



الجمهورية الجزائرية الديمقراطية الشعبية
République Algérienne Démocratique Et Populaire
وزارة التعليم العالي والبحث العلمي
Ministère De L'enseignement Supérieur Et De La Recherche Scientifique



Université Constantine 1 Frères Mentouri
Faculté des Sciences de la Nature et de la Vie

جامعة قسنطينة 1 الإخوة منتوري
كلية علوم الطبيعة والحياة

Département de Biologie appliquée

قسم: البيولوجيا التطبيقية

Mémoire présenté en vue de l'obtention du Diplôme de Master

Domaine: Sciences de la Nature et de la Vie

Filière : Sciences Biologiques / Biotechnologies / Écologie et Environnement

Spécialité : Bio-informatique

N° d'ordre :

N° de série :

Intitulé :

Approche in silico dans l'analyse épigénétique

« Le cas de l'infertilité masculine »

Présenté par : ALLECHE Imene

Le : 10/06/2024

SOUALAH Malak

Jury d'évaluation :

Président : Pr. HAMIDECHI M. A (Professeur U. KHENCHELA).

Encadrant : Dr BENZAADA Mostafa (MCA - U Constantine 1 Frères Mentouri).

Examineur : Dr AMINE KHODJA Ihsene (MAB - U Constantine 1 Frères Mentouri).

Année universitaire
2023 - 2024

Remerciements

Nous exprimons nos profonds remerciements à notre encadrant docteur BENSAADA Mostafa pour son soutien et son encadrement.

Nos vifs remerciements vont également aux membres du jury Pr. HAMIDECHI M. A. et Dr AMINE KHOUDJA. qui auront à prendre la responsabilité d'examiner et évaluer notre modeste travail et dont nous souhaitons donner satisfaction et répondre à leurs exigences.

Nous sommes profondément reconnaissantes au Dr BENSAADA Mostafa pour sa gentillesse, sa disponibilité, sa patience, afin de mener à bien ce travail. Ses précieux conseils et orientations nous ont poussées toujours le plus loin possible, rendant cette discipline une véritable passion.

Un grand merci pour le doctorant Abdelhafedh Ben Dahmane, pour ses conseils, son partage et ses encouragements.

Dédicaces

*À mon ange maman chérie « **Zeineb** » qui m'a quittée trop tôt mais dont l'amour et le souvenir continuent de m'inspirer chaque jour. Ton encouragement indéfectible et ta foi en mes capacités m'ont permis de persévérer dans mes études et d'atteindre mes objectifs. C'est grâce à toi que j'ai développé la force, la persévérance et la passion qui m'ont permis de réaliser ce travail. Je te dédie ce travail avec tout mon amour et ma reconnaissance éternelle. Repose en paix, maman.*

*À mon papa chéri « **Mokhtar** » mon pilier, mon héros. Merci pour ton aide inestimable et ton soutien indéfectible tout au long de ma vie. Tes encouragements m'ont donné la force de persévérer dans les moments difficiles et tes conseils avisés m'ont permis de prendre les bonnes décisions. Je suis tellement reconnaissante d'avoir un papa comme toi. Tu es mon modèle et mon inspiration. Je t'aime très fort.*

À ma chère famille, Je tiens à vous dédier ce travail en signe de mon amour. Merci pour votre soutien, vos encouragements.

*A mon binôme « **Malak** » je tiens à dédier ce travail en signe de mon amitié sincère et de ma reconnaissance infinie. Merci pour ton soutien indéfectible, ta collaboration exemplaire et ton amitié précieuse. Travailler avec toi a été une expérience formidable et enrichissante.*

*A mes plus chères sœurs **Dounya, Nourhen, Amira, Khawla** et **Yasmine** je vous adore et merci pour tout.*

À moi-même, pour la force, le courage et la persévérance dont j'ai fait preuve tout au long de mon parcours. Je suis fierai de tout ce que j'ai accompli et je suis reconnaissante pour les leçons que j'ai apprises en cours de route.

Imene

Dédicaces

Grâce à l'aide de DIEU tout puissant, qui m'a tracé le chemin de ma vie, j'ai Pu, Réaliser ce travail que je dédie :

*À la femme la plus chère du monde, ma très chère mère « **Sabrina** » :*

La source de tendresse qui a tout donné sans rien recevoir, je ne pourrais exprimer la profondeur des sentiments que j'éprouve pour vous.

*À mon très cher père « **Djamel** » :*

Pour vos encouragements durant mes nombreuses années d'études et votre confiance. J'espère que vous en tirez un sentiment de fierté. En signe de l'affection et du grand amour que je vous porte.

*À mes chères et adorables « **sœurs** » :*

Qui comptent beaucoup pour moi et sur qui je suis certaine de toujours pouvoir m'appuyer « Bouchera, Aya, Hadil, Ouissel, Maya, Alaa, Mirel et Meriem ».

*À ma très chère grand-mère paternelle « **kharoufa** » :*

Vous m'avez accompagné toujours par vos prières, puisse Allah vous préserver et vous accorder santé, bonheur et longue vie.

*À mon très cher tonton « **RAMI** » et mes très chers (oncles, tantes maternels, paternelle et leurs enfants) :*

Pour l'amour, le soutien, les efforts, les conseils et les encouragements permanents Merci d'être toujours là pour moi. Je vous aime infiniment.

*À ma partenaire **IMENE***

Pour ta patience et ton dévouement dans ce travail, je te dédie le fruit de nos efforts et je te remercie pour les moments que nous avons vécus ensemble et les souvenirs inoubliables. Je te souhaite une vie pleine de santé, de bonheur et de réussite.

*À mes meilleures amies « **Malak, Ikram, Amira et Ahlam** » :*

Vous êtes les plus chères à mon cœur, vous êtes pour moi des sœurs sur qui je peux compter.

Et enfin À moi-même, je mérite un peu de reconnaissance.

Malak

Résumé

L'objectif de ce travail est d'explorer in Silico les effets possibles de l'épigénétique que l'infertilité masculine. Nous avons identifié 21 gènes impliqués dans l'infertilité masculine décrits comme sous l'effet de l'épigénétique. La méthylation des îlots CpG joue un rôle majeur dans l'expression des gènes, donc l'épigénétique peut être une cause sérieuse de l'infertilité masculine, en particulier par l'effets de nombreux facteurs environnementaux. Les résultats obtenus montrent une variété dans la distribution des CpG au sein de nos gènes, ce qui signifie une forte activité de méthylation de l'ADN. Nos résultats, mêmes, préliminaires peuvent fournir un éclairage pour identifier des causes épigénétique dans l'infertilité masculine et donc dans le traitement de cette maladie.

Abstract

The aim of this work is to explore in Silico the possible effects of epigenetics on male infertility. We have identified 21 genes involved in male infertility described as being under the effect of epigenetics. Methylation of CpG islands plays a major role in gene expression, so epigenetics may be a serious cause of male infertility, particularly through the effects of numerous environmental factors. The results obtained show a variety in the distribution of CpGs within our genes, signifying strong DNA methylation activity. Even our preliminary results may shed light on the identification of epigenetic causes of male infertility, and hence on the treatment of this disease.

ملخص

الهدف من هذا العمل هو استكشاف الآثار المحتملة لعلم التخلق *in Silico* في عقم الذكور. حددنا 21 جينًا متورطًا في عقم الذكور الموصوف بأنه تحت تأثير علم التخلق. يلعب مثيلة جزر CpG دورًا رئيسيًا في التعبير الجيني، لذلك يمكن أن يكون علم التخلق سببًا خطيرًا لعقم الذكور، خاصة من خلال تأثيرات العديد من العوامل البيئية. تظهر النتائج التي تم الحصول عليها تنوعًا في توزيع CpG داخل جيناتنا، مما يعني نشاط مثيلة قوي للحمض النووي. يمكن أن توفر نتائجنا، حتى الأولية، نظرة ثاقبة لتحديد الأسباب اللاجينية في عقم الذكور وبالتالي في علاج هذا المرض..

Liste des figures

Figure 1: Processus d'approche in silico.....	18
Figure 2: Interface de NCBI.....	19
Figure 3: Interface de GenBank	19
Figure 4: Résultat des gènes Homo sapiens	20
Figure 5: Séquences des variants du gène DMRT1 sur sublime Texet.....	20
Figure 6: CDS de variant.....	21
Figure 7: Localisation de la région promotrice en amont de la CDS du gène DMRT1	21
Figure 8: Interface MEGA11.....	22
Figure 9: Alignement réalisé du gène DMRT1	22
Figure 10: Nettoyage des séquences du gène DMRT1	23
Figure 11: Séquence promotrice sélectionnée du gène DMRT1	23
Figure 12: Résultat de l'exécution du gène DMRT1	24
Figure 13: Interface du site DBCAT	24
Figure 14: Analyse du gène DMRT1	25
Figure 15: Interface WEBLOGO	26
Figure 16: Interface du site Dotlets JS beta.....	26
Figure 17: Résultat graphique pour les motifs CG du gène DMRT1	27
Figure 18: Arbre phylogénétique de 21 gènes impliqués dans l'infertilité masculine.....	31
Figure 19: Logo du premier groupe de gènes (ADAM2, DAZL, FANCM, HDAC1, ACE, EGR4).....	32
Figure 20: Logo du deuxième groupe des gènes (DDX4, EZH2, PPP1CC, DMRT1, CTCF, MCPH1)	32
Figure 21: Logo 4ème groupe des gènes (BCL2, KITLG, FANCA, APAF1).....	33
Figure 22: Logo du 5ème groupe des gènes (CPEB1, HAT1, MLH1, BRDT)	34
Figure 23: Le logo entier des régions promotrices des 21 gènes	34
Figure 24: Les CpG trouvé dans la Rp du gène EGR4.....	37
Figure 25: Les CpG trouvé dans la Rp du gène HDAC1	37
Figure 26: Les CpG trouvés dans la Rp du gène BCL2	37
Figure 27: Résultat obtenu du Dotlet Js des 21 gènes	38

Liste des tableaux

Tableau 1: la configuration du matériel informatique utilisé lors de l'apprentissage	16
Tableau 2: les outils bio-informatique et leurs principes	16
Tableau 3: les résultats obtenus sur les régions promotrices de chaque gène.....	35

Liste des abréviations et acronymes

ADN : acide désoxyribonucléique

ARN : acide ribonucléique

CDS : la séquence codante d'un gène

CGI: CpG island

CpG: Cytosine Phosphate Guanine

CpGo/e : Rapport entre le nombre de CpG observés et le nombre de CpG attendus

DBCAT: DataBase of CpG Islands & Analytical Tools

HTF: HpaII Tiny Fragments

MEGA11: Molecular Evolutionary Genetics Analysis Version 11

NCBI: The National Center for Biotechnology Information

Pb : paire de base

Rp : la région promotrice

TABLE DES MATIÈRES

Remerciements	i
Dédicaces.....	ii
Dédicaces.....	iii
Résumé	iv
Abstract.....	v
Liste des figures.....	vii
Liste des tableaux	viii
Liste des abréviations et acronymes	ix
Introduction générale.....	2
Chapitre 01 : Synthèse Bibliographique.....	3
Introduction :	5
1. Définition de l'infertilité masculine :	5
1 Classification d'infertilité masculine :	5
2 Causes physiologiques et génétiques :	6
3 Les mutations génétiques :	6
4 Causes épigénétiques de l'infertilité masculine :	7
4.1 Age :	7
4.2 Médicament :	7
4.3 Tabac :	7
5 Définition de l'épigénétique :	8
5.1 Mécanisme épigénétique :	8
5.1.1 Modification des histones :	9
5.1.2 Méthylation de l'ADN :	9
6 Rôle de l'épigénétique dans l'infertilité masculine :	10
7 La génétique de l'infertilité	10
8 L'infertilité et épigénétique	10
9 Îlots CpG	11

9.1	Critère de détection :	11
9.2	Importance des îlots CpG dans l'épigénétique et l'infertilité masculine :	12
9.2.1	Régulation de l'expression génique:.....	12
9.2.2	Qualité du sperme :.....	12
9.2.3	Transmission intergénérationnelle :	12
10	La recherche <i>in silico</i> :	13
10.1	Les domaines d'application :.....	13
Chapitre 02 : Matériel et Méthodes		14
Introduction		15
1	Matériels :	15
1.1	Données biologiques :	15
1.2	Choix des gènes :.....	15
1.3	Outils bio-informatiques :	16
1.4	Définitions :.....	16
2	Méthodes	18
2.1	Collection des données :.....	18
2.2	Préparation des données :.....	20
2.3	Alignement des séquences	21
2.4	Nettoyage des séquences :.....	22
2.5	Extraction de la région promotrice :.....	23
2.6	Analyse des séquences obtenus :.....	23
2.7	Visualisation des îlots CpG :.....	24
2.8	Classement des résultats dans un tableau (tableau 4) :.....	25
2.9	Création de logos (WebLogo) :.....	25
2.10	Utilisation du site Dotlets JS beta	26
Chapitre 03 : Résultats et Discussion		28
1	Résultat de la construction de l'arbre phylogénétique des 21 gènes étudiés.....	30
2	Visualisation des résultats dans les logos :.....	32

3	Classement et analyse des résultat du calcul du tableau :	35
4	Visualisation des résultats obtenus du DBCAT :	37
5	Observation des séquences sur DOLETE JS.....	37
6	Discussion des résultats.....	38
7	Conclusion et perspective.....	39
	Liste des références bibliographiques.....	41

Introduction

Introduction générale

Au milieu du XXe siècle, l'épigénétique émerge comme une branche de la génétique et change la donne de l'hérédité et de la recherche en biologie. Elle étudie les interactions entre les gènes et l'environnement et explore comment, il interagit lors de la fusion d'un spermatozoïde et d'un ovule (*E. Gnansia, 2017*).

