1 µl de l'eau pure et 1 µl d'ADN à 10 ng/µl de concentration. Ce mélange subit une PCR selon le programme suivant :

- 99°C durant 2 min
  - 99°C durant 15 s
  - 60°C durant 4 min
  - 10°C
- } 

#### ❖ Digestion :

C'est une étape qui permet la fragmentation de l'ADN. Pour cela, nous ajoutons 1µl de l'enzyme FuPaReagent par tube, suivi par une PCR de 40 min selon les températures et les durées suivantes :

- 50°C pendant 10 min
- 55°C pendant 10 min
- 60°C pendant 20 min
- 10°C pendant  $\infty$

❖ **Ligation :**

Les fragments obtenus sont ensuite liés aux adaptateurs P1 et A couplés aux barcodes contenus dans le kit Ion Xpress Barcode Adaptors par une ligase pendant 30 min à 22°C puis son inactivation à 72°C pendant 10 min.

Les fragments sont liés aux adaptateurs P1 et A pour permettre par la suite l'amplification simultanée de tous les fragments à partir d'amorces universelles à l'aide d'amorces complémentaires aux adaptateurs A et P1.

Avant de commencer l'étape de la ligation des barcodes, la dilution de ces derniers est nécessaire, en mélangeant 2  $\mu$ l de l'eau pure, 1  $\mu$ l de Ion Xpress Bar Code x et 1  $\mu$ l de Ion Xpress P1 Adapter.

L'étape de la ligation comprend le mélange de 2  $\mu$ l de Switch solution, 1  $\mu$ l de barcode dilué, 1  $\mu$ l de DNA Ligase et 11  $\mu$ l du produit digéré en suivant le programme ci-dessous :

- 22°C durant 30 min
- 68°C durant 5 min
- 72°C durant 5 min
- 10°C durant  $\infty$

### ❖ Purification :

Après l'étape de la digestion et de la ligation des codes-barres, la purification des produits obtenus par billes magnétiques est obligatoire.

L'ADN est d'abord sélectionné en fonction de la taille et purifié avec des billes magnétiques de l'Agencourt® AMPure. Les fragments d'ADN de ~200 et/ou ~400 pb sont sélectionnés et tous les fragments plus courts ou les enzymes restants sont éliminés. En parallèle, un lavage à l'éthanol est réalisé pour éliminer les contaminants.

Après transfert du contenu de chaque tube précédent dans un tube magnétique 1,5 ml, on ajoute dans chacun 22,5 µl de l'Agencourt, on mélange le tout et on le laisse 5 min à température ambiante. Ces tubes spéciaux sont ensuite placés dans le portoir magnétique. Après 2 min, on enlève le surnageant et on ajoute 150 µl de l'éthanol 70%. Dans le portoir magnétique, on fait tourner les tubes pour bien laver les billes. On jette le surnageant et on laisse évaporer les particules de l'éthanol dans la température ambiante pendant 5 min. On répète l'étape de lavage par l'éthanol trois fois.

### ❖ PCR universelle :

Dans chacun des tubes précédents, on ajoute 25 µl de Ion AmpliSeqPlatinum PCR Super Mix Hi-Fi et 1 µl de Library amplification Primer Mix, on vortexe le tout puis on le remet dans le portoir magnétique. On récupère cette fois-ci le surnageant qui va subir une PCR selon le programme suivant :

- 98°C durant 2 min
  - 98°C durant 15 s
  - 64°C durant 1 min
  - 10°C durant ∞
- } x21 cycles

### ❖ Purification de la librairie amplifiée :

Cette étape comprend deux purifications. La première a pour but de se débarrasser des longs fragments et la deuxième permet d'éliminer les amorces en surplus.

On transfère le produit de la PCR dans un tube 1,5 ml spécial (magnétique) puis on ajoute 12,5 µl de l'Agencourt. On laisse incuber ce mélange à température ambiante pendant 5 min, on remet ensuite les tubes dans le portoir magnétique afin de récupérer le surnagent et le transférer dans un autre tube spécial et cette fois-ci on ajoute 30 µl de l'Agencourt. Une 2ème incubation à température ambiante durant 5 min est réalisée, puis on enlève le surnagent et on fait laver les billes magnétiques deux fois par 150 µl d'éthanol à 70%. A la fin, on ajoute 30 µl de Tampon d'élution (TE) dans le portoir magnétique pendant 2 min puis on récupère le surnagent.

A ce stade du processus, nous avons la librairie prête à utiliser pour la PCR à émulsion, mais avant l'amplification, il est nécessaire de quantifier le produit obtenu par le Qubit pour déduire le facteur de la dilution afin de fusionner volume à volume les échantillons avec une concentration égale afin de pouvoir être soumises à l'étape suivante.

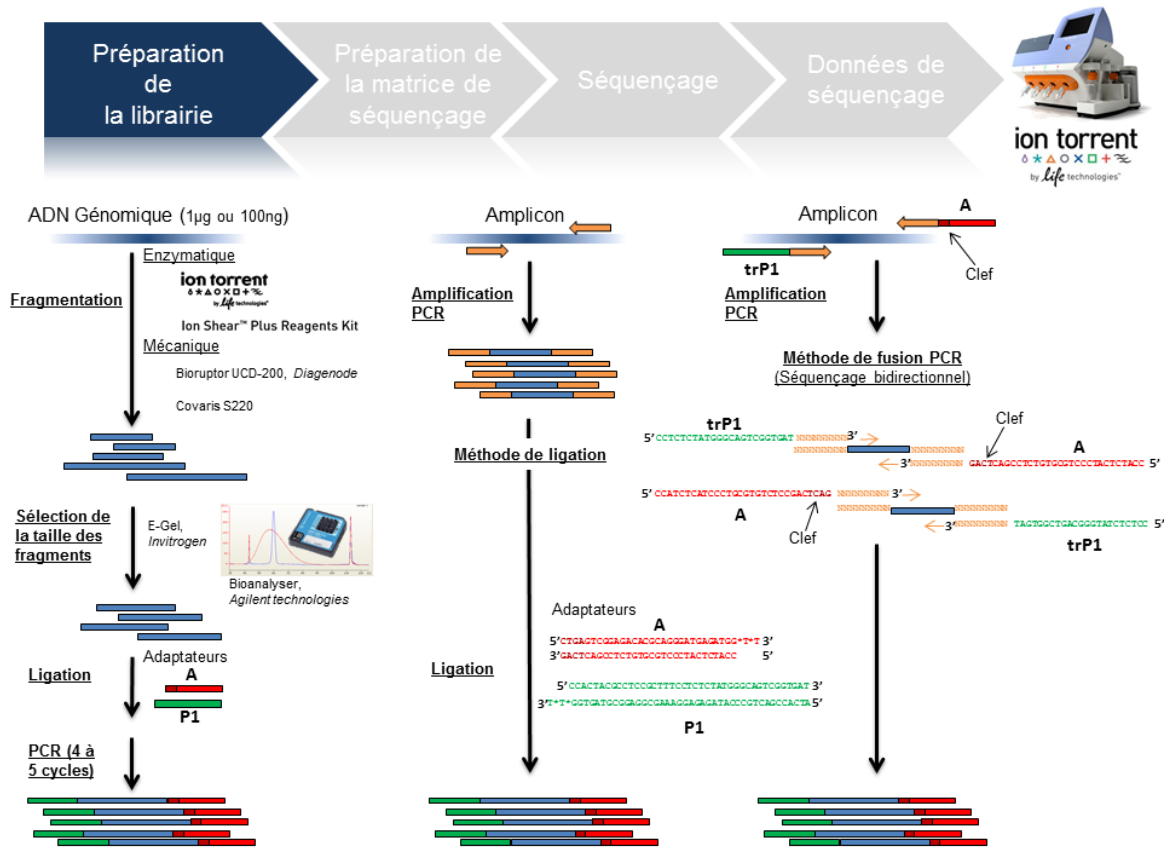


Figure 9: Etapes de la préparation de la librairie (R. BLERVAQUE et al. 2013)



### ❖ PCR à émulsion :

Cette étape automatisée permet l'amplification clonale sur l'appareil OneTouch V2 en utilisant le kit Ion PGM™ Hi-Q™ View OT2 Kit. En effet, ce kit contient des sphères à la surface desquelles un fragment de librairie sera amplifié. A cette étape, chaque fragment d'ADN est lié à une seule particule de sphère ionique (ISP) et subit une amplification clonale de sorte que chaque ISP soit recouverte de nombreuses copies du même fragment d'ADN (Figure 11). Cette amplification clonale contribue à atteindre un seuil de détection du signal nécessaire et suffisant au moment du séquençage. Au sein des gouttelettes, plusieurs configurations de microréacteurs sont envisageables et ceci malgré une optimisation du ratio 1:1 (1 ISP/1 Fragment d'ADN). Seule la configuration de monoclonalité sera source de données de séquençage. Les données générées par les autres configurations seront filtrées lors de l'analyse primaire effectuée par le « Torrent suite » (serveur du séquenceur PGM permettant l'analyse bio-informatique des données brutes). Le produit résultant du OneToucheV2 subit une purification en utilisant les billes magnétiques biotinylés couplées à la streptavidine afin de n'avoir par la suite que les ISPs contenant un fragment (Figure 10).

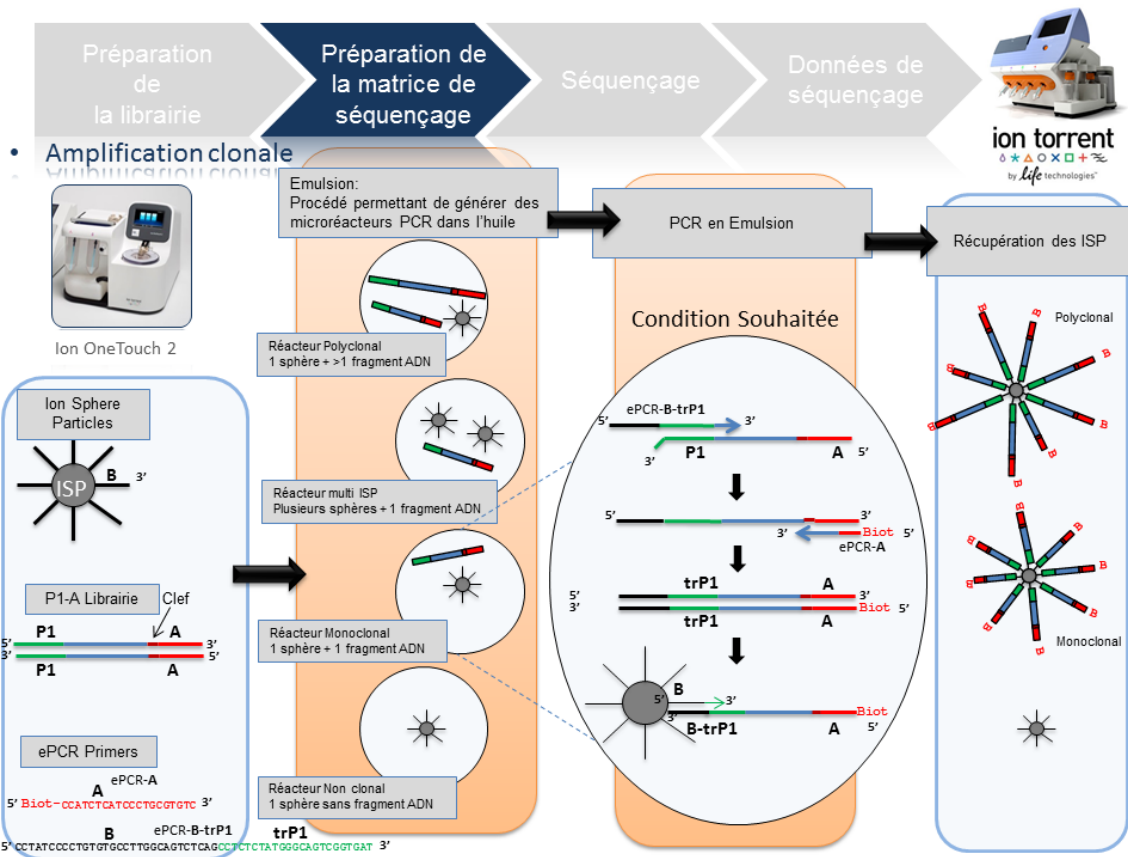


Figure 10: Etapes de la PCR à émulsion (R. BLERVAQUE et al. 2013)

### ❖ **Enrichissement :**

Dans un tube spécial 1,5 ml, on dépose 13 µl Dyna Beads et on le place dans un portoir magnétique. Après avoir éliminé le surnageant, on ajoute 130 µl de My One Beads Wash Solution. Au sein d'un strip de 8, on met la totalité de la librairie, totalité de Dyna Beads préparé (précédemment cité), 300 µl d'Ion One Touch Wash Solution et 300 µl Solution Melt-Off (qui se compose de 280 µl de Twin solution et 40 µl NaOH 1M). On récupère le produit final dans un tube contenant 10 µl de Neutralisation Solution (Figure 11).

### ❖ **Séquençage :**

On ajoute 5 µl de Control Ion Sphère à notre produit final et après une centrifugation de 15500x g pendant 1min 30sec, on ne garde que 15 µl de produit auquel on va ajouter 12 µl des amorces de séquençage suivi par une PCR de 4 min (2 min à 95°C et 2 min à 37°C).

On fait laver l'appareil de séquençage (PGM Ion Torrent) avant et après chaque manipulation. Après le lavage qui dure 15 min, on lance l'initialisation pour ajuster le PH qui ne doit pas dépasser 7,8. Ensuite, on fait le ChipCheck de la puce (afin de vérifier si elle a été reconnue par l'appareil de séquençage) puis on la remplit par le produit à séquencer après avoir ajouté 3 µl de Taq polymérase. Enfin, on démarre le séquençage qui dure 3 heures pour une puce de 314 et 5 heures pour une puce de 316 ou 318 (Figure 12).

Une fraction de billes « tests de fragments » est également ajoutée aux ISP purifiées (elles permettent un contrôle de qualité du séquençage). Les puces se déclinent selon 3 capacités de séquençage (Chip 314 >10Mb, Chip 316 >100Mb, Chip 318 >1Gb).

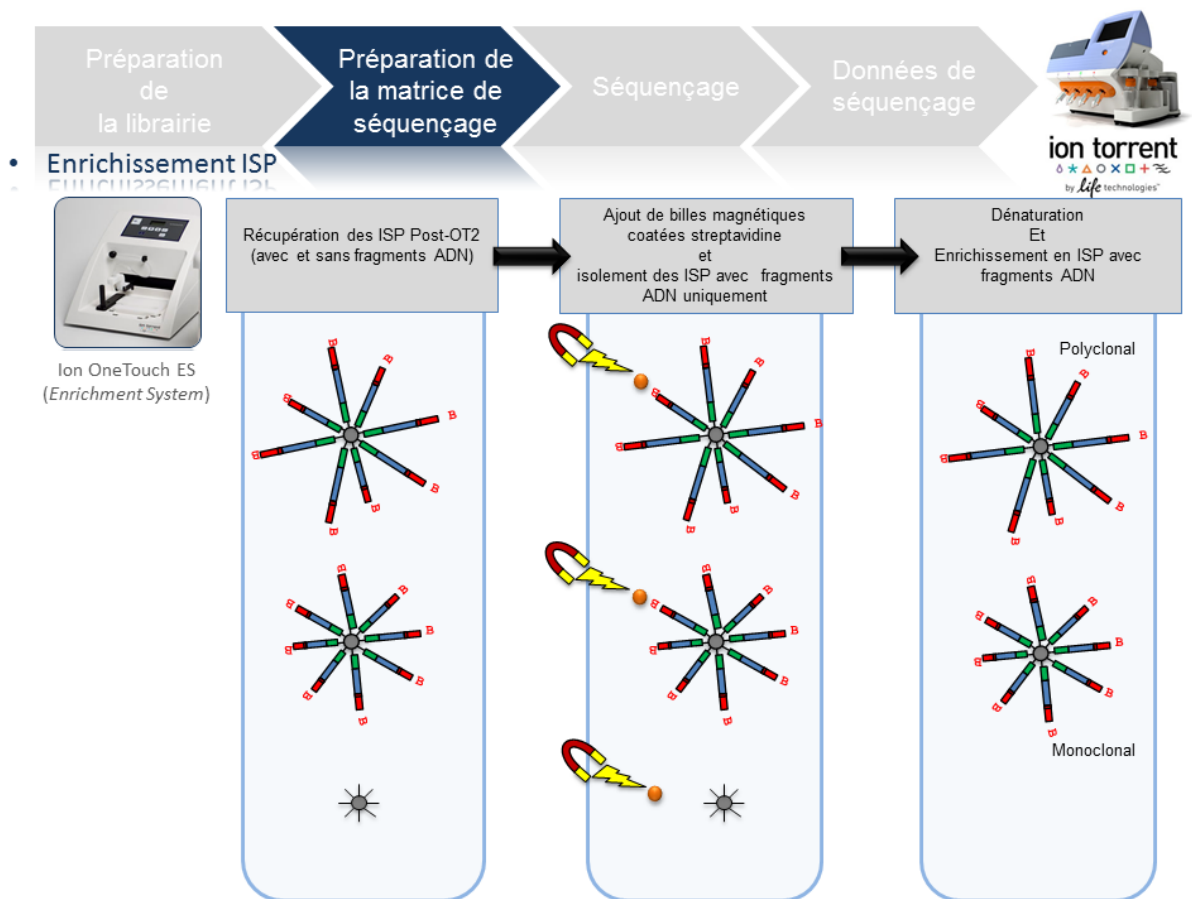


Figure 11: Etape de l'enrichissement (R. BLERVAQUE et al. 2013)



### 2.6.2.3. *Analyse Bio-informatique :*

A l'issue du « run » de séquençage, l'ensemble des données brutes est regroupé dans des fichiers qui seront transférés depuis le Ion PGM vers le Torrent Server, où ils seront traités par le programme Torrent Suite.

L'algorithme de « base calling » permet la conversion des données sous forme de lettres en séquences (A,T,C,G) formant le « read » (séquence au format .FASTA), et un score de qualité pour chaque base incorporée (Score Phred codé en ASCII). Les deux types de données étant associés dans un fichier .FASTQ (FASTA+Qualité). Les scores de qualité Phred sont reliés de façon logarithmique à la probabilité d'erreur d'identification d'une base.

Un prétraitement est également appliqué sur la base des « read » générés équivalent au nombre d'ISPs vivantes ou « Live ISPs » (On parle d'ISPs vivantes pour les ISPs associées à la clé) :

- trimming : élimination des adaptateurs et/ou portions de read de mauvaise qualité (quality value QV>12).
- filtres : élimination des « read » de petites tailles, de mauvaise qualité, des polyclonaux. L'ensemble de ces informations est repris à travers un « report » généré à l'issue du séquençage et du prétraitement. Le nombre de read générés ainsi que leur taille moyenne y sont également renseignés.

Le logiciel qu'on utilise dans notre service pour analyser les données brutes est Ion Reporter™ Software (<https://ionreporter.thermofisher.com/ir/>). Il s'agit d'un outil d'analyse de données hébergé et sécurisé qui simplifie l'informatique associée aux tests de routine autour de la variation de l'ADN.

Afin d'automatiser nos besoins en analyse de données et exploiter la puissance du logiciel Ion Reporter <sup>TM</sup> directement à partir du logiciel Torrent Suite <sup>TM</sup>, il suffit de sélectionner « Créer un flux de travail » Ion Reporter <sup>TM</sup> lors de la planification des opérations de séquençage. Le logiciel automatise le reste de l'analyse principale au transfert de données, en passant par l'analyse secondaire dans Ion Reporter <sup>TM</sup> Software.

L'intégration directe à Torrent Suite <sup>TM</sup> signifie qu'il n'est pas nécessaire de saisir les informations deux fois. Toutes les données sont automatiquement transférées après le séquençage. Ion Reporter <sup>TM</sup> fournit des outils simples pour gérer, suivre et mettre en évidence tout échantillon ou analyse, facilitant ainsi l'organisation et permettant d'accéder à nos données depuis n'importe quel ordinateur.

Le logiciel Ion Reporter <sup>TM</sup> utilise les protocoles de sécurité les plus récents, notamment le cryptage 256 bits et le transfert sur SSL, garantissant la sécurité des données. Ion Reporter <sup>TM</sup> Software repose sur le fournisseur de cloud leader du secteur, Amazon AWS, qui dessert de nombreuses agences académiques, commerciales et gouvernementales. Les systèmes de Ion Reporter <sup>TM</sup> ont été revus et testés par des professionnels de sécurité indépendants (Figure 13).

Ion Reporter <sup>TM</sup> simplifie l'annotation des variants de fichiers VCF en exploitant une multitude de contenus accessibles au public. On télécharge nos fichiers format VCF et on annote les variants avec les informations provenant de bases de données connues, telles que dbSNP, 1000 génomes, OMIM, COSMIC, etc. Mais généralement dans notre département et spécifiquement dans notre étude, on a téléchargé les annotations personnalisées propres à nous et on les a exploitées dans le logiciel Ion Reporter <sup>TM</sup>, nous disposons également dans notre laboratoire d'une base de données « *in house* » que nous interrogeons pour les variants détectés.



• **Obtention des données et du rapport de séquençage**



**Library Summary**  
Based on Predicted Per-Base Quality Scores - Independent of Alignment

Total Number of Bases [Mbp]	114.37
- Number of Q20 Bases [Mbp]	91.27
Total Number of Reads	527,111
Mean Length [bp]	216
Longest Read [bp]	575

**Ion Sphere™ Particle (ISP) Identification Summary**

Total Addressable Wells	1,262,519	Count	Percentage
- Wells with ISPs	1,115,206		88%
- Live ISPs	1,085,018		97%
- Test Fragment ISPs	10,329		<1%
- Library ISPs	1,074,456		99%

**Plugin Summary**

FastQC - v1.0.1	Completed
FastQC_reports.html	Completed
FastqCreator - v1.4.49884	Completed

« Q » correspond au score de qualité *phred*, qui est relié de façon logarithmique à la probabilité d'erreur d'identification d'une base P

$$Q = -10 \log_{10} P$$

Score de qualité phred	Probabilité d'une identification incorrecte	Précision de l'identification d'un base
10	1 pour 10	90 %
20	1 pour 100	99 %
30	1 pour 1000	99.9 %
40	1 pour 10000	99.99 %
50	1 pour 100000	99.999 %

**% enrichissement =**

$$\frac{\text{Library ISPs}}{\text{Wells with ISPs} - \text{TF ISPs}}$$

Library ISPs: 1,074,456  
Wells with ISPs: 1,115,206  
TF ISPs: 10,557

```
AATGTTGGGAGG...
GAGTGGTGGG...
GGTGGTGGG...
GGTGGTGGG...
GGTGGTGGG...
GGTGGTGGG...
GGTGGTGGG...
GGTGGTGGG...
GGTGGTGGG...
GGTGGTGGG...
```

Figure 13: Analyse des données brutes



## ❖ Outils bioinformatiques :

- **Ensembl** : (<http://www.ensembl.org/index.html>)

C'est un navigateur de génome pour les génomes de vertébrés qui soutient la recherche en génomique comparative, évolution, variation de séquence et régulation de la transcription. Ensembl annote les gènes, calcule plusieurs alignements, prédit la fonction de régulation et collecte des données sur les maladies. Les outils Ensembl incluent BLAST, BLAT, BioMart et VEP (Variant Effect Predictor) pour toutes les espèces prises en charge.

- **UCSC** : (<https://genome.ucsc.edu>)

C'est un navigateur en ligne de génome et il est hébergé par l'Université de Californie à Santa Cruz (UCSC). Ce site web interactif offre un accès aux données de séquence du génome d'une variété d'espèces de vertébrés et d'invertébrés et d'organismes modèles majeurs. Il est intégré à une vaste collection d'annotations alignées. Ce navigateur est une visionneuse graphique optimisée pour la prise en charge de performances interactives rapides. Il s'agit d'une suite d'outils web à source ouverte, construite à partir d'une base de données MySQL, qui permet une visualisation, un examen et une interrogation rapides des données à plusieurs niveaux. Sur le site Web de UCSC Genome Bioinformatics. On y retrouve également la base de données du navigateur Genome, les outils de navigation, les fichiers de données téléchargeables ainsi que la documentation.

- **Primer3** : ([www.primer3.ut.ee](http://www.primer3.ut.ee))

C'est le logiciel que nous avons utilisé pour désigner les amorces de la séquence cible, ainsi que toutes les informations relatives aux amorces synthétisées (taille des primers et le pourcentage en GC, la Tm et la taille de l'amplicon).

- **SIFT** (Sorting Intolerant from Tolerant) :

Il s'agit d'un outil bioinformatique basé sur l'alignement de séquences. Il permet d'élaborer un score qui classe une variation comme : intolérée, potentiellement intolérée, limite ou tolérée.

Un score SIFT prédit si une substitution d'acide aminé affecte la fonction protéique. Il varie de 0,0 (délétère) à 1,0 (toléré). Le score peut être interprété comme suit :

- 0,0 à 0,05 : Les variantes dont les scores se situent dans cet intervalle sont considérées comme délétères. Les variantes avec des scores plus proches de 0,0 sont prédites avec plus de certitude comme délétères.
- 0,05 à 1,0 : Les variantes avec un score compris dans cette marge devraient être tolérées (bénignes). Les variantes avec des scores très proches de 1,0 devraient être tolérées avec plus de confiance.

Remarque : les scores PolyPhen-2 et SIFT utilisent les mêmes limites, de 0,0 à 1,0, mais avec des significations opposées. Une variante avec un score PolyPhen de 0,0 devrait être bénigne. Une variante avec un score SIFT de 1,0 devrait être bénigne.

