

الجمهورية الجزائرية الديمقراطية الشعبية
وزارة التعليم العالي و البحث العلمي

REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE

UNIVERSITE 8 MAI 1945 GUELMA
FACULTE DES SCIENCES DE LA NATURE ET DE LA VIE ET SCIENCES DE LA
TERRE ET DE L'UNIVERS
DEPARTEMENT DE BIOLOGIE



Mémoire de Master

Domaine : Sciences de la Nature et de la Vie

Filière : Sciences Biologiques

Spécialité/Option : Biochimie Appliquée

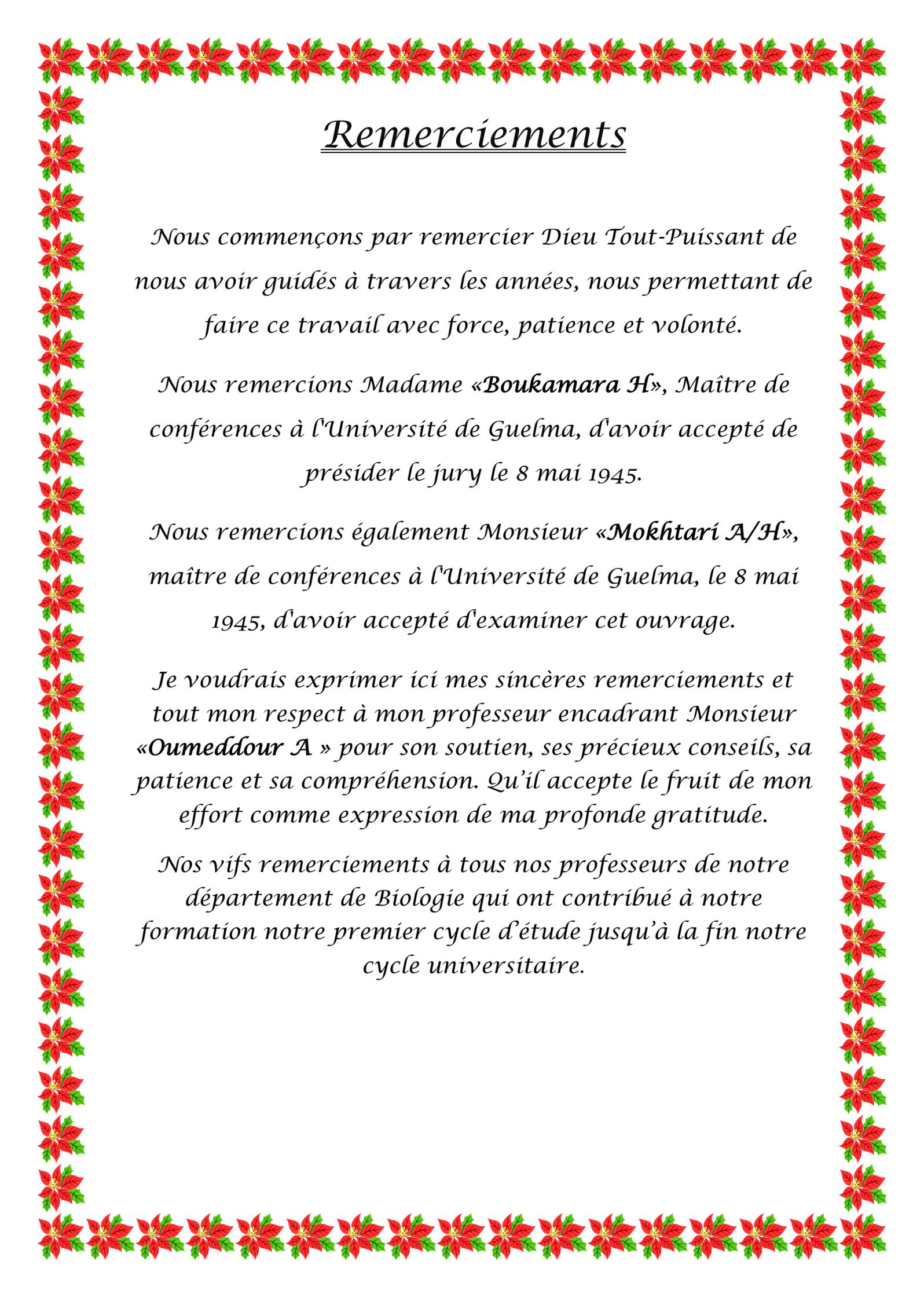
**Thème : Analyse de l'expression différentielle des gènes dans le cancer
colorectal : Etude bioinformatique**

Présenté par : Mme. Benchabane Bouthaina – Melle. Boutabet Nada – M. Moussaoui Ziad.

Devant le jury composé de :

Président (e) : Dr. BOUKAMARA H.	M.C.B	Université de Guelma
Examineur : Dr. MOKHTARI A/H.	M.C.B	Université de Guelma
Encadreur : Dr. OUMEDDOUR A.	M.C.A	Université de Guelma

Juin 2022



Remerciements

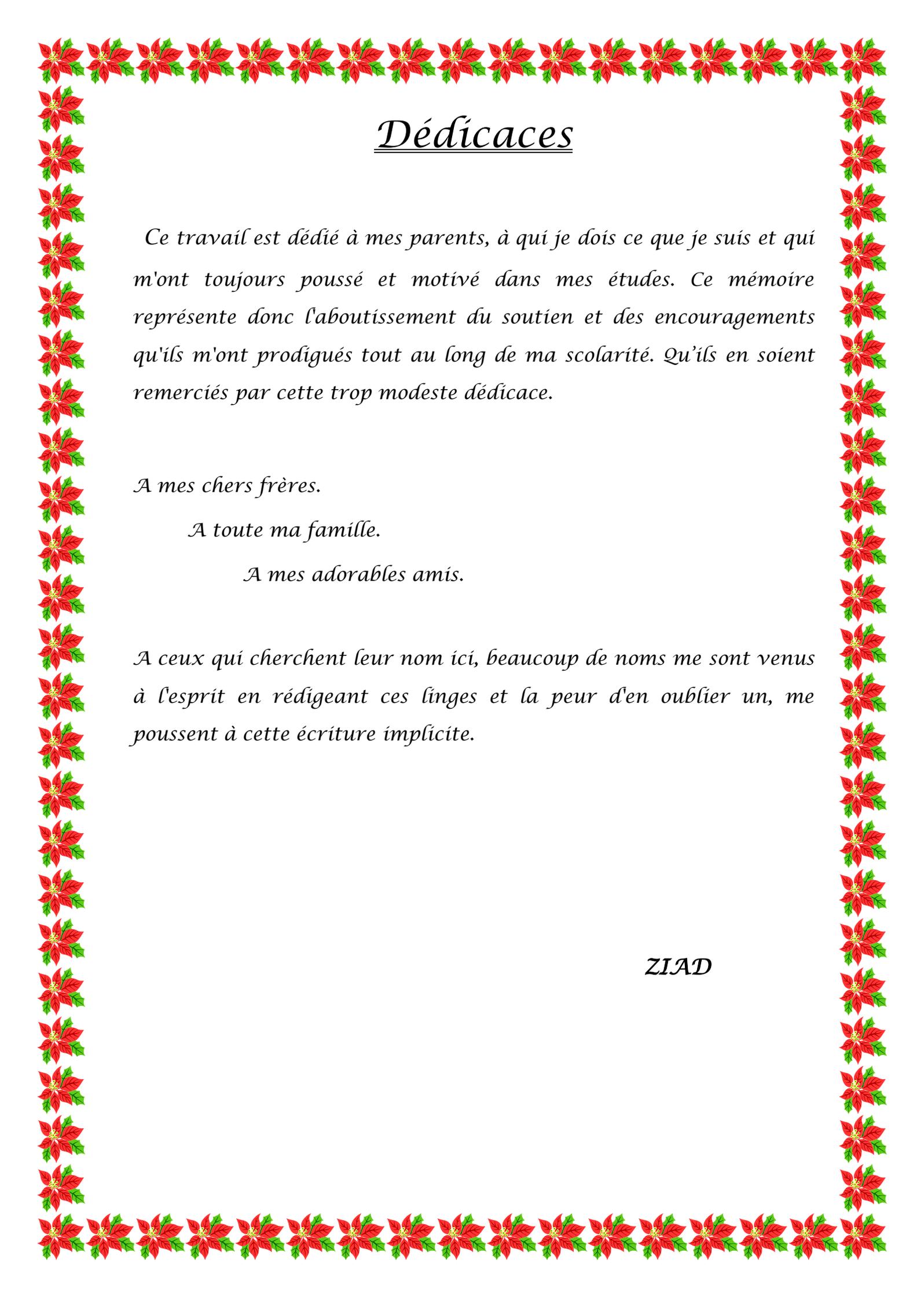
Nous commençons par remercier Dieu Tout-Puissant de nous avoir guidés à travers les années, nous permettant de faire ce travail avec force, patience et volonté.

*Nous remercions Madame «**Boukamara H**», Maître de conférences à l'Université de Guelma, d'avoir accepté de présider le jury le 8 mai 1945.*

*Nous remercions également Monsieur «**Mokhtari A/H**», maître de conférences à l'Université de Guelma, le 8 mai 1945, d'avoir accepté d'examiner cet ouvrage.*

*Je voudrais exprimer ici mes sincères remerciements et tout mon respect à mon professeur encadrant Monsieur «**Oumeddour A** » pour son soutien, ses précieux conseils, sa patience et sa compréhension. Qu'il accepte le fruit de mon effort comme expression de ma profonde gratitude.*

Nos vifs remerciements à tous nos professeurs de notre département de Biologie qui ont contribué à notre formation notre premier cycle d'étude jusqu'à la fin notre cycle universitaire.