La méthylation de l'ADN est l'un des facteurs épigénétiques du génome associé aux événements de développement et à la régulation des gènes. L'hyperméthylation d'un gène peut réduire son expression, perturber la signalisation différentielle cellulaires spermatiques et diminuer la production des spermatozoïdes (*André, Sabrina et Henry, 2008*). La méthylation de l'ADN se produit sur le carbone 5 d'une cytosine dans un dinucléotide CpG (Cytosine-Phosphate-Guanine) chez les spermies. (*Amira et Djihad, 2021*).

Les îlots CpG jouent un rôle important dans la méthylation de l'ADN. La plupart des îlots CpG sont des sites d'initiation de la transcription. L'hyperméthylation des îlots CpG situés dans les régions promotrices des gènes infertiles. Les méthodes de détection des îlots CpG sont principalement basées sur les critères proposés par Gardiner-Garden et Frommer (GGF) (*Yasmine et Ferial, 2021*) la distance, la longueur y compris le contenu GC, le rapport O/E et le seuil de longueur de l'îlot.

Notre travail de recherche consiste à :

-Déterminer le nombre d'îlots CpG dans les régions promotrices des gènes impliqués dans l'infertilité masculine.

-Les changements épigénétiques qui interfèrent avec l'infertilité masculine.

- Le manuscrit est organisé en trois chapitres :

Le premier concerne :

- Synthèse Bibliographique : présente le contexte biologique de notre travail à savoir.

-Le deuxième chapitre aborde Matériel et Méthodes.

- Le troisième chapitre est une discussion des résultats obtenus.

Chapitre 01 : Synthèse Bibliographique

Introduction :

L'infertilité Masculine est un problème de santé publique affectant la capacité d'un couple à procréer. Elle peut résulter de diverses causes, tant féminines que masculines, et peut impacter la qualité de vie. Environ 20 à 70% des cas d'infertilité sont attribués à des facteurs masculins, soulignant l'importance d'une évaluation approfondie de l'homme (*Jacques 2016*). L'évaluation de l'infertilité masculine implique des examens spécifiques comme le spermogramme, qui analyse des paramètres tels que la concentration spermatique, la mobilité des spermatozoïdes et leur morphologie (*Florence 2014*). Bien que les causes de l'infertilité masculine soient multiples, environ 30 à 40 % des cas restent inexplicables par les facteurs génétiques ou hormonaux classiques, c'est là qu'intervient l'épigénétique (*Samir et al., 2022*).

1. Définition de l'infertilité masculine :

D'après l'OMS (Organisation Mondiale De La Santé), l'infertilité d'un couple est l'incapacité à avoir une grossesse après douze mois de rapports sexuels réguliers non protégés. Selon la Haute Autorité de Santé (HAS), la durée de deux ans est à respecter pour parler d'infertilité et ne pas surestimer la fréquence de l'infertilité parce que la moitié des femmes doivent attendre au moins deux ans pour une conception. Pour d'autres comme (*Gnoth et al. 2005 ; Evers 2002*), on parle de l'infertilité après un an de rapport sexuels réguliers efficaces et non protégés et un bilan d'infertilité devrait être entrepris après 6 mois (*Gnoth et al. 2005*). Une cause exclusive d'infertilité masculine est identifiée dans un couple sur cinq, tandis que dans 40% des cas, l'infertilité est le résultat d'une responsabilité partagée entre l'homme et la femme. Cependant, de nombreuses causes d'infertilités restent inexplicables, représentant 20 à 30% des cas (*Thonneau et al. 1991*).

1 Classification d'infertilité masculine :

Il y a infertilité primaire s'il n'y a jamais eu de grossesses, et infertilité secondaire, si elle a eu lieu après une grossesse ou plus (*Geidam et al, 2008*). L'infertilité masculine peut être expliquée par des problèmes hormonaux, immunologiques, d'acheminement de sperme, environnementaux et de la structure physique des organes reproducteurs et peut être classée en 3 groupes principaux

- **non obstructive (60%)** : qui se définit par une production insuffisante de spermatozoïdes par les testicules.

- **obstructive (38%)** : la production de spermatozoïdes est normale mais il y a un blocage dans le tractus génital.

- **coïtal (2%)** : la production de spermatozoïdes est normale ainsi que les voies génitales, mais l'infertilité est secondaire à une dysfonction sexuelle (*Abdel Raheem et al, 2012*).

2 Causes physiologiques et génétiques :

Les causes d'infertilité masculine sont nombreuses et multifactorielles. On distingue des causes physiologiques et génétiques. (*Coutton et al., 2012*).

Les positions des anomalies qui causent l'infertilité masculine sont subdivisées en :

- **Pré-testiculaire** : stimulation insuffisante du testicule par gonadotrophine (5–10 %),

- **Testiculaire** : les maladies du testicule (65–75 %), Le testicule peut être soumis à des lésions traumatiques et ischémiques (torsion du cordon spermatique, chirurgie, varicocèle) pouvant altérer définitivement sa fonction exocrine. Cette dernière est également sensible à l'hyperthermie, aux rayonnements ionisants, aux toxiques environnementaux voire médicamenteux, au stress oxydant associé à de nombreuses situations pathologiques (varicocèle, infection, inflammation génitale) (*Schlosser et al., 2007*).

-**Et post-testiculaire** : Elles sont représentées par les obstructions post chirurgicales et infections, auto-immunisations, obstructions tumorales des voies séminales : kystes épидидymaires, utricule Prostatique, kyste prostatique (*Pasualotte et al., 2004*). Qui conduit à obstruction des voies séminales, les troubles de l'éjaculation, dysfonction érectile (10–20 %) (*Lawrence et Chandrasekar, 2015, Turchi, 2015*).

3 Les mutations génétiques :

Touchant des gènes intervenant dans la production ou la fonction de cellules germinales. Par ailleurs, 30 % des infertilités restent inexplicées et près de 40 % ont des causes incertaines. Ainsi, l'infertilité masculine d'origine génétique pourrait concerner près de 1 homme sur 40. Parmi les causes génétiques actuellement bien établies, on trouve les anomalies chromosomiques, les micro délétions du chromosome Y et les mutations du gène CFTR (Cystic Fibrosis Transmembrane Conductance Regulator)

4 Causes épigénétiques de l'infertilité masculine :

Outre les causes connues de l'infertilité masculine, de nombreuses études ont montré l'implication de différents facteurs liés à l'environnement et aux conditions de vie :

4.1 Age :

Le taux de paternité chez les hommes de plus de 35 ans est en augmentation de 16% selon une étude américaine menée de 1980 à 1995 (*Wagner, 2004*).

4.2 Médicament :

Les médicaments peuvent exercer une action à différents niveaux : l'aspect qualitatif ou quantitatif du sperme, (*Olivennes et al., 2006*) :

- **Inhibition hypophysaire** : testostérone, analogues GnRH, stéroïdes anabolisants.
- **Effets anti-androgéniques** : cimétidine, spironolactone.
- **Anéjaculation** : antidépresseurs, phénothiazines, bêtabloquants.
- **Dysfonctionnement érectile** : bêtabloquants, diurétiques thiazidiques, métoclopramide, hypocholestérolémiants, anxiolytiques, antidépresseurs, antiépileptiques.
- **Altération qualitative ou quantitative de la spermatogénèse** : nitrofuranes, salazopyrine, kétoconazole, médicaments anticancéreux (alkylants)
- **Trouble de la libido** : antidépresseurs (IMAO, ISRS, tricycliques), neuroleptiques (sulpiride), thymorégulateur (lithium), bêta-bloquants, diurétiques, antiépileptiques et anxiolytiques (*Hazard et al., 2000*).

4.3 Tabac :

La cigarette contient plus de 4000 composants parmi lesquels on peut citer le monoxyde de carbone, des alcaloïdes comme la nicotine, des hydrocarbures polyaromatiques, des métaux lourds comme le cadmium. On note une tendance à l'oligospermie et une diminution relative de la vitalité des spermatozoïdes. La mobilité spermatique semble être altérée comme la morphologie des gamètes qui apparaissent microcéphales (*Sepaniak et al., 2004*).

5 Définition de l'épigénétique :

Le terme « épigénétique » a été introduit pour la première fois en 1942 par Conrad H. Waddington, un chercheur en biologie évolutive. Il a utilisé ce terme pour désigner les processus biologiques qui réagissent la transformation du génotype en phénotype durant le développement et la vie. Selon Waddington, englobe les mécanismes causaux par lesquels les gènes du génotype influent sur les caractéristiques phénotypiques. (Waddington, 1942,). Ces mécanismes auraient émergé de l'évolution des espèces et jouent un rôle majeur dans la division du travail entre les différents types cellulaires d'un organisme. En d'autres termes, l'étude de l'« épigénèse » et de l'« épigénotype » permettrait de mieux comprendre par quels procédés les cellules sont orientées vers une fonction précise, entre d'autres termes « canalisées », lors de la différenciation cellulaire. Cette évolution s'est ajoutée une meilleure compréhension des mécanismes épigénétiques, comme la méthylation de l'ADN, les modifications de la chromatine et l'ARN non codant. Ces derniers ont un impact significatif sur l'expression des gènes. Ces avancées ont ouvert la voie à de nouvelles perspectives dans de nombreux domaines, tels que la médecine, la biologie du développement et l'agriculture.

5.1 Mécanisme épigénétique :

L'ADN est le support de l'information génétique, avec une séquence identique dans toutes les cellules du même organisme vivant. Cette dernière codée par un ordre spécifique des quatre bases nucléiques : Adénine, Thymine, Cytosine et Guanine. Le décryptage de ce code à quatre lettres ne permet pas d'expliquer comment une même succession peut donner autant de combinaisons. Comme ces cellules partagent le même code génétique, elles se différencient par l'activation des mécanismes épigénétiques variés qui expliquent ce code de manière nette. Chaque cellule utilise différents types de mécanismes épigénétiques indépendants, dont le mode d'action fait intervenir l'ADN et des protéines.

Certains mécanismes épigénétiques modifient la structure de la chromatine, en la faisant d'un état condensé à un état décondensé, ou inversement, selon qu'un gène a besoin d'être exprimé ou réprimé. D'autres agissent directement sur les séquences régulatrices de l'ADN situées à proximité des gènes. Ces séquences spécifiques n'ont de fonction codante pour les protéines, mais régulent où et quand les gènes sont exprimés.

Quelques-uns des mécanismes épigénétiques impliqués dans l'infertilité masculine :

5.1.1 Modification des histones :

L'ADN s'enroule autour des histones, pour former des structures en petites boulets. Les modifications chimiques sur ces histones agissent sur la compaction de ces enroulements et peuvent affecter l'accessibilité de l'ADN et réguler négativement l'expression des gènes associés à la fertilité masculine.

5.1.2 Méthylation de l'ADN :

Parmi les divers types de mécanismes épigénétiques existants, la méthylation de l'ADN est le plus étudié et le mieux compris à ce jour. La méthylation est une réaction biochimique qui implique la formation d'un lien covalent. Lorsque ce lien est formé, il se rend difficile à le rompre par apport à son énergie d'activation élevée ce qui lui confère une grande stabilité. La méthylation se produit directement sur le brin d'ADN, plus précisément sur les bases nucléotidiques cytosines (C) formant la 5-méthylcytosine (5mC). En générale la méthylation de l'ADN se produit au niveau du promoteur plutôt que sur le gène lui-même plus précisément dans les régions riches en séquences CpG. Quand le promoteur est méthylé, l'accès des facteurs d'activation et de la machinerie transcriptionnelle au promoteur devient très difficile, ce qui entraîne une forte répression de l'expression génique. En d'autres termes, la méthylation du promoteur d'un gène le rend "silencieux" (*Robertson et Jones, 2000, Jones, 2012*). Chez l'humain, environ 70-80% des sites CpG sont naturellement méthylés, ce qui permet de limiter la production des protéines à celles qui sont nécessaires à la fonction propre de chacune des cellules somatiques (*Ziller et al., 2013*). Des altérations de la méthylation de l'ADN dans les gènes impliqués dans la spermatogenèse peuvent perturber la production normale de spermatozoïdes.

a) ARN non codants (ARNnc) :

ARNnc sont des molécules d'ARN qui ne sont pas traduites en protéine tels que (les microARN) mais peuvent réguler l'expression des gènes, ces altérations dans les ARNnc impliqués dans la spermatogenèse peuvent contribuer à l'infertilité masculine.

b) Micro (miARN) :

Les miARN sont des petits ARN qui régulent l'expression des gènes en inhibant la traduction de l'ARNm ou dégradant l'ARNm cible. Les anomalies dans le miARN impliqués dans la spermatogenèse peuvent mener à des troubles de la fertilité masculine.