- ***PolyPhen-2*** (Polymorphism Phenotyping) <http://genetics.bwh.harvard.edu/pph2>

Il utilise des données des séquences et de structures, quand la structure expérimentale de la protéine d'intérêt ou d'un homologue partageant plus de 50% d'identité est disponible.

- ***MutationTaster*** :

logiciel qui évalue le potentiel pathogène des altérations de la séquence d'ADN. Il est conçu pour prédire la fonction conséquence des substitutions d'acides aminés des courtes insertions et/ou délétions (indel) et des variantes au niveau des introns et des jonctions introns-exons.

### **III. RESULTATS**

## 1. Résultat de Famille avec XP :

Le séquençage Sanger a conduit à l'identification d'une nouvelle insertion homozygote de quatre bases CATG en position 1644 dans le neuvième exon du gène *XPC*.

- Sur le plan protéique, cette insertion entraîne un changement d'acide glutamique en Alanine du codon 550 avec un décalage du cadre de lecture conduisant à un codon stop prématuré après 25 Acides aminés.

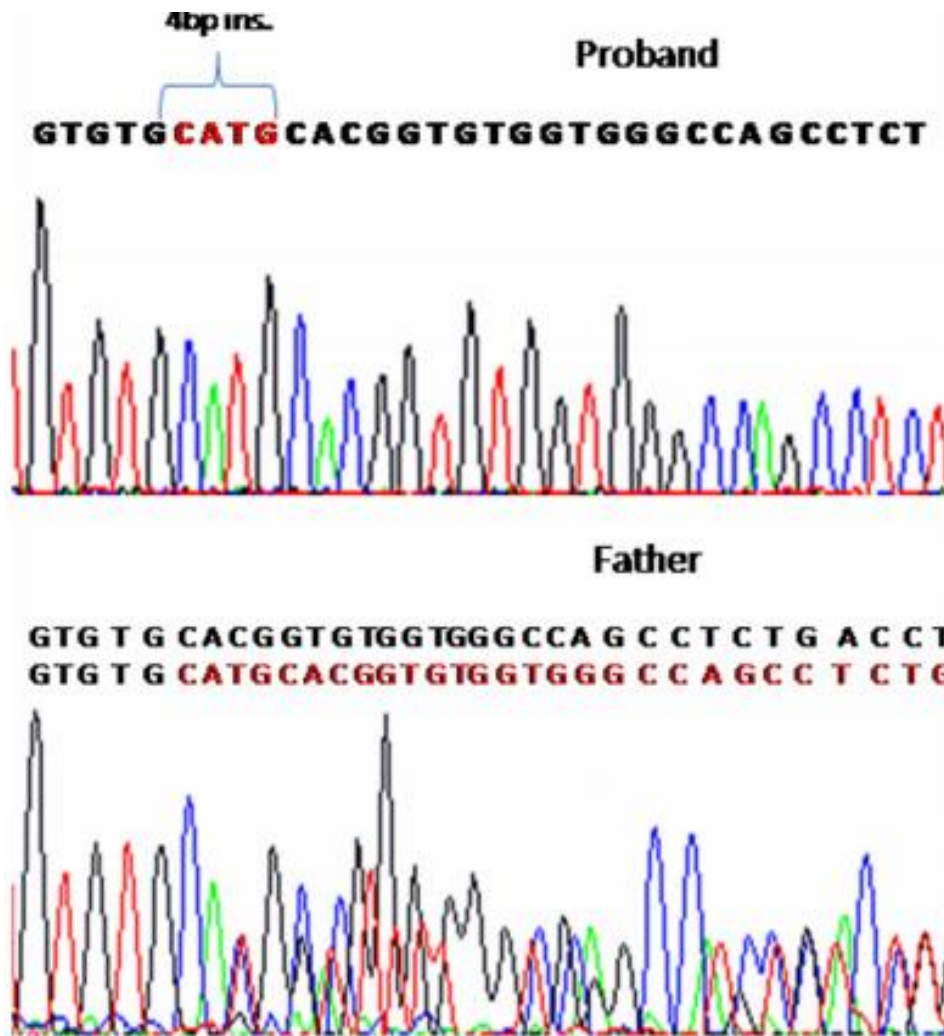
Le nom de la mutation est rédigé ainsi : c.1644\_1645insCATG (p.G550Afs\*25) (NM\_004628).

- Il s'agit d'une nouvelle mutation jamais rapportée dans les bases de données ExAC et 1000G.

Les parents ont été testés et se sont avérés hétérozygotes pour la même mutation (Figure 14).

### Confirmation de la pathogénicité du variant :

- Il s'agit d'une mutation de type insertion entraînant la formation d'un codon stop prématuré et par conséquent une protéine tronquée.
- Sur les sites de prédiction protéique Mutationtaster :
  - Il s'agit d'une mutation dont la position est :
    - chr3:14199738\_14199739insCATG au niveau chromosomique
    - g.20545\_20546insCATG au niveau génomique
  - Elle se trouve dans une région hautement conservée entre les différentes espèces et elle est prédite comme étant pathogène.



*Figure 14: Electrophorégramme de l'exon 9 montrant une insertion CATG à l'état homozygote chez le propositus et à l'état hétérozygote chez le père*

## 2. Résultats des familles avec AF :

L'analyse bioinformatique a montré la présence de variants au niveau des gènes étudiés.

Après filtration des variants, nous avons pu identifier des variations pouvant être potentiellement causales de l'AF chez nos patients :

- Une mutation nonsens au niveau du gène *FANCG* chez le premier patient avec une couverture de 174.
- Une mutation nonsens à l'état homozygote au niveau du gène *FANCG* chez le deuxième patient avec une couverture de 466.
- Une délétion homozygote d'un exon au niveau du gène *FANCA* chez le troisième patient.
- Aucune mutation au niveau des 3 gènes chez le quatrième patient.
- Une mutation nonsens homozygote au niveau gène *FANCA* chez le cinquième patient.
- Deux mutations au niveau du gène *FANCA* à type de délétions d'une base et de trois bases chez le sixième patient avec une couverture de 492 et de 288.
- Une mutation faux-sens à l'état homozygote au niveau du gène *FANCG* avec une couverture de 344 chez le septième patient.
- Une mutation faux-sens homozygote au niveau du gène *FANCA* chez le huitième patient avec une couverture de 559.
- Une mutation homozygote faux-sens au niveau du gène *FANCA* chez le neuvième patient.

Par conséquent, l'analyse moléculaire a montré des variations dans deux des trois gènes étudiés chez huit des neuf patients testés (88% des cas).

Le séquençage massif a révélé neuf variations de séquence répondant tous à un mode de transmission autosomique récessive dont trois nouvelles. Tous les patients étaient homozygotes sauf un qui était hétérozygote composite (Tableau 3).

### a. Gène *FANCA*:

62,5% des patients portent des mutations du gène *FANCA* :

- Le 3<sup>ème</sup> patient a une délétion homozygote de l'exon 15, une mutation déjà connue pathogène.
- Le 5<sup>ème</sup> patient a une substitution non-synonyme au niveau de l'exon 13. C'est une substitution de la cytosine en thymine en position 1126 : c.1126C>T.

Il s'agit d'une mutation non trouvée ni dans ExAC, ni dans 1000 Genomes Browsers. Cette variation crée un codon stop prématuré en position 376 dans la protéine FANCA (p.Gln376\*) et peut entraîner un produit protéique absent ou perturbé. Cette mutation a été déjà rapportée chez un patient atteint d'AF.

- Le 6<sup>ème</sup> patient est hétérozygote composite pour deux mutations :
  - c.3520\_3522delTGG : variante rapportée comme pathogène chez les personnes atteintes d'anémie de Fanconi (Leiden Open-source Variation Database et ClinVar). Cette mutation respectant le cadre de lecture entraîne la délétion d'un acide aminé de la protéine FANCA en position 1174 (p.Trp1174del).
  - c.2189delT : il s'agit d'une délétion de la thymine en position 2189 au niveau de l'exon 24. C'est une nouvelle mutation qui n'a pas été retrouvée dans les bases de données de variantes. Cette délétion entraîne un décalage du cadre de lecture aboutissant à la création d'un codon stop prématuré après 21 acides aminés (p.Leu730Argfs\*21).
- Le 8<sup>ème</sup> patient est porteur de la mutation faux-sens au niveau de l'exon 14 du gène *FANCA*. C'est une substitution de guanine en adénine en position 1304 (c.1304G>A, p.Arg435His). Cette mutation a été déjà rapportée comme étant pathogène dans la base de données clinvar (rs1060501879).
- Le 9<sup>ème</sup> patient est porteur d'une variation au niveau de l'exon 29. Il s'agit d'une substitution faux-sens de cytosine en thymine qui entraîne au niveau protéique un changement d'Arginine en tryptophane (c.2851C> T, p.Arg951Trp). Cette variation a déjà été rapportée dans clinvar comme étant probablement pathogène (rs755546887).

#### **b. Gène *FANCG* :**

Des variations du gène *FANCG* ont été trouvées chez 37,5% des patients dont une mutation connue pathogène et deux nouvelles variations jamais décrites auparavant.

- La 1<sup>er</sup> patient est porteur d'une variation de séquence localisée dans l'exon 11. Il s'agit d'une substitution de la Guanine par Thymine en position 1474 (c.1474G>T) conduisant à la création d'un codon stop prématuré au niveau du codon 492 (p.Glu492\*). Cette variante n'est pas présente dans les bases de données de population (ExAC et 1000G) ni dans les bases clinvar et GNOMAD.
- Le 2<sup>ème</sup> patient a une mutation dans l'exon 13. C'est une substitution de cytosine en thymine en position 1642 qui aboutit au niveau protéique à un codon stop prématuré au codon 548 (c.1642C>T / p.Arg548\*).

Cette variante est présente dans les bases de données de population (rs779834525). Cette variante a été rapportée chez des personnes atteintes d'anémie de Fanconi (PMID: 11093276, 11438206). Les variants entraînant une perte de fonction dans le gène *FANCG* sont connus pour être pathogènes (PMID: 12552564). Pour ces raisons, cette variante a été classée comme pathogène.

- Le 7<sup>ème</sup> patient est porteur d'une mutation faux-sens à l'état homozygote dans l'exon 8. Il s'agit d'une substitution d'adénine en thymine (c.1034A>T). Sur la protéine FANCG, ce changement de séquence se traduit par le remplacement de la Glutamine par la Leucine en position 345 (p.Gln345Leu). Basée sur des outils et logiciels de prédiction d'acides aminés, cette substitution est probablement pathogène selon Polyphen2, intolérée selon SIFT et il s'agit d'une mutation causale de la maladie dans MutationTaster.



Gène	Patient	Exon	Type de mutation	Variation au niveau du cDNA	Variation protéique	Génotype
FANCA	P3	15	Délétion	délétion de l'Exon 15		homozygote
	P5	13	Substitution nonsens	c.1126C>T	p.Gln376*	homozygote
	P6	36	délétion non frameshift	c.3520_3522delTGG	p.Trp1174del	Hétérozygote composite
		24	délétion frameshift	c.2189delT	p.Leu730fs	Hétérozygote composite
	P8	14	Substitution faux sens	c.1304G>A	p.Arg435His	homozygote
	P9	29	Substitution faux sens	c.2851C>T	p.Arg951Trp	homozygote
FANCG	P1	11	Substitution nonsens	c.1474G>T	p.Glu492*	homozygote
	P2	13	Substitution nonsens	c.1642C>T	p.Arg548*	homozygote
	P7	8	Substitution faux sens	c.1034A>T	p.Gln345Leu	homozygote

**Tableau 3: Mutations des gènes FANC identifiées par NGS chez huit patients**

## **IV. DISCUSSION**

# 1. Techniques de séquençage de l'ADN :

## 1.1.Histoire du séquençage de l'ADN :

La détermination de la séquence ou de l'ordre d'enchaînement des bases d'acide nucléique dans les échantillons biologiques fait partie intégrante d'une grande variété d'applications de recherche. Au cours des cinquante dernières années, un grand nombre de chercheurs se sont intéressés à produire des techniques et des technologies pour séquencer des molécules d'ADN et d'ARN. Grâce à ces progrès, on est passé du séquençage d'un nombre court d'oligonucléotides à des millions de bases, de la lutte pour la déduction de la séquence codante d'un seul gène au séquençage rapide du génome entier [6].

Le début du séquençage de l'ADN remonte aux années 1970 par la découverte de deux méthodes développées indépendamment, l'une aux États-Unis par l'équipe de Walter Gilbert, et l'autre au Royaume-Uni par l'équipe de Frederick Sanger en 1977. Il s'agit de deux méthodes fondées sur des principes différents et pour lesquelles Gilbert et Sanger ont été récompensés par le prix Nobel de chimie en 1980 [7,8] :

- L'approche de Maxam et Gilbert est une méthode basée sur la **dégradation chimique**.
- La méthode de Sanger est une approche par **synthèse enzymatique** appelée terminaison de chaîne.

Le premier organisme biologique dont le génome a été séquencé est le virus bactériophage  $\phi$ X174 [9]. La première méthode d'ADN semi-automatisée a été développée par Lorey Smith en 1986 [10].

Après l'achèvement de la détermination de la séquence de référence du génome humain euchromatique et haploïde en 2003 [11], l'ère génomique humaine personnalisée est devenue

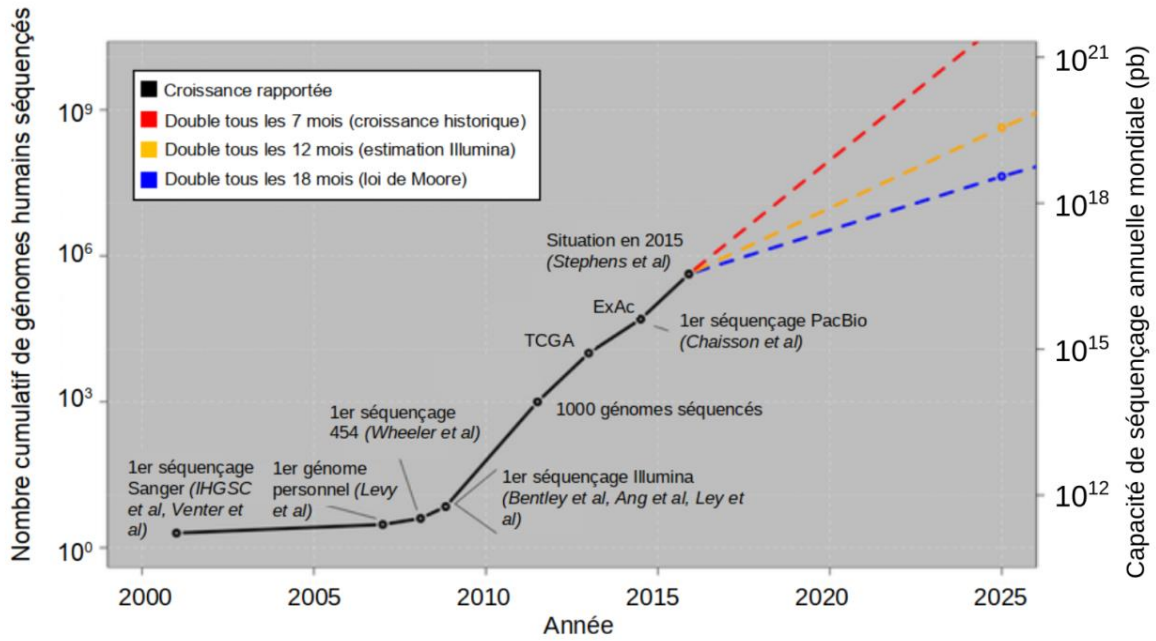
possible grâce au développement au cours des deux dernières décennies, de nouvelles technologies de séquençage plus rapides et moins onéreuses. La diminution du prix du séquençage du génome humain s'est accompagnée d'une élévation du nombre de génomes séquencés [12,13] (Figures 15 et 16). Ceci a permis non seulement de générer des séquences diploïdes personnelles du génome humain contribuant à une meilleure compréhension de la variation du génome humain mais également d'analyser à l'échelle du génome les mécanismes épigénétiques.

Le terme « Next generation » a impliqué une prochaine étape dans le développement de la technologie de séquençage de l'ADN. Ce terme a été remplacé par le terme seconde génération, troisième génération, etc.

Peu de temps après l'apparition de NGS (séquençage 2<sup>ème</sup> génération), les technologies troisième génération (Third Generation Sequencing : TGS) ont émergé. La définition des différents générations de technologie de séquençage de l'ADN a été largement discutée, en particulier en ce qui concerne la distinction entre la deuxième et la troisième génération [14,15,16,17].

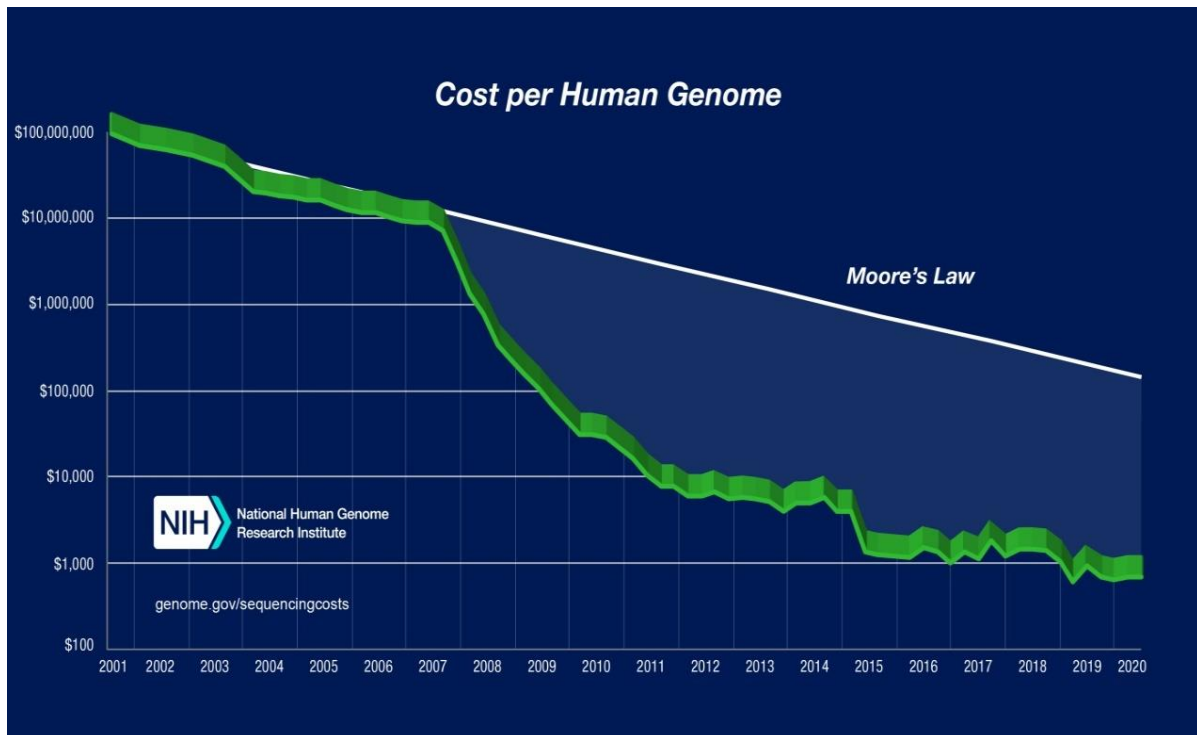
Il est actuellement admis que les caractéristiques distinctives du TGS sont le séquençage d'une molécule unique (SMS) et le séquençage en temps réel (par opposition au NGS, où le séquençage est interrompu après chaque incorporation de base) [18].

Ces méthodes de troisième génération permettent la lecture de fragments de grandes tailles et des éléments répétitifs du génome produisant ainsi des reconstructions plus contiguës du génome.



**Figure 15: Evolution des capacités de séquençage**

(adaptée de Stephens et coll. par Erwin van Dijk) [15]



**Figure 16: Evolution du coût du séquençage [15]**

## 1.2.Séquençage 1ère génération : Séquençage Sanger

Depuis sa description et jusqu'à ce jour, elle reste la plus utilisée dans les laboratoires, notamment en milieu hospitalier. Son principe repose sur l'incorporation par l'ADN polymérase de manière aléatoire et inconstante de didéoxyribonucléotides triphosphates (ddNTP: ddATP, ddGTP, ddCTP et ddTTP) qui sont marqués à l'aide de fluorophores différents pour chaque ddNTP (fluorophores « quatre couleurs »). Lors de la synthèse de l'ADN complémentaire à partir d'une amorce spécifique, quand un ddNTP est incorporé à la place d'un dNTP, la réaction d'extension s'arrête car l'ADN polymérase ne peut plus continuer sa polymérisation. En effet, le didéoxynucléotide ne possède pas de groupe 3 - hydroxyle indispensable à la réaction de polymérisation de l'enzyme. Statistiquement, un ddNTP complémentaire sera incorporé au moins une fois à la place d'un dNTP pour chaque « base » de l'ADN cible aboutissant, à la fin de la réaction, à des fragments de taille différente. L'analyse de la réaction est ensuite effectuée par électrophorèse capillaire. Lors de la migration, chaque fragment (contenant un ddNTP marqué par un fluorophore) sera excité par un laser et le signal obtenu sera analysé par un logiciel spécifique. L'analyse informatique des signaux permet d'obtenir la séquence étudiée sous forme d'un électrophorégramme. En général, cette technique permet d'obtenir des séquences de longueur comprise entre 400 et 850 pb [19] (Figure 17).

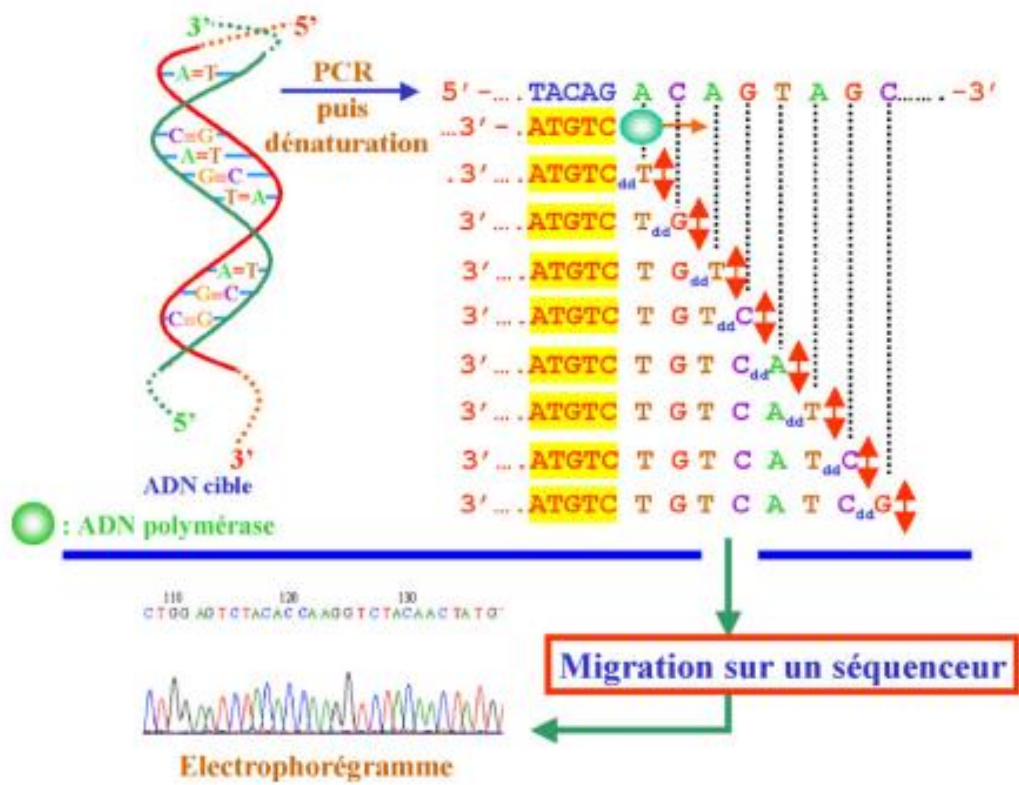


Figure 17: Principe du séquençage Sanger [19]

### **1.3.Séquençage deuxième génération (Next Generation Sequencing) :**

Les méthodes de séquençage de deuxième génération peuvent être regroupées en deux grandes catégories, le séquençage par hybridation et le séquençage par synthèse (SBS) [20].

- **Séquençage par hybridation :** Cette méthode a été développée dans les années 1980, en utilisant des oligonucléotides d'ADN en réseau de séquence connue sur des filtres qui ont été hybridés à des fragments marqués de l'ADN à séquencer. En s'hybridant à plusieurs reprises et en éliminant l'ADN non hybridé indésirable, il était possible de déterminer si les fragments marqués s'hybridaient correspondaient à la séquence des sondes d'ADN sur le filtre. En se basant sur des informations de chevauchement provenant des spots d'hybridation de sonde, il était ainsi possible de construire de plus grandes informations de séquence contiguë. Le séquençage par hybridation a été largement relégué à des technologies qui dépendent de l'utilisation de sondes spécifiques pour interroger des séquences, comme dans les applications de diagnostic pour identifier les SNP liés à la maladie (polymorphismes mononucléotidiques) dans des gènes spécifiques ou identifier des anomalies chromosomiques macroscopiques (réarrangements, délétions, duplications, variantes du nombre de copies, CNV) [21, 22, 23, 24, 25] (Figure 18).