Dédicaces

Ce travail est dédié à mes parents, à qui je dois ce que je suis et qui m'ont toujours poussé et motivé dans mes études. Ce mémoire représente donc l'aboutissement du soutien et des encouragements qu'ils m'ont prodigués tout au long de ma scolarité. Qu'ils en soient remerciés par cette trop modeste dédicace.

A mes chers frères.

A toute ma famille.

A mes adorables amis.

A ceux qui cherchent leur nom ici, beaucoup de noms me sont venus à l'esprit en rédigeant ces lignes et la peur d'en oublier un, me poussent à cette écriture implicite.

ZIAD



Dédicaces

Au Bon Dieu

El hamdoulillah pour m'avoir donné la force et la patience afin de réaliser ce modeste travail que je dédie particulièrement à :

À mes chers parents « Mahmoud et Samia » pour vos sacrifices pour mon bonheur, votre encouragement, votre prière tout au long de ma vie, votre amour et votre soutien vers le chemin de la réussite et que dieu vous garde pour moi en bonne santé « je vous aime »

A mon cher frère et mes chères sœurs « Mohamed, Leïla, Meriem, Omayma et Amina », pour votre intérêt et votre encouragement

A mes petits « Djana et Kousay »

À mon cher Mehdi merci pour ton soutien, ton intérêt et tes encouragements

A toute la famille grande et petite

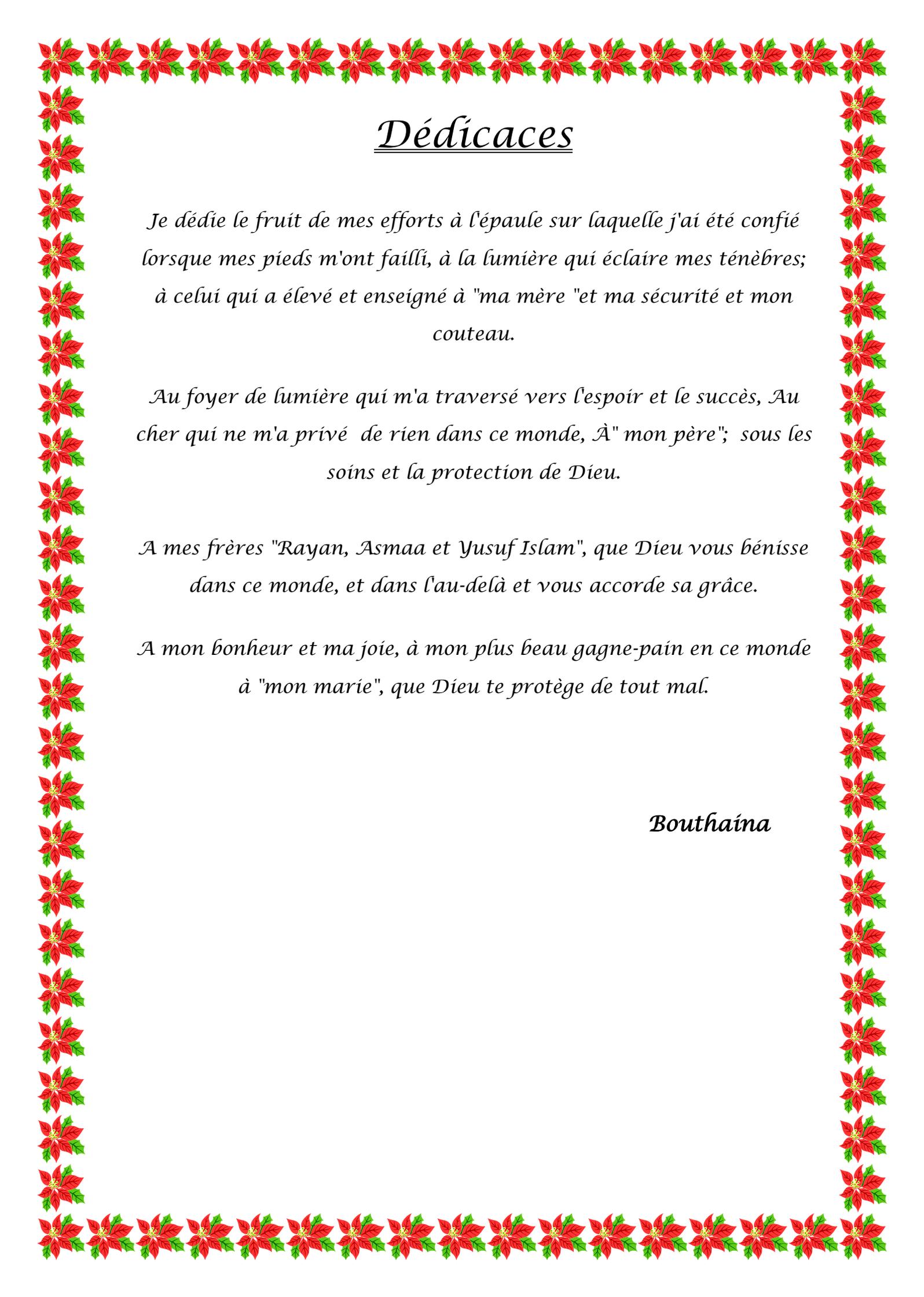
A mon intime Imene Qui a toujours était à mes côtés, merci pour votre aide et votre encouragement. Merci d'être le meilleur ami du monde

A mes chères amis« Amina, Karima, Bassma, Marwa et Malika »

A mon trinome « Bouthaina et Ziad » pour leurs efforts et pour leur patience tout au long l'année

A tous ceux qui, par un mot, m'ont donné la force de continuer

Nada



Dédicaces

Je dédie le fruit de mes efforts à l'épaule sur laquelle j'ai été confié lorsque mes pieds m'ont failli, à la lumière qui éclaire mes ténèbres; à celui qui a élevé et enseigné à "ma mère "et ma sécurité et mon couteau.

Au foyer de lumière qui m'a traversé vers l'espoir et le succès, Au cher qui ne m'a privé de rien dans ce monde, À " mon père"; sous les soins et la protection de Dieu.

A mes frères "Rayan, Asmaa et Yusuf Islam", que Dieu vous bénisse dans ce monde, et dans l'au-delà et vous accorde sa grâce.

A mon bonheur et ma joie, à mon plus beau gagne-pain en ce monde à "mon marie", que Dieu te protège de tout mal.

Bouthaina

Résumé

Le but de notre étude était d'identifier les gènes clés communs liés au cancer colorectal et à son diagnostic. Des données de trois profils ont été téléchargées à partir de la base de données GEO, puis analysées en utilisant plusieurs outils bioinformatiques. Les gènes à expression différentiels communs (DEGs) ont été déterminés par la base de données Venn diagram. L'analyse *Gene Ontology* (GO) et l'analyse *Kyoto Encyclopedia of Genes and Genomes* (KEGG) ont été réalisées à l'aide du site DAVID, le réseau d'interaction protéine-protéine (PPI) a été construit à l'aide du site STRING et visualisé par le logiciel Cytoscape. En fin, la survie totale et l'expression des gènes clés ont été vérifiées à l'aide de la base de données GEPIA2.

Nous avons obtenu 222 DEGs, dont 47 gènes surexprimés et 175 gènes sous-exprimés. L'analyse du GO et KEGG ont montré que les gènes étaient enrichis dans de nombreux processus biologiques, cellulaires et moléculaires associés au cancer colorectal. En outre, 10 gènes clés ont été trouvés, dont trois ont été identifiés associés à la mauvaise survie totale des patients et exprimés de manière significative dans les tissus cancéreux par rapport aux tissus normaux.

En conclusion, les trois gènes (ZG16-AQP8-CXCL8) peuvent être un bon bio-indicateur pour le diagnostic et la thérapie génique ciblée du cancer colorectal.

Mots clés: cancer colorectal, bioinformatique, gènes à expression génique différentielle, développement du cancer, polypes, adénocarcinome.

Abstract

The aim of our study was to identify common key genes related to colorectal cancer and its diagnosis. Data from three profiles were downloaded from the GEO database and then analyzed using several bioinformatics tools. Common differentially expression genes (DEGs) were determined by the Venn diagram database. Gene Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analysis were performed using the site web (DAVID). Protein-protein interaction network (PPI) was constructed using the STRING database and visualized by Cytoscape software. Finally, total survival and expression of key genes were verified using the GEPIA2 database.

A total of 222 DEGs were identified, including 47 *upregulated genes* and 175 down-regulated *genes*. Analysis of GO and KEGG showed that genes were enriched in many biological, cellular and molecular processes associated with colorectal cancer. In addition, 10 key genes were found, three of which were identified to be associated with poor overall patient survival and significantly expressed in cancerous tissues compared to normal tissues.

In conclusion, the three genes (ZG16-AQP8-CXCL8) can be a good bio-indicator for the diagnosis and targeted gene therapy of colorectal cancer.

Keywords: colorectal cancer, bioinformatics, genes with differential gene expression, cancer development, polyps, adenocarcinoma.

ملخص

كان الهدف من دراستنا هو تحديد الجينات الرئيسية الشائعة المتعلقة بسرطان القولون والمستقيم وتشخيصه. تم تنزيل البيانات من ثلاثة ملفات تعريف من قاعدة بيانات GEO ثم تحليلها باستخدام العديد من أدوات المعلوماتية الحيوية. تم تحديد جينات التعبير التفاضلي الشائعة (DEGs) بواسطة قاعدة بيانات مخطط Venn diagram. تم إجراء تحليل الإثراء الوظيفي GO والتحليل KEGG باستخدام الموقع DAVID ، وتم إنشاء شبكة تفاعل البروتين-البروتين (PPI) باستخدام الموقع STRING وتصور بواسطة برنامج Cytoscape. أخيرًا ، تم التحقق من البقاء الكلي والتعبير عن الجينات الرئيسية باستخدام قاعدة بيانات GEPIA2.