6 Rôle de l'épigénétique dans l'infertilité masculine :

Comme on a déjà mentionner l'épigénétique est un ensemble des modifications héréditaires de l'expression des gènes sans altération de la séquence de l'ADN, cette dernière a un rôle majeur dans l'infertilité masculine par apport aux modifications épigénétiques comme la méthylation de l'ADN, modification des histones et ARN non codant affectent la production des spermatozoïdes, leur fonction et leur capacité à fertiliser un ovule.

7 La génétique de l'infertilité

L'essor de la biologie moléculaire et de la génétique médicale ont permis d'explorer de nouvelles pistes et la découverte de nouvelles étiologies de l'infertilité masculine d'ordre génétique. Depuis une dizaine d'années, nous observons l'émergence d'un nouveau domaine de recherche, la « génétique de l'infertilité ». Les causes génétiques de l'infertilité sont suspectées dans 30% des cas (*Smith, Pfeifer, & Collins, JAMA.,2003*). Celles-ci font actuellement l'objet de recherches intenses à travers le monde. Naturellement, parmi ces causes, on distingue l'infertilité d'origine chromosomique ou génique. Ces anomalies peuvent s'exprimer de plusieurs manières et elles sont source de l'échec de la reproduction dans un nombre de cas non négligeable. Dans une proportion importante, une base génétique chromosomique ou génique connue est parfois transmise par les parents. L'avènement des nouvelles techniques de cytogénétique et de biologie moléculaire a récemment permis d'impliquer de nouveaux facteurs dans des phénotypes spécifiques d'infertilité (*Karger ; 2017*). La recherche d'une cause génétique de l'infertilité, qu'elle soit masculine ou féminine, est maintenant une démarche courante. Ainsi, l'étude génétique doit trouver sa place comme examen de première intention devant toute infertilité d'origine inconnue, à la recherche d'anomalies chromosomiques. (*Belmokhtar ,2020*).

8 L'infertilité et épigénétique

L'épigénétique réfère à « un changement héritable mais réversible de l'activité du génome qui est dû à un mécanisme autre que l'altération dans la séquence primaire de l'ADN » (*Jaenisch et Bird, 2003*). Les modifications épigénétiques caractérisées par la méthylation de l'ADN, les modifications pos-traductionnelles des histones, et le remodelage de la chromatine sont des

régulateurs importants d'un certain nombre de processus biologiques, y compris la spermatogenèse. Plusieurs gènes dans les testicules sont régis par des mécanismes épigénétiques, ce qui indique une influence directe de ces derniers sur le processus de la spermatogenèse. Les changements épigénétiques, à la différence des aberrations génétiques, peuvent être moins nuisibles parce qu'ils sont potentiellement réversibles. D'autres recherches pourraient identifier certains médicaments capables d'inverser les changements épigénétiques (*Rajender et al, 2011*).

9 Îlots CpG

Les îlots CpG (ou CGI pour CpG Island) ont été très étudiés depuis leur découverte, particulièrement chez les mammifères, en raison de leur implication potentielle dans la régulation de l'expression des gènes, ont été observés par Edwin Cooper et ses collègues au début des années 1980, connus par des régions typiques du génome définies par une fréquence élevée de dinucléotides CpG par rapport au reste du génome, sont des séquences d'ADN où une cytosine (C) est suivie d'une guanine (G) par liaison phosphodiester, formant ainsi un site CpG. Ces régions sont généralement non méthylées dans les cellules saines, ce qui signifie que la cytosine (C) n'est pas marquée par un groupe méthyle (-CH₃). Les îlots CpG sont souvent situés à proximité des promoteurs de gènes et peuvent jouer un rôle crucial dans la régulation de l'expression génique. Une étude approfondie de certains fragments d'ADN qui a approuvé que certaines régions riches en dinucléotides CpG, appelées "îlots HTF", étaient effectivement non méthylées. Ces îlots CpG, riches en G et C, ont été reconnus comme des régions spécifiques du génome où les sites CpG restent normalement non méthylés. Cette découverte a ouvert la voie à une meilleure compréhension de l'épigénétique et de l'importance des îlots CpG dans la régulation de l'expression des gènes (*Deltour, Chopin et Leprince, 2005*).

9.1 Critère de détection :

On désigne plusieurs méthodes pour la détection des îlots CpG, mais la plus courante basée sur l'utilisation d'un algorithme informatique. Le critère de détection des îlots CpG s'appuie sur plusieurs attributs, spécialement la longueur des séquences, le taux de G + C, et le rapport CpG observé/CpG attendu (CpGo/e). L'utilisation de ces critères peut parfois entraîner la détection d'éléments transposables riches en G et C, et en CpG, ce qui peut mener à des résultats non spécifiques. Les (CGI) les îlots CpG) ont été initialement identifiés par (*Tykocinski and Max. 1984*) sous forme de petites zones qui contiennent les HpaII d'enzyme de restriction dans le génome et ont donc été appelés (HTF). Ensuite, Gardiner-Garden et Frommer (GGF) (*G.G.M ET*

F.M 1987), ont défini les CGIs comme un segment d'ADN a une longueur supérieure à 200 Pb, une teneur en GC plus de 50% et le rapport CpG observé/ attendu ou CpG OE pas moins de 0,6. Le rapport est calculé de la manière suivante :

$$\text{CpG Observé/attendu} = \frac{NCpG}{Nc * Ng}$$

Ou Nc , Ng et $NCpG$ sont respectivement des nombres de C, G et dinucléotides CpG dans la région de longueur N . Par la suite, (*D. Takai et P.A. Jones, 2021*) ont réévalué ces trois paramètres et ont proposé un nouvel ensemble de critères ($GC_{ratio} \geq 55\%$, $CpG_{O/E} \geq 0,65$, longueur $\geq 500Pb$)

9.2 Importance des îlots CpG dans l'épigénétique et l'infertilité masculine :

Dans le cadre de l'infertilité masculine, les îlots CpG ont une importance particulière pour les raisons suivantes :

9.2.1 Régulation de l'expression génique:

Les îlots CpG, sont des régions du génome où les sites CpG sont normalement non méthylés. La méthylation de l'ADN au niveau de ces îlots CpG peut perturber l'expression des gènes impliqués dans la spermatogenèse et le développement embryonnaire, sans modifier la séquence d'ADN, en affectant la qualité et la mobilité des spermatozoïdes comme (DAZL, ADAM2, FANCM, HDAC1, DDX1...) (*Peter, 2015*)

9.2.2 Qualité du sperme :

La méthylation des îlots CpG agit sur et la capacité de conception. Des altérations de la méthylation des îlots CpG dans l'ADN des spermatozoïdes agissent sur l'intégrité du génome ce qui conduit à une diminution de la mobilité et provoque une morphologie anormale des spermatozoïdes, en réduisant ainsi les chances de féconder l'ovocyte tels que (ADAM2, BCL2, FANCM...) (*Dr Peter, 2015*).

9.2.3 Transmission intergénérationnelle :

Les altérations épigénétiques des îlots CpG peuvent être transmises à la génération suivante, ce qui peut affecter la qualité du sperme et le développement embryonnaire précoce. Ces modifications épigénétiques jouent un rôle majeur dans l'infertilité masculine (*Boitrelle, 2016*).

10 La recherche *in silico* :

En la littérature "*en silice*" est un terme utilisé pour définir les expériences utilisées en bio-informatique pour analyser des données à l'aide d'outils informatiques en particulier dans les domaines de la biologie et la médecine.

10.1 Les domaines d'application :

Comme la bio-informatique est un domaine très large, on s'intéresse sur l'application de ce domaine dans le cas de l'infertilité masculine permet de :

- Identifier les gènes impliqués dans l'infertilité masculine et affectés par la méthylation
- Analyser l'expression des gènes
- Identifier les ilots CpG dans les régions promotrices des gènes
- Etudier les bases moléculaires de l'infertilité masculine
- Comprendre les interactions entre les protéines
- Développer de nouveaux diagnostics et traitements

Chapitre 02 : Matériel et Méthodes

Introduction

Cette section décrit les outils, les techniques, et les procédures utilisés pour mener à bien notre étude. De la description du matériel expérimental à l'explication des méthodes d'analyse, en passant par la justification des choix effectués, ce chapitre offre une vision élargie du processus d'analyse des effets épigénétique sur l'infertilité masculine.

1 Matériels :

1.1 Données biologiques :

Nous avons utilisé comme données, les séquences nucléiques des 21 gènes impliqués à l'infertilité masculine en choisissant plusieurs variants Homo sapiens de chaque gène.

1.2 Choix des gènes :

Le choix des gènes étudiés dans cette recherche sur l'infertilité masculine a été effectué après une analyse approfondie de la littérature scientifique. Nous avons effectué des recherches pour accéder à des bases de données bibliographiques comme PubMed et Google Scholar pour identifier les articles traitant de l'implication des gènes soumis à la régulation épigénétique dans les processus de reproduction masculine. Notre étude a porté sur l'analyse des îlots CpG de chaque gène, qui jouent un rôle crucial dans la régulation de l'expression génétique et la méthylation de l'ADN. Nous avons sélectionné ces gènes en fonction de leur implication dans la spermatogenèse, la maturation des spermatozoïdes et la fécondation, ainsi que de leur présence d'îlots CpG. Cette approche nous a permis d'identifier un ensemble de gènes en relation avec l'infertilité masculine.

1.3 Outils bio-informatiques :

-Quelques d'outils bio-informatiques ont utilisés

Tableau 1: la configuration du matériel informatique utilisé lors de l'apprentissage

Ordinateur	Caractéristiques
Processeur	Intel(R) Core (TM) I5-6300U CPU @ 2.40ghz 2.50 Ghz
Mémoire Installée RAM	8,00 Go (7,88 Go Utilisable)
Type De Système	Système D'exploitation 64 Bits
Version Du Système	22H2

Tableau 2: les outils bio-informatique et leurs principes

Outil	Principe
GenBank	Collection des séquences
Sublim textet	Editeur de texte
Mega11	Alignement des séquences et construction d'arbre phylogénétique
DBCAT	Détection des ilots CpG
Weblogo	Visualisation des séquences alignées
Dotlet JS	Visualisation des motifs CG
Python	Analyser les données génomique
Outil	Principe

1.4 Définitions :

La base de données NCBI :

Chaque base de données du **NCBI** est spécialisée :

- GenBank qui offre un accès libre et ouvert à un vaste éventail de séquences d'ADN et de protéines.

MEGA11 (Molecular Evolutionary Genetics Analysis) est un logiciel d'analyse bio-informatique gratuit et complet pour les séquences moléculaires. Il est largement utilisé par les chercheurs en biologie, en génétique et en évolution pour une large gamme d'analyses principalement :

Alignement des séquences il dispose également d'outils pour éditer et visualiser les alignements des séquences des ADN, des ARN et des protéines.

ClustalW est un programme informatique utilisé en bio-informatique pour réaliser des alignements multiples de séquences. Un alignement multiple permet de comparer et d'aligner des séquences d'acides aminés (protéines) ou de nucléotides (ADN, ARN) provenant de différents organismes.

- Phylogénétique : MEGA11 peut être utilisé pour construire des arbres phylogénétiques à partir d'alignements de séquences.

Un arbre phylogénétique (arbre évolutif) : est une représentation graphique schématique des relations de parenté entre des groupes d'êtres vivants. Il permet de visualiser l'histoire évolutive de ces groupes en montrant comment ils ont divergé d'un ancêtre commun au fil du temps.

CpG Island Finder est un outil en ligne qui identifie les îles CpG dans les séquences d'ADN. Ces régions sont souvent associées aux promoteurs des gènes, et leur méthylation peut affecter l'expression des gènes. Il permet de définir plusieurs paramètres pour identifier les îles CpG, comme la longueur minimale de l'île, la teneur minimale en GC et le nombre minimal de dinucléotides CpG.

WebLogo est un outil en ligne pour la représentation de motifs de séquences biologiques. Il contribue à créer des représentations graphiques informatives des données de séquences d'ADN, d'ARN et de protéines, il peut être utilisés pour :

- Identification des motifs conservés dans les séquences.
- Comparaison des séquences homologues.
- Communication les résultats de recherche.

Dotlet JS beta est un outil en ligne open-source pour la visualisation et l'analyse de données génomiques. Il s'agit d'une version bêta d'un projet en cours de développement, mais il offre plusieurs fonctionnalités tels que ; visualisation de gènes et d'annotations, analyse de variants génétiques, exploration des données d'expression génique, analyse de l'évolution des gènes.

Python est un langage de programmation open source multi-plateformes et orienté objet. Il est utilisé dans de nombreux domaines, notamment le machine learning, la data science, le développement logiciel, l'analyse de données et la gestion d'infrastructures.