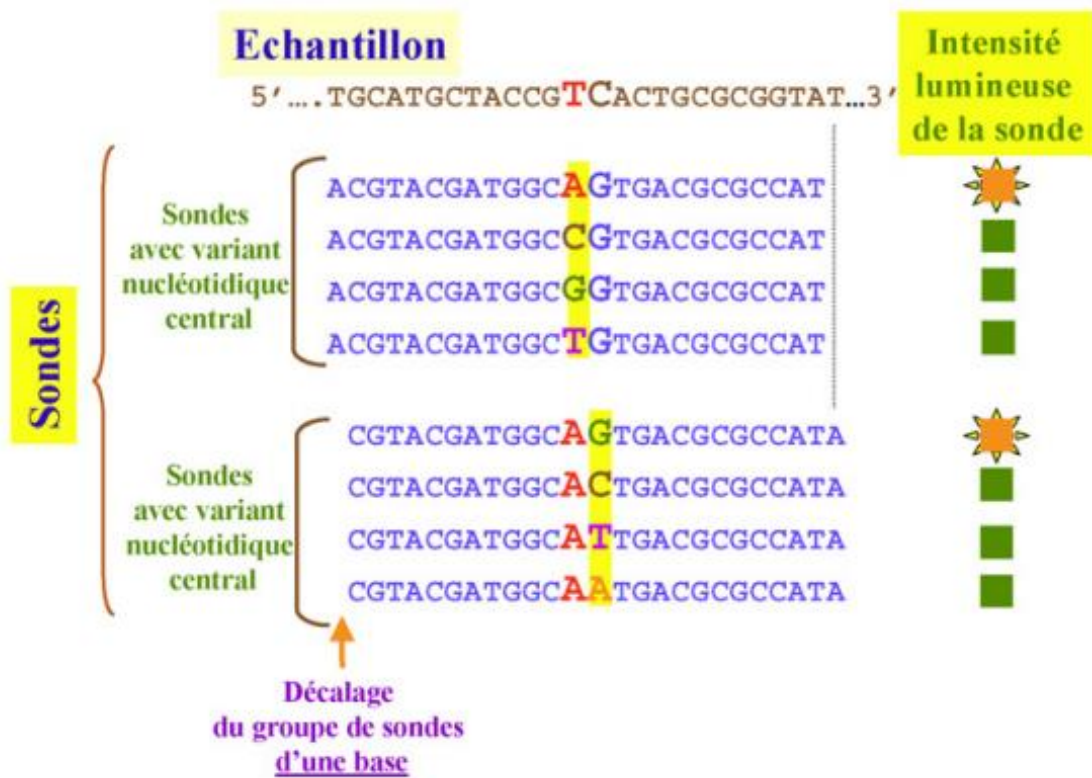


Figure 18: Principe général du séquençage par hybridation (SBH) [19]

- Séquençage par synthèse (SBS) : Les méthodes de séquençage par synthèse (SBS) ont adopté plusieurs approches distinctes [26, 20, 27, 28, 29].

Les méthodes de « deuxième génération » utilisent généralement un support solide contenant des microcanaux ou puits dans lesquels se produisent les réactions de séquençage. En général, la plupart de ces nouvelles méthodes SBS n'utilisent pas de terminateurs didésoxy, bien que des terminateurs « réversibles » soient utilisés dans certaines technologies, qui permettent aux réactions d'incorporation de nucléotides de se dérouler normalement tout en imaginant les nucléotides incorporés, puis en supprimant les fragments bloquant la synthèse sur le marqueur pour permettre l'incorporation de la base suivante dans la séquence. Les méthodes SBS actuelles diffèrent de l'approche du séquençage Sanger original par le fait qu'elles reposent sur des lectures beaucoup plus courtes (actuellement jusqu'à environ 300 à 500 bases). En outre, ils ont généralement un taux d'erreur intrinsèquement plus élevé par rapport au séquençage de Sanger et reposent sur une couverture de séquence élevée « séquençage massivement parallèle » de millions à des milliards de lectures de séquences d'ADN courtes (50 à 300 nucléotides). Cependant, pour certaines technologies, des erreurs de séquence se produisent et ne peuvent pas toujours être corrigées en augmentant le nombre de lectures. Par exemple, des séquences d'homopolymères se produisent lorsque l'ADN contient des multiples consécutifs de la même base, tels que AAAAAAAAAA. Dans ce cas, les plates-formes de séquençage sont limitées pour déterminer avec précision le nombre de bases consécutives. Chaque plate-forme génère son propre ensemble d'erreurs potentielles de séquence et les utilisateurs doivent être conscients de ces limitations. L'utilisation de plusieurs plates-formes avec des technologies différentes est souvent réalisée pour contourner ces problèmes.

En 2011, trois nouvelles plates-formes de séquençage majeures ont vu la lumière dont chacune possède sa propre technologie de séquençage :

- PGM d'Ion Torrent-Life Technologies : séquençage par mesure ionique
- Roche 454-Life Science : pyroséquençage
- Illumina MiSeq : mesure de fluorescence

Deux étapes sont communes à tous les SGS : la préparation des échantillons (template) et le séquençage. La préparation des échantillons comporte trois étapes supplémentaires : l'extraction d'acide nucléique, la préparation de la librairie et l'amplification de librairie. Le seul fragment d'ADN à séquencer est soit lié à des billes, à des surfaces ioniques ou à une cuve à circulation.

En effet, chaque plate-forme a ses spécificités et le choix de telle ou telle plateforme dépend de plusieurs paramètres qui sont propres à chaque laboratoire qu'elles soient d'ordre financier (le budget) ou d'ordre technique (les applications du NGS et le débit des tests).

#### *a. Pyroséquençage de Roche-Life Science :*

La plateforme Roche-Life Science est le premier système de NGS à avoir connu un succès commercial. Cette plateforme utilise une technique de pyroséquençage qui repose sur la détection du pyrophosphate libéré lors de l'incorporation de nucléotides [3,30].

Après préparation de la matrice par emPCR [31], les billes amplifiées par clonage sont piégées dans les puits de la plaque PTP (Pico Titer Plate), le pyrophosphate (PPi) libéré lors de la réaction de polymérisation de l'ADN est détecté et utilisé comme indicateur d'incorporation de base spécifique. Au cours de l'ajout des quatre bases de manière séquentielle, l'un des dNTP (dATP, dGTP, dCTP, dTTP) sera complémentaire aux bases du brin matrice et lorsqu'une base est incorporée dans l'ADN complémentaire par l'ADN polymérase, un pyrophosphate (PPi) est libéré de manière égale à la quantité de nucléotide

incorporé. En présence de l'ATP sulfurylase, la luciférase, la luciférine et l'adénosine 5' phosphosulfate (APS), le pyrophosphate sera utilisé dans une série de réactions enzymatiques aboutissant à la production de la lumière qui sera mesurée (figure 19).

En même temps, les bases incomparables sont dégradées par l'apyrase. Ensuite, un autre dNTP est ajouté dans le système réactionnel et la réaction de pyroséquençage est répétée [20,32]. Cette technique est assez robuste mais très onéreuse.

La limite du pyroséquençage réside dans le séquençage d'homopolymères car il ne peut détecter plus de cinq nucléotides identiques [33].

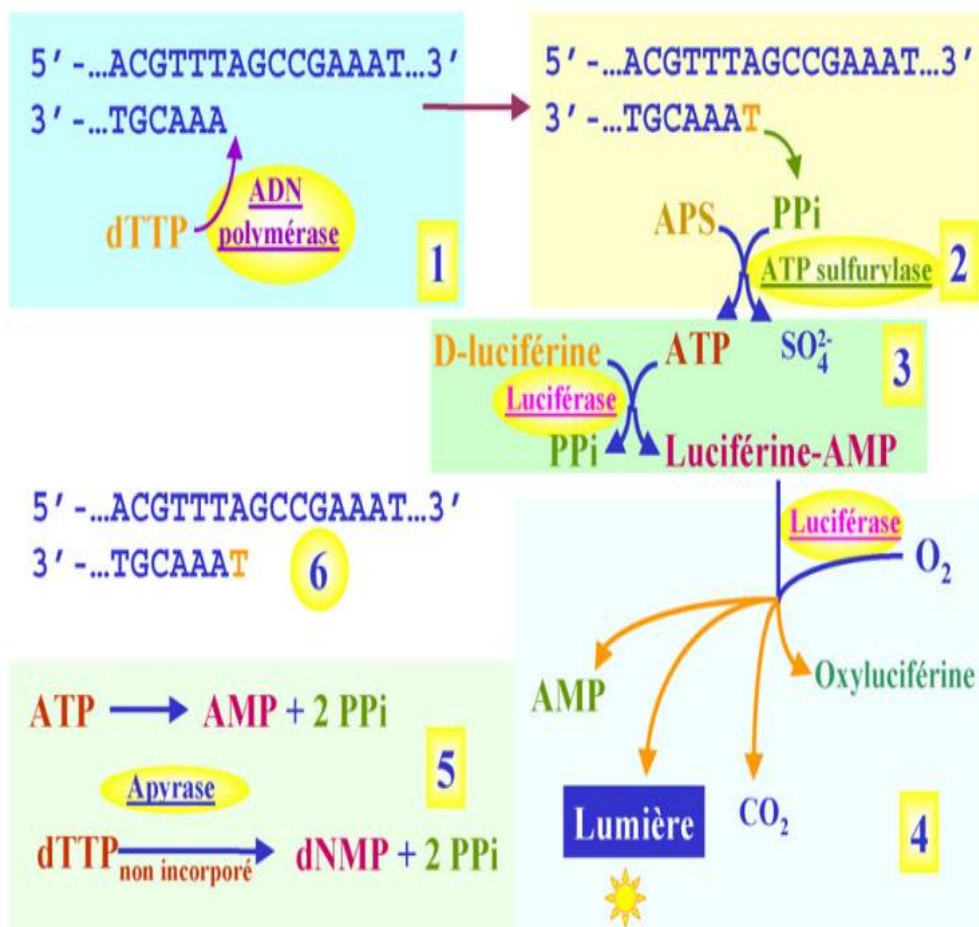


Figure 19: Principe du pyroséquençage [19]

***b. Séquençage par terminaison réversible de Illumina :***

Développée pour la première fois par Solexa et Lynx Therapeutics, le séquençage Illumina est basé sur une technique connue sous le nom d' « amplification par pont » sur un support solide. Il s'agit d'une lame de verre appelée Flowcell (FC) couverte par des amorces à la surface, les molécules d'ADN (environ 500 pb), liées à chaque extrémité avec des adaptateurs appropriés, sont utilisées comme substrats pour des réactions de synthèse d'amplification répétées. Les adaptateurs vont leur permettre de se fixer de façon aléatoire sur la FlowCell par hybridation sur les amorces. Lorsqu'un nouveau brin est synthétisé par une polymérase, il se fixe de façon covalente à la FC. Le brin d'origine est alors éliminé par dénaturation, et l'extrémité libre du brin restant s'hybride à une amorce adjacente pour former un pont. La polymérase synthétise à nouveau le brin complémentaire pour former un pont d'ADN double brin puis les copies sont libérées par dénaturation. Le cycle d'amplification en pont recommence pour former à terme un regroupement d'ADN clonal en une zone appelée cluster. Les brins antisens sont ensuite clivés. Il s'agit de la linéarisation [14,34].

L'amplification est suivie d'un séquençage appelé Séquençage par Synthèse (SBS) en utilisant un nucléotide marqué par fluorescence. A chaque cycle de séquençage, il y a émission d'un signal lors de l'incorporation, par l'ADN polymérase, d'un nucléotide complémentaire au brin d'ADN attaché au Flowcell. La mesure de la fluorescence est ensuite convertie en données de séquence [35] (figure 20).

L'erreur de séquençage des homopolymères (comme dans cas du pyroséquençage) est surmontée par cette technique en raison à l'incorporation d'une seule base à la fois. Le séquençage par paires permet aux utilisateurs de séquencer le fragment d'ADN des deux extrémités avec une couverture élevée, un nombre élevé de lectures et plus de données par rapport aux systèmes de séquençage à une seule extrémité. Le séquençage génère des données

de séquence de haute qualité grâce à l'augmentation de la probabilité d'alignement sur une référence. Il facilite la détection des réarrangements génomiques (insertions, suppressions et inversions), éléments de séquence répétitifs, fusions de gènes et nouveaux transcrits. De plus, il offre une couverture très uniforme de l'ensemble du génome humain avec des données de haute précision.

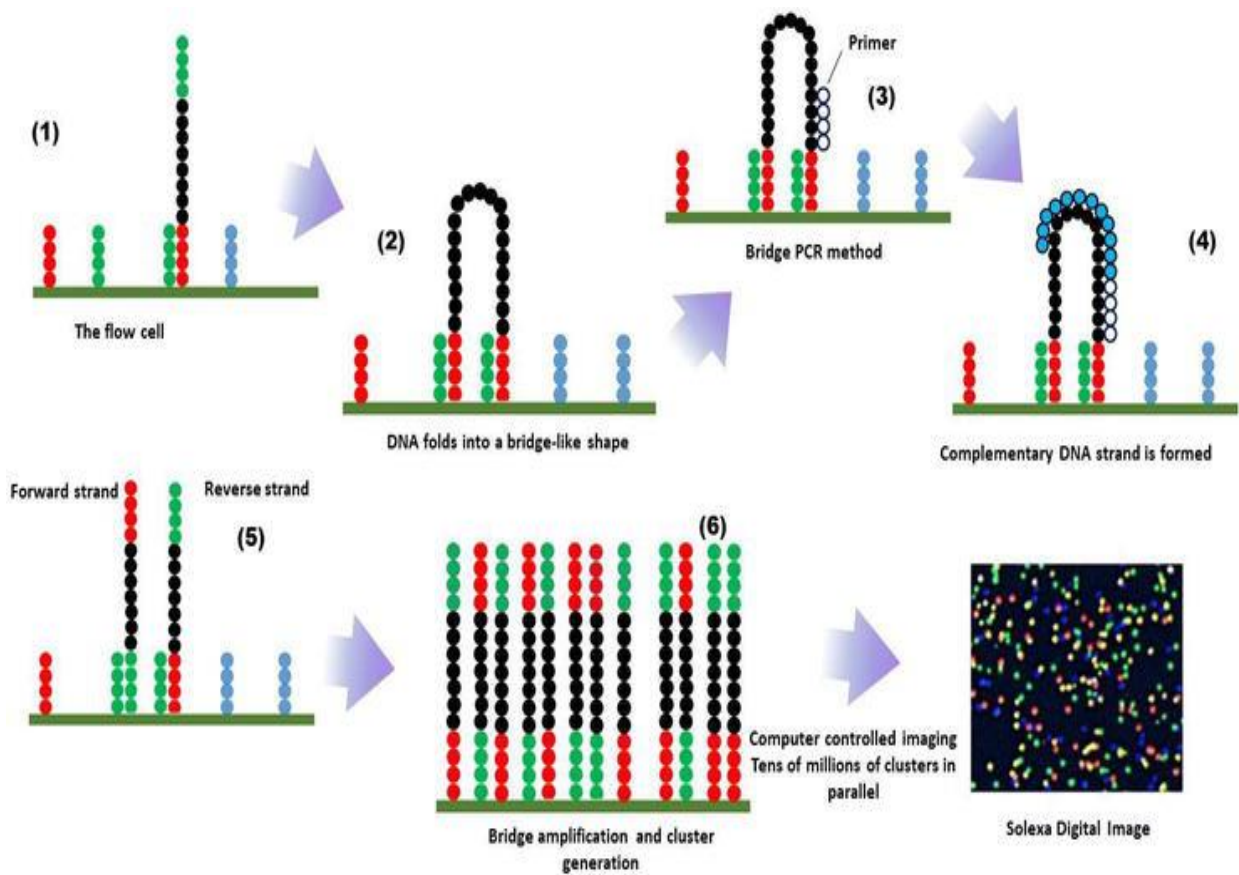


Figure 20: Principe du séquençage par Illumina/Solexa (1).



Les matrices d'ADN liées à des séquences adaptatrices vont se fixer sur une surface en verre (flowcell) via les séquences complémentaires d'oligos (2, 3, 4). Les molécules d'ADN se replient en pont et une amplification PCR en pont est réalisée. (5) Amplification en pont et formation de millions de copies ou de clusters. (6) Le séquençage des clusters est réalisé par la méthode de terminaison réversible cyclique. Enfin, les lectures résultantes (des dizaines de millions) sont analysées et la séquence d'ADN est recodée [36].

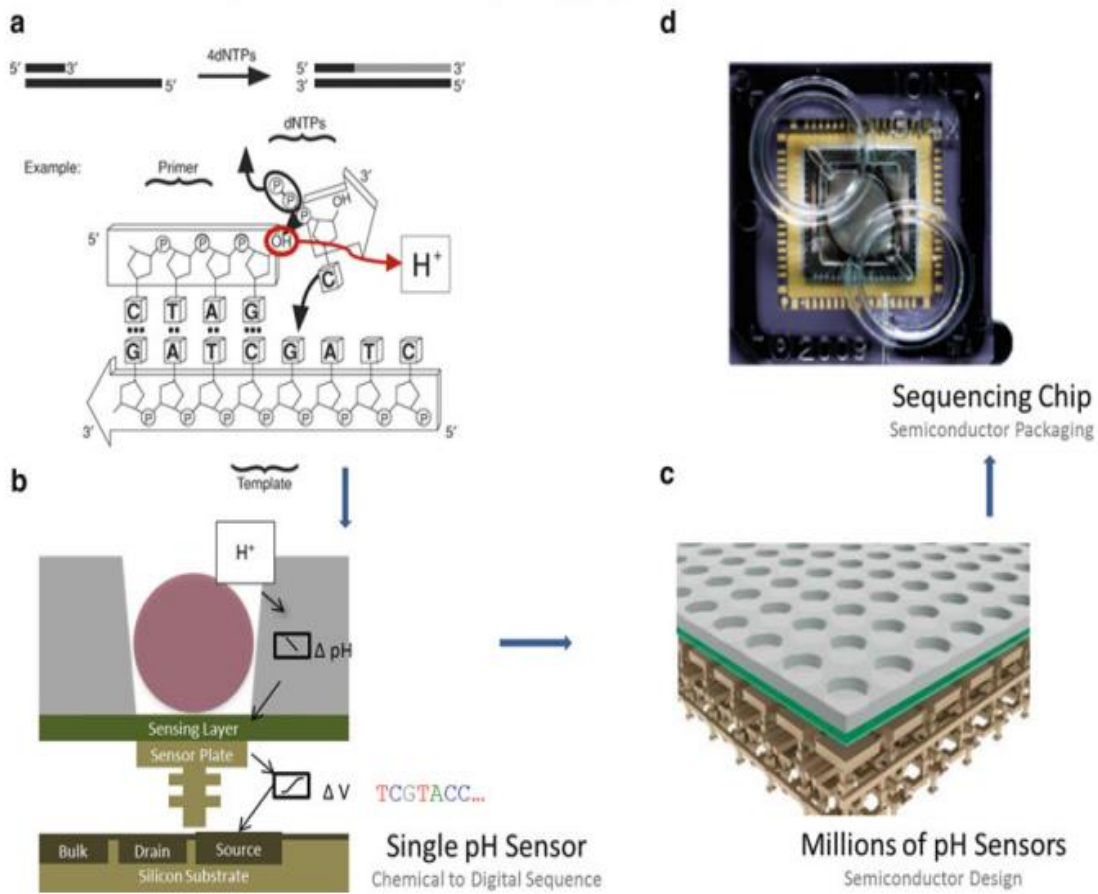
***c. Séquençage par mesure ionique de Ion Torrent-Life Technologies :***

Ion-PGM (Personnel Genome Machine) a été mis sur le marché commercial en fin 2010 par Ion Torrent-Life Technologies (Thermo Fischer Scientific). Le PGM utilise la technologie de séquençage à semi-conducteurs offrant ainsi un autre type de plateforme par rapport aux pyroséquençage et Illumina. La technologie convertit directement la séquence nucléotidique en informations numériques sur une puce semi-conductrice [4]. Dans une réaction de synthèse d'ADN, lorsqu'un nucléotide correct est incorporé en face de sa base complémentaire dans une chaîne d'ADN en croissance, un ion hydrogène est libéré. Cela modifie le pH de la solution qui peut être enregistré comme un changement de tension par un capteur d'ions, comme un pH-mètre. Si aucun nucléotide n'est incorporé, aucun pic de tension ne se produit. En inondant et en lavant séquentiellement une « chambre de séquençage » avec des réagents de séquençage qui n'incluent qu'un seul des 4 nucléotides à la fois, des changements de tension se produisent lorsque le nucléotide approprié est incorporé. Lorsque deux nucléotides adjacents incorporent le même nucléotide, deux hydrogènes sont libérés et la tension double. Cependant, les grandes chaînes d'homopolymères du même nucléotide sont parfois difficiles à discerner. Les réactions de séquençage Ion Torrent se produisent dans des millions de puits qui recouvrent une puce semi-conductrice contenant des millions de pixels qui convertissent les informations chimiques en informations de séquençage [20,21].

Le PGM est le premier appareil de séquençage commercial à ne pas nécessiter de balayage par fluorescence ni par caméra, ce qui se traduit par une vitesse plus élevée, un coût inférieur et une taille d'instrument inférieure par rapport aux autres plate-formes [21].

Le système Ion Torrent est vendu par Thermo-Fisher et plusieurs versions de la plate-forme sont disponibles, notamment le système Ion Personal Genome Machine™ (PGM™), le système Ion Proton™, le système Ion S5 et le système ION S5 XL, chacun avec des débits caractéristiques.

Actuellement, il permet une lecture de 200 pb en 2 heures. Pour commencer le processus, l'ADN est fragmenté en 200-1500 fragments de base qui sont liés à des adaptateurs. Les fragments d'ADN sont attachés à une bille par des séquences complémentaires (sur les billes et sur les adaptateurs). L'amplification clonale des bibliothèques se fait ensuite par émulsion PCR (emPCR) sur la bille (Ion Sphere Particules : ISPs) avec le système OT2/ES (OneTouch2™) ou le Ion Chef™. Le ratio idéal de la PCR en émulsion est 1/1 (Fragment d'ADN de la bibliothèque/ISP). Ce processus permet à des millions de billes d'avoir chacune plusieurs copies d'une séquence d'ADN. Les IPS sont ensuite chargées sur la puce (314, 316 ou 318), contenant des puits de telle sorte qu'une seule bille peut pénétrer dans un puits. Les réactifs de séquençage sont ensuite distribués à travers les puits. Lorsque le nucléotide approprié est incorporé, un ion hydrogène est émis. Cette libération entraîne une modification du pH, et donc du voltage, qui est enregistrée au niveau des senseurs de la puce et convertie en données brutes sous forme de ionogramme, avant d'être transformée en données de séquences. En effet, l'incorporation de nucléotides est directement convertie en tension qui est enregistrée instantanément, accélérant considérablement le processus (Figure 21)[37].



*Figure 21 : Principe du séquençage à semiconducteurs [38].*

Le Département de Génétique Médicale à l'Institut National d'Hygiène s'est équipé de plateforme PGM-Ion Torrent-Life Technologies en mai 2016. Le choix du PGM s'est fait pour deux principales raisons :

1. La rapidité d'avoir des résultats par rapport aux autres plate-formes. En effet, notre laboratoire offre principalement des tests moléculaires pour des maladies génétiques à visée diagnostique et non seulement dans un cadre de recherche. Par conséquent, le facteur de temps est un élément très important afin de rendre des résultats aux patients dans des délais raisonnables.
2. Le prix de la plateforme doit être proche au budget alloué par notre système de santé au DGM avec un souci de meilleur rapport coût-bénéfice.

Dans un objectif de développement des technologies au sein d'une structure publique, le PGM était la plate-forme la plus appropriée pour notre laboratoire.

- La qualité du séquençage est très stable par rapport aux autres technologies de NGS. Par exemple, le HiSeq 2000 de Illumina diminue sensiblement après 50 cycles, ce qui peut être causé par la décroissance du signal fluorescent lorsque la lecture augmente.

- Le débit de mapping de Ion Torrent est correct mais il est difficilement comparable aux autres technologies puisque chacune utilise sa propre méthode d'alignement.

- Malgré la différence significative entre les données, notamment le taux de mésappariement, le taux d'insertion et le taux de suppression, HiSeq 2000 et Ion Torrent restaient incomparables du fait des principes de séquençage différents. Par exemple, le site polynucléotidique ne pourrait pas être identifié facilement dans Ion Torrent, mais il est démontré que l'Ion Torrent a une qualité stable tout au long des lectures de séquençage et une

bonne performance en termes de précision des mésappariements, mais il est limité pour la détection des indels (faux positifs).

- La distribution de profondeur GC est meilleure dans Ion Torrent. Dans ce dernier, la profondeur de séquençage est similaire alors que le contenu de GC est compris entre 63% et 73%. Toutefois, dans HiSeq 2000, la profondeur de séquençage moyenne est de 4x lorsque le contenu du GC est à 60%, alors qu'elle est de 3x avec un contenu de 70% de GC.

- Ion Torrent possède trois différentes puces, 314, 316 et 318, ce qui permet une production plus élevée au séquençage. Par exemple, la puce 318 permet la production de 1Gb en 2 à 3 heures [28, 37]. Le choix de ces puces dans notre laboratoire se fait selon le nombre des patients à analyser et le type de panel (la taille des amplicons est différente pour chaque panel).

## **2. Applications du séquençage en Génétique médicale :**

### **2.1. Les différents niveaux de séquençage :**

Actuellement, le séquençage personnel du génome est devenu un outil de la pratique médicale courante. Il permet non seulement d'identifier la cause de certaines maladies génétiques, mais peut également suggérer des thérapies adjuvantes. Il s'agit de la mise en œuvre clinique d'informations personnelles sur le génome.