حصلنا على 222 DEGs ، بما في ذلك 47 جينة مفرطة التعبير و 175 جينة ناقصة التعبير. أظهر تحليل GO و KEGG أنه تم إثراء الجينات في العديد من العمليات البيولوجية والخلوية والجزئية المرتبطة بسرطان القولون والمستقيم. بالإضافة إلى ذلك ، تم العثور على 10 جينات رئيسية ، تم تحديد ثلاثة منها على أنها مرتبطة بضعف بقاء المريض بشكل عام وتم التعبير عنها بشكل كبير في الأنسجة السرطانية مقارنة بالأنسجة الطبيعية.

في الختام ، يمكن أن تكون الجينات الثلاثة (ZG16-AQP8-CXCL8) مؤشرًا بيولوجيًا جيدًا للتشخيص والعلاج الجيني المستهدف لسرطان القولون والمستقيم.

الكلمات المفتاحية: سرطان القولون والمستقيم ، المعلوماتية الحيوية ، الجينات ذات التعبير الجيني التفاضلي ، تطور السرطان ، الأورام الحميدة ، السرطان الغدي.

Sommaire

Remerciements

Dédicaces

Résumé

Abstract

ملخص

Liste des Figures

Liste des Tableaux

Liste des abréviations

Introduction 1

I. Etude bibliographique

1. Le cancer colorectal 3

1.1. Rappels structuraux du colon et du rectum 3

1.1.1. Colon 3

a. Le cæcum 4

b. Le colon ascendant (droit) 4

c. Le colon transverse 4

d. Le colon descendant (gauche) 4

e. Le colon sigmoïde (pelvien) 4

f. Rectum 5

1.1.2. Les parois du colon et du rectum 5

1.2. Définition du cancer colorectal 6

1.3. Développement du cancer colorectal 7

1.3.1. Polypes 7

1.3.2. Progression du polype au cancer : changements morphologiques 8

1.3.3. Progression du polype au cancer : changements génétique 9

1.4. Epidémiologie 11

1.4.1. Dans le monde 11

1.4.2. En Algérie 11

1.5. Les facteurs de risque 12

1.6. Les traitements du cancer colorectal 13

1.6.1. La chirurgie.....	13
1.6.2. La radio thérapie.....	13
1.6.3. La chimiothérapie.....	14
1.6.4. La thérapie ciblée	14
1.7. Les stades du cancer colorectal.....	14
1.7.1. Le systèmes tumeur, ganglion lymphatique, métastase(TNM)	15
1.7.2. La classification Duke.....	17
1.7.3. La classification d'astler-coller	17
II. Partie pratique	
2. Méthodes	19
2.1. L'objectif de travail.....	19
2.2. Méthodes d'analyse	19
2.2.1. Source de données et informations sur les puces à ADN	19
2.2.2. Traitement des données et identification des DEG	19
2.2.3. Analyse de l'ontologie des gènes et de l'enrichissement des voies	19
2.2.4. Analyse du réseau et des modules PPI	20
2.2.5. Sélection des gènes clés	20
2.2.6. Analyse de survie globale et validation de l'expression des gènes clés	20
3. Résultats et discussion	21
3.1. Résultats.....	21
3.1.1. Traitement des données et identification des DEG	21
3.1.2. Analyse de l'ontologie des gènes et de l'enrichissement des voies	24
3.1.3. Analyse du Kyoto Encyclopedia of Genes and Genomes (KEGG).....	27
3.1.4. Analyse de l'interaction Protéine-Protéine	28
3.1.5. Sélection des gènes clés	29
3.1.6. Analyse de survie globale et validation de l'expression des gènes clés	30
3.2. Discussion.....	32
Conclusion	38
Références bibliographiques	
Glossaire	

Liste de Figure

Figure 01 : Les différents segments du colon	3
Figure 02 : Les différents segments du rectum	5
Figure 03 : Les couches du colon et du rectum	6
Figure 04 : Schéma simplifié de l'évolution d'un cancer colorectal à partir d'un polype	9
Figure 05 : Méthodes de développement et de dépistage du CCR, (A) Modifications histologiques du cancer colorectal, (B) modifications génétiques acquises du CCR, (C) des tests de dépistage efficaces pour le CCR	10
Figure 06 : Classification TNM et stades UICC des carcinomes coliques. Exemple illustré des carcinomes coliques. Version modifiée d'après Johns Hopkins médecine colorectal cancer	15
Figure 07 : Les tracés volcaniques (<i>Volcano plots</i>) des trois ensembles de données GEO, le bleu représente les gènes sous-exprimés, le rouge représente les gènes surexprimés, ($P < 0:05$ et $\log FC > 1$).	21
Figure 08 : DEG communs aux trois ensembles de données GEO. (A) 47 gènes surexprimés. (B) 175 gènes sous-exprimés.	22
Figure 09 : Le réseau PPI, les gènes surexprimés en rouge, les gènes sous-exprimés en bleu.....	28
Figure 10 : Les dix principaux gènes clés classés par les algorithmes topologiques <i>Maximal Clique Centrality</i> (MCC) et Degree.....	29
Figure 11 : Les cinq gènes exprimés de manière significativement différentielle entre les tissus cancéreux et les tissus normaux, les tissus cancéreux en rouge, les tissus normaux en gris. (COAD : <i>Colon adenocarcinoma</i>).....	30
Figure 12 : La vérification de la survie globale des trois gènes (AQP8, CXCL8, ZG16). 31	

Liste des tableaux

Tableau 01 : Facteurs de risque du cancer colorectal	12
Tableau 02 : Les stades du cancer colorectal	16
Tableau 03: Selon la classification d'astler-coller des cancers du côlon et du rectum	18
Tableau 04 : Analyse des DEG courants dans le cancer colorectal	23
Tableau 05 : Analyse GO des DEG courants dans le cancer colorectal.	25
Tableau 06 : Analyse KEGG des DEG courants dans le cancer colorectal.....	27

Liste des abréviations

ADN: Acide Désoxyribonucléique.

AJCC: *American Joint Committee on Cancer.*

APC: *Adenomatous Polyposis Coli.*

AQP8: *Aquaporin 8.*

BB: Bordure en brosse.

BP: *Biological Process.*

BRAF: *V-raf murine Sarcoma Viral oncogene Homolog B.*

cAMP: *Cyclic Adenosine Monophosphate.*

CC: *Cellular Component.*

CCR: Cancer Colorectal.

CD44: *Cluster of Differentiation.*

CDx2: *Caudal-type homeobox transcription factor 2.*

CIRC: Centre international de recherche sur le cancer.

CLCA4: *Chloride channel accessory.*

COAD: *Colon Adenocarcinoma.*

CSC: Cellules Souches Cancéreuses.

CXCL8: *C-X-C Motif chemokine Ligand 8.*

DAVID: *Database for Annotation Visualization and Integrated Discovery.*

DEG: *Differentially Expressed Genes.*

EMT: Epithéliale Mésoenchymateuse.

FAS: *Fas Receptor.*

FASL: *Fas Ligand.*

FC: *Fold change.*

GEO: *Gene Expression Omnibus.*

GEPIA2: *Gene Expression Profiling Interaction Analysis.*

GI: Gastro Intestinal.

GO: *Gene Ontology.*

GPL: *Gene Platform.*

GSE: *Gene Sample Expression.*

GUCA2A: *Guanylate Cyclase-Activating factor 2A.*

GUCA2B: *Guanylate Cyclase Activator 2 B.*

HNPCC: *Hereditary Nonpolyposis Colorectal Cancer.*

HR: *Hazard Ratio.*

KEGG: *Kyoto Encyclopedia of Genes and Genomes.*

KRAS: *Kirsten Ras.*

MCC: *Maximal Clinique Centrality.*

MCODE: *Molecular Complex Detection.*

MF: *Molecular Function.*

MLH 1: *Homologue Mut « L ».*

MS4A12: *Membrane Spanning 4-domains A12.*

MSH 2: *Homologue Mut « S ».*

NF- κ B: *Nuclear Factor Kappa-lightchain-enhancer of activated B cells.*

OS: *Overall Survival.*

P: *Valeur de probabilité.*

P53: Protéine 53.

PAF: Polypose Adénomateuse Familiale.

PI3K/aKT: *Phosphatidylinositol-3-Kinase / Protein Kinase B.*

PMS 2: Segregation Post-méiotique 2.

PPI: *Protein-Protein Interaction.*

PSS: Polypes dentelés sessiles.

PTPRC: *Protein Tyrosine Receptor Type C.*

SLC26A3: *Solute Carrier family 26 member 3.*

SSA: *Sessile Serrated Adenomas.*

SSP: *Sessile Serrated Polyps.*

Tj: *Tight junction.*

TNF- α : *Tumor Necrosis Factor Alpha.*

TNM: Tumeur, Ganglion Lymphatique, Métastase.

TSA: Traditional Serrated Adenoma.