Sublime Text est un éditeur de texte et de code source multiplateforme populaire, connu pour sa rapidité, sa flexibilité et son large éventail de fonctionnalités. Il est utilisé par des millions de développeurs, d'ingénieurs et de rédacteurs dans le monde entier pour écrire et éditer du code, du texte brut et des balises.

FASTA (ou FastA) est un format de fichier texte largement utilisé en bio-informatique pour stocker des séquences d'acides aminés (protéines) ou de nucléotides (ADN, ARN). Il permet de représenter une ou plusieurs séquences de manière simple et lisible.

2 Méthodes

Notre approche *in silico* comprend plusieurs étapes :

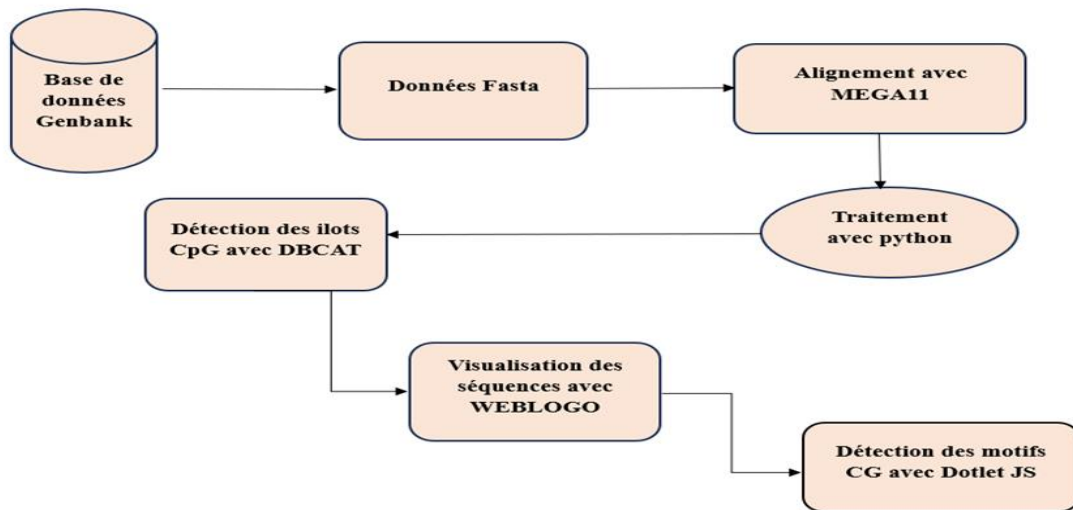


Figure 1: Processus d'approche *in silico*

2.1 Collection des données :

- Sélection des gènes soumis à l'épigénétique et affectés par l'infertilité masculine : Il s'agit d'étudier les séquences de chaque gène ciblé.

-Récupération des séquences génétiques : Nous avons utilisé la base de données GenBank gérée par le NCBI pour un accès libre aux séquences.

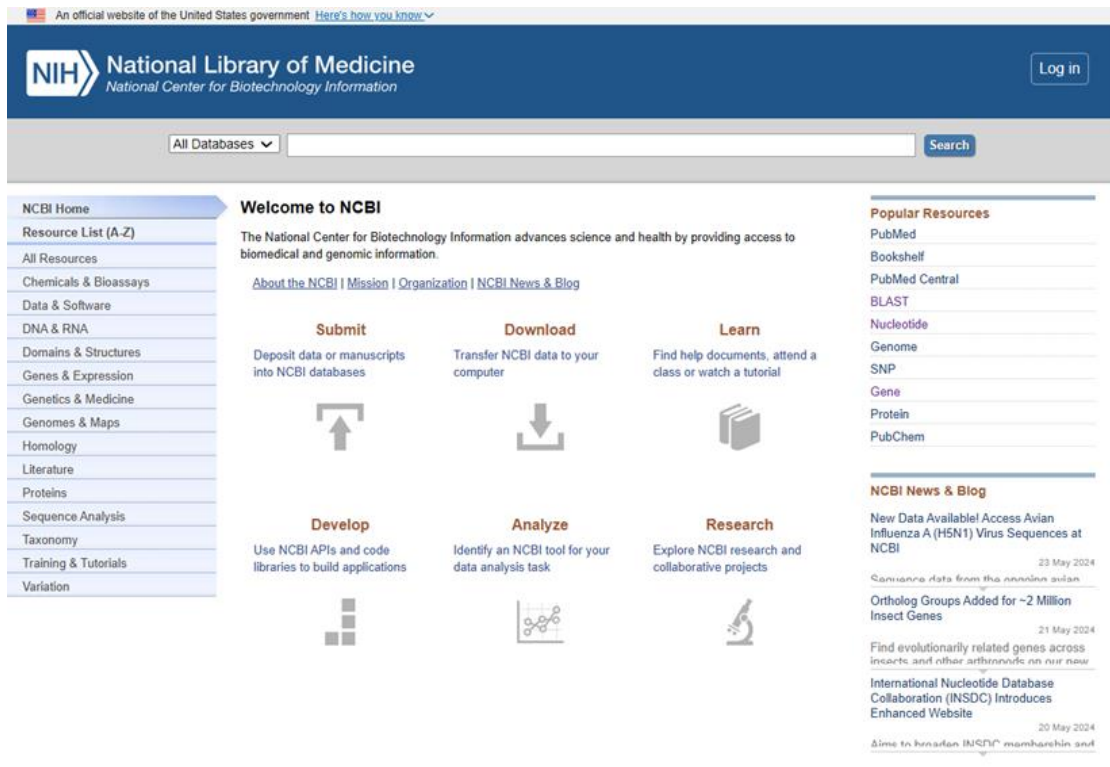


Figure 2: Interface de NCBI

-Recherche des gènes : le nom du gène a été saisi dans la barre de recherche de GenBank.

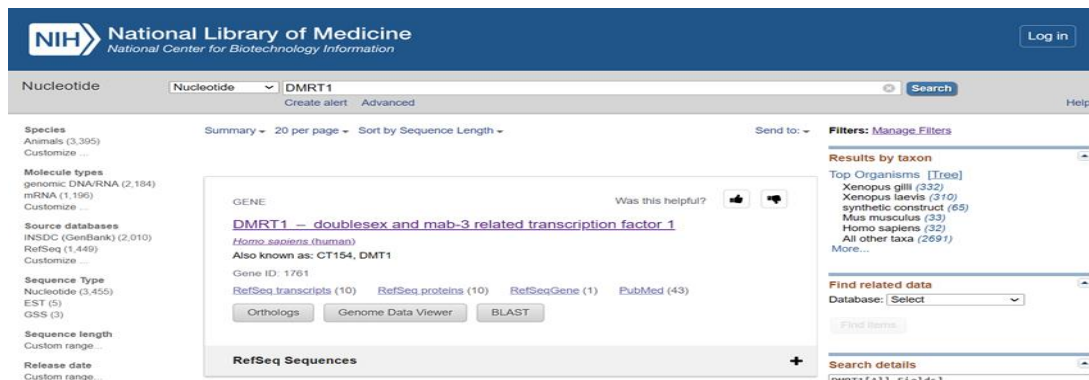


Figure 3: Interface de GenBank

-Sélection des gènes humains : l'icône "Homo sapiens" a été cliquée pour limiter la recherche aux gènes humains et étudier les variants chez l'homme



Figure 4: Résultat des gènes *Homo sapiens*

-Exportation des séquences : les séquences des variants sélectionnés ont été exportées au format FASTA.

-Enregistrement des fichiers : les fichiers FASTA contenant les séquences des gènes ont été enregistrés.

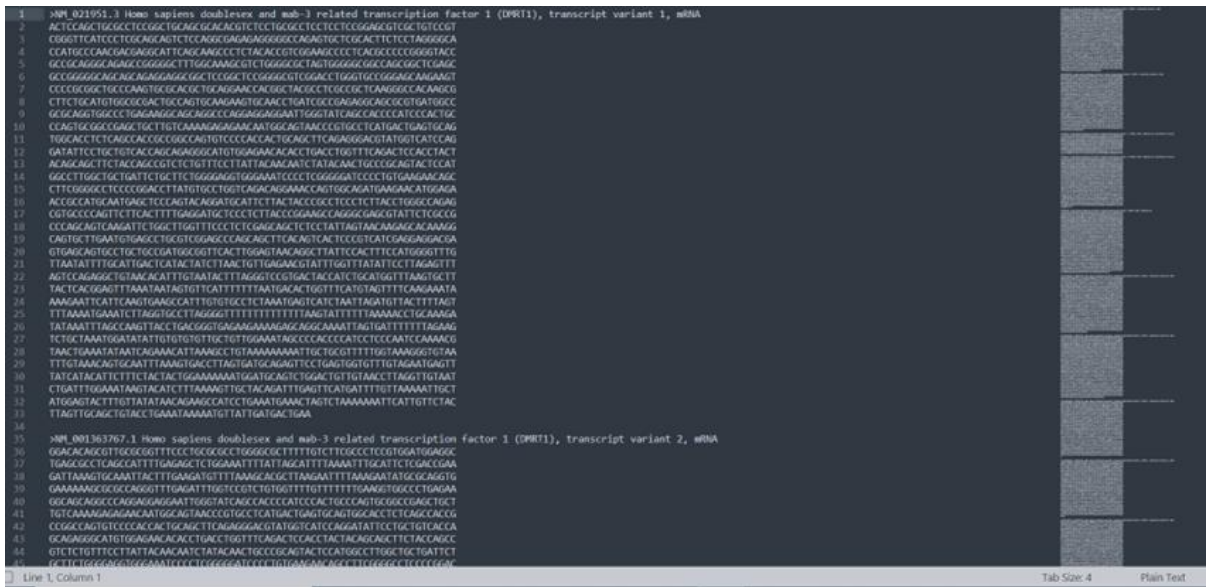


Figure 5: Séquences des variants du gène DMRT1 sur sublime Textet

2.2 Préparation des données :

Après avoir collecté et sélectionné les variants de chaque gène, nous avons vérifié le CDS (Coding Séquence) de chaque variant. Pour ce faire, nous avons identifié la position du CDS sur chaque séquence.


```

/db_xref="HGNC:HGNC:2934"
/db_xref="MIM:602424"
CDS 363..1010
/gene="DMRT1"
/gene_synonym="CT154; DMT1"
/codon_start=1
    
```

Figure 6: CDS de variant

```

ORIGIN
1  cagagttgcg ggcttcgctc ggctcggggg tccccacgat ctctcacctt gtgtttcggc
61  ggagttgchg gggttcctc tgctcccggg tccctgagat ccctgatccc gctctcaggg
121  caggaatttc cggggttcgc ccctctgggc ttccccgcga tcccagagcc cgcgttccaa
181  cagaattggg agggtttatt gcgctagagg agtcttcagg atcttcgctc cgcacctccg
241  aggtggccct gagaaggcag caggcccagg aggaggaatt gggatcagc caccatcc
301  cactgcccag tgcggccgag ctgcttgta aaagagagaa caatggcagt aaccctgtcc
361  tcatgactga gtgcagtggc acctctcagc caccgccggc cagtgtcccc accactgcag
421  ctccagaggg acgtatggc atccaggata ttctgtctgt caccagcaga gggcatgtgg
481  agaacacacc tgacctgggt tcagactcca cctactacag cagcttctac cagccgtctc
541  tgtttcctta ttacaacaat ctatacaact gcccgagta ctccatggcc ttggctgctg
601  attctgcttc tggggagggtg ggaatcccc tcgggggatc cctgtgaag aacagccttc
661  ggggctccc cggaccttat gtgctgggtc agacaggaaa ccagtggcag atgaagaaca
721  tggagaaccg ccatgcaatg agctcccagt acaggatgca ttcttactac ccgctccct
781  cttacctggg ccagagcgtg ccccagttct tcacttttga ggatgctccc tcttaccggg
841  aagccagggc gaggctattc tcgcccga gcagtcaaga ttctggcttg gtttccctct
901  cgagcagctc tcctattagt aacaagagca caaaggcagt gcttgaatgt ggcctgcgt
961  cggagcccag cagcttcaca gtcactccg tcatcgagga ggacgagtga gcagtgcctg
    
```

Figure 7: Localisation de la région promotrice en amont de la CDS du gène DMRT1

2.3 Alignement des séquences

-Nous avons utilisé le logiciel MEGA11 pour réaliser un alignement correct de chaque gène.

Nous avons suivi les étapes suivantes : sélection de la barre **Edit/Build Alignement**, affichage des séquences non alignées de plusieurs variants du même gène.