Au cours de la dernière décennie, la recherche est passée de l'étude de gènes et de protéines individuels à l'analyse des génomes et des protéomes entiers et ceci grâce au développement de technologies à haut débit pour l'analyse de l'ADN et des protéines. Les chercheurs peuvent désormais étudier les différents aspects de la cellule à l'échelle du génome, à savoir les mutations d'ADN, les modifications d'histones, la méthylation de l'ADN, la structure de la chromatine, la transcription et la traduction. L'intégration et l'analyse des informations provenant des différents processus se produisant au niveau cellulaire va aider à mieux comprendre la biologie des organismes et la physiopathologie de la maladie [39].

L'utilisation du NGS en pratique clinique a été rapportée pour la première fois à la fin de la dernière décennie et a permis de pallier aux limites du séquençage de Sanger. Elle comprend plusieurs applications telles que le diagnostic des maladies héréditaires, la recherche de maladies complexes, l'assistance aux protocoles de traitement du cancer par l'étude du profil tumoral et autres [39]. En règle générale, devant une suspicion de maladie génétique, un test de « phénotype à génotype » s'impose. L'analyse génique est réalisée afin de déterminer la cause du tableau clinique du patient [41]. Cela peut impliquer une approche ciblée du séquençage, dans laquelle une région spécifique du génome est étudiée pour identifier les variants génétiques qui produisent le phénotype à l'étude.

### **2.1.1. Le séquençage Sanger**

Il était pendant des années et est toujours la méthode de choix pour le diagnostic moléculaire de certaines maladies génétiques. D'ailleurs et même après l'avènement des nouvelles technologies, il fallait confirmer la présence d'un variant par séquençage. La méthode de séquençage de Sanger a trouvé d'immenses applications en raison de sa simplicité et elle est plus susceptible d'être étendue [14].

En effet, un séquenceur est un appareil indispensable pour tous les laboratoires de génétique. Actuellement, on peut lancer jusqu'à 96 séquences par run avec une taille moyenne de 800 à 1000 pb de fragments d'ADN [14, 40, 41]. Bien que le séquençage de Sanger était le pilier du projet original du génome humain, cette approche présente encore certaines limites notamment un coût élevé et un débit faible. Elle est inadéquate pour étudier des grands gènes [25].

### **2.1.2. Panel de gènes d'intérêt :**

En pratique clinique, le panel est constitué d'un groupe de gènes déjà décrits dans la littérature comme étant impliqués dans une pathologie donnée ou dans les diagnostics différentiels de cette dernière (par exemple : l'ensemble des gènes impliqués dans les dystrophies musculaires). Cette approche permet d'obtenir de très bons résultats en termes de couverture et de profondeur.

Les panels sont soit prêts à l'emploi et commercialisés par la plate-forme de NGS (comme le kit des gènes de prédispositions héréditaires au cancer du sein et de l'ovaire (BRCA1 et BRCA2)), ou conçu sur commande (design). Il est donc possible de personnaliser les panels en fonction des besoins et de la demande des cliniciens. La liste est alors établie par le laboratoire lui-même [42]. Cependant, l'inconvénient dans le séquençage par panel de

gènes réside dans la définition de la liste de gènes, ce qui implique qu'une mauvaise prestation clinique ou une erreur diagnostique peut entraîner le choix d'un panel inadéquat. Un autre point limitant est la nécessité de réactualiser à chaque fois la liste de gènes inclus dans le panel, surtout en cas d'identification de nouveaux gènes impliqués dans une maladie génétique.

### **2.1.3. Séquençage de l'Exome clinique (large panel) :**

Une recherche plus large des mutations peut être effectuée en utilisant le séquençage par panel de gènes (Clinical Exome Sequencing : CES), commercialisé par plusieurs fournisseurs, et comportant la majorité des gènes répertoriés comme étant impliqués en pathologie humaine (Plus de 5000 gènes OMIM associés à des maladies génétiques connues).

Cette approche présente plusieurs avantages notamment une bonne flexibilité. En effet, en cas de réorientation diagnostique ou de résultat négatif, on peut procéder à une évaluation d'autres gènes pertinents ou faire une réanalyse des nouveaux gènes impliqués en fonction des avancées scientifiques [43]. L'inconvénient principal est que les séquences ciblées étant plus nombreuses que lors d'un panel, la profondeur est moins importante et peut être insuffisante dans certaines zones des régions cibles par rapport au panel de gènes. De même, les gènes récemment décrits et non intégrés dans la base OMIM ne sont pas encore inclus dans ces kits commerciaux. Le coût de cette stratégie est un peu plus avantageux que celui des deux autres techniques.

### **2.1.4. Le séquençage de l'exome entier (WES/Whole Exome Sequencing):**

Cette approche offre l'avantage d'analyser toute la région codante dans laquelle la majorité des mutations est décrite et qui représente environ 2% du génome. Cette technique nous offre la possibilité d'analyser des panels virtuels et d'étendre cette analyse à d'autres



gènes si nécessaire. Les régions d'intérêt étant plus grandes, les données de couverture et de profondeur sont moindres.

L'avantage de cette technique est d'éviter de réaliser successivement des examens génétiques ciblés gène par gène, et de permettre d'augmenter le nombre de diagnostics. En effet, le séquençage d'un exome révèle, selon la technologie utilisée, entre 20000 variations nucléotidiques (chez un caucasien) et 25000 variations (chez un africain). Après filtrage informatique, il subsiste en moyenne entre 150 et 500 variants privés non synonymes ou ayant un effet potentiel sur l'épissage et qui peuvent être impliqués dans la pathologie [44]. L'interprétation clinico-biologique de ces données, visant à confirmer l'implication de l'un ou l'autre de ces variants, représente encore un travail colossal.

L'analyse des données du WES peut également identifier des anomalies dans des gènes sans rapport avec l'indication initiale, en particulier des gènes de prédisposition aux cancers, ce qui multiplie les questions éthiques quant à l'utilisation de cette technique. Actuellement, la prescription de ces examens doit rester réservée à des personnes capables d'interpréter les résultats, d'expliquer les avantages et inconvénients.

### **2.1.5. Le séquençage du génome complet (WGS/Whole Genome Sequencing) :**

Cette stratégie ne nécessite pas de technique d'enrichissement puisqu'elle consiste à séquencer l'intégralité du génome. Cependant, elle reste très coûteuse et elle génère un volume très important de données.

### **2.2.Choix du test de séquençage :**

Le choix du test approprié requiert une description précise du phénotype avec une expertise clinique préalable à la prescription du test surtout par des spécialistes en génétique

médicale. L'enjeu de l'étape clinique est à la fois médical (faire le bon diagnostic) et économique (ne pas prescrire de multiples analyses coûteuses inutiles).

Le séquençage de Sanger est recommandé lorsque la pathologie est bien étiquetée et les signes cliniques sont spécifiques, orientant vers un syndrome ou une maladie précise, et surtout si cette dernière est associée à une mutation récurrente ou région hotspot.

Si le gène a une grande taille ou comprend plusieurs exons, ou si on est devant une maladie où plusieurs gènes sont impliqués (sans hotspot mutationnel), l'approche par séquençage Sanger devient laborieuse et coûteuse [45] et on a recours dans ces cas à une analyse d'un panel de gènes, incluant le(s) gène(s) candidat(s). L'avantage d'atteindre une grande profondeur de lecture fait de cette technique un excellent outil de recherche des mutations somatiques.

Alternativement, le séquençage ciblé de panels de gènes (Exome clinique) est indiqué lorsque le généticien clinicien évoque le diagnostic d'un syndrome précis ou définit une maladie spécifique avec hétérogénéité génétique.

En revanche, devant une forte présomption de maladie monogénique avec un tableau clinique peu spécifique et des investigations para-cliniques non concluantes, le WES est le plus approprié pour avoir enfin une chance de poser un diagnostic.

Finalement, le séquençage du génome dans son intégralité (WGS) est encore loin d'être utilisé en routine diagnostique et reste pour l'instant limité à la recherche principalement lorsque les investigations initiales s'avèrent peu informatives [39].

Dans la famille avec XP, nous avons procédé à un séquençage ciblé à la recherche de la mutation récurrente au niveau de l'exon 9 du gène XPC. Tandis que pour les familles avec AF, nous avons conçu dans notre laboratoire un panel ciblé par l'analyse des gènes en cause

pour certaines maladies héréditaires à l'aide de Ion AmpliSeq On-Demand panel de Ion Torrent. Parmi les 22 gènes, ce panel contient 3 gènes de l'AF (*FANCA*, *FANCC* et *FANCG*) qui représentent 84% de tous les gènes. Donc, un résultat normal avec ce panel n'exclue pas la présence d'une mutation dans les autres gènes impliqués dans AF.

D'ailleurs toutes les études moléculaires de la présente étude entrent dans un cadre de diagnostic clinique et les résultats sont rendus aux patients lors d'une consultation génétique accompagnés d'un conseil génétique.

### **2.3. Bioinformatique et interprétation des variants dans les plateformes**

#### **NGS :**

Les données brutes sont lues à partir de l'un des types d'instruments NGS et elles sont soumises à une série de processus bioinformatiques (également appelés «pipeline») pour délivrer à la fin un *variant call file* (VCF) (Tableau 4) [46,47]. Ces processus comprennent le démultiplexage, l'analyse de la qualité, la cartographie des lectures sur un génome de référence et l'identification/annotation de variants [48]. Pour configurer et gérer un service clinique de NGS et en raison de ces processus spécialisés, il est nécessaire d'avoir un personnel spécialisé en bioinformatique (bioinformaticiens).

Nous disposons, au sein de notre laboratoire, d'une équipe expérimentée en bioinformatique chargée de l'interprétation des résultats de NGS, qu'ils soient faits au niveau de notre plate-forme de PGM ou sous-traités à d'autres plateformes externes.

Dans le NGS, plusieurs patients sont regroupés et séquencés à la fois grâce à l'utilisation de codes-barres pour marquer les fragments d'ADN des échantillons, ce qui permet de réduire les coûts de séquençage. Pour cette raison, une étape de démultiplexage est nécessaire au cours de laquelle toutes les lectures (*reads*) sont triées par un code

barre/échantillon avant toute analyse ultérieure [49]. Le fichier démultiplexé avec des lectures brutes est appelé FASTQ fichier. Les lectures individuelles d'un échantillon sont ensuite mappées sur un génome de référence (fichier BAM) et toute différence entre la lecture de référence et celle de séquençage est notée. Le fichier de sortie qui définit tous les variants détectés d'un patient et leurs fractions alléliques est appelé *variant call file* [50]. Cette liste des variants est interprétée. Des outils bioinformatiques (précédemment cités dans la partie matériels et méthodes) peuvent être utilisés pour filtrer les variants répondant à certains critères (par exemple, fréquence d'allèle mineur supérieure à un seuil ou variants précédemment identifiés comme bénins ou synonyme au niveau des bases de données) [51].

L'interprétation des variants détectés par le NGS devient plus complexe lorsqu'elle est appliquée à la totalité des gènes (WES ou WGS) ou à un grand nombre de gènes (grand panel de gènes). Il est évident que plus la surface du génome séquencée est grande, plus la probabilité de trouver des variants rares ou nouveaux augmente, nécessitant ainsi une interprétation très minutieuse.

Afin de pouvoir interpréter des variants germinaux, des recommandations ont été publiées par l'American College of Medical Genetics and Genomics (ACMG) et l'American College of American Pathologists (CAP). Ces recommandations assignent la force de la preuve pour les différents critères concernant un variant particulier et des règles pour combiner tous les critères afin de classer ce variant comme pathogène, probablement pathogène, de signification incertaine, probablement bénin ou bénin [52]. Les critères incluent la fréquence des allèles mineurs des bases de données de populations et la prévalence d'un variant chez les individus affectés, les données de ségrégation, les études fonctionnelles, le type de mutation et son effet prévu, la similarité de la mutation avec les mutations connues, des modèles informatiques d'effet et des facteurs de transmission [52,53].

Il existe des bases de données sur les populations (Tableau 4) qui contiennent désormais des informations sur plus de 120000 personnes [54,55]. Ces normes sont un début et vont permettre de faire des comparaisons entre les laboratoires et les études de recherche au niveau mondial [55]. Des critères similaires pour les tests somatiques sont aujourd'hui disponibles. L'utilité de ces consignes pour la normalisation de l'interprétation des variants somatiques et de la communication d'informations entre les laboratoires reste à évaluer [56,57]. Bien que certains outils soient disponibles pour faciliter la mise en œuvre de ces recommandations pour la classification des variants, leur utilisation nécessite beaucoup de travail afin d'évaluer ces critères et d'appuyer le processus [58].

Type de base de données	Nom de base de données	Si-Web
Bases de données de population	Exome aggregation consortium (Exac)	<a href="http://exac.broadinstitute.org/">http://exac.broadinstitute.org/</a>
	gnomAD browser	<a href="http://gnomad.broadinstitute.org/">http://gnomad.broadinstitute.org/</a>
	1000 Genomes	<a href="http://www.internationalgenome.org/">http://www.internationalgenome.org/</a>
	Exome server project	<a href="http://evs.gs.washington.edu/EVS/">http://evs.gs.washington.edu/EVS/</a>
Bases de données des maladies héréditaires	ClinVar	<a href="https://www.ncbi.nlm.nih.gov/clinvar/">https://www.ncbi.nlm.nih.gov/clinvar/</a>
	dbSNP	<a href="https://www.ncbi.nlm.nih.gov/projects/SNP/">https://www.ncbi.nlm.nih.gov/projects/SNP/</a>
	NCBIb genetic testing registry	<a href="https://www.genetests.org">https://www.genetests.org</a>
	Leiden open variant database (links to many locus-specific databases)	<a href="http://www.lovd.nl/3.0/home">http://www.lovd.nl/3.0/home</a>
	HGMD (The Human Gene Mutation Database)	<a href="http://www.hgmd.cf.ac.uk/ac/index.php">http://www.hgmd.cf.ac.uk/ac/index.php</a>
Bases de données d'oncologie	Catalogue of somatic mutations in cancer (COSMIC)	<a href="http://cancer.sanger.ac.uk/cosmic">http://cancer.sanger.ac.uk/cosmic</a>
	The cancer genome atlas (TCGA)	<a href="http://cancergenome.nih.gov/">http://cancergenome.nih.gov/</a>
	OncoKB (annotated TCGA data)	<a href="http://oncokb.org/#/">http://oncokb.org/#/</a>
	JAX CKB	<a href="https://www.jax.org/clinical-genomics/ckb">https://www.jax.org/clinical-genomics/ckb</a>
	My cancer genome	<a href="https://www.mycancergenome.org/">https://www.mycancergenome.org/</a>

**Tableau 4: Bases de données publiques utilisées dans l'interprétation des données de NGS [59,60].**

Généralement au sein de notre laboratoire d'accueil DGM-INH, en cas d'identification d'un nouveau variant après l'analyse de NGS (panel de gènes ou WES), nous utilisons uniquement les sites connus de prédiction, tels que : PolyPhen, MutationTaster ou autres (voir partie matériel et méthodes) car on ne dispose pas de moyens pour réaliser les études fonctionnelles des variants détectés.

Concernant notre étude portant sur l'AF, 6 variants détectés chez 6 patients de notre cohorte ont déjà été rapportés au niveau des bases de données (ClinVar, HGMD, etc.) (Tableau 3). Pour les 3 nouveaux variants : Nous avons une mutation décalant le cadre de lecture avec formation d'un codon stop prématuré et une mutation nonsense ; ces dernières sont en général des mutations délétères.

Pour le 3<sup>ème</sup> nouveau variant, il s'agit d'une mutation faux-sens pour laquelle nous avons utilisé les sites de prédiction et elle est prédite comme étant probablement pathogène et pathogène.

#### **2.4.Limites et difficultés du NGS :**

Bien que le NGS soit une approche diagnostique révolutionnaire et les laboratoires de génétique s'équipent aujourd'hui de ces technologies partout dans le monde, il reste encore des limites techniques et des difficultés d'interprétation à surpasser, sans oublier toute la problématique liée aux aspects éthiques et aux consensus universels pour cadrer les applications du NGS.

Plusieurs aspects sont encore à discuter et à régler dans le cadre de l'application du NGS en pratique médicale [61] :

- les programmes de contrôle qualité/assurance qualité de tout le processus de la technique initiale à la validation clinique [62] ;

- la gestion et le stockage des données révélées (analyse, gestion et instrumentation pour le stockage) qui nécessitent des dispositifs électroniques d'une capacité extrêmement élevée [63] ;
- les défis de l'analyse des données de séquence pour l'interprétation clinique (telles que les variants génétiques précédemment inconnus) sont des problèmes majeurs à résoudre [64] ;
- le signalement de résultats complexes peut être extrêmement difficile en ce qui concerne les implications cliniques dans le diagnostic, le pronostic et le traitement de précision de la maladie [65] ;
- les découvertes fortuites ayant des implications biomédicales significatives (exemple : mutation *BRCA*) peuvent poser des responsabilités éthiques aux généticiens (devoir de déclaration) [66] ;
- la contrefaçon de brevet peut affecter des laboratoires utilisant du NGS et des gènes ou des séquences d'ADN protégées par un brevet [67] ;
- et enfin, la manière dont les NGS peuvent être correctement remboursés exige que les établissements universitaires, les laboratoires et les organismes de réglementation soient en collaboration avec les cliniciens, fournisseurs de services et assurances.

## **2.5.Considérations éthiques :**

Un certain nombre de considérations éthiques doivent être prises en compte principalement lors de l'analyse du séquençage de l'exome en milieu clinique. Par exemple, en comparaison avec la plupart des approches de diagnostic traditionnelles, la distinction entre diagnostic et recherche est assez floue dans le cas de NGS. Cela est principalement dû au fait que, contrairement aux autres approches diagnostiques, les méthodologies de test et de validation de NGS ne sont pas encore bien établies. De plus, comme la plupart des cliniciens utiliseraient le NGS pour comprendre les maladies rares, la précision du diagnostic ne peut



pas être établie dans de nombreux cas en raison du nombre de cas et de la nature unique de chaque patient. Il convient également de garder à l'esprit que les principes fondamentaux des enquêtes en génétique doivent être fondés sur les principes forts de bienfaisance, de réciprocité, de justice et de responsabilité professionnelle.

Trois domaines principaux sont couverts dans ce qui suit :

- Eduquer et informer le patient
- Obtenir un consentement éclairé
- Confidentialité des informations du patient et des membres de sa famille

Le récent avis du Comité Consultatif National d'Éthique (CCNE) pour les sciences de la vie et de la santé sur la problématique du NGS propose qu'il est urgent pour la communauté médicale de prendre la mesure de la révolution annoncée par ces technologies globales de séquençage (CCNE- avis N°124). Dans un futur proche, chaque individu peut avoir son génome séquencé dès la naissance, ces données l'accompagneront tout au long de sa vie, comme une carte d'identité. En effet, contrairement aux informations épigénétiques ou aux mutations acquises, l'information génétique constitutionnelle est par nature propre à un individu et fixée pour la vie. Ces considérations pourraient nous amener à vouloir définir une norme comme référence absolue, qui figerait la « séquence normale de référence de l'espèce ».

Parler de gènes normaux et de gènes mutés est une aberration, puisque tous nos gènes, produits de cette évolution, sont, par définition, des gènes mutés. Par ailleurs, il est absolument nécessaire que la pratique du NGS soit très clairement encadrée. D'après le CCNE, « l'analyse globale du génome demande une justification en termes de nécessité (besoin de résoudre un problème clinique) et de proportionnalité (équilibre des avantages et inconvénients pour le patient). Quand un séquençage global d'ADN est décidé, un protocole rigoureux doit guider la conduite à tenir pour les découvertes non sollicitées ». En effet, lors

du séquençage à haut débit, on se trouve confronté à des variants décrits comme étant pathogènes, mais dans des gènes que l'on ne souhaitait pas forcément explorer. On comprend intuitivement que, plus on analyse de larges portions d'ADN, plus on identifie de ces variations nucléotidiques dites de « découvertes fortuites » (*incidental findings*). Que faire dans ce cas ? Un argument fréquemment avancé est que la connaissance du statut de porteur d'une « anomalie » génétique pourrait inciter l'individu à modifier ou à adapter son comportement. Ainsi, la maladie ne serait plus perçue comme une simple fatalité, mais un état contre lequel il serait possible de se prémunir en adaptant ses comportements.

Enfin, le NGS fait bouger la frontière entre la pratique diagnostique, visant à identifier la cause moléculaire d'une pathologie chez un patient, et la recherche qui tend à décrire de nouvelles causes de la pathologie pour en comprendre les mécanismes. Jusqu'ici, le système de santé était organisé autour de cette stricte séparation, que les approches globales ont tendance à remettre en cause.

Les conditions de stockage et le respect de la confidentialité des données représentent d'autres problèmes qui ne sont pas encore parfaitement encadrés. Qui est propriétaire des données générées au cours des analyses de NGS ? Combien de temps conserver les données de séquence ? Comment assurer la sécurité des fichiers et des serveurs ? Autant de questions sans réponse nécessitant une réflexion profonde, pluridisciplinaire et mesurée, et ceci indépendamment des contraintes temporelles liées à la poursuite de nos pratiques.

### **3. Exemples de maladies génétiques avec risque de cancer :**

#### **3.1.La Xéroderma Pigmentosum :**

Le nom de la maladie est issu du terme latin « Xeroderma Pigmentosum » qui signifie « derme sec et pigmenté », elle fut décrite pour la première fois par un dermatologue Hongrois Moritz Kaposi en 1874. Sa prévalence est estimée à environ 3 sur 1 million [[http://www.orpha.net/orphacom/cahiers/docs/FR/Prevalence\\_des\\_maladies\\_rares\\_par\\_ordre\\_alphabetique.pdf](http://www.orpha.net/orphacom/cahiers/docs/FR/Prevalence_des_maladies_rares_par_ordre_alphabetique.pdf)] [68] avec une fréquence plus haute au Japon [69] et dans les pays avec un taux de consanguinité élevé comme le Maroc où la prévalence a été estimée à 1/80,504 [70].

Cette génodermatose est caractérisée par une sensibilité accrue aux UV et à la lumière du soleil. Les cellules photo-exposées vont subir un vieillissement accéléré avec apparition de lésions au niveau de la peau et des yeux, celles-ci peuvent dégénérer conduisant à de multiples cancers cutanéomuqueux [71,72].

A l'heure actuelle, il n'existe pas encore de traitement curatif pour soigner les malades souffrant de XP. La photoprotection reste le seul moyen disponible pour prévenir ou au moins retarder l'apparition de néoplasies cutanées et oculaires, d'où la nécessité d'un diagnostic précoce de la maladie [73].

#### **3.1.1. Diagnostic clinique et paraclinique :**

En dehors d'un contexte familial évocateur, le diagnostic est assez difficile au début. En effet, les premières manifestations peuvent être interprétées comme un simple « coup de soleil » ou comme une photosensibilisation ou même une dermatite atopique. La répétition et la persistance de l'érythème anormalement intense au niveau des régions photo-exposées, doivent faire suspecter une hypersensibilité anormale à la lumière solaire. La présence d'antécédents familiaux de XP peut orienter le diagnostic [71].

Le diagnostic de XP peut souvent être posé cliniquement dans les premières années de vie, devant l'atteinte de la peau, des yeux et du système nerveux [74].

➤ ***La peau***

Environ 60% des personnes atteintes d'XP ont des antécédents de réaction aiguë aux coups de soleil lors d'une exposition minimale aux UV. Les symptômes cutanés apparaissent entre un et deux ans. Il s'agit de macules hyperpigmentées ressemblant à des taches de rousseur (lentigos) qui se limitent uniquement aux zones exposées au soleil [75](Figure 22).

Une exposition continue au soleil rend la peau sèche avec une pigmentation accrue. La plupart des personnes atteintes d'XP développent une xérose (peau sèche) et une poikilodermie (constellation d'hyper et d'hypopigmentation, d'atrophie et de télangiectasie). Les kératoses actiniques précancéreuses se développent à un jeune âge. La XP est un exemple de photo-vieillessement accéléré.



*Figure 22: Photo d'un enfant atteint de XP montrant les lésions cutanées  
(<https://dermatomaroc.com/>)*

➤ ***Les yeux :***

Les anomalies oculaires sont presque aussi fréquentes que les anomalies cutanées en raison de leur exposition à la lumière du soleil. Les lésions peuvent apparaître au cours de la première décennie de la vie et se limitent généralement aux parties antérieures des yeux exposées aux UV: conjonctive, cornée et paupières.

Les patients souffrent souvent de photophobie qui peut être associée à un écoulement conjonctival important. Des kératites sévères pouvant elles-mêmes entraîner une opacification de la cornée.

Par ailleurs, les paupières développent une pigmentation accrue et perdent leurs cils. L'atrophie de la peau peut entraîner des ectropions, des entropions, ou dans les cas les plus graves, la perte complète de la paupière [76,77].

➤ ***Le système nerveux :***

Lors de l'étude de 2011 menée sur 106 patients, 25% des individus présentaient des manifestations neurologiques [74]. Ces troubles sont en fait progressifs et peuvent apparaître tardivement après les manifestations cutanées. Le début peut être dans la petite enfance ou, chez certains individus, retardé jusqu'à la deuxième décennie ou plus tard [78]. Les anomalies neurologiques peuvent être légères (hyporéflexie isolée) ou graves, avec microcéphalie, déficience intellectuelle progressive, perte auditive neurosensorielle, spasticité, ataxie et/ou convulsions. Des examens complémentaires peuvent être réalisés en fonction du type d'atteinte neurologique :

- L'électromyogramme (EMG) et les tests de vitesse de conduction nerveuse à la recherche d'une neuropathie axonale ou mixte.