UICC: *Union for International Cancer Control.*

ZG16: *Zymogen Granule Protein 16.*

miR-92a : *specific microRNAs-92a.*

Introduction

Introduction

Le cancer colorectal (CCR) est l'un des cancers les plus étudiés dans les pays modernes et le plus courant de toutes les tumeurs malignes gastro-intestinales (GI). Malgré la sensibilisation aux facteurs de risque et aux mécanismes pathogènes, et la détection précoce de cette maladie, elle reste la troisième cause de décès par cancer [1]. En 2020, le cancer colorectal représentait 10 % de l'incidence mondiale du cancer et 9,4 % des décès par cancer, un peu moins que le cancer du poumon qui représente 18 % des décès. Les hommes sont en effet plus touchés que les femmes, avec un âge moyen de diagnostic entre 65 et 70 ans [2]. Les premiers cas de CCR apparaissent dès l'âge de 25 ans, avec une augmentation quasi anormale de l'incidence, qui ne se stabilise qu'à l'âge de 75-80 ans [3]. Le nombre mondial de nouveaux cas de cancer colorectal devrait atteindre 3,2 millions en 2040, sur la base des projections du vieillissement, de la croissance démographique et du développement humain [2].

Comme d'autres tumeurs malignes, le CCR est considéré comme une maladie polyphasée à différents stades de l'initiation, de la progression et de la métastase de la tumeur, qui s'accompagne de nombreux changements, notamment des mutations génétiques, des changements épi-génétiques, des contextes cellulaires et des influences environnementales [4].

Certaines études se sont concentrées sur la compréhension du développement du cancer en identifiant les différences dans l'expression des gènes entre les stades normaux et bénins de l'adénome et du cancer. Un point de départ commun pour ces études est la résection chirurgicale des tissus cancéreux et normaux des patients, suivie d'un profilage d'expression global pour déterminer les gènes exprimés de manière différentielle. Ces études peuvent aboutir à des dizaines, voire des milliers de ces gènes, dont seule une petite partie peut réellement avoir une utilité clinique [5].

Notre étude s'appuie sur l'utilisation de plusieurs outils bioinformatiques ainsi que sur des bases de données telles que *Gene Expression Omnibus* (GEO) pour extraire et analyser un ensemble de données et identifier les gènes communs à expression différentielle (DEG) associés au cancer colorectal. Elle permet également l'analyse d'enrichissement des gènes

(*Gene Ontology* : GO), ainsi que l'analyse de la *Kyoto Gene and Genomic Encyclopedia (KEGG)* afin d'étudier leur implication dans le développement du cancer colorectal. Le réseau d'interaction protéine-protéine (*protein-protein interaction* : PPI) a été construit par STRING et visualisé par le Cytoscape, puis le *plugin* Cytohubba a été utilisé pour déterminer les gènes clés associés au cancer colorectal.

Enfin, ces gènes ont été vérifiés à l'aide de la base de données GEPIA2 pour déterminer les chances de survie (*Overall Survival* : OS) et leur expression chez les patients atteints du cancer colorectal.

Notre travail sera divisé en deux parties:

- ✓ Une partie théorique sera consacrée au cancer colorectal : Rappels structuraux du colon et du rectum, développement du cancer colorectal, épidémiologie, facteurs de risque, traitements et les stades.
- ✓ Une partie pratique sera consacrée à l'analyse bioinformatique des données de patients atteints de cancer colorectal.

Etude bibliographique

1. Le cancer colorectal

1.1. Rappels structuraux du colon et du rectum

Pour comprendre le cancer colorectal il est utile de connaître la structure du colon et du rectum.

1.1.1. Colon

Le colon est l'organe principal du tractus gastro-intestinal, il joue un rôle vital dans la digestion des aliments, l'absorption d'eau et d'électrolytes, l'absorption des nutriments, la propulsion du contenu intestinal et finalement l'expulsion des déchets [6]. Aussi est un tube musculaire d'environ 1,5 mètre (5 pieds) de long et 5 centimètres (2 pouces) de diamètre [7].

Le colon commence au cæcum et se termine par le rectum, il comprend six sections (Figure 01) [6] :

- ✓ Le cæcum.
- ✓ Le colon ascendant (droit).
- ✓ Le colon transverse.
- ✓ Le colon descendant (gauche).
- ✓ Le colon sigmoïde (pelvien).
- ✓ Le rectum.

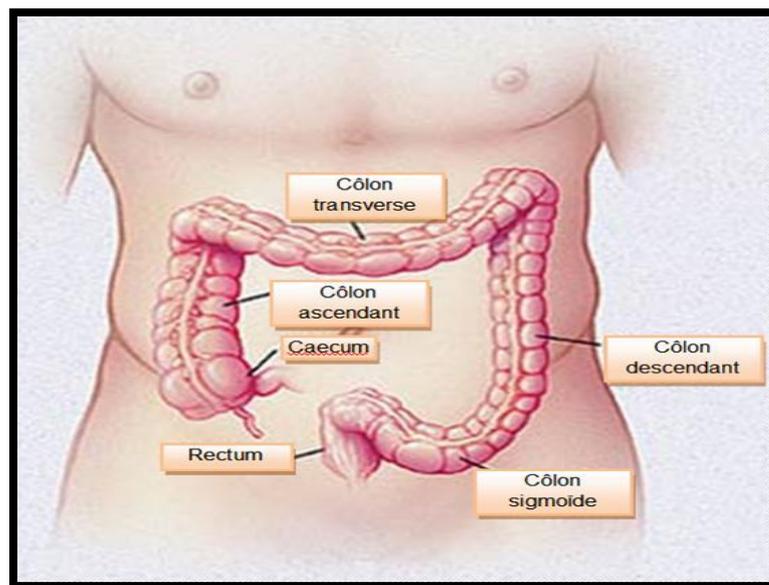


Figure 01: Les différents segments du colon [8].

a) Le cæcum

Le cæcum partie initial du côlon, il est situé en dessous de l'abouchement iléal [9], mesure environ 6 cm de long et s'étend vers le bas dans l'appendice, un sac tubulaire sinueux contenant du tissu lymphoïde. **L'appendice** est considéré comme le vestige d'un organe redondant; sa forme étroite et torsadée en fait un site attractif pour l'accumulation et la multiplication des bactéries intestinales [10].

b) Le colon ascendant (droit)

Le côlon ascendant est une continuation du cæcum, celui-ci s'étend vers le haut le long du côté latéral de la paroi abdominale vers le quadrant supérieur droit de l'abdomen. Il mesure environ 15 cm de long et est recouvert sur ses faces antérieure et latérale par du péritoine (donc considéré comme rétro péritonéal). Une fois qu'il a atteint la surface inférieure du lobe droit du foie, il tourne pour former le côlon transverse [6].

c) Le colon transverse

Le côlon transverse est le segment le plus long du côlon avec des longueurs mesurées de 40 à 50 cm. Il se situe en avant de l'estomac et de l'intestin grêle. Il est entièrement recouvert de mésentère appelé méso-côlon transverse. C'est le segment le plus mobile du côlon [11].

d) Le colon descendant (gauche)

Après la flexion colique gauche, le côlon se déplace vers le bas vers le bassin et s'appelle le côlon descendant. Il est rétro-péritonéal chez la majorité des individus, mais il est situé en avant du rein gauche, passant sur son bord latéral et lorsque commence à tourner médialement, il devient le côlon sigmoïde [12].

e) Le colon sigmoïde (pelvien)

Le côlon sigmoïde, il s'agit d'un segment du côlon en forme de S, de longueur variable très tortueuse et mobile [11]. Le côlon sigmoïde est attaché à la paroi pelvienne postérieure par un mésentère « le méso-côlon sigmoïde ». La grande longueur du mésentère permet à cette partie du côlon d'être particulièrement mobile [12].

f) Rectum

Le rectum est le dernier segment du gros intestin qui occupe le bassin postérieur le long de la concavité de l'os sacré. Cet organe peut mesurer de 12 à 18 cm de long[13]. Le rectum suit le côlon sigmoïde et débute à la charnière recto sigmoïde et se termine au canal anal. Il est divisé en rectum supérieur, moyen et inférieur (Figure 02) [14].

La longueur du rectum peut varier et dépend de sa distension, et de la quantité de fixation méso-rectale. C'est un organe avec une capacité d'accommodation accrue et une capacité à changer de taille [13]. Il se compose de deux parties totalement différentes[15]:

- ✓ Le rectum pelvien ou ampoule du rectum.
- ✓ Le rectum périnéal ou canal anal.

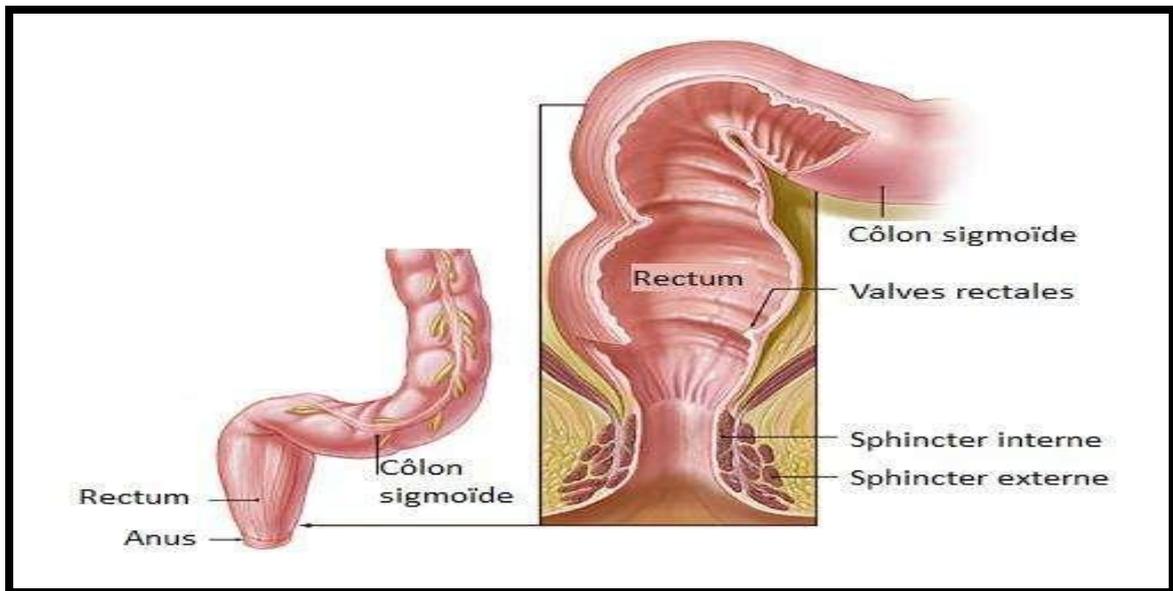


Figure 02 : Les différents segments du rectum [16].