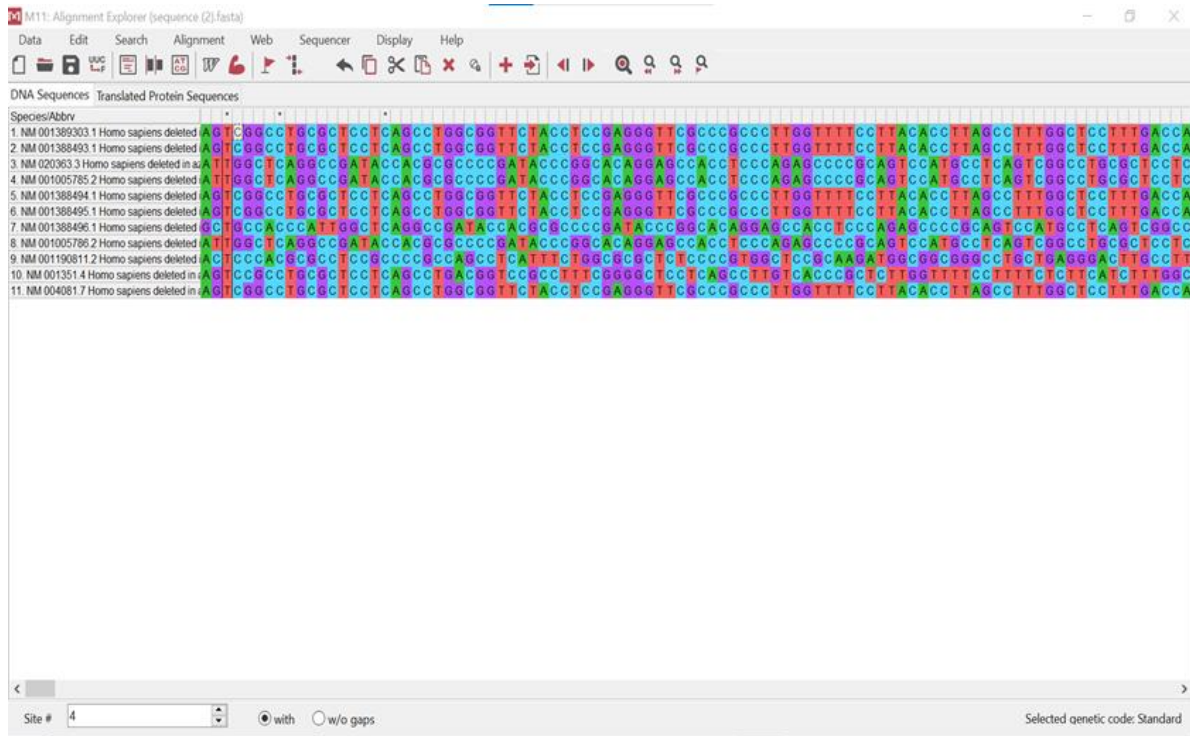


Figure 8: Interface MEGA11

-Sélection de la barre Align DNA.

-Lancement de l'alignement.

2.4 Nettoyage des séquences :

-Filtrage des séquences : après avoir obtenu les séquences alignées, nous avons filtré les séquences non identiques.



Figure 9: Alignement réalisé du gène DMRT1

-Seuls les variants présentant une certaine identité ont été conservés pour faciliter l'analyse.

2.5 Extraction de la région promotrice :

En utilisant les positions CDS, nous avons supprimé la région qui commence par les 3 nucléotides ATG jusqu'à la terminaison de chaque séquence.

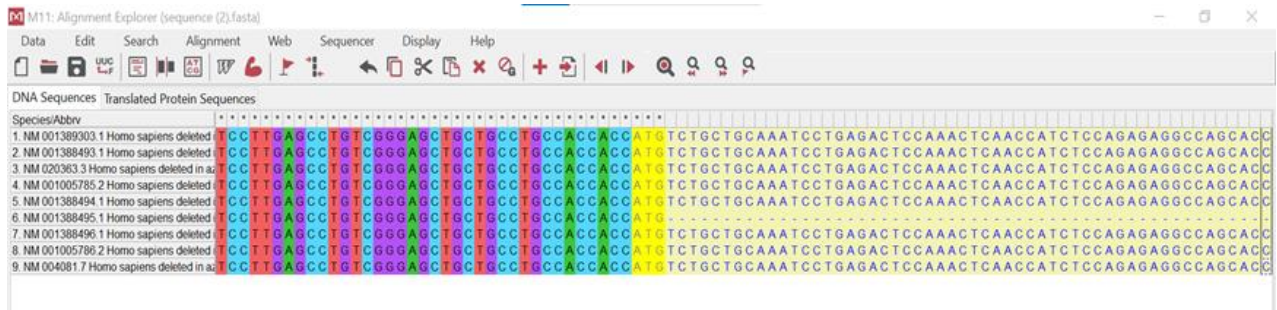


Figure 10: Nettoyage des séquences du gène DMRT1

-Cette opération a permis d'extraire la région promotrice de chaque gène.

-Obtention de la région promotrice.

-Les régions promotrices extraites des séquences nettoyées ont été enregistrées au format FASTA.

2.6 Analyse des séquences obtenus :

Sélection d'une séquence représentative étant donné que les séquences promotrices des variants sont identiques, nous avons choisi une seule séquence représentative pour l'analyse.

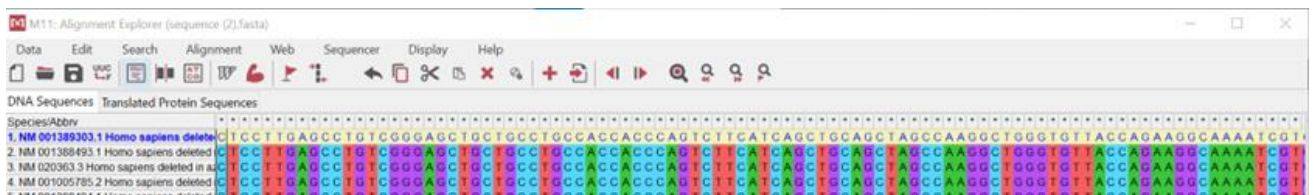
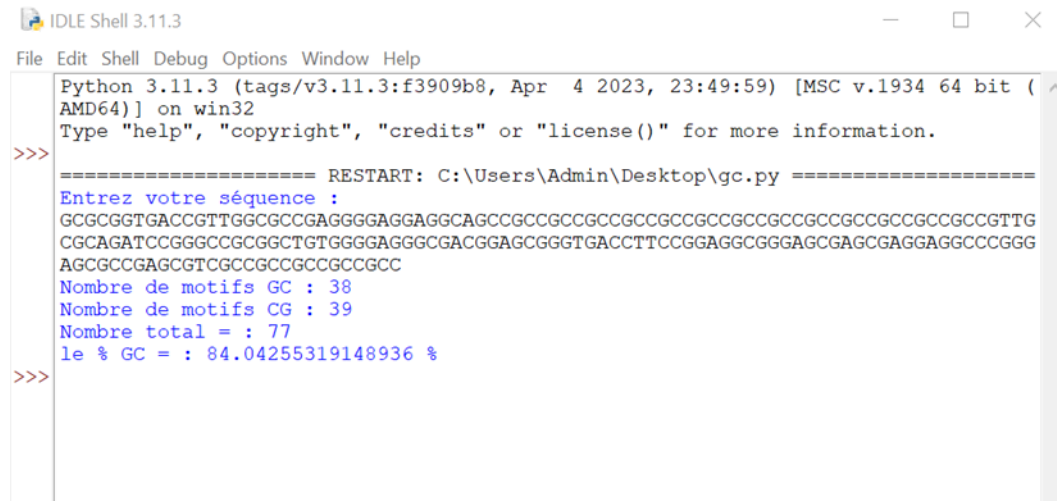


Figure 11: Séquence promotrice sélectionnée du gène DMRT1

Analyse de la séquence promotrice_un script Python a été utilisé pour calculer le pourcentage de CG (%CG) et le nombre de motifs CG et GC dans la séquence promotrice sélectionnée. La taille de la région promotrice a également été prise en compte dans l'analyse



```
Python 3.11.3 (tags/v3.11.3:f3909b8, Apr 4 2023, 23:49:59) [MSC v.1934 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license()" for more information.
>>>
===== RESTART: C:\Users\Admin\Desktop\gc.py =====
Entrez votre séquence :
GCGCGGTGACCGTTGGCGCCGAGGGGAGGAGGCAGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGCCGTTG
CGCAGATCCGGGCCGCGGCTGTGGGGAGGGGACGGAGCGGGTGACCTTCCGGAGCGGGAGCGAGCGAGGAGCCCGGG
AGCGCCGAGCGTCGCCGCCGCCGCCGCCGCC
Nombre de motifs GC : 38
Nombre de motifs CG : 39
Nombre total = : 77
le % GC = : 84.04255319148936 %
>>>
```

Figure 12: Résultat de l'exécution du gène DMRT1

2.7 Visualisation des îlots CpG :

Le site web "CpG Island Finder" a été utilisé pour visualiser les différents types d'îlots CpG (CpG intermédiaires, CpG isolés, CpG clusters) présents dans la séquence promotrice.



The image shows the DBCAT website interface. On the left is the 'CpG island Finder' tool with the following parameters:
parameter set:
o/e ratio: 0.60, 0.65
minimal length: 200, 300, 500
paste DNA sequence below: (1,000,000 bases in maximum)
Buttons: GO!, clear
Note: Please display the sequence in **FASTA** format.
Note: Any CpG island which has GC content less than 50% is filtered out.

The right panel features a banner 'Welcome to DBCAT' with an image of a computer monitor displaying a CpG island visualization. Below the banner is the title 'DataBase of CpG Islands & Analytical Tools' and a section titled 'WHAT IS DBCAT' containing the following text:
DataBase of CpG islands and Analytical Tool (DBCAT) is developed in order to recognize comprehensive methylation profiles of DNA alteration in human cancer.
DBCAT is an online methylation analytical tool composed of three parts: a CpG Island Finder, a genome query browser and an analytical tool for methylation microarray data. The analytical tool can analyze raw data generated from scanners and search genes with methylated regions which could affect gene expression regulation.
DBCAT not only identifies the regions of methylation but also searches the database to pick up genes with methylated regions of functional meanings.

Figure 13: Interface du site DBCAT

- La taille et le contenu en GC de chaque type d'îlot CpG ont également été analysés.

Possible CpG islands shown in graph:

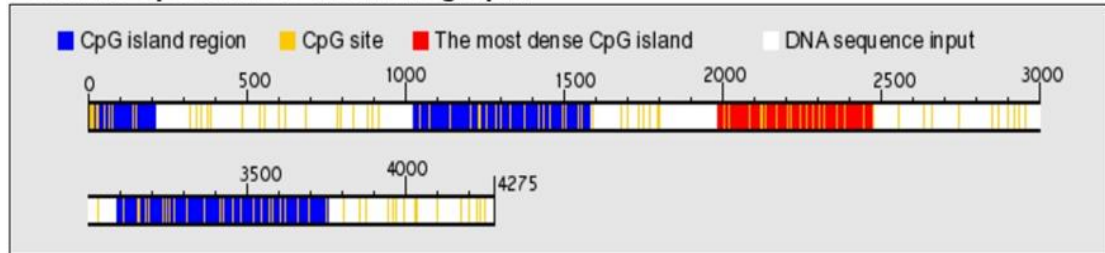


Figure 14: Analyse du gène DMRT1

2.8 Classement des résultats dans un tableau (tableau 4) :

-Les résultats obtenus ont été organisés dans un tableau récapitulatif (tableau 4).

-Ce tableau comprend les informations suivantes pour chaque gène analysé :

-Nom du gène.

-Taille de la région promotrice.

-Pourcentage de CG (%CG) dans la région promotrice.

-Nombre de motifs CG.

-Nombre de motifs GC.

-Nombre de CpG dans la région promotrice.

-Structure des CpG.

2.9 Création de logos (WebLogo) :

-Il y a deux types de logos :

Logo individuel : le site web "WebLogo" a été utilisé pour créer des logos représentant chaque région promotrice étudiée. Ce logo décrit la distribution des motifs de nucléotides au sein de chaque séquence promotrice.

Logo global : un logo global a également été créé pour regrouper les régions promotrices de tous les gènes étudiés. Ce logo permet d'identifier les motifs de nucléotides communs à l'ensemble des séquences promotrices analysées.

-Pour création séquences logos :

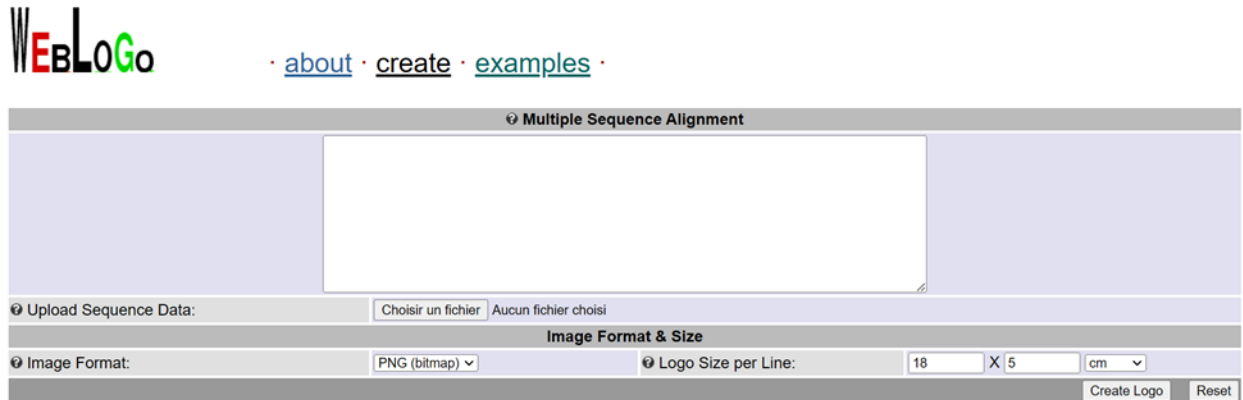


Figure 15: Interface WEBLOGO

-**Sélection du fichier FASTA:** le fichier FASTA contenant les séquences promotrices a été sélectionné comme entrée pour la création des logos.