- La tomodensitométrie (TDM) et l'Imagerie par Résonance Magnétique (IRM) du cerveau.
- L'audiométrie permet de révéler relativement tôt une perte auditive [79].

➤ **Cancers :**

En l'absence de protection solaire, les patients avec XP de moins de 20 ans présentent un risque accru de développer des cancers principalement au niveau zones photo-exposées de la peau. En effet, les patients XP ont un risque 10000 fois plus élevé par rapport à la population générale de développer au cours de leur vie un cancer non mélanocytaire (carcinomes squameux et baso-cellulaires) et 2000 fois un risque de survenue de mélanome [74].

Des néoplasies au niveau de la cavité buccale ont été observées chez certains patients principalement le carcinome épidermoïde de la pointe de la langue [80,81].

D'autres cancers ont été rapportés chez quelques individus atteints d'XP notamment des gliomes du cerveau et de la moelle épinière, une leucémie, des tumeurs du poumon, de l'utérus, du sein, du pancréas, de l'estomac, des reins et des testicules [74, 82-84].

Chez notre patient, en plus de la photophobie et des anomalies cutanées sur les régions photo-exposées, le jeune homme présentait une atteinte oculaire sévère avec des lésions inflammatoires au niveau de la cornée et un ectropion bilatéral. Par ailleurs, il n'avait pas de signes neurologiques apparents ni de lésions précancéreuses.

Il existe des tests cellulaires mesurant la réparation défectueuse de l'ADN comme la synthèse non programmée d'ADN (UDS, pour *unscheduled DNA synthesis*) qui s'effectue sur les fibroblastes de peau en culture afin de mesurer l'hypersensibilité aux UV. Le diagnostic

d'XP est confirmé par la réduction d'UDS et l'hypersensibilité à la destruction induite par les UV. Le diagnostic d'XP-V est confirmé par l'UDS normale et la sensibilité spécifique aux UV en présence de caféine.

Actuellement, l'UDS est en phase d'être supplanté par les nouvelles techniques de biologie moléculaire recherchant les mutations causales de la maladie.

Des techniques d'immunohistochimie proposées permettent d'identifier le groupe génétique mais elles ne sont pas encore de pratique courante [85].

En effet, le diagnostic de XP est dans un premier temps clinique, la confirmation par biologie moléculaire peut ne pas être facilement disponible et être réalisée plus tardivement après avoir effectué des tests fonctionnels de réparation de l'ADN.

Quand la dyschromie est évidente, associée à la photophobie, le diagnostic clinique de XP devient aisé.

Le diagnostic différentiel peut se poser avec les autres génodermatoses notamment le syndrome de Cockayne, la trichothiodystrophie, le syndrome cérébro-oculo-facio-squelettique (COFS : cerebro-oculo-facio-skeletal syndrome) ou le syndrome de Bloom. Chacune de ces affections a des caractères distinctifs. Les tests de réparation de l'ADN après exposition aux UV ou UDS permettent alors de confirmer le diagnostic de XP (Figure 23).

Dans le syndrome de Cockayne, en plus de la photosensibilité il y a un retard de croissance post-natal associé à une microcéphalie et une atteinte progressive du système nerveux. Les patients peuvent présenter également une atteinte oculaire (rétinopathie, cataracte) et une perte auditive neurosensorielle.



La forme photosensible de la trichothiodystrophie est caractérisée par des cheveux épars et cassants, souvent accompagnés d'une ichtyose congénitale, d'un déficit intellectuel, d'un retard de croissance et d'ongles anormaux.

Le COFS (type 2 et 3) est caractérisé par un retard de croissance intra-utérin, une microcéphalie, des contractures articulaires congénitales avec une atteinte du système nerveux central.

Le syndrome de Bloom associe un retard de croissance pré et post-natal avec des difficultés pour s'alimenter, une dysmorphie faciale et des infections à répétition en rapport avec un déficit immunitaire.



**Figure 23: Diagnostic différentiel de la XP. A: patiente avec TTD âgée de 3 ans présentant des cheveux cassants [86]. B: patiente atteinte de syndrome de Cockayne avec un aspect cachectique, des yeux enfoncés et des lésions cutanées. C: enfant avec dysmorphie fac**

### 3.1.2. La génétique :

La XP est une maladie héréditaire de transmission autosomique récessive appartenant aux groupes de maladie liée à un défaut dans les gènes du système de réparation.

En effet, chaque cellule normale est capable de reconnaître et de réparer la ou les lésions de l'ADN au fur et à mesure qu'elles se produisent afin de préserver l'intégrité du génome humain. Il existe plusieurs voies de réparation de l'ADN dont le système NER (Nucleotide excision Repair).

Ce trouble autosomique récessif est génétiquement hétérogène et dû à huit groupes de complémentation avec NER défectueux et un groupe de variantes (XP-V) avec NER normal. À ce jour, neuf gènes ont été associés avec XP: *XPA*, *ERCC3 (XPB)*, *XPC*, *ERCC2 (XPD)*, *DDB2 (XPE)*, *ERCC4 (XPF)*, *ERCC5 (XPG)*, *ERCC1*, et *POLH (XP-V)*.

Il est à noter qu'un schéma de nomenclature spécifique est proposé pour les gènes humains de la réparation de l'ADN [88]:

- Pour les gènes de Xeroderma Pigmentosum, le C final pour «Complémentation» est omis alors qu'il est maintenu pour les gènes de l'Excision Repair Cross Complementing (*ERCC*).
- Lorsqu'un gène *ERCC* est identifié comme étant strictement identique à un gène de XP, du syndrome de Cockayne (CS) ou de la Trichothiodystrophy (TTD), le nom est éventuellement remplacé par le nom de gène XP, CS ou TTD. C'est le cas par exemple des gènes *ERCC2*, *ERCC3* et *ERCC6* qui sont respectivement devenus *XPD*, *XPB* et *CSB*.

Le tableau 5 présente les différents groupes de complémentation de XP en fonction du gène muté.

<b>Groupe de Complémentation</b>	<b>Gène</b>	<b>Locus</b>	<b>% des XP avec mutations dans ce gène</b>	<b>Protéine</b>
<b>A</b>	<i>XPA</i>	9q22.33	25%	Protéine de reconnaissance des lésions et excision (réparation de l'ADN)
<b>B</b>	<i>ERCC3</i>	2q14.3	Rare	XPB : sous-unité du facteur de transcription <i>IIIH</i> ( <i>GTF2H1</i> ) (transcription et réparation de l'ADN)
<b>C</b>	<i>XPC</i>	3p25.1	25%	Protéine de reconnaissance des lésions (réparation de l'ADN)
<b>D</b>	<i>ERCC2</i>	19q13.32	15%	Hélicase, sous-unité XPD du facteur de transcription <i>IIIH</i> ( <i>GTF2H1</i> )
<b>E</b>	<i>DDB2</i>	11p11.2	Rare	Protéine de liaison aux dommages à l'ADN
<b>F</b>	<i>ERCC4</i>	16p13.12	6%	Endonucléase XPF (Protéine de réparation de l'ADN)
<b>G</b>	<i>ERCC5</i>	13q33.1	6%	Endonucléase (Protéine de réparation de l'ADN)
<b>*</b>	<i>ERCC1</i>	19q13.32	**	Endonucléase (Protéine de réparation de l'ADN)
<b>Variant</b>	<i>POLH</i>	6p21.1	21%	ADN polymérase eta

**Tableau 5: Les groupes de complémentation XP et les gènes de la XP [89]**  
(<https://www.ncbi.nlm.nih.gov/books/>).

\* Auparavant dénommée groupe H, cette désignation a été retirée.

\*\* Seule une personne ayant une mutation dans ERCC1 a été rapportée [90].

### **3.1.2.1. Les gènes de la XP**

#### **a. Le gène XPA :**

Le groupe de complémentation A (XP-A) correspond à une forme sévère de Xeroderma Pigmentosum. Sa prévalence globale est inconnue mais il représente 25% de tous les cas de XP et il s'agit de la forme la plus courante de XP au Japon.

Le gène *XPA* est localisé en 9q22.3q31, contient 6 exons. La protéine XPA forme en fait un complexe trimère avec les facteurs ERCC1 et ERCC4 (XPF), qui participe à la fois à la reconnaissance du dommage et à l'excision du brin endommagé [91,92].

#### **b. Le gène XPB :**

Egalement nommé (*ERCC3*), localisé en 2q14.3, comporte 15 exons, code pour une protéine de 782 AA.

La protéine XPB est une sous-unité du facteur de transcription *IIH* (*GTF2H1*), ce dernier est à la fois impliqué dans la régulation de la transcription de l'ADN des gènes actifs, et dans le mécanisme de réparation de l'ADN par excision de nucléotides (NER).

#### **c. Le gène XPC :**

Le groupe de complémentation C (XP-C) est le sous-type le plus fréquent de XP dans la population caucasienne. Les patients XPC représentent environ 25% de tous les cas de XP.

Le gène *XPC* est localisé en 3p25, comprend 16 exons, code pour une protéine de 940AA dont la fonction est de détecter les dommages de l'ADN lors de la première étape de l'une des deux voies de réparation de l'ADN par excision de nucléotide. Il s'agit de la voie dite de GG-NER pour « Global Genome Nucleotide Excision Repair » [93].

**d. Le gène XPD :**

Le groupe de complémentation D (XP-D) est un sous-type de XP qui représente environ 15% des cas de XP.

Le gène *XPD* ou *ERCC2*, localisé en 19q13.32, comporte 23 exons, code pour une protéine de 86,9 kDa de 760 acides aminés. La protéine ERCC2 (XPD) est une hélicase qui déroule la molécule d'ADN dans la direction 5'-3', dès les phases précoces de la réparation de l'ADN par excision de nucléotide. En effet, elle est en charge du maintien de l'ouverture de la double hélice d'ADN pour permettre à la machinerie de réparation d'accomplir sa tâche. La protéine ERCC2 (XPD) fait partie du facteur de transcription TFIIH.

**e. Le gène XPE :**

Le groupe de complémentation E (XP-E) est un sous-type extrêmement rare de XP, moins de 10 cas ont été rapportés dans la littérature.

Gène *XPE* ou *DDB2* (damage specific DNA binding protein 2), localisé en 11p12-p11, comprend 10 exons. Il code pour une protéine de 48kDa pour 427 acides aminés. La protéine DDB2 combinée avec DDB1 forme un hétérodimère, qui, avec XPC, est impliqué dans la reconnaissance initiale des dommages à l'ADN induite par les UV dans des parties non transcrites du génome. Une mutation de cette protéine impacte le processus de reconnaissance des dommages et donc la réparation de l'ADN.

**f. Le gène XPF :**

Le groupe de complémentation F est un sous-type XP de sévérité généralement moyenne. Les patients XPF représentent 6% de tous les cas de XP. Il s'agit d'un groupe rare, dont la majorité des cas a été découverte au Japon.

Le gène *XPF* ou *ERCC4*, locus 16p13.12, comprend 11 exons, code pour la protéine ERCC4 de 905 acides aminés. Il s'agit d'une endonucléase qui coupe la lésion en 5' du brin endommagé lors du mécanisme du NER.

***g. Le gène XPG :***

Le groupe de complémentation G (XPG) est un sous-type extrêmement rare. Les patients souffrant de XPG représentent 6% de tous les cas de XP, la plupart en Europe (84).

Le gène *XPG* ou *ERCC5*, localisé en 13q33, comprend 15 exons. La protéine ERCC5 est une protéine de 112 kDa qui fonctionne comme une endonucléase et dont le rôle dans le NER est de couper l'ADN en 3' de la lésion.

***h. Le gène ERCC1 :***

Le gène ERCC1 code pour une protéine impliquée dans la réparation par excision de nucléotides (NER) et la réparation du pontage inter-brin de l'ADN (Interstrand DNA Crosslink : ICL). ERCC1 interagit avec ERCC4 pour former une endonucléase qui excise l'ADN. ERCC1 est nécessaire pour stabiliser et améliorer l'activité ERCC4 [90].

***i. Le gène POLH (XPV)***

XP Variant (XPV) est un sous-type de XP à symptomatologie légère. Pour cette raison, il est souvent diagnostiqué plus tardivement, à la fin de l'adolescence ou vers l'âge de 20 ans. Il est observé chez environ 20% des patients XP.

Le gène *POLH* (XPV), localisé en 6p21.1-p12, comportant 11 exons et codant pour l'ADN-polymérase eta (pol-eta) faite de 713 acides aminés. Cette protéine est chargée de resynthétiser la zone excisée et restaurer la continuité de l'ADN lors de la réparation par NER. Les cellules XP-V sont en effet capables de mener à bien le processus de réparation par

excision de nucléotides mais présentent un défaut dans la réplication des ADN endommagés par les ultraviolets. Certains patients atteints de XP présentent donc une réparation de l'ADN par NER normale, mais une réplication post-réparation défectueuse [94]. C'est pourquoi, ces patients sont attribués à une classe spécifique de XP. Il s'agit de la classe XP «variante», XP-V.

### **3.1.2.2. Diagnostic moléculaire :**

Le diagnostic moléculaire par les techniques de séquençage est possible et ceci à la recherche de la mutation causale, mais il est préférable de le réaliser après avoir caractérisé le bon gène à étudier par la détermination du groupe de complémentation ou par une étude de liaison dans les cas familiaux.

Ces techniques sont devenues très utiles non seulement pour le diagnostic mais essentiellement pour l'identification des sujets hétérozygotes et pour le diagnostic anténatal.

Tester dans un premier temps les mutations les plus fréquentes en fonction de la zone géographique peut parfois être pertinent pour les patients de certaines parties du monde, comme le Japon [95] et l'Afrique du Nord [96]

- *XPA* : - Inde: c.335\_338delTTATinsCATAAGAAA [97]
  - Japon: c.390-1G>C [98]
  - Tunisie: p.Arg228Ter [99]
- *XPC* : -North Africa: c.1643\_1644delTG [96,100,101]
- *ERCC2* : -Iraqi Jewish: p.Arg683Gln [102].
- *POLH* : -Tunisie/ North Africa: Del exon 10 [103].
  - Japon: c.490G>T; p.Ser242Ter; p.Glu306Ter et c.1661delA [104]
  - Basque / Nord d'Espagne: c.764+1G>A [105]



Soufir et coll. a rapporté l'analyse génétique de 66 familles maghrébines avec XP (Algérie, Maroc et Tunisie). Dans cette cohorte, le groupe XP-C et le groupe XP-A étaient les plus prédominants avec des fréquences respectives de 85% et 12%. Le gène XPC est le principal gène causant la maladie en Afrique du Nord avec une mutation récurrente c.1643\_1644delTG (p.Val548AlafsX25) qui est responsable de 74% de tous les cas XP. Concernant le gène XPA, et selon la même étude, la mutation c.682C>T p.Arg228\* est la seconde mutation la plus fréquente représentant environ 87% des mutations de ce gène et 10% de l'ensemble des patients XP. Ces données suggèrent la possibilité du diagnostic moléculaire des patients marocains atteints de XP en recherchant la mutation à effet fondateur du gène XPC suivi de la mutation récurrente du gène XPA [96].

Plusieurs études se sont succédées démontrant que le gène XPC est le principal gène causant la XP avec une mutation récurrente en Méditerranée région [103,104].

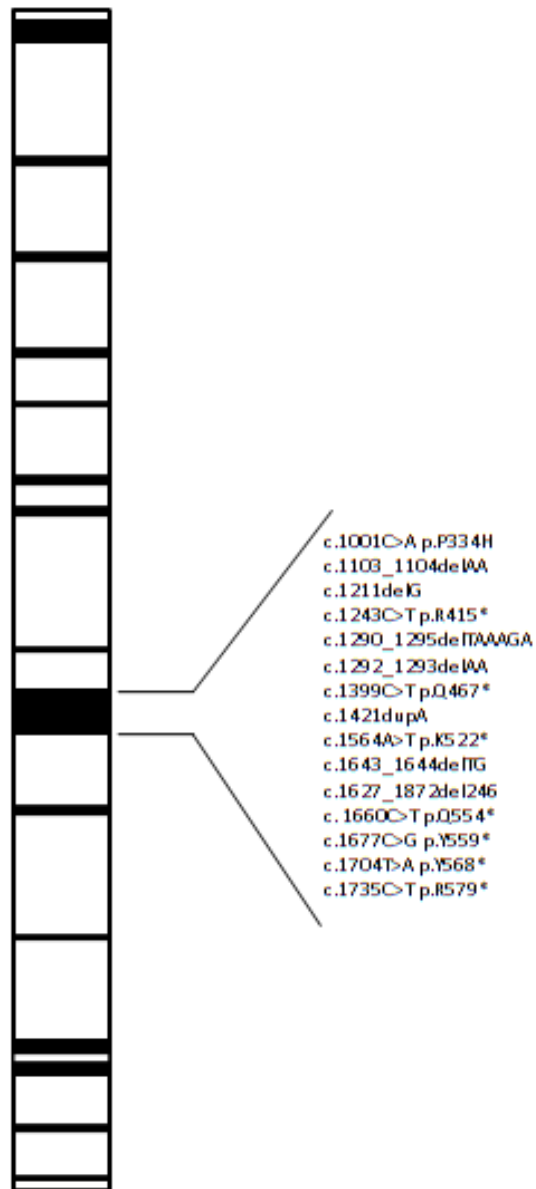
Suite à ces constatations, le Département de Génétique Médicale à Rabat offre depuis plus de dix ans, le diagnostic moléculaire de la XP par dépistage de la mutation récurrente: c.1643\_1644delTG (p.Val548AlafsX25) dans le gène XPC.

L'insertion de 4 bases paires CATG entre les positions 1644\_1645 dans l'exon 9, retrouvée chez notre patient, entraîne un décalage de cadre de lecture avec formation d'un codon stop prématuré après 25 acides aminés se trouve dans la région d'interaction avec RAD23B, qui est hautement conservé parmi les espèces selon Mutation Taster Server. Par conséquent, ce variant est très probablement la mutation causale de XP chez notre patient. Environ 60 mutations délétères du gène XPC ont été rapportées. La plupart d'entre elles entraînent un décalage du cadre de lecture (délétion et insertion) ou sont des mutations non-sens (substitution), conduisant à une protéine XPC tronquée. Quelques mutations d'épissage et faux-sens sont également décrites. Selon la base de données des mutations du gène humain

(HGMD), 25% de toutes les mutations du gène *XPC* sont situées dans le neuvième exon (Figure 24). Il s'agit de la première mutation de type insertion dans l'exon 9 dont la protéine la séquence correspond au domaine d'interaction avec le RAD23B impliqué dans la voie NER. Ces données suggèrent que le séquençage de l'exon 9 pourrait être une première étape dans une stratégie de diagnostic moléculaire de la XP en plus des mutations fondatrices spécifiques en fonction des populations. Plusieurs études de mutations fondatrices spécifiques à une population ont été publiées. A côté de la mutation nord-africaine identifiée chez 87% des patients avec complémentation XP groupe C [5], d'autres mutations fondatrices ont également été identifiées dans la population japonaise, un représentant 90% des allèles du gène *XPA* [98] et quatre autres représentant 87% du type de variante XP au Japon [104].

En plus du séquençage ciblé par la méthode de Sanger, un panel comprenant les neuf gènes *DDB2*, *ERCC1*, *ERCC2*, *ERCC3*, *ERCC4*, *ERCC5*, *POLH*, *XPA*, *XPC* ainsi que d'autres gènes d'intérêt (pour le diagnostic différentiel) peut être réalisé.

Si on n'arrive pas à confirmer le diagnostic par ces deux méthodes; on peut avoir recours au séquençage de tout l'exome (WES) ou du génome (WGS). Les données résultant de ces derniers peuvent suggérer un diagnostic non évoqué au paravent (mutation dans un ou des gènes différents qui entraîne une présentation clinique similaire)



*Figure 24: Les mutations décrites au niveau de l'exon 9*

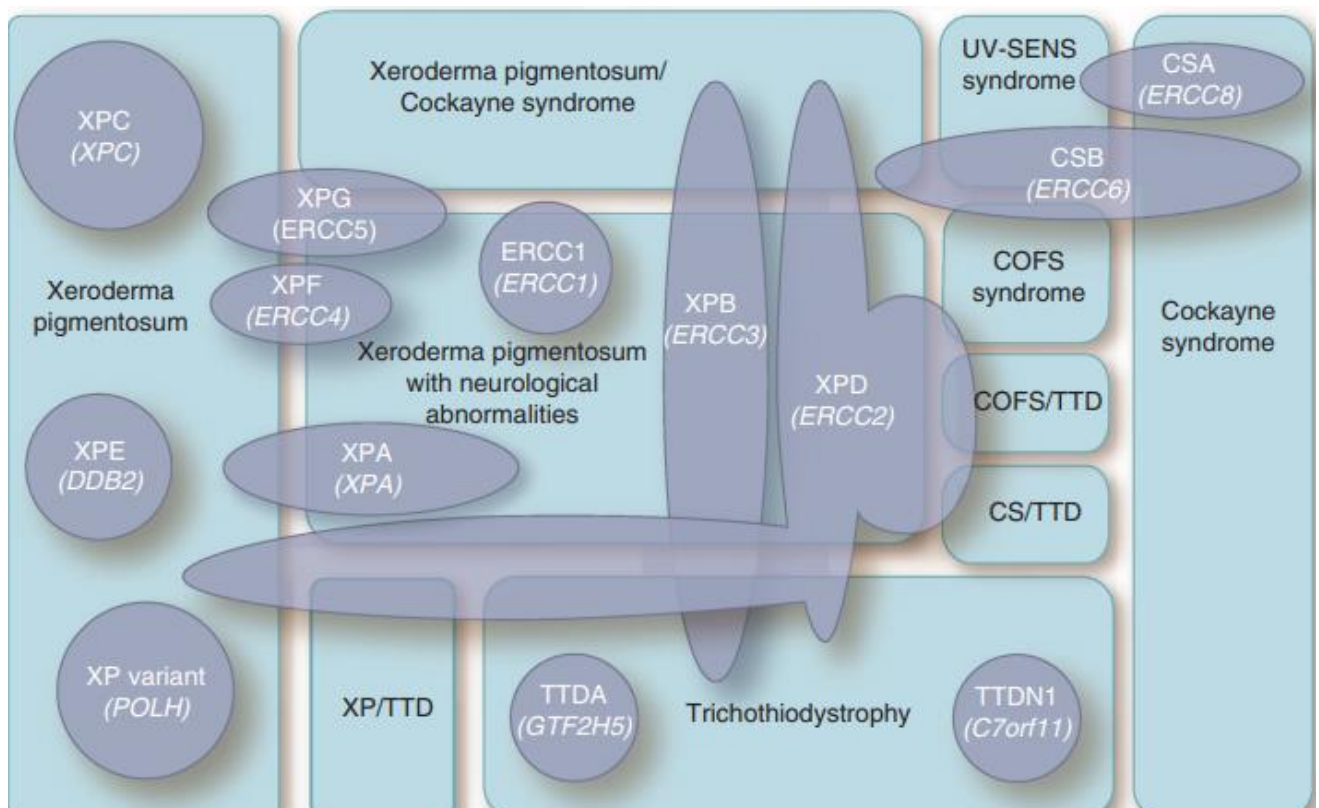
### **3.1.2.3. *Corrélation génotype-phénotype :***

En fonction du gène touché, des tableaux cliniques différents sont observés. La fréquence d'apparition d'une mutation varie également en fonction des gènes. En effet, à titre d'exemple, XPA est un groupe relativement fréquent alors que XPE est plutôt rare [106]. De la même manière, en fonction du groupe, le tableau clinique et la sévérité de la pathologie sont variables (XPG est une atteinte sévère alors que XPF est plutôt moyenne).

Les manifestations neurologiques sont associées plutôt à la XP type A (retard mental et/ou spasticité et/ou hyporeflexie et/ou ataxie). Certes, les types XP-A and C sont tous les deux caractérisés par une photosensibilité importante, les cancers cutanées sont plus fréquents chez patients XP-C (75%) que les patients XP-A (30%) [96].

### **3.1.2.4. *Pathologies alléliques:***

En plus des phénotypes XP, des variantes pathogènes dans les gènes responsables de la XP sont associées à d'autres phénotypes notamment le syndrome cérébro-oculo-facio-squelettique (COFS), le syndrome de Cockayne, l'anémie de Fanconi (FA), la trichothiodystrophie (TTD) et le syndrome progéroïde XFE (figure 25).



**Figure 25: Pathologies dues à un défaut de réparation de l'ADN**

Relations entre les entités cliniques (rectangles) et les défauts moléculaires (ovales). Une pathologie peut être due à des mutations dans différents gènes. Inversement, différentes mutations dans un gène peuvent entraîner plusieurs tableaux cliniques (Modifié à partir de Kraemer (2004) et Kraemer et al. (2007)). COFS syndrome : cerebro-oculo-facial-skeletal syndrome ; CS : Cockayne syndrome ; DDB2 : double-strand DNA-binding protein 2 ; TTD : trichoiodystrophy ; XP : xeroderma pigmentosum [107].