1.1.2. Les parois du colon et du rectum

La paroi du colon et du rectum est composée des quatre couches (Figure 03) [11]:

- ✓ **Muqueuse :** La muqueuse est la couche la plus interne du côlon. Il est en outre subdivisé en trois sous-couches distinctes : l'épithélium glandulaire, la lamina propria et la musculature muqueuse, la sous-couche la plus externe. Comparée à l'intestin grêle, la muqueuse du côlon est dépourvue de villosités.

- ✓ **Sous-muqueuse** : a une riche alimentation vasculaire et un drainage lymphatique, et en raison de sa proximité avec la muqueuse, elle favorise son absorption d'électrolytes, de fluides et de nutriments.
- ✓ **Musculaire externe** : est composée de cellules musculaires lisses, disposés en deux couches distinctes : la couche longitudinale externe et la couche circulaire interne, Entre les deux couches il existe un plexus de fibres nerveuses de cellules ganglionnées, appelé plexus d'auerbach.
- ✓ **Séreuse** : La séreuse, prolongement du péritoine viscéral, est la couche la plus externe du côlon.

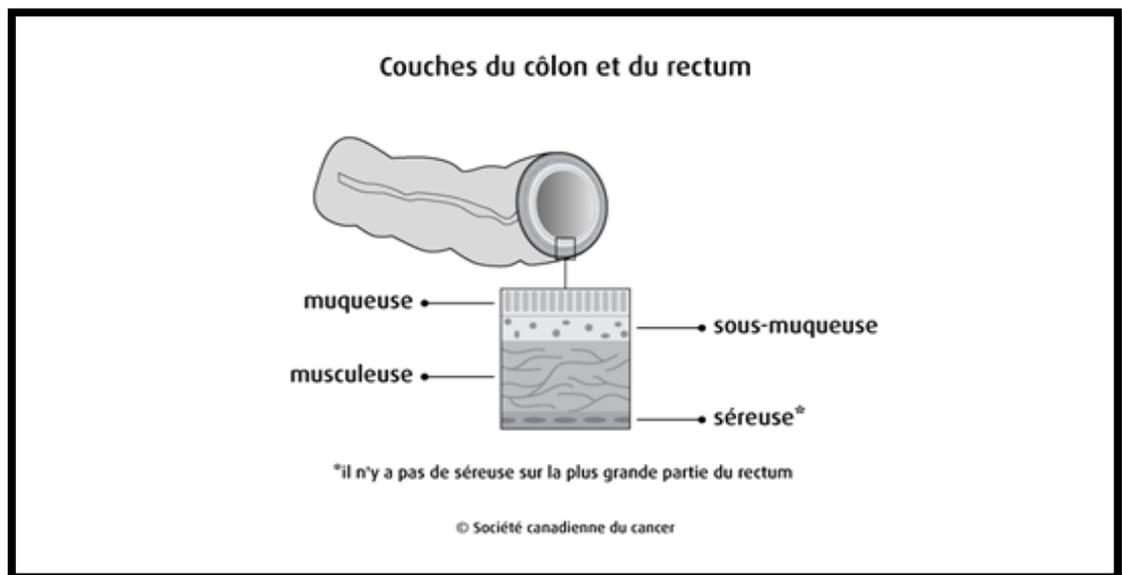


Figure 03: Les couches du colon et du rectum [17].

1.2. Définition du cancer colorectal

Les cancers du côlon et du rectum (cancer colorectal) commencent lorsque le processus de remplacement normal des cellules de la muqueuse tourne mal. Des erreurs dans la division cellulaire des muqueuses sont fréquentes. Pour des raisons mal comprises, il arrive parfois que des erreurs se produisent qui échappent à nos systèmes. Lorsque cela se produit, ces cellules commencent à se diviser indépendamment des freins et contreponds normaux qui contrôlent la croissance. Au fur et à mesure que ces cellules anormales se développent et se divisent, elles peuvent entraîner des excroissances dans le côlon appelées polypes [18].

Les polypes varient en type, mais beaucoup sont des tumeurs précancéreuses qui se développent lentement au fil des ans et ne se propagent pas. Au fur et à mesure que les polypes se développent, des mutations génétiques supplémentaires déstabilisent davantage les cellules et peuvent rendre les cellules plus bizarres. Lorsque ces tumeurs précancéreuses changent de direction (se développant à travers le tube plutôt qu'au milieu de celui-ci) et envahissent d'autres couches du gros intestin (telles que la sous-muqueuse ou la couche musculaire), le polype précancéreux devient cancéreux. Dans la plupart des cas, ce processus est lent, prenant au moins 8 à 10 ans pour passer de ces premières cellules aberrantes à un cancer franc [18].

1.3. Développement du cancer colorectal

Le CCR se développe souvent lentement et ne produit généralement pas de symptômes jusqu'à ce qu'il atteigne une taille considérable de plusieurs centimètres, ce qui peut bloquer le passage des matières fécales et entraîner des crampes, des douleurs ou des saignements qui peuvent se présenter sous la forme de saignements visibles lors des selles ou rarement de sombres selles « goudronneuses ». Lorsque le cancer se développe, les cellules cancéreuses sont d'abord peu nombreuses et confinées à la première couche de la paroi du côlon, la muqueuse. Au fil du temps et en l'absence de traitement, la tumeur s'étend plus profondément dans la paroi à travers d'autres couches. C'est ce qu'on appelle le carcinome invasif [19].

Des cellules cancéreuses peuvent également se détacher de la tumeur et emprunter les vaisseaux lymphatiques ou sanguins pour envahir d'autres parties du corps [20] :

- Les ganglions lymphatiques proches du côlon.
- Le foie, les poumons, le péritoine, le cerveau ou les os.

La plupart des tumeurs du côlon se développent via un processus en plusieurs étapes impliquant une série de modifications histologiques, morphologiques et génétiques qui s'accumulent avec le temps [19]. Les nouvelles tumeurs qui se forment alors s'appellent des métastases[20].

1.3.1. Polypes

Polype est un terme dérivé du mot grec *polypes*, qui signifie « masse morbide ». Généralement, ce terme décrit toute masse faisant saillie dans la lumière d'un vaisseau

creux, n'importe où dans les voies gastro-intestinales, génito-urinaires ou respiratoires. Habituellement, les polypes proviennent de la couche muqueuse de ces organes et ne présentent pas nécessairement un comportement néoplasique. Les polypes colorectaux peuvent également être histologiquement classés comme: néoplasiques ou non néoplasiques.

Les polypes néoplasiques, sont de première importance car ils recèlent un potentiel malin, ce qui représente une étape dans le développement du cancer colorectal, et bien que la plupart des polypes néoplasiques n'évoluent pas en cancer, il est bien admis que la majorité des carcinomes colorectaux évoluent à partir de polypes adénomateux; la séquence d'événements conduisant à cette transformation est appelée séquence adénome-carcinome [21].

Les polypes non néoplasiques, sont le type le plus courant de polypes colorectaux. On pense aujourd'hui qu'il s'agit d'un sous-groupe de polypes dentelés qui ne partagent pas le potentiel malin des autres sous-groupes les SSA et les TSA. Ces polypes ont généralement un diamètre inférieur à 5 mm (polypes diminutifs) sessiles et se trouvent le plus souvent côlon et rectum. Ils présentent des caractéristiques histologiques d'hyperplasie sans dysplasie; pour cette raison, ils ne sont pas considérés comme pré-malignes. Malheureusement, les polypes hyperplasiques ne peuvent pas toujours être distingués des polypes adénomateux à l'endoscopie et ils sont donc souvent retirés [21].

1.3.2. Progression du polype au cancer : changements morphologiques

Au fur et à mesure que les cellules du polype prolifèrent, la taille du polype augmente et les mutations génétiques et les changements épi-génétiques peuvent commencer à s'accumuler. Cela se traduit par une dysplasie cytologique et histologique. À mesure que les dommages à l'ADN cellulaire augmentent avec le temps, les caractéristiques d'une dysplasie de haut grade peuvent se développer, ce qui implique un risque très élevé de progression vers un carcinome invasif. S'ils ne sont pas retirés, ces polypes peuvent se développer la capacité d'envahir les tissus voisins et de se développer dans et au-delà de la paroi du côlon et du rectum [19].

Cette croissance maligne localisée peut devenir néo-vascularisée, offrant ainsi un accès facile aux systèmes lymphatique et circulatoire et favorisant ainsi la propagation des cellules cancéreuses vers des organes distants. La détection et la résection précoces des

polypes précancéreux sont essentielles pour interrompre la séquence adénome-carcinome et prévenir le développement et la propagation du CCR (Figure 04)[19].