-**Paramétrage du site:** les paramètres du site WebLogo ont été ajustés pour obtenir des logos clairs et lisibles. Cela inclut notamment le choix de la hauteur des caractères, de l'échelle de couleur et des options de mise en page.

2.10 Utilisation du site Dotlets JS beta

Visualiser les motifs CG et détecter les microsatellites

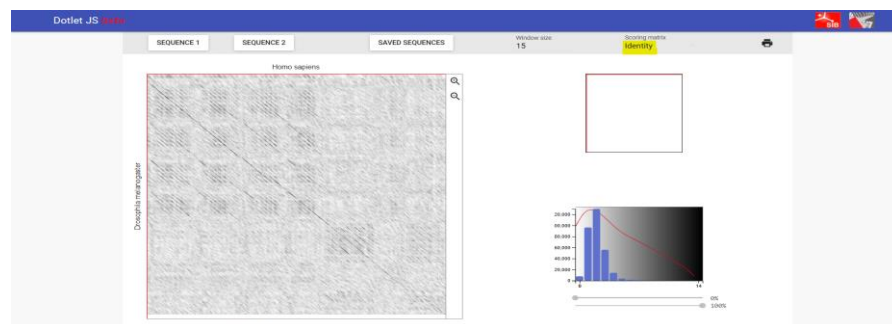


Figure 16: Interface du site Dotlets JS beta

Chapitre 03 : Résultats et Discussion

1 Résultat de la construction de l'arbre phylogénétique des 21 gènes étudiés

Nous avons construit un arbre phylogénétique à partir des séquences des régions promotrices des 21 gènes impliqués dans l'infertilité masculine. Les 21 gènes sont regroupés dans l'arbre sous forme de 5 clades importants :

- Clade1: ADAM2, DAZL, FANCM, HDAC1, ACE, EGR4
- Clade2: DDX4, EZH2, PPP1CC, DMRT1, CTCF, MCPH1
- Clade3: HMGB2
- Clade4: BCL2, KITLG, FANCA, APAF1
- Clade5: CPEB1, HAT1, MLH1, BRDT

Ce groupement en clades suggère que chaque groupe est impliqué dans des processus biologiques proches et essentiels à la spermatogenèse et à la fertilisation masculine. De même qu'il indique une corrélation fonctionnelle entre eux dans le contexte de l'infertilité masculine

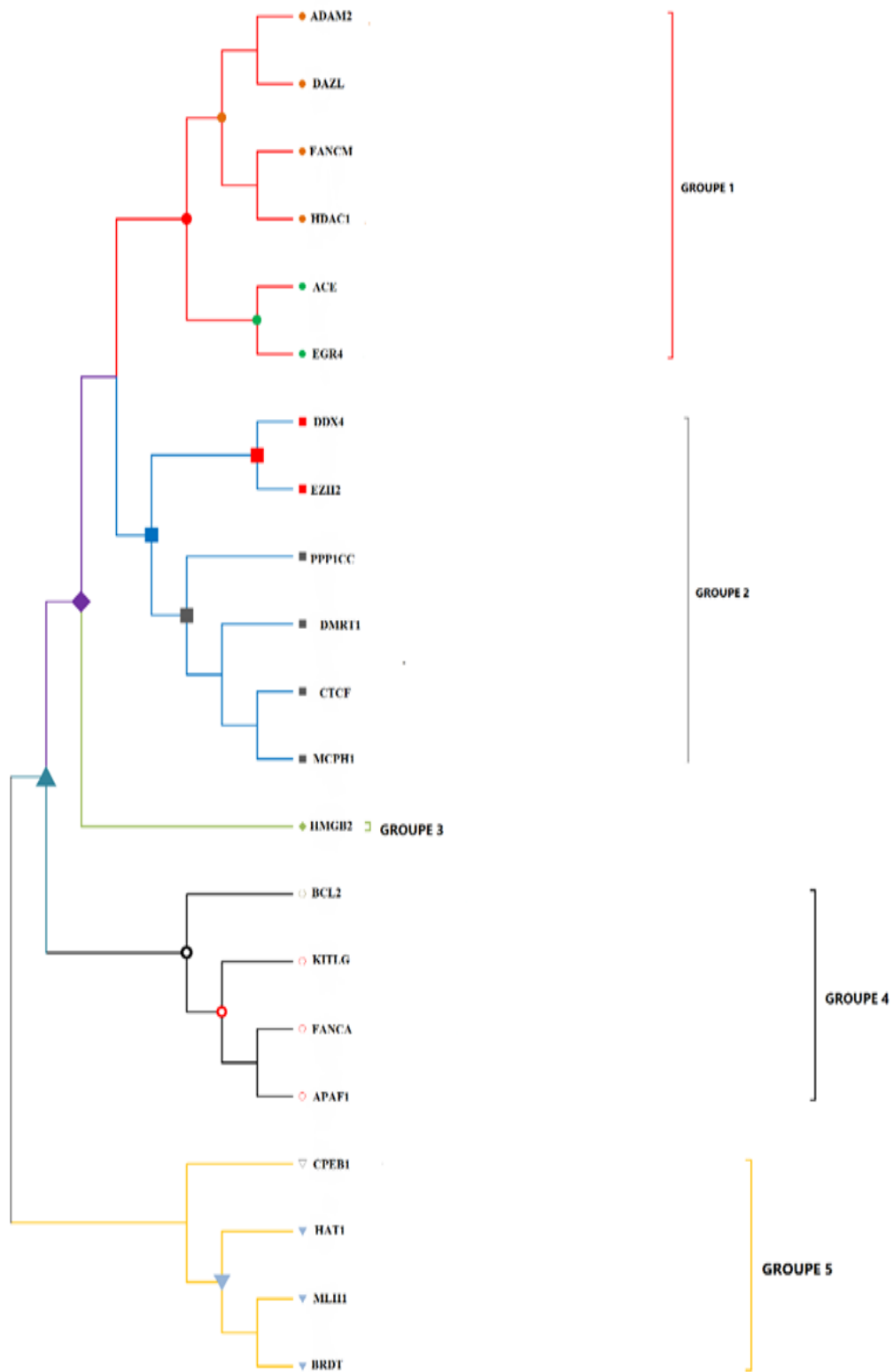


Figure 18: Arbre phylogénétique de 21 gènes impliqués dans l'infertilité masculine

2 Visualisation des résultats dans les logos :

D'après le logo du premier groupe

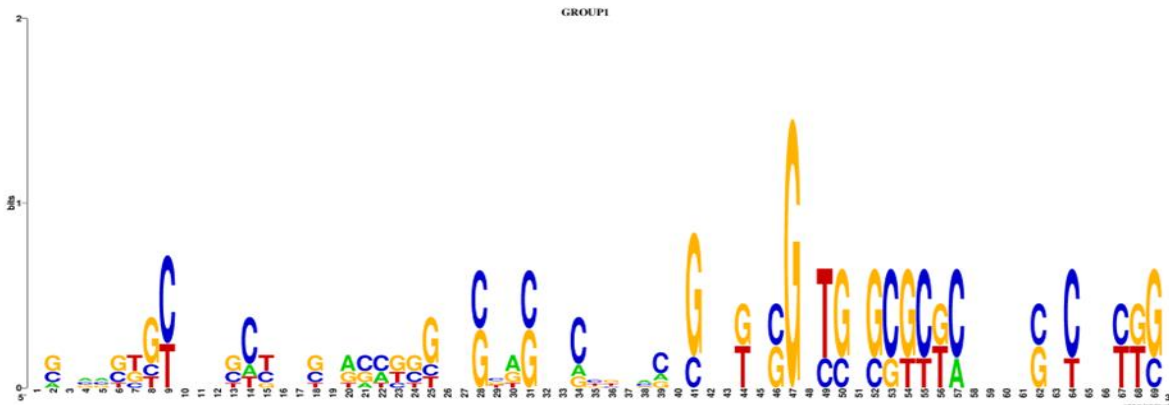


Figure 19: Logo du premier groupe de gènes (ADAM2, DAZL, FANCM, HDAC1, ACE, EGR4)

Nous constatons, la présence de motifs similaires CG pour tout le groupe dans les positions (28,31,41). Ce motif est connu pour être associé aux îlots CpG, le fait qu'il soit présent suggère que la méthylation de l'ADN pourrait affecter l'expression des gènes de ce groupe. La présence d'un motif GC dans les positions (50,52,69) pourrait suggérer que ces gènes sont soumis à une régulation épigénétique. Comme ils sont impliqués dans la production et la fonction des spermatozoïdes, des profils de méthylation anormaux pourraient contribuer aux fonctions reproductrices et donc à l'infertilité. La présence de la base G conservée dans la position 47 sur l'ensemble de ces gènes, peut être importante dans le contrôle de l'expression génétique en fixant des facteurs de transcription spécifiques (figure19).

-D'après le logo du deuxième groupe

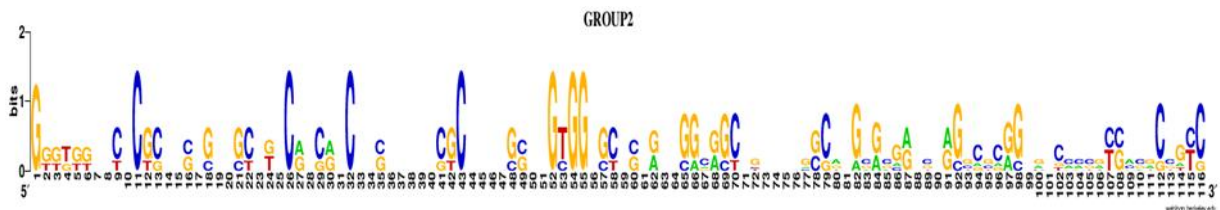


Figure 20: Logo du deuxième groupe des gènes (DDX4, EZH2, PPP1CC, DMRT1, CTCF, MCPH1)

Nous constatons que l'enchaînement des motifs CG et GC dans le logo de ce groupe est un indicateur de la présence d'îlots CpG dans les régions promotrices de ces gènes. Des anomalies dans la méthylation des îlots CpG provoque une atténuation de l'activité des gènes ce qui conduit à de de fertilité. La présence de séquence conservé G dans les positions 1, 54 et 55 indique que ces nucléotides sont importants et ont été conservés au cours de l'évolution de la séquence, ces positions pourraient être impliquées dans la liaison des protéines ou le repliement de l'ADN. La Présence de séquence conservée C dans les positions 11,26 et 43 pourraient avoir une indication de la présence de motifs fonctionnels (*Delphine, 2017*).

-D'après le 3ème groupe

La position unique du gène HMGB2 dans l'arbre phylogénétique suggère qu'il s'agit d'un gène évolutif distinct, il peut avoir divergé des autres gènes du groupe plus tôt dans l'évolution, de même que ça peut être lié à sa fonction distincte de celle des autres gènes. Selon les données bibliographique ce gène est impliqué dans le fonctionnement de la fonction spermatique (*Sachin, Neha and Martin,2013.*) HMGB2 est abondamment exprimé dans les spermatozoïdes humains, où il se localise principalement dans le noyau et le cytoplasme. Son expression est régulée par divers facteurs transcriptionnels, notamment p53 et SP1(*Maria et al.,2018*).

-D'après le logo du 4ème groupe :

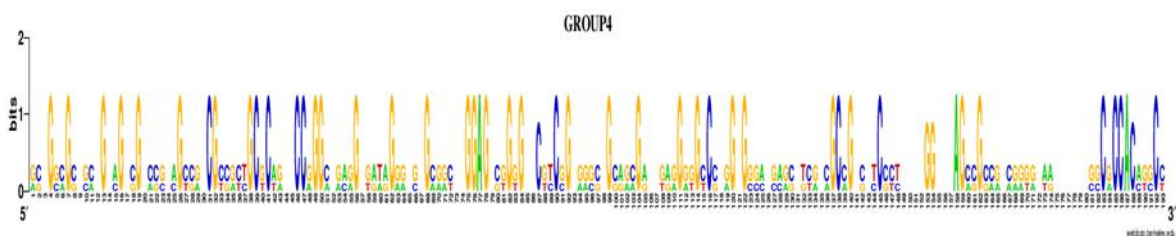


Figure 21: Logo 4ème groupe des gènes (BCL2, KITLG, FANCA, APAF1)

La présence de nombreuses bases G conservées dans la région promotrice suggère la présence de motifs transcriptionnels importants, peut aussi indiquer que cette région promotrice est fortement régulée par des facteurs de transcription. La présence de motifs CG dans les positions 31 et 32 suggère la présence d'un îlot CpG. La présence de motifs GC dans les positions 38 et 39 peut avoir plusieurs significations : faire partie de sites de liaison pour des protéines spécifiques, peuvent contribuer à la stabilité de l'ADN, influencer la structure de l'ADN, ou simplement refléter une variation aléatoire de la séquence d'ADN (*ZAHIA, 2016*).