### **3.1.2.5. Gènes XP et la voie de réparation NER :**

Le NER est le mécanisme défectueux chez les patients souffrant de XP. Ce mécanisme est assuré par un ensemble de protéines assurant chacune une fonction spécifique permettant le déroulement de chaque étape du processus. Lorsque l'un des maillons de cette chaîne de réparation est défaillant, c'est tout le système qui est affecté. La conséquence d'un tel défaut est une mauvaise ou une absence de réparation de l'ADN endommagé par les rayons UV et donc les symptômes associés à cette mauvaise réparation : pathologies cutanées (zones particulièrement exposées aux UV) et les troubles neurologiques (dans certains groupes).

Les neuf gènes responsables de XP codent pour des protéines qui sont requises pour la réparation de l'ADN endommagé et jouent un rôle clé dans l'une ou l'autre des voies du NER, GG-NER (Global Genome - NER) ou TC-NER (Transcription Coupled - NER).

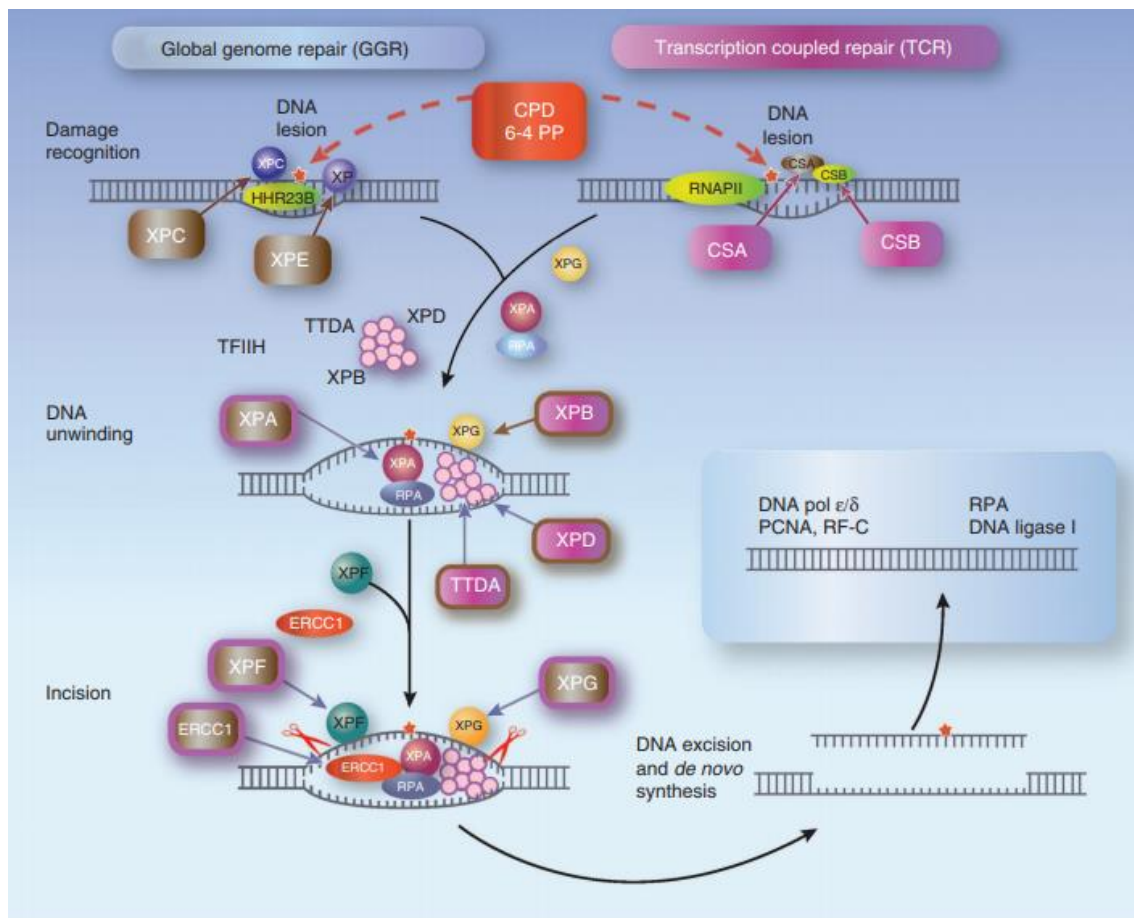
La réparation couplée à la transcription (TCR) supprime les dommages au niveau des gènes activement transcrits, tandis que la réparation globale du génome (GGR) corrige les dommages dans le reste du génome.

- Dans la GGR, des dommages tels que les dimères de cyclobutane pyrimidine (CPD) induits par le rayonnement UV ou 6-4 photoproduits (6-4 PP) sont reconnus par des protéines, notamment les produits des gènes *XPE (DDB2)* et *XPC*.
- Dans la TCR, la lésion semble bloquer la progression de l'ARN polymérase II dans un processus impliquant les produits des gènes *CSA* et *CSB*.
- Après la reconnaissance initiale des dommages, les voies convergent. Les hélicases *XPB (ERCC3)* et *XPB (ERCC2)* déroulent la région entourant la lésion avec les produits des gènes *XPA* et *XPG (ERCC5)* et la protéine de réplication A (RPA). Les endonucléases *XPF* et *XPG (ERCC5)* effectuent des incisions pour enlever la lésion dans un fragment d'environ 30 nucléotides. La lacune résultante est comblée par la

synthèse d'ADN de novo. Le produit du gène *XPV*, pour XP Variant, également appelé DNA pol $\eta$ , n'est pas impliqué dans le NER directement mais dans le mécanisme de resynthèse par réplication de l'ADN endommagé. En effet, la polymérase *XPV* (ou *POLH*) est chargée de synthétiser le nouveau brin après excision du brin endommagé.

(Figure 26) [107].

Ce système est coordonné de sorte que si une partie de la voie est mutée, la voie entière ne fonctionne pas normalement.



**Figure 26: Voie de réparation Nucleotide excision repair (NER)**

Diagramme modifié à partir de Van Steeg and Kraemer (1999) et Kraemer et al. (2007). PCNA: proliferating cell nuclear antigen; RF-C: replication factor-C; RPA: replication protein A [107].



### **3.1.3. Prise en charge**

Il n'existe aucun traitement curatif de la maladie. Seules les mesures préventives contre l'exposition solaire sont indiquées chez les enfants atteints de la XP (crèmes et vêtements protecteurs, écrans UV domestiques).

Cependant, même avec la plus stricte protection, l'apparition de lésions cutanées au cours de la vie des patients demeure quasi incontournable. Il existe des traitements symptomatiques disponibles pour la prise en charge de ces lésions (chirurgie, radiothérapie et traitements topiques). Un examen dermatologique et ophtalmologique régulier avec le traitement de toute lésion cancéreuse sont essentiels. La carence vitaminique D, fréquente, doit être supplémentée. Ces différentes mesures thérapeutiques permettent malgré tout une amélioration de la qualité de vie des patients.

### **3.2.L'Anémie de Fanconi :**

L'anémie de Fanconi fait partie des syndromes d'insuffisance médullaire héréditaire (Inherited bone marrow failure syndromes, IBMFS). Il s'agit d'un groupe hétérogène de maladies génétiques caractérisées par une insuffisance médullaire, des anomalies congénitales et un risque accru de développement de cancers.

Décrite pour la première fois par Guido Fanconi en 1927, l'AF est considérée comme la plus fréquente cause des aplasies médullaires constitutionnelles. Sa prévalence varie de 1 à 9 / 1000000. La fréquence des porteurs hétérozygotes de FA est estimée à 1 sur 181 aux États-Unis [108]. Dans les populations ayant des mutations avec un effet fondateur, la fréquence peut être plus élevée, comme dans les populations juives ashkénazes dont la fréquence est estimée à 1 sur 77-108 [108].

En raison de la variabilité phénotypique observée chez les patients AF, le diagnostic d'AF ne peut pas être basé sur les données cliniques uniquement. Même en l'absence d'anomalies physiques, le diagnostic ne doit pas être exclu car elles peuvent être présentes chez 60 à 75% des individus.

### **3.2.1. Diagnostic clinique:**

Bien que l'AF soit principalement une maladie pédiatrique, les patients adultes atteints d'AF âgés de plus de 18 ans représentent une proportion croissante de l'ensemble de la population de patients atteints d'AF. Cela peut être attribué à une meilleure prise en charge des jeunes patients atteints d'AF et à des tests diagnostiques rigoureux chez les adultes.

Les principales caractéristiques cliniques de l'AF sont:

- les anomalies congénitales
- une insuffisance progressive de la moelle osseuse se manifestant par une pancytopenie
- une prédisposition au cancer

Cependant, certaines personnes atteintes d'AF n'ont ni anomalies physiques ni insuffisance médullaire.

#### ***a. Les anomalies congénitales :***

Elles sont présentes chez environ 75% des personnes atteintes d'AF, environ 25 à 40% des patients atteints d'AF sont physiquement normaux [109]. On peut retrouver :

- Un retard de croissance prénatal et/ou postnatal (attribué à une production insuffisante d'hormone de croissance et à une hyperthyroïdie).

- Une pigmentation cutanée anormale (40%) : hyperpigmentation généralisée, taches café au lait, hypopigmentation.
- Des malformations squelettiques des membres supérieurs, unilatérales ou bilatérales (35%), les malformations des membres inférieurs sont moins fréquentes (5%) :
  - Les pouces sont généralement plus petits (hypoplasiques), bifides, dupliqués, triphalangiens, longs, placés proximalelement ou absents.
  - Le rayon de l'avant-bras est réduit ou absent.
  - Mains: éminence thénar plate, absente du premier métacarpien, clinodactylie, polydactylie.
  - Cubitus dysplasique ou court.
- Une microcéphalie (20%).
- Des anomalies ophtalmiques (20%) : Microphthalmie, cataractes, astigmatisme, strabisme, plis épicaux, hypotélorisme, hypertélorisme, ptose.
- Des anomalies des voies génito-urinaires :
  - Anomalies rénales (20%) : fer à cheval, rein ectopique, hypoplasique, dysplasique ou absent; hydronéphrose.
  - Les anomalies des organes génitaux males sont observées dans 25% des cas (hypospadias, micropénis, cryptorchidie, anorchidie, hypo- ou azoospermie, fertilité réduite). Seules 2% des femmes peuvent présenter une malposition utérine, un utérus bicorne ou des petits ovaires.
- Retard de développement et/ou déficience intellectuelle (10%).
- Autres : Endocrinien, perte auditive, malformation cardiaque congénitale, gastro-intestinale, système nerveux central, Anomalies de la colonne vertébrale et du cou.
- Une dysmorphie faciale caractéristique (2%) : un visage triangulaire, micrognathie, hypoplasie médiane.

***b. Les anomalies hématologiques :***

Il s'agit de la manifestation pathologique la plus prédominante de l'AF. 75 à 90% des patients atteints d'AF développent une insuffisance médullaire au cours de la première décennie de vie.

La plupart des patients atteints d'AF risquent de développer divers degrés d'hémopathie maligne, notamment le syndrome myélodysplasique (SMD) ou la leucémie aiguë myéloïde (LAM). En effet, le risque de survenue d'une LAM est d'environ 800 fois plus élevé que celui de la population générale. Des études récentes ont révélé un certain nombre d'anomalies chromosomiques récurrentes chez les patients atteints d'AF avec SMD ou LAM (par exemple, gain de 1q23-32, 3q26 ; monosomie 7...), suggérant que ces anomalies peuvent être des marqueurs prédictifs valables [110,111].

***c. Prédilection aux cancers :***

En plus des cancers hématologiques, les tumeurs solides, principalement les cancers à cellules squameuses (SCC) de la tête et du cou et les cancers cervicaux/gynécologiques, surviennent à des taux évidemment plus élevés chez les patients atteints d'AF [112]. Environ un tiers des patients atteints d'AF auraient développé une tumeur solide au cours de leur quatrième décennie de vie [113].

Pour résumer, un déficit de croissance, une pigmentation anormale de la peau (taches café au lait ou une hypopigmentation) et des malformations du système squelettique (anomalies du pouce) sont les caractéristiques physiques les plus fréquentes. Dans notre étude, 77% des patients avaient une petite taille, suivis d'anomalies cutanées et squelettiques présentes dans 44% des cas. Dans la première décennie de vie, les patients peuvent présenter des manifestations hématologiques avec production anormale de cellules sanguines. L'anémie

peut être isolée ou associée à l'atteinte de l'autre lignée cellulaire hématopoïétique. Tous les patients ont présenté une anémie. Une bicytopenie (anémie avec thrombocytopenie ou leucopénie) a été observée chez 66% des cas, tandis qu'une pancytopenie a été retrouvée chez 33% des patients entraînant une insuffisance médullaire progressive. Cependant, aucun des patients n'avait de preuve clinique de tumeurs hématologiques ou solides.

Le diagnostic différentiel peut se poser avec les autres causes d'aplasie médullaire sans la mais avec absence d'instabilité chromosomique. L'anémie de Diamond-Blackfan; caractérisée par la présence d'une anémie macrocytaire sans autre cytopénie significative avec des anomalies congénitales et la dyskératose congénitale associant une atteinte cutanéomuqueuse (pigmentation anormale de la peau, dystrophie des ongles et leucoplasie de la muqueuse), une atteinte multi-systémique (fibrose pulmonaire ou hépatique) et une prédisposition aux cancers.

Le diagnostic différentiel se pose également avec d'autres syndromes associés à des cassures chromosomiques spontanées notamment l'ataxie télangiectasie et le syndrome de Bloom.

### **3.2.2. Etude cytogénétique :**

L'analyse cytogénétique en présence d'un clastogène, tel que la mitomycine C ou le diépoxybutane, a été la première étude génétique réalisée chez des patients suspects d'AF [114,1155]. Ce test est positif et confirme le diagnostic dans plus de 90% des cas. Pour les cas restants, l'absence de cassures chromosomiques sur une culture des lymphocytes sur sang périphérique peut s'expliquer par un mosaïcisme somatique. Dans ce dernier cas, le caryotype sous agents alkylants doit être réalisé sur les fibroblastes, en particulier chez les patients présentant des arguments cliniques pertinents évocateurs d'AF.

Toutes les cellules des patients inclus dans notre étude ont montré une instabilité significativement accrue par rapport aux témoins avec un taux de cassures chromosomiques et des formes radiales plus élevés.

### **3.2.3. La Génétique de l'AF**

#### **3.2.3.1. Les gènes de l'AF**

Jusqu'au jour de la rédaction de cette thèse, des variantes pathogènes dans au moins 23 gènes différents se sont avérés provoquer l'AF (*FANCA*, *FANCB*, *FANCC*, *FANCD1/BRCA2*, *FANCD2*; *FANCE*; *FANCF*; *FANCG*; *FANCI*; *FANCJ/BRIP1*, *FANCL*, *FANCM*, *FANCN/PALB2*, *FANCO/RAD51C*, *FANCP/SLX4*, *FANCQ/ERCC4*, *FANCR (RAD51)*, *FANCS (BRCA1)*, *FANCT (UBE2T)*, *FANCU (XRCC2)*, *REV7 / MAD2L2*, *RFWD3*). Ils sont hérités selon trois modes de transmission (autosomique récessive, autosomique dominant et lié à l'X) (tableau 6).

La variabilité phénotypique observée dans l'AF reflète l'hétérogénéité génétique.

Gène <sup>1,2</sup>	Groupe de Complémentation <sup>3</sup>	% des AFA attribués aux variants de ce gène <sup>4</sup>
<i>BRCA1</i>	FA-S	<1%
<i>BRCA2</i>	FA-D1	2%
<i>BRIP1</i>	FA-J	2%
<i>ERCC4</i>	FA-Q	<1%
<i>FANCA</i>	FA-A	60%-70%
<i>FANCB</i>	FA-B	2%
<i>FANCC</i>	FA-C	14%
<i>FANCD2</i>	FA-D2	3%
<i>FANCE</i>	FA-E	3%
<i>FANCF</i>	FA-F	2%
<i>FANCG (XRCC9)</i>	FA-G	10%
<i>FANCI</i>	FA-I	1%
<i>FANCL</i>	FA-L	<1%
<i>FANCM</i>	FA-M	<1%
<i>FAAP100</i>	FA-Y	1 individu <sup>8</sup>
<i>PALB2</i>	FA-N	<1%
<i>RAD51</i>	FA-R	2 cas rapportés
<i>RAD51C</i>	FA-O	<1%
<i>REV7 (MAD2L2)</i>	FA-V	1 cas rapporté
<i>RFWD3</i>	FA-W	1 cas rapporté
<i>SLX4</i>	FA-P	<1%
<i>UBE2T</i>	FA-T	<1%
<i>XRCC2</i>	FA-U	1 cas rapporté
Unknown	NA	<5%

**Tableau 6: Groupes de complémentation et gènes FANC responsables de l'AF (Genereviews).**

Les mutations dans ces gènes sont héritées selon un mode autosomique récessif sauf pour le gène *FANCB* qui est lié à l’X.

Les gènes les plus fréquemment mutés sont *FANCA*, *FANCG* et *FANCC*, qui représentent ensemble entre 84 et 94% de l’ensemble des gènes impliqués dans l’AF. De plus, la plupart des mutations décrites sont des mutations privées, incluant de grandes délétions intragéniques qui représentent environ 20% de tous les allèles du gène *FANCA* (Fanconi Database, The Rockefeller University, <http://www.rockefeller.edu/fanconi>).

Au cours des dernières années, plusieurs laboratoires ont proposé le séquençage massivement parallèle pour le diagnostic moléculaire de l’AF en utilisant des panels ciblés qui comprennent tous les gènes de l’AF [116]. D’autres gènes connus pour être associés à d’autres troubles de la moelle osseuse ou de l’instabilité chromosomique peuvent également être inclus dans certains panels.

Les trois gènes ont été inclus dans notre panel afin de définir le profil moléculaire chez les patients marocains. Dans notre étude, nous avons pu confirmer le diagnostic chez 8 des 9 patients testés issus de familles non apparentées. Parmi eux, 62,5% portaient des mutations dans le gène *FANCA*, ce qui est similaire aux résultats précédents suggérant que *FANCA* est le premier gène impliqué dans l’AF, représentant 60 à 70% [117,118]. Jusqu’au 28 Septembre 2021, plus de 1296 variantes ont été rapportées comme des mutations pathogènes dans la base de données Human Mutation Gene Database (HGMD) (<http://www.hgmd.cf.ac.uk/ac/index.php>).

Les types de mutations les plus fréquentes dans le gène *FANCA* sont les mutations faux-sens et nonsens et les grandes délétions [117,119]. La plupart des mutations sont des mutations privées propre à chaque famille.



La délétion de l'exon 15 n'a été retrouvée que chez un patient et ne représente que 20% des mutations *FANCA*. Il ne s'agit pas de mutation récurrente chez les patients marocains testés avec AF contrairement à la population tunisienne qui a une fréquence plus élevée de délétion de l'exon 15, et qui constitue une mutation récurrente avec un effet fondateur en Tunisie [120].

La mutation faux-sens c.1304G>A, p.Arg435His a déjà été rapporté à l'état hétérozygote composite avec la mutation c.2854C>T, p.Q952X [121].

De même que la mutation c.2851C>T, p.Arg951Trp, rapportée dans plusieurs études et elle connue comme étant pathogène. Concernant la mutation c.1126C>T, p.Gln376\*, elle fut décrite une fois comme étant associée à un phénotype d'AF [122].

Un seul patient était hétérozygote composite pour deux mutations du gène *FANCA* (c.3520\_3522delTGG, p.Trp1174del et c.2189delT, p.Leu730Argfs\*21).

Dans une étude récente publiée par Ben Haj Ali A. et al, l'équipe a présenté le résultat de l'étude moléculaire du gène *FANCA* chez 29 patients originaires du Nord Afrique (Tunisie, Libye et Algérie). Trois mutations à effet fondateur ont été identifiés dans 86.7% des 22 familles: la délétion de l'exon 15, chez 36.4% patients (8/22), une délétion des exons 4 and 5 dans 23% (5/22) et une mutation intronique c.2222+166G>A, dans 27.3% (6/22) [123].

Certes ; nous n'avons pas retrouvé de mutations récurrente du gène *FANCA* dans notre étude, mais deux des cinq mutations identifiés ont été déjà décrite dans les populations maghrébines. En plus de la population Tunisienne, la délétion de l'exon 15 a été également retrouvée chez des patients originaires d'Algérie et de la Libye. La mutation c.1304G>A, p.Arg435His ; retrouvée chez un patient Marocain ; a été déjà rapporté chez un patient d'origine Tunisienne et un autre d'origine Libyenne (Figure 27)



*Figure 27: Distribution géographique des mutations communes du gène FANCA en Afrique du Nord (Modifié à partir de Ben Haj Ali A. et al ;2021).*

Le gène *FANCG* est le deuxième gène muté chez nos patients, il représente 37,5% et il est relativement plus élevé en le comparant à d'autres études rapportant une fréquence de 10% [124,125]. 178 mutations ont été décrites à ce jour dans la base de données HGMD de type faux-sens et nonsens, mutations d'épissage, petite délétion et insertion et des insdel. Contrairement au gène *FANCA* ; il n'y a pas de grands réarrangements dans le gène *FANCG*.

La mutation c.1642C>T, p.Arg548\* retrouvée dans notre série a été déjà rapportée chez un patient d'origine Turque [126].

Les deux autres mutations retrouvées dans le gène *FANCG* sont des nouvelles mutations qui n'ont jamais été décrites. En effet, la mutation c.1034A>T, p.Gln345Leu n'est pas rapportée dans la base de données ClinVar, elle n'est pas présente dans les bases de données des populations et elle n'a pas été retrouvée dans la population marocaine sur 200 chromosomes (Base de données locale). Selon les sites de prédiction des variations de séquence sur la protéine Polyphen2 et SIFT, elle est prédite comme étant probablement pathogène et pathogène respectivement. Par conséquent, on peut dire que ce résultat est compatible avec le diagnostic de l'AF.

Concernant la nouvelle mutation délétère c.1474G>T (p.Glu492\*), elle n'est pas rapporté dans la base de données ClinVar ni dans la base de données GNOMAD, la formation d'un codon stop prématuré entrainera la formation d'une protéine tronquée et par conséquent un phénotype de l'AF.

Comme dans les populations iranienne et coréenne, aucune mutation du gène *FANCC* n'a été retrouvé dans la présente étude [127,128].

L'étude moléculaire n'est pas seulement importante pour établir un diagnostic génétique du propositus, mais a également un intérêt pronostique et des implications thérapeutiques. L'identification de la mutation causale permet de prodiguer un conseil génétique adéquat, de réaliser un diagnostic prénatal et de détecter les personnes hétérozygotes dans une famille à risque.

#### **3.2.3.2. *Corrélation génotype-phénotype***

Le spectre clinique de l'AF reste très hétérogène. Il n'y a pas de corrélations génotype-phénotype claires. En général, les variantes nulles conduisent à un phénotype plus sévère (par exemple, anomalies congénitales, insuffisance médullaire précoce et SMD/AML que les variantes hypomorphes.

Les patients AF du sous-type A ont une maladie plus bénigne et un début tardif d'insuffisance médullaire. Les patients des sous-types *FANCC* et *FANCG* ont généralement un phénotype plus sévère et nécessitent une intervention précoce qui comprend une greffe de moelle osseuse.

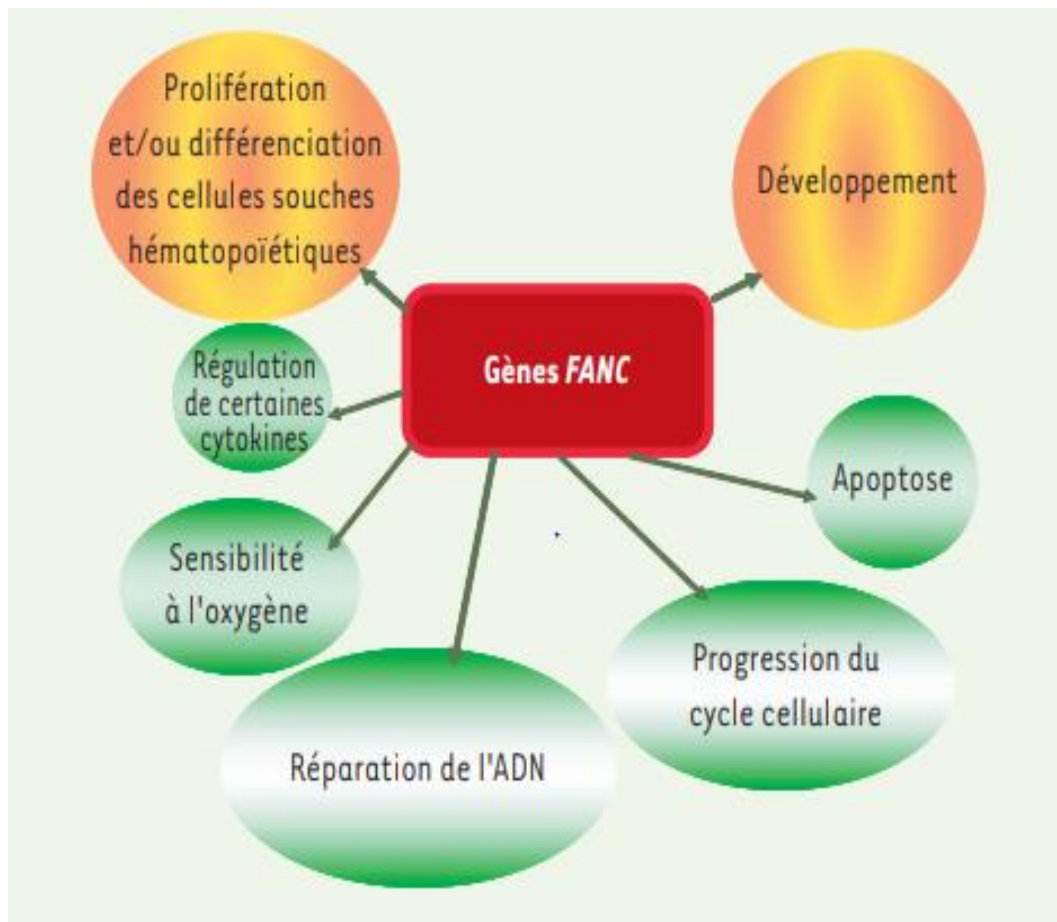
Les variants pathogènes bialléliques du gène *BRCA2* sont associés à des leucémies aiguës précoces et à des tumeurs solides [129-131]. La probabilité cumulée de toute malignité était de 97 % à l'âge de six ans, y compris la LAM, le médulloblastome et la tumeur de Wilms [132].