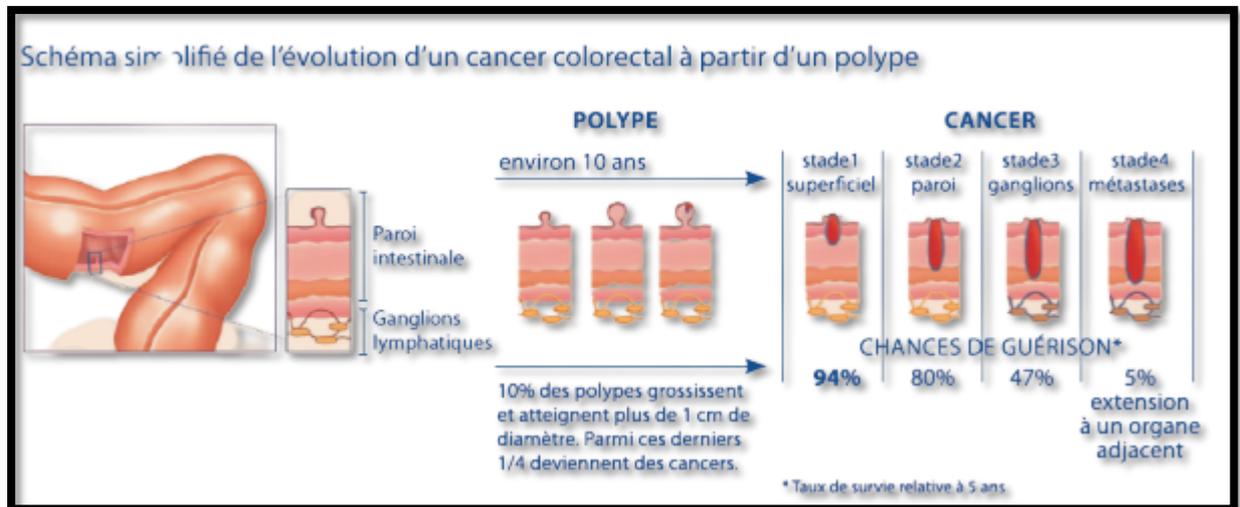


Figure 04 : Schéma simplifié de l'évolution d'un cancer colorectal à partir d'un polype [22].

1.3.3. Progression du polype au cancer : changements génétiques

La progression histologique du polype au cancer est le résultat d'une série et/ou d'une accumulation de modifications génétiques et épi-génétiques. Les mutations de l'ADN peuvent être acquises ou héritées. Les véritables mutations héréditaires associées au CCR telles que les mutations des gènes MLH1, MSH2, PMS2 et APC, sont rares et représentent environ 5 % des CCR. Cependant, l'étude de ces mutations héréditaires en plus des mutations sporadiques de réparation des mésappariements d'APC et d'ADN a fourni des informations clés sur la progression génétique par étapes des polypes précancéreux au cancer (Figure 05)[19].

Il existe deux principales voies génétiques qui conduisent au développement du CCR. Celles-ci correspondent généralement aux deux types de polypes à partir desquels le CCR se développe : les adénomes et les polypes dentelés sessiles (PSS). La voie de l'instabilité chromosomique généralement associée aux adénomes traditionnels est observée dans 65 à 70 % de tous les cancers sporadiques et se caractérise par une cascade de mutations qui s'accumulent. En règle générale, les premières mutations qui se développent se situent dans

le gène APC, qui affecte la ségrégation des chromosomes lors de la division cellulaire. Des mutations ultérieures se développent alors dans l'oncogène KRAS, qui a des effets en aval sur la croissance, la différenciation, la motilité et la survie des cellules. Au fil du temps, ces mutations peuvent entraîner une perte de fonction du gène p53, qui est un maître régulateur de la transcription et de l'apoptose, impactant ainsi un large éventail de fonctions cellulaires qui aboutit finalement à la carcinogenèse[19].

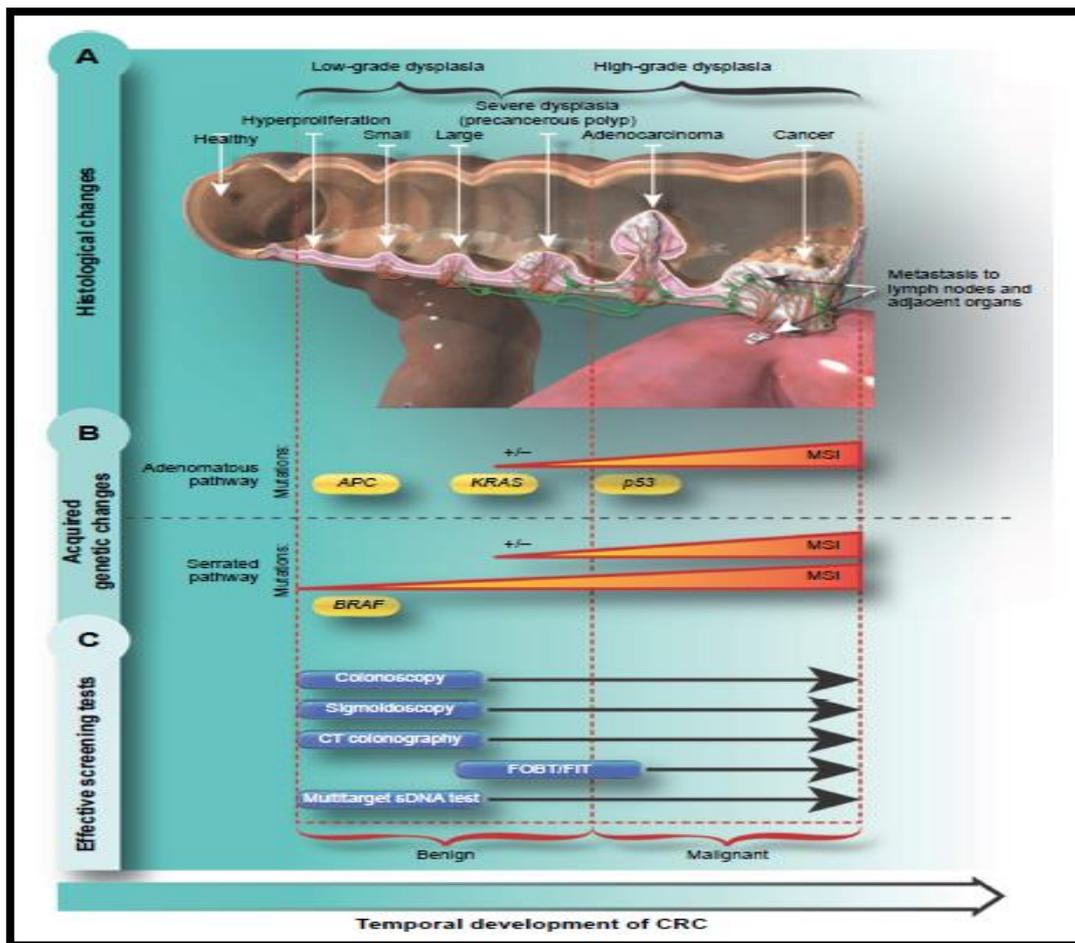


Figure 05: Méthodes de développement et de dépistage du CCR, (A) Modifications histologiques du cancer colorectal, (B) modifications génétiques acquises du CCR, (C) des tests de dépistage efficaces pour le CCR [19].

En revanche, le développement des SSP a tendance à commencer par des mutations du gène BRAF, ce qui entraîne une altération de la signalisation de croissance et une perte d'apoptose. Des mutations de KRAS peuvent également survenir dans les SSP, mais elles sont beaucoup moins fréquemment associées aux SSP que les polypes adénomateux [19].

1.4. Epidémiologie

1.4.1. Dans le monde

La prévalence du CCR a considérablement augmenté à un rythme alarmant dans le monde ces dernières années. On estime à 1,93 million le nombre de nouveaux cas diagnostiqués et 0,94 million de décès causés par le CCR en 2020. Le CCR est la troisième cause de décès liés au cancer chez les deux sexes dans le monde, avec 515 637 décès estimés chez les hommes et 419 536 décès chez les femmes en 2020, et selon les estimations de GLOBOCAN 2020, il y a 1,15 million de nouveaux cas de cancer du côlon, 0,7 million de nouveaux cas de cancer rectal et 50 000 nouveaux cas de cancer anal dans le monde. Avec des progrès continus, ces chiffres devraient passer à 1,92 million, 1,16 million et 78 000 en 2040 [2].

Le diagnostic aux différents stades cliniques du CCR peut expliquer en partie les différences marquées dans les taux de survie, donc le diagnostic à des stades avancés est l'un des déterminants de la disparité de survie et d'un grand nombre de décès par CCR dans le monde. Ainsi, des programmes de dépistage basés sur la population ont été largement proposés et mis en œuvre dans certains pays hautement développés depuis plus de 10 ans [2].

1.4.2. En Algérie

L'Algérie a enregistré 43 920 nouveaux cas de cancer en 2017, soit une incidence de 112,2 nouveaux cas pour 100 000 habitants (93,7 nouveaux cas pour 100 000 chez les hommes et 131,7 nouveaux cas pour 100 000 chez les femmes). Le cancer colorectal est le cancer le plus fréquent chez les hommes, suivi du cancer du poumon (13,5 nouveaux cas pour 100 000 hommes), du cancer de la prostate (11,1 nouveaux cas/100 000 hommes), du cancer de la vessie (10,16 nouveaux cas/100 000 hommes) et du cancer de l'estomac (8,4 nouveaux cas/100 000 hommes). L'âge d'apparition est respectivement de 57, 62, 72, 65 et 60 ans [23].