-D'après le logo du 5^{ème} groupe

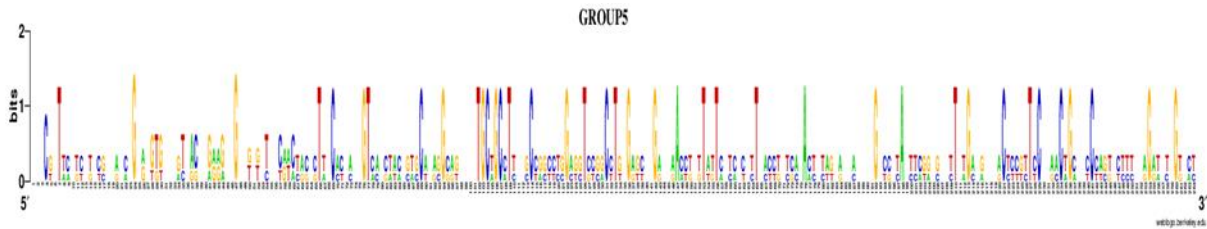


Figure 22: Logo du 5^{ème} groupe des gènes (CPEB1, HAT1, MLH1, BRDT)

La présence de bases conservées G et C à plusieurs positions peut indiquer la présence de sites de liaison pour des facteurs de transcription spécifiques, les séquences riches en G conservées peuvent être importantes pour le contrôle de l'expression des gènes en attirant des facteurs de transcription spécifiques. De même que la présence de motifs GC et CG, mais leur distribution peut varier d'une région à l'autre.

-D'après le logo entier

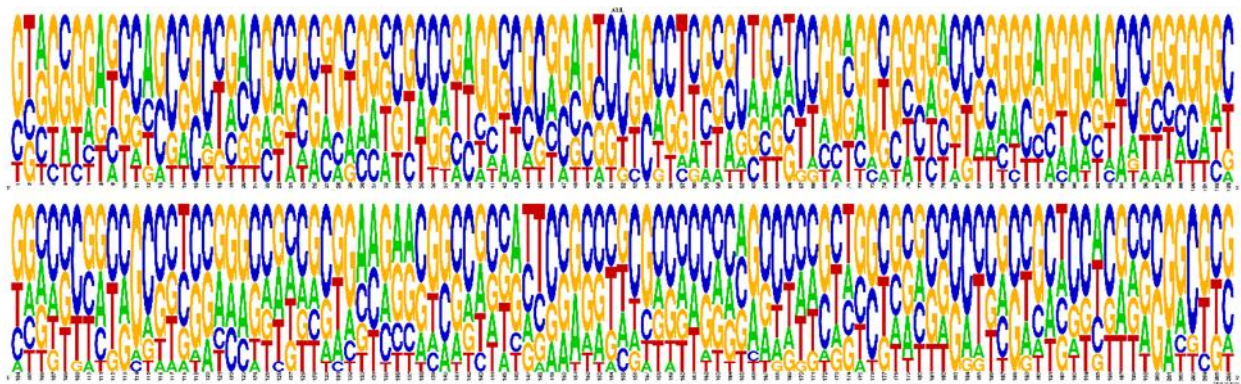


Figure 23: Le logo entier des régions promotrices des 21 gènes

Les motifs GCCG palindromes est une séquence d'ADN qui se lit de la même manière dans les deux sens. Ainsi nous avons remarqué une présence d'un palindrome parfait CCGGCC dans la position [129-125]. Les motifs CCGG et GGCC peuvent servir de sites de liaison pour des facteurs de transcription spécifiques. La présence des poly G et poly C peuvent avoir plusieurs fonctions dans les régions promotrices tels que Sites de liaison pour des protéines, influencer la structure de l'ADN, ou stabilité de l'ARN.

3 Classement et analyse des résultat du calcul du tableau :

-D'après les résultats

Nous avons remarqué que les gènes ACE et EGR4 ont un % CG $73.5 \approx 79.6$ proche et cela peut expliquer pourquoi ils sont positionnés sur le même clade. Aussi, cela peut indiquer que ces deux gènes partagent des fonctions biologiques similaires ou qu'ils sont régulés par des mécanismes similaires. La même chose pour les gènes FANCM et HDAC1 qui ont un % CG très proche $56 \approx 57.2$. Comme pour les gènes ACE et EGR4, cette observation indique que les gènes FANCM et HDAC1 partagent des fonctions biologiques similaires ou qu'ils sont régulés par des mécanismes similaires.

Nos résultats montrent un %CG dans les régions promotrices de chaque gène est élevé dans l'intervalle ($63 \approx 75$) % (tableau 3) cela signifie leur régulation épigénétique. Les gènes EZH2 et PPP1CC aient un %CG = 75% et les nombres des motifs CG et GC pour DMRT1, PPP1CC, EZH2 proches, suggèrent les régions promotrices sont soumises à la régulation de l'expression des gènes, car la distribution des îlots CpG et donc aux effets épigénétique (tableau3).

Tableau 3: les résultats obtenus sur les régions promotrices de chaque gène

Nom du gene	Taille de la région RP	CG% dans RP	Nombre de motif CG	Nombre de motif GC	Total	Nbr CpG dans la RP	Structure CpG
ACE	34	73.5	5	6	11	0	CpG isolés
ADAM2	42	57.1	1	3	4	0	CpG isolés
CTCF	124	71	15	21	36	0	CpG isolés
DAZL	208	61.54	10	19	29	0	CpG isolés
DMRT1	362	63.25	23	30	53	1	-CpG intermédiaire 356p - CpG isolés
FANCA	32	81.25	4	7	11	0	CpG isolés
KITLG	188	84.04	39	38	77	0	CpG isolés
MCPH1	41	70.73	5	6	11	0	CpG isolés
MLH1	36	50	1	1	2	0	CpG isolés
PPP1CC	184	75	22	23	45	0	CpG isolés

DDX4	55	67.3	4	8	12	0	CpG isolés
APAF1	535	69	45	53	98	0	CpG isolés
BRDT	298	57.3	21	32	53	0	CpG isolés
CPEB1	269	50.5	3	20	23	0	CpG isolés
EGR4	231	79.6	32	41	73	1	CpG intermédiaire 225p -CpG isolés
EZH2	135	75	27	20	47	0	-CpG isolés
FANCM	34	73.5	5	6	11	0	CpG isolés
HMGB2	133	62.4	12	10	22	0	CpG isolés
HDAC1	400	57.2	13	27	40	1	-CpG isolés -CpG Cluster
HAT1	54	51	15	16	31	0	-CpG Isolés
BCL2	702	53	86	112	198	1	-CpG isolés -CpG intermédiaire

Les nombres des motifs CG et GC pour DMRT1, PPP1CC, EZH2, sont aussi proches et ils sont bien répartis dans les régions promotrices. Le %CG élevé dans KITLG = 84.04 et FANCA =81.25, signifie que leurs régions promotrices sont riches en îlots CpG. Nous constatons un nombre de motifs GC élevé dans BCL2 égale à 112. Le gène FANCA dont la région promotrice fait 32pb et un % CG =81,25 suggère que la majorité de la région promotrice est composée d'îlots CpG. Cela renforce l'idée que ce gène est étroitement régulé par la méthylation de l'ADN. La petite taille de la région promotrice peut suggérer que les éléments de régulation sont concentrés dans cette région. Le nombre de motifs CG et GC presque identique dans KITLG $38 \approx 39$ suggère que sa région promotrice est équilibrée en CG et GC (Tableau 3).

Le nombre de motifs CG supérieur à GC dans le gène CPEB1, suggère une présence d'un îlot CpG, de même pour le %CG entre 50 et 62 dans les gènes CPEB1, HAT1, MLH1 et BRDT c'est cohérent avec la présence d'îlots CpG dans ces gènes. Nous constatons que souvent le % de CG est supérieure à 50% ce qui renforce l'idée que ces gènes sont susceptibles d'être régulés par la méthylation de l'ADN dans les îlots CpG (Tableau 3).

4 Visualisation des résultats obtenus du DBCAT :

D 'après les résultats obtenus

Possible CpG islands shown in graph:

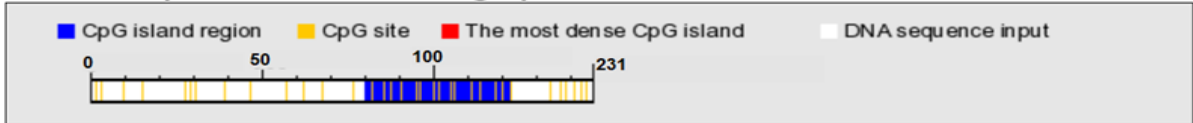


Figure 24: Les CpG trouvé dans la Rp du gène EGR4

Possible CpG islands shown in graph:

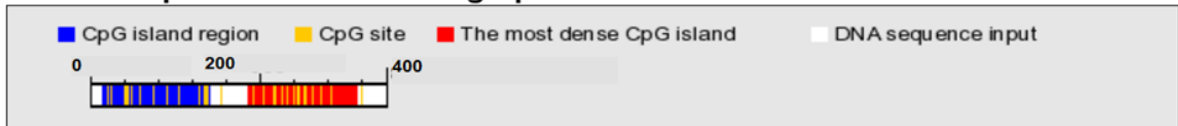


Figure 25: Les CpG trouvé dans la Rp du gène HDAC1

Possible CpG islands shown in graph:

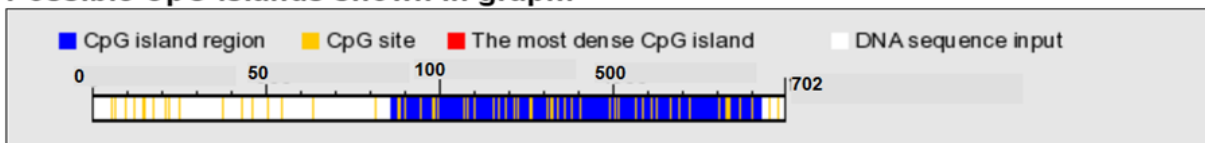


Figure 26: Les CpG trouvés dans la Rp du gène BCL2

Nous avons obtenu comme résultats, une présence de différents types de CpG notamment CpG isolés (en jaune), CpG cluster (en rouge) et CpG intermédiaire (en bleu), ainsi nous avons remarqué une variété dans les positions des CpG dans les différentes régions promotrices des gènes ces résultats suggèrent que ces régions peuvent avoir des fonctions différentes et jouer des rôles distincts dans la régulation de l'expression génique.

5 Observation des séquences sur DOLETE JS

-D 'après les résultats obtenus

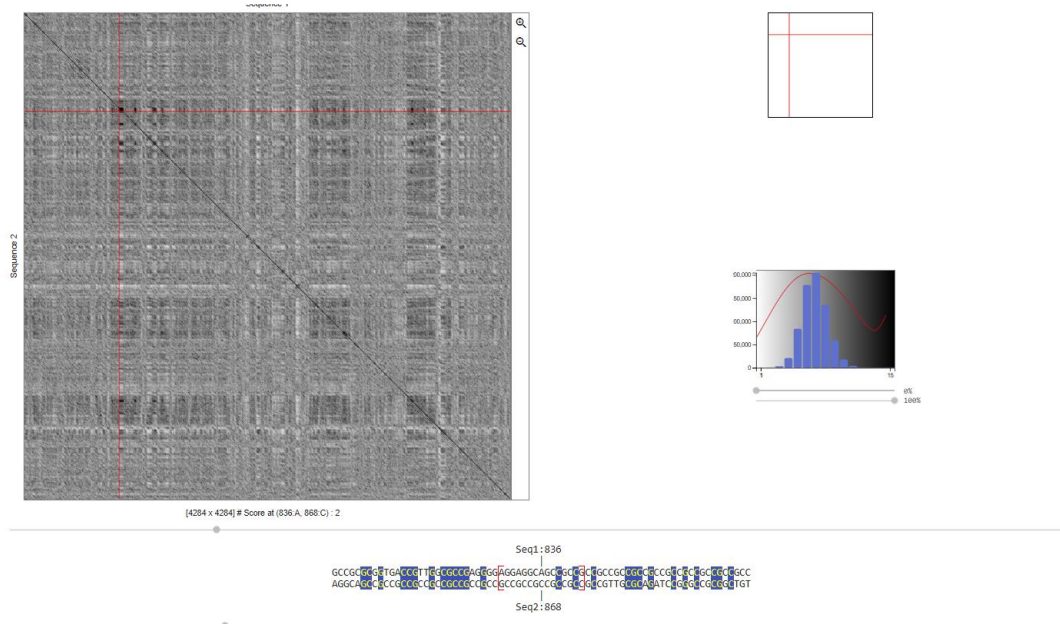


Figure 27: Résultat obtenu du Dotlet Js des 21 gènes

Présence des motifs CG et des microsatellites pourrait être pertinente pour l'analyse de l'infertilité masculine, îlots CpG peut affecter l'expression des gènes, microsatellites sont des segment ADN répètes en tandem peut être instable

6 Discussion des résultats

Dans le travail que nous venons d'effectuer, nous avons dans un premier temps sélectionné des gènes impliqués dans l'infertilité masculine et qui jouent des rôles majeurs dans la spermatogenèse et la fertilité masculine. Nous avons constaté que 2 gènes comme, ADAM2 et DAZL forment un groupe monophylétique, leurs fonctions dans la division cellulaire et la spermatogenèse, sont essentielles pour le spermatozoïde. Plus largement ; le gène EGR4 essentiel pour la formation des spermatozoïdes, et occupe une position proche de ADAM2 et DAZL sur l'arbre phylogénétique. Nos résultats montrent aussi que ces gènes renferment des % de CG proches et donc montrent l'effet épigénétique dans la régulation de ces gènes.