Plus récemment, il a été démontré que les variantes pathogènes du gène *FANCB* provoquent surtout une insuffisance médullaire à début précoce et des anomalies congénitales graves. [133].

Concernant le gène *FANCG*, les variantes pathogènes peuvent être associées à une insuffisance médullaire sévère et à une incidence plus élevée de leucémie que le gène *FANCC* [Faivre et al 2000]. Les tumeurs solides (médulloblastome, tumeur de Wilms) sont associées plutôt à des variantes pathogènes du gène *PALB2* [134].

### **3.2.3.3. Gènes *FANC* et la voie *FA-BRCA***

Les produits de ces gènes *FANC* (-A, B, C, D1 (*BRCA2*) / D2, E, F, G (*XRCC9*), I, J (*BRIP1*), L (*PHF9*), M, N (*PALB2*), O (*RAD51C*), P (*SLX4*), Q (*XPF*), R (*RAD51*), S (*BRCA1*), T (*UBE2T*), U (*XRCC2*), V (*REV7 / MAD2L2*), W (*RFWD3*)) interagissent dans une voie de signalisation commune qui est la voie *FA-BRCA*, impliquée dans le maintien de l'intégrité du génome à travers le contrôle de la réparation de l'ADN (Figure 28).



**Figure 28: Principaux processus cellulaires affectés dans l'anémie de Fanconi.**

Tous les groupes génétiques de complémentation présentent un défaut de réparation de l'ADN lésé par des agents de pontages de l'ADN, des anomalies de la régulation du cycle cellulaire et de l'homéostasie de facteurs de croissance, de l'apoptose et de la détoxification d'intermédiaires actifs de l'oxygène [135].

La perturbation de cette voie conduit au phénotype cellulaire et clinique commun observé dans l'AF.

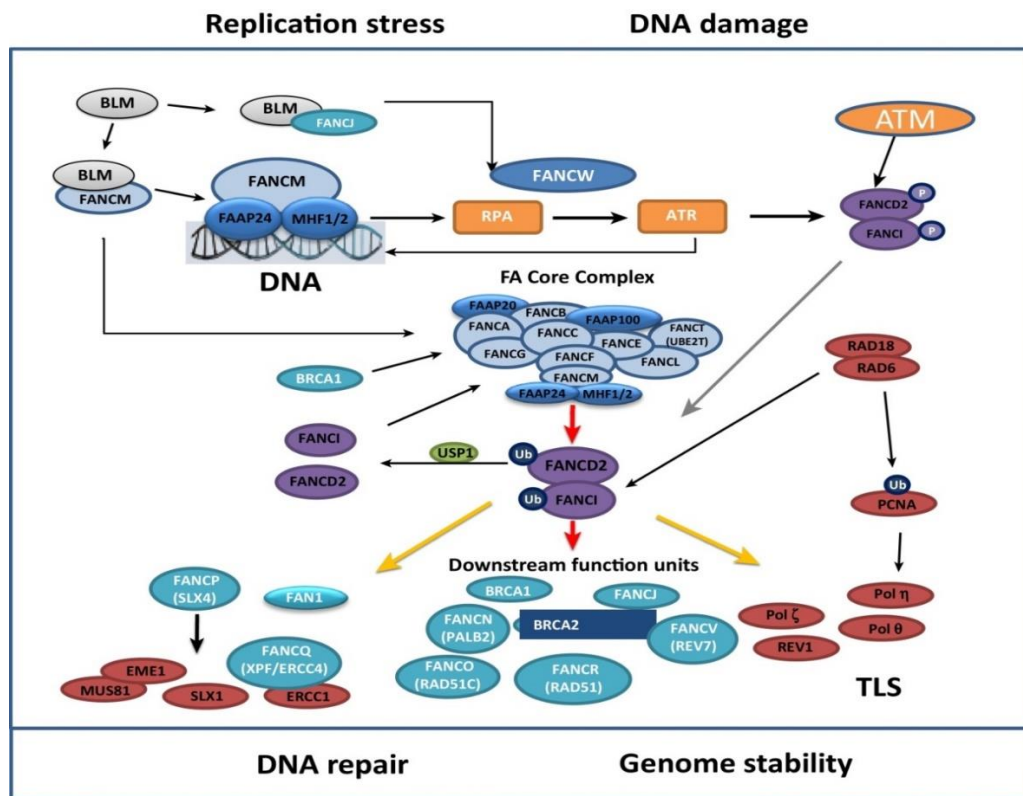
L'ADN est répliqué par l'ADN polymérase au niveau de plusieurs fourches de réplication. Toutes les lésions qui perturbent la structure de la chromatine, telles que les réticulations d'ADN induites chimiquement, vont entraîner un blocage de la réplication à la position de la lésion. La voie de l'AF agit principalement en réponse à des dommages à l'ADN qui entraînent le blocage des fourches de réplication de l'ADN pendant la phase S du cycle cellulaire. L'activation du complexe central FA, composé d'au moins 8 des protéines FA connues (A, B, C, E, F, G, L et M), se produit par le recrutement d'ATR (ataxie télangiectasie liée à Rad3) à l'ADN simple brin au niveau de la fourche de réplication bloquée [137,138]. Le complexe FA activé fonctionne comme une ligase d'ubiquitine qui mono-ubiquitine FANCD2. FANCL est l'élément catalytique important du complexe central. La monoubiquitination de FANCD2 entraîne sa relocalisation du compartiment nucléaire soluble vers la chromatine et par la suite son association avec d'autres protéines de réparation, telles que BRCA1, RAD51, NBS1 et PCNA et avec FANCD1 [136,139].

Les cellules qui ne peuvent pas former le complexe 1 ou le complexe 2 sont généralement hypersensibles aux agents alkylants tels que le MMC et ces cellules présentent des cassures chromosomiques et des formes radiales caractéristiques de l'AF. Les mutations dans huit sous-types d'AF (FANCA, FANCB, FANCC, FANCE, FANCF, FANCG, FANCL et FANCM) entraînent la perte de la monoubiquitylation FANCD2 et FANCI, qui est l'étape réglementaire centrale de la voie FA. On rapporte que 90% des individus atteints de FA présentent des mutations enregistrées dans ces huit gènes de FA.

Une association étroite a été établie entre la voie FA et les protéines BRCA dans les mécanismes de réparation de l'ADN (Figure 29).

La voie de réparation de l'ADN FA/BRCA intervient dans la réparation des lésions de l'ADN double brin (Interstrand-crosslinks (ICL)). Il s'agit de la formation de liens covalents entre brins d'ADN qui créent un obstacle à la réplication de l'ADN et la transcription. Les ICL sont d'origine endogène ou exogène les produits chimiques, tels que les aldéhydes réactifs, et les expositions, y compris les psoralènes naturels et la réticulation de l'ADN par des médicaments comme le cisplatine. FANCA, -B, -C, -E, -F, -G, -L et -M comprennent le « complexe de base » FA qui fonctionne comme une multi-sous-unité E3 ubiquitine ligase. Lorsque des dommages à l'ADN surviennent, le noyau complexe interagit avec FANCL pour catalyser la monoubiquitination de FANCD2 et FANCI. Le FANCD2 omniprésent et l'hétérodimère FANCI est un composant clé du processus de réparation de l'ADN. Autres composants de la voie FA / BRCA ont des fonctions distinctes mais liées dans la réparation de l'ADN. FANCP (également appelé SLX4) et FANCQ (ERCC4) créent le complexe nucléase qui est nécessaire pour l'incision à la réticulation de l'ADN des sites. FANCD1, FANCI (BRIP1), FANCN (PALB2) et FANCO (RAD51C) en recombinaison homologue et réparation via BRCA / RAD51 [140].





**Figure 29: Voie de signalisation FA. Les protéines FANC**

(FANCA, B, C, E, F, G, L, M, T et éventuellement I) ainsi que les FAAP (FAAP 20/24/100 et MHF1/2) forment le complexe central qui assure l'activité de l'ubiquitine E3 ligase nécessaire pour l'ubiquitination des protéines FANCI et FANCD2, qui orchestrent les acteurs en aval, y compris le reste des protéines FA et d'autres protéines non-FA pour réparer les dommages à l'ADN via des mécanismes de BER, NER, TLS, HR et NHEJ pour maintenir la stabilité du génome. De plus, USP1 peut désubiquitiner FANCD2/I, inactivant ainsi la voie FA-BRCA. Plus important encore, ATM, ATR et HHR6, ainsi que d'autres connus (FANCM et FAAP) ou encore à identifier, peuvent agir en tant que régulateurs en amont pour les réponses de point de contrôle bien exécutées en cas de dommages à l'ADN ou de stress réplcatifs. La détection de la protéine FANCM (en amont), incluant éventuellement la modulation FANCW des fonctions RPA, serait un exemple typique de signalisation FA pour montrer des rôles indépendants de la voie [136].

#### **3.2.3.4. *Prise en charge :***

La prise en charge des patients atteints d'AF est pluridisciplinaire (Hématologue, endocrinologue, ophtalmologue, traumatologue.....) avec un suivi régulier afin de réaliser un traitement adéquat des symptômes manifestes

Le plus souvent, il s'agit de traitement symptomatique (transfusion sanguine, corticothérapie...). La greffe de cellules souches hématopoïétiques (GCSH) est le seul traitement curatif des manifestations hématologiques. Idéalement, elle doit être réalisée avant l'apparition du SMD/LAM et avant les transfusions multiples. Les personnes atteintes d'AF sont sensibles à la chimiothérapie et à la radiothérapie, elles ont besoin de régimes de transplantation spéciaux et doivent être soignées et transplantées dans des centres expérimentés.

#### **3.2.3.5. *Conseil génétique :***

Le conseil génétique des familles dépend du mode de transmission de la maladie. Dans notre série, il s'agit de gènes obéissant au mode de transmission autosomique récessif.

Les parents d'un enfant atteint d'AF autosomique récessif sont des hétérozygotes obligatoires. Les tests génétiques moléculaires sont recommandés pour les parents d'un proposant pour confirmer que les deux parents sont hétérozygotes pour un variant pathogène lié à l'AF et pour permettre une évaluation fiable du risque de récurrence. Si les deux parents sont hétérozygotes pour la mutation, le risque de survenue de la maladie est de 25% pour chaque grossesse ultérieure.

Les hétérozygotes ne sont pas à risque d'AF autosomique récessive. Cependant, la mutation hétérozygote d'un sous-ensemble de gènes liés à l'AF (*BRCA1*, *BRCA2*, *PALB2*, *BRIP1* et *RAD51C*) est associée à un risque accru de cancer du sein et d'autres cancers.

## **Conclusion et respectives**

Certes le séquençage à haut débit montre aujourd'hui des résultats très prometteurs dans plusieurs maladies génétiques en analysant plusieurs gènes à la fois, la méthode de Sanger gardera toujours sa place dans le diagnostic moléculaire ciblé des anomalies de développement. Le rôle du généticien est de proposer l'analyse adéquate aux familles en prescrivant un test génétique adapté en fonction des données génétiques actualisées (nombre de gènes, types de mutations, présence de mutations récurrentes, etc.).

Les technologies innovantes de séquençage du génome et ceux axés sur les maladies génétiques ajouteront de plus en plus des variantes aux bases de données actuelles. Le défi qui reste est l'analyse de ces informations et les connaissances à acquérir concernant la biologie de notre propre génome. La variation structurelle est toujours difficile à évaluer en utilisant uniquement les plates-formes NGS. Ceci requiert la comparaison et la standardisation des technologies de séquençage avec une amélioration des algorithmes de cartographie et d'assemblage de novo afin de pouvoir réaliser une prédiction précise des nouveaux variants surtout de type indels.

Un autre défi technique est le stockage et l'accès aux vastes quantités de données produites par les projets génomiques. Doit-on stocker les fichiers de données produits par les machines de séquençage ou uniquement les données traitées et analysées ?

À notre connaissance, il s'agit de la première étude marocaine à effectuer une analyse moléculaire à l'aide des méthodes NGS pour des patients marocains atteints d'AF.

Nos résultats confirment que l'approche NGS est une méthode efficace pour la détection de la mutation causale dans les gènes responsables de l'anémie de Fanconi. Nous projetons dans un avenir très proche, d'élargir l'étude à d'autres patients atteints d'AF par l'analyse des trois gènes les plus fréquemment mutés afin mieux caractériser le profil moléculaire des patients Marocains dans l'espoir de retrouver une ou plusieurs mutations récurrentes. Ces

mutations ainsi identifiées, permettront la mise en place en santé publique d'un test génétique peu coûteux avec un rendement raisonnable. Concernant les patients qui n'ont aucun variant dans les gènes étudiés, nous leur proposerons ; dans un cadre de recherche de réaliser un panel élargi comprenant les gènes de l'anémie de Fanconi dans l'objectif d'identifier une mutation causale. Si aucune mutation n'est retrouvée au niveau des gènes connus, on peut le proposer une étude de l'exome total.

La détermination des gènes impliqués dans les maladies héréditaires avec risque de cancer notamment celles avec un défaut de réparation de l'ADN ; la connaissance des mécanismes moléculaires à l'origine du phénotype et la compréhension des voies de signalisations sous-jacentes conduiraient inéluctablement au développement de nouvelles thérapies contre le cancer. Certaines ont déjà vu le jour comme les inhibiteurs PARP, les inhibiteurs HDAC1 [141].

# Références

1. Sanger F, Nicklen S and Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proc. Natl. Acad. Sci. USA* 1977; 74: 5463-5467
2. Lapunzina P, López RO, Rodríguez-Laguna L, García-Miguel P, Rojas Martínez A and Martínez-Glez V. Impact of NGS in the medical sciences: Genetic syndromes with an increased risk of developing cancer as an example of the use of new technologies. *Genetics and Molecular Biology* 2014; 1 suppl 37: 241-249
3. Yao Y, Dai W. Genomic instability and cancer. *J Carcinog Mutagen* 2014; 5:1000165..
4. Dutrillaux B, Couturier J. *La Pratique de L'analyse Chromosomique*. Paris: Masson; 1981.
5. Lorenz TC. Polymerase Chain Reaction: Basic Protocol Plus Troubleshooting and Optimization Strategies. *J. Vis. Exp.* 2012 (63), e3998.
6. Sanger F. Determination of nucleotide sequences in DNA. *Biosci Rep* 2004 Aug Oct; 24(4-5):237-253.
7. Rothberg JM et al. An integrated semiconductor device enabling non-optical genome sequencing. *Nature* 2011 Jul 20; 475(7356):348-52.
8. Liu L, Li Y, Li S, Hu N, He Y, Pong R, Lin D, Lu L, Law M. Comparison of Next-Generation Sequencing Systems. *BioMed Research International* 2012; 2012: 11 pages.
9. Heather JM, Chain B. The sequence of sequencers: The history of sequencing DNA. *Genomics* 2016; 107:1–8
10. Maxam M, Gilbert W. A new method for sequencing DNA. *Proc. Natl. Acad. Sci. USA*, 1977; 74: 560-564.
11. Sanger F, Air GM, Barrell BG, Brown NL, Coulson AR, Fiddes CA, Hutchison CA, Slocombe PM, Smith M. Nucleotide sequence of bacteriophage phi X174 DNA. *Nature* 1977; 265: 687-695.
12. Smith LM, Sanders JZ, Kaiser RJ, Hughes P, Dodd C, Connell CR, Heiner C, Kent SB, Hood LE. Fluorescence detection in automated DNA sequence analysis. *Nature* 1986; 321:674–679.
13. Collins FS, Morgan M, Patrinos A. The Human Genome Project: lessons from large-scale biology. *Science* 2003 Apr 11; 300(5617):286-290.
14. Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, et al. Big Data: Astronomical or Genomical?. *PLoS Biol* 2015; 13(7): e1002195.
15. Wetterstrand KA. The cost of sequencing a Human Genome 2021: Available from URL:<https://www.genome.gov/sequencingcosts/>

16. Schadt EE, Turner S, Kasarskis A. A window into third-generation sequencing. *Hum. Mol. Genet.* 2010; 19(R2): R227–R240.
17. Niedringhaus TP, Milanova D, Kerby MB, Snyder MP, Barron AE. Landscape of next-generation sequencing technologies. *Anal. Chem.* 2011; 83: 4327–4341.
18. Pareek CS, Smoczynski R, Tretyn A. Sequencing technologies and genome sequencing. *J. Appl. Genet.* 2011; 52: 413–435.
19. Gut IG. New sequencing technologies. *Clin. Transl. Oncol.* 2013; 15(11): 879–881.
20. Van Dijk EL, Jaszczyszyn Y, Naquin D, Thermes C. The Third Revolution in Sequencing Technology. *Trends Genet* 2018 Sep; 34(9): 666-681.
21. Lamoril J, Ameziane N, Deybacha JC, Bouizegarènea P, Bogard M. Les techniques de séquençage de l'ADN: une révolution en marche. *Immuno-analyse et Biologie Spécialisée* 2008; 23(5) : 260-279.
22. Mardis ER. The impact of next-generation sequencing technology on genetics. *Trends Genet* 2008 Mar; 24(3):133-141.
23. Church GM. Genomes for all. *Sci Am.* 2006; 294(1):46–54.
24. Drmanac R, Drmanac S, Chui G, Diaz R, Hou A, Jin H, Jin P, Kwon S, Lacy S, Moer B, Shafto J, Swanson D, Ukrainczyk T, Xu C, Little D. Sequencing by hybridization (SBH): advantages, achievements, and opportunities. *Adv Biochem Eng Biotechnol* 2002; 77:75– 101.
25. Hanna GJ, Johnson VA, Kuritzkes DR, Richman DD, Martinez-Picado J, Sutton L, Hazelwood JD, D'Aquila RT. Comparison of sequencing by hybridization and cycle sequencing for genotyping of human immunodeficiency virus type 1 reverse transcriptase. *J Clin Microbiol* 2000; 38(7): 2715–2721.
26. Mirzabekov AD. DNA sequencing by hybridization- a megasequencing method and a diagnostic tool? *Trends Biotechnol* 1994; 12(1):27–32.
27. Qin Y, Schneider TM, Brenner MP. Sequencing by hybridization of long targets. *PLoS One* 2012; 7(5):e35819.
28. Fuller CW et al. The challenges of sequencing by synthesis. *Nat Biotechnol* 2009; 27(11):1013–1023.
29. Metzker ML. Sequencing technologies - the next generation. *Nat Rev Genet* 2010; 11(1):31–46.
30. Quail MA et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* 2012; 13:341.



31. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, Waterston RH. DNA sequencing at 40: past, present and future. *Nature* 2017; 550(7676):345–353.
32. Berka J, Chen YJ, Leamon JH, et al. Bead emulsion nucleic acid amplification. U.S. Patent Application, 2011.
33. Shao K, Ding W, Wang F, Li H, Ma D, Wang H. Emulsion PCR: a High Efficient Way of PCR Amplification of Random DNA Libraries in Aptamer Selection. *PLoS ONE* 2011; 6:e24910.
34. Foehlich T, et al. High-throughput nucleic acid analysis. U.S. Patent, 2010.
35. Fakruddin M, Chowdhury A, Hossain M, Mannan KSB, Mazumdar RM. Pyrosequencing-principles and applications. *Life* 2012; 50:65.
36. Slatko BE, Gardner AF, Ausubel FM. Overview of Next Generation Sequencing Technologies. *Curr Protoc Mol Biol* 2018 April ; 122(1): e59.
37. Mardis ER. Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 2008; 9:387-402.
38. Shuikan A, Alharbi SA, Alkhalifah DHM and Hozzein WN. High-Throughput Sequencing and Metagenomic Data Analysis, *Metagenomics - Basics, Methods and Applications* 2019 IntechOpen.
39. Boland JF, Chung CC, Roberson D, Mitchell J, Zhang X, Im KM, He J, Chanock SJ, Yeager M, Dean M. The new sequencer on the block: comparison of Life Technology's Proton sequencer to an Illumina HiSeq for whole-exome sequencing. *Hum Genet* 2013; 132(10):1153-1163.
40. Trygve O. Tollefsbol (ed.). *Single-Cell Semiconductor Sequencing Biological Aging: Methods and Protocols, Methods in Molecular Biology*. August 2013; vol. 1048.
41. Evelien M Bunnik, Karine G Le Roch. *An Introduction to Functional Genomics and Systems Biology*. *Adv Wound Care (New Rochelle)* 2013 Nov; 2(9):490-498.
42. Hert DG, Fredlake CP, Barron AE. Advantages and limitations of next-generation sequencing technologies: A comparison of electrophoresis and non-electrophoresis methods. *Electrophoresis* 2008; 29:4618-4626.
43. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, et al. The sequence of the human genome. *Science* 2001; 291:1304-1351.
44. Bell, C. J. et al. Carrier testing for severe childhood recessive diseases by next-generation sequencing. *Sci Transl Med* 2011; 3(65): 65ra64.

45. Majewski J, Schwartzentruber J, Lalonde E, Montpetit A, Jabado N. What can exome sequencing do for you? *J Med Genet* 2011 Sep; 48(9):580-589.
46. Xue, Y, Ankala A, W. R. Wilcox and M. R. Hegde. Solving the molecular diagnostic testing conundrum for Mendelian disorders in the era of next-generation sequencing: single-gene, gene panel, or exome/genome sequencing. *Genet Med* 2015; 17(6): 444-451.
47. Metzker ML. Sequencing technologies-the next generation. *Nature Reviews. Genetics* 2010; 11:31-46.
48. Magi A, Benelli M, Gozzini A, Girolami F, Torricelli F, Brandi ML. Bioinformatics for next generation sequencing data. *Genes (Basel)* 2010; 1(2):294–307.
49. DePristo MA, Banks E, Poplin R, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet* 2011; 43(5):491–498.
50. De Leng WW, Gadellaa-van Hooijdonk CG, Barendregt-Smouter FA, et al. Targeted next generation sequencing as a reliable diagnostic assay for the detection of somatic mutations in tumors using minimal DNA amounts from formalin fixed paraffin embedded material. *PLoS One* 2016; 11(2):e0149405.
51. Tu J, Ge Q, Wang S, et al. Pair-barcode high-throughput sequencing for large-scale multiplexed sample analysis. *BMC Genomics* 2012;13:43.
52. Van der Auwera GA, Carneiro MO, Hartl C, et al. From FastQ data to highconfidence variant calls: the Genome Analysis Toolkit best practices pipeline. *Curr Protoc Bioinformatics* 2013; 43:11.10.1–33.
53. Yohe S, Thyagarajan B. Review of Clinical Next-Generation Sequencing. *Arch Pathol Lab Med* 2017; 141 (11): 1544–1557.
54. Richards S, Aziz N, Bale S, et al. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genet Med* 2015; 17(5):405–424.
55. Duzkale H, Shen J, McLaughlin H, et al. A systematic approach to assessing the clinical significance of genetic variants. *Clin Genet* 2013; 84(5):453–463.
56. Amendola LM, Jarvik GP, Leo MC, et al. Performance of ACMG-AMP variant-interpretation guidelines among nine laboratories in the Clinical Sequencing Exploratory Research Consortium. *Am J Hum Genet* 2016; 98(6):1067–1076.

57. Lek M, Karczewski KJ, Minikel EV, et al. Analysis of protein-coding genetic variation in 60,706 humans. *Nature* 2016; 536(7616):285–291.
58. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, Handsaker RE, Lunter G, Marth GT, Sherry ST, McVean G, Durbin R. The variant call format and VCF tools. 1000 Genomes Project Analysis Group.
59. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 2009 Aug 15; 25(16):2078-2079.
60. Ritter DI, Roychowdhury S, Roy A, et al. Somatic cancer variant curation and harmonization through consensus minimum variant level data. *Genome Med* 2016; 8(1):117.
61. Li MM, Datto M, Duncavage EJ, et al. Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *J Mol Diagn* 2017; 19(1):4–23.
62. Kleinberger J, Maloney KA, Pollin TI, Jeng LJ. An openly available online tool for implementing the ACMG/AMP standards and guidelines for the interpretation of sequence variants. *Genet Med* 2016; 18(11):1165.
63. International Human Genome Sequencing Consortium. Finishing the euchromatic sequence of the human genome. *Nature* 2004; 431(7011):931–945.
64. Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* 2008; 455(7216):1061–1068.
65. Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M et al. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. *Nature* 2008; 456(7218):66–72
66. Pao W, Miller V, Zakowski M, Doherty J, Politi K, Sarkaria I, Singh B, Heelan R, Rusch V, Fulton L et al. EGF receptor gene mutations are common in lung cancers from “never smokers” and are associated with sensitivity of tumors to gefitinib and erlotinib. *Proc Natl Acad Sci USA* 2004; 101(36):13306–13311.
67. « Prévalence des maladies rares: Données bibliographiques », Les cahiers d’Orphanet, Série Maladies Rares, Janvier 2020, Numéro 1: Classement par ordre alphabétique des maladies ou groupes de maladies.