Chez la femme, le cancer le plus fréquent est le cancer du sein avec une incidence de 49,3 nouveaux cas/100 000 femmes, suivi du cancer colorectal (12,9 nouveaux cas/100 000 femmes), du cancer de la thyroïde (9,3 nouveaux cas/100 000 femmes), du col de l'utérus (7,2 nouveaux cas/100 000 femmes). Et le cancer de l'estomac (5,9 nouveaux cas/100 000 femmes). L'âge médian d'apparition chez les femmes est de 47, 65, 42, 62 et 65 ans [23].

1.5. Les facteurs de risque

Contrairement à d'autres cancers comme le cancer du poumon, aucun facteur de risque n'est à lui seul responsable de la plupart des cas de cancer colorectal. Outre l'âge et le sexe masculin, les facteurs de risque suivants : antécédents familiaux de cancer colorectal, maladie inflammatoire de l'intestin, tabagisme, consommation excessive d'alcool, forte consommation de viande rouge et transformée, obésité, et diabète. L'augmentation du risque est la plus forte pour les personnes ayant des parents au premier degré atteints d'un cancer colorectal (Tableau 01) [24].

Tableau 01: Facteurs de risque du cancer colorectal [18].

les facteurs	les risques
Âge avancé	✓ Les données internationales indiquent que le risque de développer un cancer colorectal augmente avec l'âge. La plupart des cas de cancer colorectal surviennent chez des personnes âgées de 50 ans ou plus.
Antécédents familiaux	✓ Antécédents familiaux de syndromes de cancer colorectal héréditaire, tels que la polypose adénomateuse familiale (PAF) ou le cancer colorectal héréditaire sans polypose (HNPCC ou syndrome de Lynch).
Viande	✓ Manger beaucoup de viande rouge, en particulier de viande transformée, augmente le risque de cancer colorectal.
Alcool	✓ L'alcool a été classé comme cancérigène du groupe 1 par le Centre international de recherche sur le cancer (CIRC) et doit être évité pour réduire le risque de cancer colorectal.
Tabac	✓ Le tabagisme est une cause bien connue de cancer du poumon, mais certaines des substances cancérigènes contenues dans la fumée se dissolvent dans la salive et si elles sont avalées, peuvent provoquer des cancers du système digestif comme le cancer colorectal.

Le CCR peut être prévenu en minimisant l'exposition d'un individu à ces facteurs de risque en adoptant un mode de vie sain. Cela inclut d'éviter de consommer du tabac et de l'alcool, de maintenir un poids corporel sain, des activités physiques, une alimentation pauvre en viandes rouges et transformées et riche en fibres [25].

1.6. Les traitements du cancer colorectal

Il existe trois approches principales pour traiter le cancer : la chirurgie, la radiothérapie et la thérapie pharmacologique (y compris la chimiothérapie, la thérapie ciblée, l'hormonothérapie et l'immunothérapie). Certains patients reçoivent les trois modalités de traitement, tandis que d'autres reçoivent un ou deux types [26].

1.6.1. La chirurgie

La chirurgie est l'ablation de la tumeur et des tissus sains environnants au cours d'une opération. Il s'agit du traitement le plus courant du cancer colorectal et est souvent appelé résection chirurgicale. Une partie du côlon ou du rectum sain et des ganglions lymphatiques voisins sont également retirés [27].

Elle consiste à enlever la zone du côlon qui contient la tumeur et une marge de sécurité d'au moins 5 cm autour de la tumeur afin qu'aucune cellule cancéreuse ne reste. Dans certains cas très avancés, il est nécessaire d'enlever tout le côlon : c'est ce qu'on appelle une colectomie totale[28].

1.6.2. La radiothérapie

La radiothérapie stéréotaxique s'est développée rapidement. Il s'agit d'une technique qui utilise plusieurs minuscules faisceaux de lumière qui se croisent en un point spécifique (tumeur). Il peut donc délivrer de fortes doses dans un petit volume, mais la précision de ce traitement impose au patient de rester complètement immobile pendant le traitement. En conséquence, des doses plus élevées peuvent être délivrées sur plusieurs cycles, permettant la destruction complète de tumeurs qui seraient autrement inaccessibles. Par conséquent, il peut être utilisé pour traiter les métastases hépatiques, pulmonaires ou osseuses et même les récurrences locales [14].

1.6.3. La chimiothérapie

La chimiothérapie est l'utilisation de médicaments pour détruire les cellules cancéreuses, généralement en empêchant les cellules cancéreuses de se développer, de se diviser et de produire plus de cellules. La chimiothérapie est généralement administrée par un oncologue médical, un médecin spécialisé dans le traitement du cancer avec des médicaments. La chimiothérapie systémique se déplace dans la circulation sanguine pour atteindre les cellules cancéreuses dans tout le corps. Les moyens courants d'administrer la chimiothérapie comprennent un tube intraveineux (IV) placé dans une veine à l'aide d'une aiguille ou dans une pilule ou une capsule qui est avalée (par voie orale) [27].

1.6.4. La thérapie ciblée

En plus de la chimiothérapie, une nouvelle classe de médicaments fait son apparition depuis plusieurs années : la thérapie ciblée. Ces molécules sont conçues pour bloquer des mécanismes spécifiques par lesquels les cellules cancéreuses se développent ou prolifèrent. Ces traitements sont administrés en cures espacées d'une à deux semaines dans le traitement des tumeurs métastatiques, le plus souvent en complément de la chimiothérapie. Quatre médicaments sont utilisés [28]:

- Le bévacicumab (Avastin).
- Le panitumumab (Vectibix).
- Le cetuximab (Erbix).
- L'aflibercept (Zaltrap).

1.7. Les stades du cancer colorectal

Le stade du cancer est une évaluation de l'étendue du cancer dans le corps. Il est utilisé pour planifier les tests qui peuvent être nécessaires et le traitement qui vous convient le mieux. Si elles ne sont pas traitées, les cellules cancéreuses peuvent se développer à travers la muqueuse du côlon et dans l'abdomen. Le cancer peut alors envahir des structures ou des organes extérieurs au côlon. Les cellules cancéreuses peuvent également se détacher des tumeurs du côlon et se propager aux ganglions lymphatiques voisins par la lymphe ou le sang [29].

1.7.1. Le système tumeur, ganglion lymphatique, métastase (TNM)

Le système tumeur, ganglion lymphatique, métastase (TNM) de l'American Joint Committee on Cancer (AJCC) est utilisé pour stadifier le cancer du côlon. Dans le système AJCC, les informations clés suivantes sur le cancer sont utilisées pour attribuer les stades (Figure 06) [29] :

- **T** : Jusqu'où la tumeur s'est développée dans ou à travers la paroi du côlon.
- **N** : Si des ganglions lymphatiques ont un cancer en eux.
- **M** : Si le cancer s'est propagé à des zones ou à des organes en dehors du côlon (métastasé).

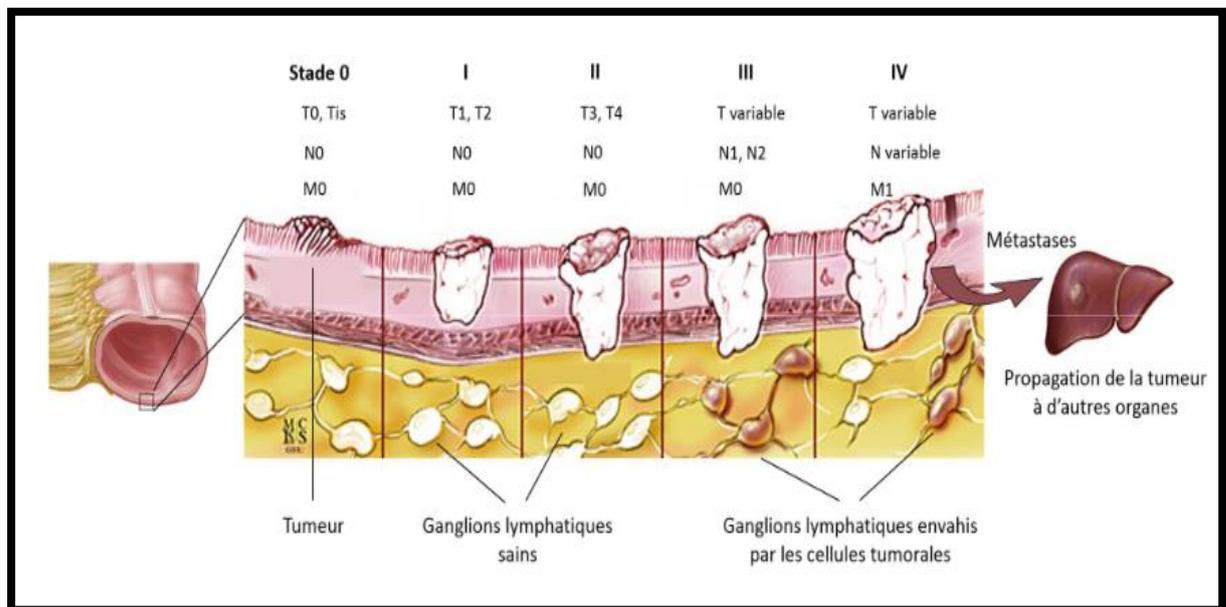


Figure 06 : Classification TNM et stades UICC des carcinomes coliques. Exemple illustré des carcinomes coliques. Version modifiée d'après Johns Hopkins médecine colorectal cancer[30].

Les scores T, N et M sont combinés pour attribuer un stade au cancer. Il existe cinq stades de cancer du côlon. Ils sont numérotés 0, I (1), II (2), III (3) ou IV (4) [29].