Pour le DMRT1 et PPP1CC ; des études ont suggéré que PPP1CC peut interagir avec des protéines telles que DMRT1 pour réguler la phosphorylation de ces protéines ce qui explique leurs positionnements dans le 2ème groupe (figure18). Les gènes CTCF et MCPH1 interprètent dans la structure de la chromatine donc ils partagent la même fonction ce qui signifie leurs regroupements

dans l'arbre. Les modifications qui peuvent subir ces gènes provoquent en premier lieu une expression réduite du gène, dysfonctionnement de la spermatogenèse, augmentation des mutations et des anomalies dans la maturation des spermatozoïdes et cela explique leurs relations directes avec l'infertilité masculine.

Nos résultats montrent que la distribution des dinucléotides CpG dans les gènes n'est pas aléatoire. Ils tendent à se regrouper dans des régions appelées îlots CpG, qui jouent un rôle important dans la régulation de l'expression génétique. La distribution des CpG dans ces gènes montre une grande variabilité entre et au sein du même clade. Cela suggère que la méthylation des CpG joue un rôle important dans la régulation de l'expression de ces gènes, et que ce rôle peut varier d'un gène à l'autre. Les analyses de la distribution des CpG et des motifs CG et GC sur les régions promotrices de chaque gène ont montré des résultats intéressants. Par exemple, les régions promotrices de BCL2, EGR4 et DMRT1 comportent des CpG intermédiaires et des motifs CG et GC, ce qui suggère une forte activité de la méthylation de l'ADN dans ces régions, ainsi que le gène HDAC1 comporte dans sa région un CpG cluster ce qui suggère d'une part une forte activité de la méthylation de l'ADN dans ces régions et d'autre part une modulation de la régulation épigénétique de ces gènes.

De nos résultats il serait donc intéressant d'élargir cette approche à d'autres gènes comme : HSP70, SIRT1, EGFR et MYC..., il serait important de déterminer le rôle de l'épigénétique dans l'infertilité masculine.

7 Conclusion et perspective

Notre travail a montré qu'une modification épigénétique d'un groupe de gènes impliqués dans la spermatogenèse peuvent entraîner une expression réduite, des dysfonctionnements de la fonction reproductrice, et/ou des anomalies dans la maturation des spermatozoïdes. Ces modifications peuvent être induites par des facteurs environnementaux, tels que l'âge, la nutrition, le stress, et les médicaments, et peuvent être transmises à la progéniture, car la méthylation des CpG est un mécanisme épigénétique clé dans la régulation de l'expression des gènes impliqués dans l'infertilité masculine.

Des recherches supplémentaires sont nécessaires pour confirmer les résultats que nous avons observés et ainsi mieux comprendre le rôle de la méthylation CpG dans l'infertilité masculine.

En résumé, notre étude fournit des preuves que la méthylation CpG joue un rôle important dans la régulation épigénétique des gènes impliqués dans l'infertilité. Ces résultats pourraient avoir des implications importantes pour le diagnostic de cette pathologie et le développement de nouvelles thérapies pour le traitement de l'infertilité masculine.

Liste des références bibliographiques

1. Abdelbasset Boukelia. (2020). Méthodes d'intelligence computationnelle pour l'analyse et l'exploration de données épigénétiques massives. <https://doi.org/10.13140/RG.2.2.23538.32965>
2. Abdel Raheem A, Ralph D, Minhas S. Male infertility. *Br J Med Surg Urol.* 2012; 5:254-268.
3. Amira ZAROOUR, & Djihad CHELIA « Etude Comparative des Méthodes in Silico pour la Détection des Ilots CpG », 2021
4. Belmokhtar R. ResearchGate , Exploration de la stérilité humaine : Recours aux techniques cytogénétiques (caryotype, FISH), cité le 2 juillet 2020.
5. B. Yasmine et M.Ferial. « Nouvelle approche basée sur l'optimisation par essaim de particule pour la détection des ilots CpG dans le génome humain », cité le 30 septembre 2021.
6. Boitrelle, F. (s. d.). Liens entre la morphologie et les marques épigénétiques, la qualité de l'ADN, le contenu chromosomique et les capacités fécondantes du spermatozoïde humain.
7. Cazin, C. (s. d.). Exploration des causes génétiques des infertilités masculines sévères et implémentation d'une stratégie diagnostique par séquençage haut débit.
8. Cederroth, C. R., Vassalli, J.-D., & Nef, S. (2007). De l'épigénétique et du développement. *Revue Médicale Suisse*, 3(100), 528-532. <https://doi.org/10.53738/REVMED.2007.3.100.0528>
9. Choucair, F. (s. d.-a). Unraveling the sperm transcriptome by next generation sequencing and the global epigenetic landscape in infertile men.
10. Choucair, F. (s. d.-b). Unraveling the sperm transcriptome by next generation sequencing and the global epigenetic landscape in infertile men.
11. Coutton, C., Satre, V., Arnoult, C., & Ray, P. (2012). Génétique de l'infertilité masculine : Les nouveaux acteurs. *médecine/sciences*, 28(5), 497-502. <https://doi.org/10.1051/medsci/2012285014>
12. Eddine, A. S., & Abdelhafedh, B. (s. d.). Nouvelle approche de prédiction des classes protéiques issues d'un séquençage NGS par Deep Learning.
13. Gnansia, E. (2017). Environnement, génétique et épigénétique. *Revue de médecine périnatale*, 9(2), 66-72. <https://doi.org/10.1007/s12611-017-0413-4>
14. Gonon, F., & Moisan, M.-P. (2013). L'épigénétique, la nouvelle biologie de l'histoire

- individuelle : Revue française des affaires sociales, 1, 21-31.
<https://doi.org/10.3917/rfas.125.0021>
15. Grandjean, V. (2010). Épigénétique et infertilité/fertilité. *Basic and Clinical Andrology*, 20(4), 282-283. <https://doi.org/10.1007/s12610-010-0105-4>
16. Guibourdenche, M. (s. d.). Effets d'une exposition chronique maternelle à un pesticide et un régime obésogène sur les fonctions physiologiques digestives : Approches in vivo et in vitro.
17. Hamamah, S. (2023). Vers une stratégie nationale de lutte contre l'infertilité. In *Infertilité* (p. 439-444). Elsevier. <https://doi.org/10.1016/B978-2-294-78134-6.00049-8>
18. Houria, T., & Somia, S. (s. d.). Analyse in silico de quatre facteurs de transcription exprimés chez *Triticum aestivum* sous conditions de stress biotique et abiotique.
19. Humbert, M. (s. d.). Diagnostic génétique de l'azoospermie et prise en charge en assistance médicale à la procréation au CHU de bordeaux, en particulier dans le contexte de l'agénésie bilatérale des canaux déférents.
20. Huyghe, E. (s. d.). INFERTILITE MASCULINE.
21. Id, H. (s. d.). Causes de l'infertilité masculine et intérêt des FIV ICSI / IMSI.
22. Jacques Young « Infertilité masculine, mécanismes, causes et exploration », VOL 80, Janvier-Février 2016.
23. Junien, C., Panchenko, P., Fneich, S., Pirola, L., Chriett, S., Amarger, V., Kaeffer, B., Parnet, P., Torrisani, J., Bolaños Jimenez, F., Jammes, H., & Gabory, A. (2016). Épigénétique et réponses transgénérationnelles aux impacts de l'environnement : Des faits aux lacunes. *médecine/sciences*, 32(1), 35-44.
<https://doi.org/10.1051/medsci/20163201007>
24. Karaouzene, T. (s. d.). Bioinformatique et infertilité : Analyse des données de séquençage haut-débit et caractérisation moléculaire du gène DPY19L2.
25. Lakhdari, N. (s. d.). Programmation néonatale de l'infertilité mâle : Rôle de la dérégulation de l'expression des microARNs dans l'apoptose des cellules germinales.
26. Marcelli, F., Robin, G., & Rigot, J.-M. (2009). Prise en charge de l'infertilité masculine. *Progrès en Urologie*, 19(4), 260-264. <https://doi.org/10.1016/j.purol.2008.10.027>
27. Morange, M. (2005). L'épigénétique : Un domaine de recherche aux multiples facettes. *médecine/sciences*, 21(4), 339-339. <https://doi.org/10.1051/medsci/2005214339>
28. Olivennes, F., Hazout, A., & Frydman, R. « Assistance médicale à la procréation. Elsevier Masson ». (2006).

29. Perrot Tafforeau, F. (2018). Épigénétique et génétique tissent le fil de soi: *Insistance*, n° 13(1), 137-152. <https://doi.org/10.3917/insi.013.0137>
30. Schlosser J, Nakib I, Carré-Pigeon F, Stearman F. 2007. Infertilité masculine : stratégie de prise en charge. *EMC Urologie* 12: 18-760 - A -.
31. SEPANIAK, S. FORGES, T. FONTAINE, B et al, « Impact négatif du tabac sur la fertilité masculine : des spermatozoïdes à la descendance ». *Journal de Gynécologie Obstétrique et Biologie de la Reproduction*, Vol 33, Issue 5, pp 384-390, 2004.
32. Sabrina Yara « Le rôle du stress oxydant dans les changements épigénétiques contribuant aux complications du syndrome métabolique », Septembre 2014.
33. Smith, S., Pfeifer, S. M., & Collins, J. A. (2003). Diagnosis and management of female infertility. *Jama*, 290(13), 1767-1770
34. Thakur, S., Garg, N., & Adamo, M. L. (2013). Deficiency of Insulin-Like Growth Factor-1 Receptor Confers Resistance to Oxidative Stress in C2C12 Myoblasts. *PLoS ONE*, 8(5), e63838. <https://doi.org/10.1371/journal.pone.0063838>
35. Wallerand, H., Bernardini, S., Chabannes, E., & Bittard, H. (2003). Infertilité masculine de cause génétique et biologie moléculaire. *Progrès en Urologie*.
36. WHO. Laboratory Manual for the Examination and Processing of Human Semen. 5th ed. Geneva: World Health Organization, 2010. « Génétique de l'infertilité masculine Les nouveaux acteurs », *médecine science*, Issue 5, Vol 28, pp 497-502, Mai 2012.
37. Young, J. (2016). Infertilité masculine : Mécanismes, causes et exploration. 80.
38. Yurgel, M. E., Shah, K. D., Brown, E. B., Burns, C., Bennick, R. A., DiAngelo, J. R., & Keene, A. C. (2018). Ade2 Functions in the Drosophila Fat Body To Promote Sleep. *G3 Genes|Genomes|Genetics*, 8(11), 3385-3395.
<https://doi.org/10.1534/g3.118.200554>
39. Zarour, A. (s. d.). Etude Comparative des Méthodes in Silico pour la Détection des Ilots CpG.
40. ZAHIA AOUABED « méthodes bio-informatiques pour analyse des mécanismes moléculaires conduisant à la résistance aux médicaments dans le cancer du sein », Juillet 2016.

Année universitaire : 2023-2024

Présenté par : ALLECHE Imene
SOUALAH Malak

Approche in silico dans l'analyse épigénétique

« Le cas de l'infertilité masculine »

Mémoire pour l'obtention du diplôme de Master en Bio-informatique

L'objectif de ce travail est d'explorer in Silico les effets possibles de l'épigénétique que l'infertilité masculine. Nous avons identifié 21 gènes impliqués dans l'infertilité masculine décrits comme sous l'effet de l'épigénétique. La méthylation des îlots CpG joue un rôle majeur dans l'expression des gènes, donc l'épigénétique peut être une cause sérieuse de l'infertilité masculine, en particulier par l'effets de nombreux facteurs environnementaux. Les résultats obtenus montrent une variété dans la distribution des CpG au sein de nos gènes, ce qui signifie une forte activité de méthylation de l'ADN. Nos résultats, mêmes, préliminaires peuvent fournir un éclairage pour identifier des causes épigénétique dans l'infertilité masculine et donc dans le traitement de cette maladie.

Mots-clefs : Epigénétique., Infertilité masculine., Ilots CpG, Méthylation.

Président du jury : Pr. HAMIDECHI M. A (Professeur U. KHENCHELA).

Encadrant : Dr BENZAADA Mostafa (MCA - U Constantine 1 Frères Mentouri).

Examineur(s) : Dr AMINE KHODJA Ihsene (MAB - U Constantine 1 Frères Mentouri)