[http://www.orpha.net/orphacom/cahiers/docs/FR/Prevalence\\_des\\_maladies\\_rares\\_par\\_ordre\\_alphabetique.pdf](http://www.orpha.net/orphacom/cahiers/docs/FR/Prevalence_des_maladies_rares_par_ordre_alphabetique.pdf)

68. Hirai Y, Kodama Y, Moriwaki S, et al. Heterozygous individuals bearing a founder mutation in the XPA DNA repair gene comprise nearly 1% of the Japanese population. *Mutat Res* 2006; 601:171–178.
69. Doubaj Y, Laarabi FZ, Elalaoui SC, Barkat A, Sefiani A. Carrier frequency of the recurrent mutation c.1643\_1644delTG in the XPC gene and birth prevalence of the xeroderma pigmentosum in Morocco. *J Dermatol* 2012; 39(4):382–384.
70. Zghal M, Fazaa B, Kamoun M R. Xérodernapigmentosum. Elsevier Masson 2006: 1-20
71. Laura J. Niedernhofer. Tissue specific accelerated aging in NER deficiency. *Mecanisms of ageing and development* 2008; 129: 408-415.
72. Fazaa B., KamounM.R..Xérodernapigmentosum. *Ann dematol venereol* 2003; 130: 69-73.
73. Bradford PT, Goldstein AM, Tamura D, Khan SG, Ueda T, Boyle J, Oh KS, Imoto K, Inui H, Moriwaki S, Emmert S, Pike KM, Raziuddin A, Plona TM, DiGiovanna JJ, Tucker MA, Kraemer KH. Cancer and neurologic degeneration in xeroderma pigmentosum: long term follow-up characterises the role of DNA repair. *J Med Genet.* 2011;48: 168–176.
74. Kraemer KH, Lee MM, Andrews AD, Lambert WC. The role of sunlight and DNA repair in melanoma and non melanoma skin cancer. The xeroderma pigmentosum paradigm. *Arch Dermatol* 1994; 130:1018–1021.
75. Dollfus H, Porto F, Caussade P, Speeg-Schatz C, Sahel J, Grosshans E, Flament J, Sarasin A. Ocular manifestations in the inherited DNA repair disorders. *Surv Ophthalmol* 2003; 48:107–122.
76. Brooks BP, Thompson AH, Bishop RJ, Clayton JA, Chan CC, Tsilou ET, Zein WM, Tamura D, Khan SG, Ueda T, Boyle J, Oh KS, Imoto K, Inui H, Moriwaki S, Emmert S, IliffNT, Bradford P, Digiovanna JJ, Kraemer KH. Ocular manifestations of xerodermapigmentosum: long-term follow-up highlights the role of DNA repair in protection from sun damage. *Ophthalmology* 2013; 120:1324–1336.
77. Rapin I, Lindenbaum Y, Dickson DW, Kraemer KH, Robbins JH. Cockayne syndrome and xeroderma pigmentosum. *Neurology* 2000; 55:1442–1449.
78. Totonchy MB, Tamura D, Pantell MS, Zalewski C, Bradford PT, Merchant SN, Nadol J, Khan SG, Schiffmann R, Pierson TM, Wiggs E, Griffith AJ, Digiovanna JJ, Kraemer

- KH, Brewer CC. Auditory analysis of xeroderma pigmentosum 1971-2012: hearing function, sun sensitivity and DNA repair predict neurological degeneration. *Brain* 2013; 136:194–208.
79. Kraemer KH, Lee MM, Scotto J. Xeroderma pigmentosum. Cutaneous, ocular, and neurologic abnormalities in 830 published cases. *Arch Dermatol* 1987;123:241–250.
  80. Butt FM, Moshi JR, Owibingire S, Chindia ML. Xeroderma pigmentosum: a review and case series. *J Craniomaxillofac Surg* 2010; 38:534–537.
  81. DiGiovanna JJ, Patronas N, Katz D, Abangan D, Kraemer KH. Xeroderma pigmentosum: spinal cord astrocytoma with 9-year survival after radiation and isotretinoin therapy. *J Cutan Med Surg* 1998; 2:153–158.
  82. Lai JP, Liu YC, Alimchandani M, Liu Q, Aung PP, Matsuda K, Lee CC, Tsokos M, Hewitt S, Rushing EJ, Tamura D, Levens DL, DiGiovanna JJ, Fine HA, Patronas N, Khan SG, Kleiner DE, Oberholtzer JC, Quezado MM, Kraemer KH. The influence of DNA repair on neurologic degeneration, cachexia, skin cancer and internal neoplasms: autopsy report of four xerodermapigmentosum patients (XP-A, XP-C and XP-D). *Acta Neuropathol Commun* 2013; 1:4.
  83. Fassihi H, Sethi M, Fawcett H, Wing J, Chandler N, Mohammed S, Craythorne E, Morley AM, Lim R, Turner S, Henshaw T, Garrood I, Giunti P, Hedderly T, Abiona A, Naik H, Harrop G, McGibbon D, Jaspers NG, Botta E, Nardo T, Stefanini M, Young AR, Sarkany RP, Lehmann AR. Deep phenotyping of 89 xerodermapigmentosum patients reveals unexpected heterogeneity dependent on the precise molecular defect. *Proc Natl Acad Sci USA* 2016; 113:E1236–45.
  84. Zghal M, Fazaa B, Abdelhak S, Mokni M. Xeroderma pigmentosum. *EMC - Dermatologie* 2014; 9(3):1-13.
  85. Liang C, Kraemer KH, Morris A, Schiffmann R, Price VH, Menefee E, DiGiovanna JJ. Characterization of tiger tail banding and hair shaft abnormalities in trichothiodystrophy. *J Am Acad Dermatol* 2005; 52:224 –232.
  86. Lekshmi N, Surya Prabha B and Prajnya R. Inherited Disorders of DNA Repair. *Genetic Clinics* 2017; 10 (4):7-14.
  87. Lehman AR, Kirk-Bell S, Arlett CF, Paterson MC, Lohman PHM, De Weerd-Kastelein EA, Bootsma D. Xeroderma pigmentosum cells with normal levels of excision repair have a defect in DNA synthesis after UV-irradiation. *Proc Nat Acad Sci* 1975; 72: 219-223.

88. Kraemer KH, DiGiovanna JJ. Xeroderma Pigmentosum. 2003 Jun 20 [Updated 2016 Sep 29]. In: Adam MP, Ardinger HH, Pagon RA, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2021. (<https://www.ncbi.nlm.nih.gov/books/NBK1397/>)
89. Kashiyama K, Nakazawa Y, Pilz DT, Guo C, Shimada M, Sasaki K, Fawcett H, Wing JF, Lewin SO, Carr L, Li TS, Yoshiura K, Utani A, Hirano A, Yamashita S, Greenblatt D, Nardo T, Stefanini M, McGibbon D, Sarkany R, Fasshi H, Takahashi Y, Nagayama Y, Mitsutake N, Lehmann AR, Ogi T. Malfunction of nuclease ERCC1- XPF results in diverse clinical manifestations and causes Cockayne syndrome, xeroderma pigmentosum, and Fanconi anemia. *Am J Hum Genet* 2013; 92:807–819.
90. Li L, Peterson CA, Lu X, Legerski RJ. Mutations in XPA that prevent association with ERCC1 are defective in nucleotide excision repair. *Molec Cell Biol* 1995; 15: 1993-1998.
91. Park CH, Sancar A. Formation of a ternary complex by human XPA, ERCC1, and ERCC4 (XPF) excision repair proteins. *Proc Nat Acad Sci* 1994; 91: 5017-5021.
92. Sugasawa K, Ng JM, Masutani C, Iwai S, van der Spek PJ, Eker APM, Hanaoka F, Bootsma D, Hoeijmakers JHJ. Xeroderma pigmentosum group C protein complex is the initiator of global genome nucleotide excision repair. *Molec Cell* 1998; 2: 223-232.
93. Lehman AR, Kirk-Bell S, Arlett CF, Paterson MC, Lohman PHM, De Weerd-Kastelein EA, Bootsma D. Xeroderma pigmentosum cells with normal levels of excision repair have a defect in DNA synthesis after UV-irradiation. *Proc Nat Acad Sci* 1975;72: 219-223.
94. Hirai Y, Kodama Y, Moriwaki S, Noda A, Cullings HM, Macphee DG, Kodama K, Mabuchi K, Kraemer KH, Land CE, Nakamura N. Heterozygous individuals bearing a founder mutation in the XPA DNA repair gene comprise nearly 1% of the Japanese population. *Mutat Res* 2006; 601:171–178.
95. Soufir N, Ged C, Bourillon A, Austerlitz F, Chemin C, Stary A, Armier J, Pham D, Khadir K, Roume J, Hadj-Rabia S, Bouadjar B, Taieb A, de Verneuil H, Benchiki H, Grandchamp B, Sarasin A. A Prevalent Mutation with Founder Effect in Xeroderma Pigmentosum Group C from North Africa Nadem Soufir. *Journal of Investigative Dermatology* 2010; 130 : 1537–1542.
96. Tamhankar PM, Iyer SV, Ravindran S, Gupta N, Kabra M, Nayak C, Kura M, Sanghavi S, Joshi R, Chennuri VS, Khopkar U. Clinical profile and mutation analysis of

- xeroderma pigmentosum in Indian patients. *Indian J Dermatol Venereol Leprol* 2015; 81:16–22.
97. Hirai Y, Kodama Y, Moriwaki S, Noda A, Cullings HM, Macphee DG, Kodama K, Mabuchi K, Kraemer KH, Land CE, Nakamura N. Heterozygous individuals bearing a founder mutation in the XPA DNA repair gene comprise nearly 1% of the Japanese population. *Mutat Res* 2006; 601:171–178.
  98. Messaoud O, Ben Rekaya M, Kefi R, Chebel S, Boughammoura-Bouatay A, Bel Hadj Ali H, Gouider-Khouja N, Zili J, Frih-Ayed M, Mokhtar I, Abdelhak S, Zghal M. Identification of a primarily neurological phenotypic expression of xeroderma pigmentosum complementation group A in a Tunisian family. *Br J Dermatol* 2010; 162:883–886.
  99. Hadj-Rabia S, Oriot D, Soufir N, Dufresne H, Bourrat E, Mallet S, Poulhalon N, Ezzedine K, Grandchamp B, Taïeb A, Catteau B, Sarasin A, Bodemer C. Unexpected extradermatological findings in 31 patients with xeroderma pigmentosum type C. *Br J Dermatol* 2013; 168:1109–1113.
  100. Jerbi M, Ben Rekaya M, Naouali C, Jones M, Messaoud O, Tounsi H, Nagara M, Chargui M, Kefi R, Boussen H, Mokni M, Mrad R, Boubaker MS, Abdelhak S, Khaled A, Zghal M, Yacoub-Youssef H. Clinical, genealogical and molecular investigation of the xeroderma pigmentosum type C complementation group in Tunisia. *Br J Dermatol* 2016; 174:439–43.
  101. Falik-Zaccai TC, Erel-Segal R, Horev L, Bitterman-Deutsch O, Koka S, Chaim S, Keren Z, Kalfon L, Gross B, Segal Z, Orgal S, Shoval Y, Slor H, Spivak G, Hanawalt PC. A novel XPD mutation in a compound heterozygote; the mutation in the second allele is present in three homozygous patients with mild sun sensitivity. *Environ.Mol.Mutagen* 2012; 53:505-514.
  102. Ben Rekaya M, Laroussi N, Messaoud O, Jones M, Jerbi M, Naouali C, Bouyacoub Y, Chargui M, Kefi R, Fazaa B, Boubaker MS, Boussen H, Mokni M, Abdelhak S, Zghal M, Khaled A, Yacoub-Youssef H. A founder large deletion mutation in Xeroderma pigmentosum-Variant form in Tunisia: implication for molecular diagnosis and therapy. *Biomed Res Int* 2014; 2014:256245.
  103. Masaki T, Ono R., Tanioka M, Funasaka Y, Nagano T, Moriwaki S, Nishigori C. Four types of possible founder mutations are responsible for 87% of Japanese patients with Xeroderma pigmentosum variant type. *J Dermatol Sci* 2008; 52:144–148.

104. Calmels N, Greff G, Obringer C, Kempf N, Gasnier C, Tarabeux J, Miguet M, Baujat G, Bessis D, Bretones P, Cavau A, Digeon B, Doco-Fenzy M, Doray B, Feillet F, Gardeazabal J, Gener B, Julia S, Llano-Rivas I, Mazur A, Michot C, Renaldo-Robin F, Rossi M, Sabouraud P, Keren B, Depienne C, Muller J, Mandel JL, Laugel V. Uncommon nucleotide excision repair phenotypes revealed by targeted high-throughput sequencing. *Orphanet J Rare Dis* 2016;11:26.
105. A Hafeez Diwan et William D James. Xeroderma Pigmentosum. Medscape Reference. Nov 29, 2011.
106. DiGiovanna JJ, Kraemer KH. Shining a light on xeroderma pigmentosum. *J Invest Dermatol* 2012; 132:785–796.
107. Rosenberg PS, Tamary H, Alter BP. How high are carrier frequencies of rare recessive syndromes? Contemporary estimates for Fanconi Anemia in the United States and Israel. *Am J Med Genet A* 2011; 155A:1877–1883.
108. D'Andrea AD. Susceptibility pathways in Fanconi's anemia and breast cancer. *N Engl J Med* 2010; 362:1909–1919.
109. Neitzel H, Kühl JS, Gerlach A, Ebell W, Tönnies H. Clonal chromosomal aberrations in bone marrow cells of Fanconi anemia patients: results and implications. In: Schindler D, Hoehn H, eds. *Fanconi Anemia: a Paradigmatic Disease for the Understanding of Cancer and Aging. Monographs in Human Genetics* 2007; Vol 15:79-94.
110. Mehta PA, Harris RE, Davies SM, Kim MO, Mueller R, Lampkin B, Mo J, Myers K, Smolarek TA. Numerical chromosomal changes and risk of development of myelodysplastic syndrome—acute myeloid leukemia in patients with Fanconi anemia. *Cancer Genet Cytogenet* 2010; 203:180–186.
111. Kutler DI, Singh B, Satagopan J, Batish SD, Berwick M, Giampietro PF, Hanenberg H, Auerbach AD. A 20-year perspective on the International Fanconi Anemia Registry (IFAR). *Blood* 2003; 101: 1249–1256.
112. Rosenberg PS, Socié G, Alter BP, Gluckman E. Risk of head and neck squamous cell cancer and death in patients with Fanconi anemia who did and did not receive transplants. *Blood* 2005; 105:67–73.
113. Eiler ME, Frohnmayer D, Frohnmayer L, Larsen K, Olsen J. Fanconi Anemia: guidelines for diagnosis and management. *Fanconi anemia Research Fund* 2008; 1(3): 30-13.



114. Esmer C, Sánchez S, Ramos S, Molina B, Frias S, Carnevale A. DEB test for FA detection in patients with atypical phenotypes. *Am J Med Genet A* 2004 Jan 1; 124A(1): 35-39.
115. Ameziane N, Sie D, Dentro S, Ariyurek Y, Kerkhoven L, Joenje H et al. Diagnosis of fanconi anemia: mutation analysis by nextgeneration sequencing. *Anemia* 2012;2012: 132856.
116. Castella M, Pujol R, Callén E, Trujillo JP, Casado JA, Gille H et al. Origin, functional role, and clinical impact of FA FANCA mutations. *Blood* 2011 Apr 7; 117(14): 3759-3769.
117. Ameziane N, Errami A, Léveillé F, Fontaine C, de Vries Y, van Spaendonk RM et al. Genetic subtyping of FA by comprehensive mutation screening. *Hum Mutat* 2008 Jan; 29(1): 159-166.
118. Levran O, Raffaella D, Kanan P, Sat DB Helmut H, Arleen DA. Spectrum of sequence variations in the FANCA gene: an international FA registry (IFAR) study. *Hum Mutat* 2005 Feb; 25(2): 142- 149.
119. Amouri A, Talmoudi F, Messaoud O, d'Enghien CD, Rekaya MB, Allegui I et al. High frequency of exon 15 deletion in the FANCA gene in Tunisian patients affected with FA disease: implication for diagnosis. *Mol Genet Genomic Med* 2014 Mar; 2(2): 160-165.
120. Moghrabi NN, Johnson MA, Yoshitomi MJ, Zhu X, J Al-Dhalimy M, Olson SB, Markus Grompe C Sue Richards. Validation of Fanconi anemia complementation Group A assignment using molecular analysis. *Genet Med* 2009 Mar; 11(3):183-192.
121. De Rocco D et al. Molecular analysis of Fanconi Anemia : the experience of the Bone Marrow Study Group of the Italian Association of Pediatric Onco-Hematology. *Heamatologica* 2014; 99 (6):1022-1031.
122. Ben Haj Ali A, Messaoud O, Elouej S, Talmoudi F , Ayed W, Mellouli F, Ouederni M , Hadiji S, De Sandre-Giovannoli A, V Delague, Lévy N, Bogliolo M, Surrallés J, Abdelhak S and Amouri A. FANCA Gene Mutations in North African Fanconi Anemia Patients. *Front Genet* 2021; 12:610050.
123. Park J, Chung NG, Chae H, Kim M, Lee S, Kim Y et al. FANCA and FANCG are the major FA genes in the Korean population. *Clin Genet* 2013 Sep; 84(3): 271-275.
124. Kee Y, D'Andrea AD. Molecular pathogenesis and clinical management of FA. *J Clin Invest* 2012 Nov; 122(11): 3799-3806.
125. Demuth I, Wlodarski M, Tipping AJ, Morgan NV, De Winter JP, Gräsl MS Thiel, , Schindler D, D'Andrea AD, Altay C, Kayserili H, Zatterale A, Kunze J, Ebell

- W, Mathew C G, Joenje H, Sperling K. Spectrum of mutations in the Fanconi anaemia group G gene, *FANCG/XRCC9*. *Eur J Hum Genet* 2000 Nov; 8(11):861-868.
126. Wegman-Ostrosky T, Savage SA. The genomics of inherited bone marrow failure: from mechanism to the clinic. *Br J Haematol* 2017 May; 177(4): 526-542.
  127. EsmailNia G, Fadaee M, Royer R, Najmabadi H, Akbari MR. Profiling FA Gene mutations among Iranian patients. *Arch Iran Med*. 2016 Apr;19(4): 236-40.
  128. Hirsch B, Shimamura A, Moreau L, Baldinger S, Hag-alshiekh M, Bostrom B, Sencer S, D'Andrea AD. Association of biallelic BRCA2/FANCD1 mutations with spontaneous chromosomal instability and solid tumors of childhood. *Blood* 2004; 103:2554–2559.
  129. Wagner JE, Tolar J, Levran O, Scholl T, Deffenbaugh A, Satagopan J, Ben-Porat L, Mah K, Batish SD, Kutler DI, MacMillan ML, Hanenberg H, Auerbach AD. Germline mutations in BRCA2: shared genetic susceptibility to breast cancer, early onset leukemia, and Fanconi anemia. *Blood* 2004; 103:3226–3229.
  130. Myers K, Davies SM, Harris RE, Spunt SL, Smolarek T, Zimmerman S, McMasters R, Wagner L, Mueller R, Auerbach AD, Mehta PA. The clinical phenotype of children with Fanconi anemia caused by biallelic FANCD1/BRCA2 mutations. *Pediatr Blood Cancer* 2012; 58:462–465.
  131. Alter BP, Rosenberg PS, Brody LC. Clinical and molecular features associated with biallelic mutations in FANCD1/BRCA2. *J Med Genet* 2007; 44:1–9.
  132. Jung M, Ramanagoudr-Bhojappa R, van Twest S, Rosti RO, Murphy V, Tan W, Donovan FX, Lach FP, Kimble DC, Jiang CS, Vaughan R, Mehta PA, Pierri F, Dufour C, Auerbach AD, Deans AJ, Smogorzewska A, Chandrasekharappa SC. Association of clinical severity with FANCB variant type in Fanconi anemia. *Blood* 2020; 135:1588–1602.
  133. Reid S, Schindler D, Hanenberg H, Barker K, Hanks S, Kalb R, Neveling K, Kelly P, Seal S, Freund M, Wurm M, Batish SD, Lach FP, Yetgin S, Neitzel H, Ariffin H, Tischkowitz M, Mathew CG, Auerbach AD, Rahman N. Biallelic mutations in PALB2 cause Fanconi anemia subtype FA-N and predispose to childhood cancer. *Nat Genet* 2007; 39:162–164.
  134. Papadopoulo D, et Moustacchi E. L'anémie de Fanconi: gènes et fonction(s) revisités. *M/S : médecine sciences* 2005; 21(8-9):730–736.
  135. Nepal M, Che R, Zhang J, Ma C, and Fei P. Fanconi Anemia Signaling and Cancer. *Trends Cancer* 2017 December; 3(12):840–856.

136. Andreassen PR, D'Andrea AD, Taniguchi T. ATR couples FANCD2 monoubiquitination to the DNA damage response. *Genes Dev.* 2004; 18(16):1958–1963.
137. Zou L, Elledge SJ. Sensing DNA damage through ATRIP recognition of RPA-ssDNA complexes. *Science* 2003; 300(5625):1542–1548.
138. Wang, X., P.R. Andreassen, and A.D. D'Andrea, Functional interaction of monoubiquitinated FANCD2 and BRCA2/FANCD1 in chromatin. *Mol Cell Biol* 2004; 24(13): 5850-5862.
139. Lopez-Martinez D, Liang CC, Cohn MA. Cellular response to DNA interstrand crosslinks: the Fanconi anemia pathway. *Cell Mol Life Sci* 2016; 73(16):3097-3114.
140. Minten EV and Yu DS. DNA Repair: Translation to the Clinic *Clin Oncol. R Coll Radiol* 2019 May; 31(5): 303–310.

## **Résumé:**

**Introduction :** De nombreuses maladies génétiques, notamment celles dues à des anomalies de réparation de l'ADN, sont associées à un risque accru de cancer. L'étude de ces syndromes a considérablement évolué et a conduit à l'identification de gènes impliqués dans la carcinogenèse. Dans ce travail, nous nous sommes intéressés à deux pathologies: La Xérodérma Pigmentosum (XP) et l'Anémie de Fanconi (AF).

**Matériels et méthodes :** Pour la famille avec XP, nous avons appliqué le Séquençage de Sanger pour la recherche de la mutation récurrente c.1643\_1644delTG au niveau de l'exon 9 du gène *XPC*. Concernant les familles atteintes de l'AF, nous avons opté pour le Séquençage haut débit pour analyser les trois gènes les plus fréquemment mutés *FANCA*, *FANCC*, et *FANCG* et qui représentent 84% de tous les gènes impliqués dans l'AF.

**Résultats :** Le séquençage ciblé de l'exon 9 du gène *XPC* a révélé la présence d'une nouvelle mutation à l'état homozygote de type insertion de 4 bases chez le propositus. Les parents ont été testés et se sont révélés porteurs de cette mutation à l'état hétérozygote. L'analyse moléculaire effectuée chez neuf patients atteints d'AF et présentant tous une grande instabilité chromosomique sous mitomycine, a confirmé le diagnostic dans 88.8% des cas avec trois nouvelles mutations prédites comme pathogènes. Des variations du gène *FANCA* et du gène *FANCG* ont été retrouvées respectivement dans 62.5% et 37.5% des cas. .

**Conclusion :** Grâce aux techniques de séquençage, le diagnostic moléculaire est actuellement possible même pour les maladies génétiques avec une grande hétérogénéité génétique. Le choix du test moléculaire dépend du profil génétique de la pathologie. L'identification de la mutation causale permet d'adapter la prise en charge des malades et de prodiguer aux familles un conseil génétique adéquat.

**Mots clés :** Cancer, Xeroderma pigmentosum, anémie de Fanconi, séquençage Sanger, séquençage à haut débit.

## **Abstract**

**Introduction:** Many genetic diseases, especially those caused by abnormalities in DNA repair, are associated with an increased risk of cancer. The study of these syndromes predisposing to cancer has evolved considerably and has led to the identification of genes involved in carcinogenesis. In this work, we were interested in two pathologies which are due to a DNA repair defect: Xeroderma Pigmentosum (XP) and Fanconi anemia (FA).

**Materials and methods:** We applied Sanger sequencing for the family with XP to search for the recurrent mutation c.1643\_1644delTG in the exon 9 of the *XPC* gene. Concerning the FA families, we opted for the Next Generation Sequencing (Ion PGM) to analyse the three most frequently mutated genes, *FANCA*, *FANCC*, and *FANCG*, which represent 84% of all genes involved in FA.

**Results:** The targeted sequencing of the exon 9 of the *XPC* gene revealed the presence of a new homozygous insertion of 4 bases in the propositus. The parents have been tested and found to be heterozygous carrier of the same mutation. Molecular analysis of nine patients with FA, sharing all a high chromosomal instability to mitomycin, confirmed the diagnosis in 88.8% of cases with three new mutations predicted as pathogens. Mutations in the *FANCA* and the *FANCG* genes were found in 62.5% and 37.5% of cases respectively.

**Conclusion:** Thanks to sequencing technologies, molecular diagnosis is currently possible even for genetic diseases with genetic heterogeneity. The choice of molecular test depends on the pathology's genetic profile. The identification of the causal mutation allows to adapt the management of patients and to provide an adequate genetic counselling to families.

**Keywords :** Cancer, Xeroderma pigmentosum, Fanconi anemia, Sanger Sequencing, Next generation sequencing.

**Structure de recherche: Centre de Génomique Structurale et Fonctionnelle des Pathologies Humaines**