Les stades du cancer colorectal sont présentés dans le tableau 02.

Tableau 02: Les stades du cancer colorectal [28].

Stades	Description
0	✓ La tumeur n'affecte que la couche de mucus de la paroi intestinale.
1	✓ La tumeur s'est propagée aux couches sous-muqueuses et musculaires.
2	✓ Correspond au moment où la tumeur a traversé toutes les couches de la paroi intestinale sans toucher les autres organes.
3	✓ Le cancer s'est propagé aux ganglions lymphatiques voisins.
4	✓ Est un stade avancé de la maladie dans lequel les cellules cancéreuses se sont propagées dans des endroits éloignés pour former des métastases (foie, poumon, ... etc).

La classification du cancer colorectal détaillée selon la nomenclature TNM [31] :

- **Tumeur (T) :**

- **Tx** : Tumeur primitive non évaluable.
- **T0** : Pas de tumeur primitive.
- **Tis** : Envahissement de l'épithélium ou de la lamina propria.
- **T1** : Envahissement de la sous-muqueuse.
- **T2** : Envahissement de la musculature.
- **T3** : Envahissement de la sous-séreuse ou des tissus péricolo-rectaux non péritonisés.
- **T4** : Atteinte de la séreuse ou des structures adjacentes.
- **T4a** : Perforation du péritoine viscéral.
- **T4b** : Adhésion ou invasion d'un organe de voisinage.

- **Ganglion (N) :**

- **Nx** : Ganglions non évalués.
- **N0** : Pas de métastase ganglionnaire régionale.
- **N1** : 1 à 3 ganglions métastatiques régionaux.
- **N1c** : Nodule(s) tumoraux, c.à.d satellite(s) dans la sous-séreuse,

ou dans les tissus non péritonéales péri-coliques ou péri-rectaux sans métastase ganglionnaire régionale.

- **N2** : 4 ou plus ganglions métastatiques régionaux.
- **N2a** : 4 à 6 ganglions métastatiques.
- **Métastase (M)** :
 - **M0** : Pas de métastase.
 - **M1** : Métastases à distance.
 - **M1 a** : Métastase localisée à un seul organe (foie, poumon, ovaire, ganglion(s) lymphatique(s), autre que régional M1b Métastases dans plusieurs organes ou péritonéales.

1.7.2. La classification Duke

En 1932, le Dr Cuthbert E. Dukes, un pathologiste londonien, a décrit un système de classification pour la stadification du carcinome rectal.

Cette classification a été divisée en :

- **Stade A** : limité à la paroi rectale.
- **Stade B** : extension extra-rectale sans atteinte ganglionnaire régionale.
- **Stade C** : métastase ganglionnaire régionale.

Bien que Dukes n'ait jamais nommé officiellement un stade D, il a décrit une quatrième étape qui comprenait la propagation de la maladie à des organes distants. L'invasion des organes adjacents et les métastases mésentériques ou péritonéales sont généralement également considérées comme le stade D. Malgré que l'article original n'inclue pas les carcinomes du côlon, Dukes a affirmé que tous les carcinomes gastro-intestinaux pouvaient également être mis en scène de cette manière [32].

1.7.3. La classification d'astler-coller

Cette classification, également connue sous le nom de classification Dukes modifiée, divise le cancer colorectal en (Tableau 03) [33] :

Tableau 03: Selon la classification d'astler-coller des cancers du côlon et du rectum [33].

Stades	Description
Stade A	✓ Tumeur limitée à la muqueuse.
Stade B1	✓ Tumeur limitée à la sous-muqueuse, pas d'envahissement ganglionnaire.
Stade B2	✓ Tumeurs confinées à la couche musculaire, pas d'envahissement des ganglions lymphatiques.
Stade C1	✓ La tumeur n'a pas dépassé la paroi intestinale, métastase ganglionnaire.
Stade C2	✓ Tumeur dépassant la paroi intestinale et métastases ganglionnaires.

Méthodes

2. Méthodes

2.1. L'objectif de travail

L'objectif de ce travail est de l'analyse de l'expression différentielle des gènes associés au cancer colorectal à l'aide des outils bioinformatique intégrée.

2.2. Méthodes d'analyse

2.2.1. Source de données et informations sur les puces à ADN

Les données des puces à ADN utilisées dans cette étude ont été téléchargées à partir du référentiel public de données de génomique fonctionnelle : la base de données *Gene Expression Omnibus* (GEO) (<http://www.ncbi.nlm.nih.gov/gds/>). Trois profils d'expression génique (GSE4107, GSE8671, GSE113513) étudiant le CCR et des tissus normaux ont été choisis. Les données de matrice de GSE4107 consistaient en 12 tissus de CCR et 10 tissus normaux basés sur la plate-forme GPL570 ([HG-U133_Plus_2] *Affymetrix Human Genome U133 Plus 2.0 Array*). Les données de la matrice de GSE8671 referment 32 tissus de CCR et 32 tissus normaux basés sur la plate-forme de GPL570 ([HG-U133_Plus_2] *Affymetrix Human Genome U133 Plus 2.0 Array*). Les données de la matrice de GSE113513 comprennent 14 tissus CCR et 14 tissus sains, plate-forme : GPL15207 ([*PrimeView*] *Affymetrix Human Gene Expression Array*).

2.2.2. Traitement des données et identification des DEG

Les trois profils d'expression d'ARNm obtenus à partir de la base de données GEO ont été analysés par les outils en ligne GEO2R (<https://www.ncbi.nlm.nih.gov/geo/geo2r/>). Le $|\log_2\text{-fold change (FC)}| > 1$ et une valeur $P < 0,05$ ont été utilisés comme critères de sélection pour l'identification des gènes différentiellement exprimés (DEG). Ensuite, les données brutes au format TXT ont été traitées par l'outil de *Venn Diagram* (<http://bioinformatics.psb.ugent.be/webtools/Venn/>) pour obtenir les DEG communs entre les trois profils. Les DEG avec un $\log FC > 0$ ont été considérés comme des gènes surexprimés (*upregulated genes*), tandis que les DEG avec $\log FC < 0$ ont été considérés comme des gènes sous-exprimés (*down-regulated genes*).

2.2.3. Analyse de l'ontologie des gènes et de l'enrichissement des voies

L'analyse de l'ontologie des gènes (*Gene ontology*, GO) est une approche couramment utilisée dans les études fonctionnelles de données transcriptome ou de génome à haut débit.

L'analyse d'enrichissement des voies de GO (*Biological process* : BP, *Cellular component* : CC, et *Molecular function* : MF) et *Kyoto Encyclopedia of Genes and Genomes* (KEGG) a été réalisée à l'aide d'un outil bioinformatique en ligne, le logiciel DAVID (version 6.8, <https://david.ncifcrf.gov/>). $P < 0,05$ a été défini comme critère de sélection.

2.2.4. Analyse du réseau et des modules PPI

Pour explorer les interactions entre les DEG, nous avons utilisé une base de données en ligne, STRING (version 11.5, <https://string-db.org/>). La confiance élevée (0,400) a été choisie comme score d'interaction minimum requis. Ensuite, après avoir supprimé les nœuds déconnectés, un réseau PPI a été visualisé à l'aide du logiciel Cytoscape 3.9.0. De plus, le *plug-in Molecular Complex Detection* (MCODE) du logiciel Cytoscape a été utilisé pour cribler les modules du réseau PPI (seuil de degré = 2, seuil de score de nœud = 0,2, k-core = 2 et profondeur max = 100).

2.2.5. Sélection des gènes clés

Les gènes ont été extraits à l'aide du *plug-in cytoHubba* du logiciel Cytoscape. Ensuite, nous avons sélectionné les dix principaux gènes clés classés par les algorithmes topologiques *Maximal Clique Centrality* (MCC) et *Degree*.

2.2.6. Analyse de survie globale et validation de l'expression des gènes clés

L'outil *Gene Expression Profiling Interactive Analysis* GEPIA2 (<http://gepia2.cancer-pku.cn/>), a été utilisée pour évaluer l'effet de dix gènes clés sur la survie globale (*Overall survival*, OS) des patients atteints de CCR. Pour estimer le taux d'OS des patients atteints de CCR, les échantillons ont été divisés en groupes d'expression élevée et faible en fonction de l'expression médiane. Le rapport de risque (*Hazard ratio*, HR), les intervalles de confiance à 95 % correspondants et les valeurs P du *log-rank* ont été calculés et affichés sur le graphique.

En plus, nous avons utilisé la même outille pour analyser les données d'expression de séquençage d'ARN des patients atteints de CCR en comparaison avec des échantillons témoins ($|\text{Log}_2\text{FC}|$ Cutoff = 1 et p-valeur seuil : 0,05).

Résultats et discussion

3.1. Résultats

3.1.1. Traitement des données et identification des DEG

Les trois profils d'expression d'ARNm obtenus à partir de la base de données GEO ont été analysés par l'outil en ligne GEO2R avec une valeur $P < 0,05$ et $|\log_2\text{-fold change (FC)}| > 1$. Il y avait 4149 DEG dans l'ensemble de données GSE4107, dont 3133 gènes surexprimés et 1016 gènes sous-exprimés. 2042 DEG dans l'ensemble de données GSE8671, dont 519 gènes surexprimés et 1523 gènes sous-exprimés. 582 DEG dans l'ensemble de données GSE113513, dont 311 gènes surexprimés et 271 gènes sous-exprimés. Les tracés volcaniques (*Volcano Plot*) des trois ensembles de données GEO sont illustrés dans la figure 07.

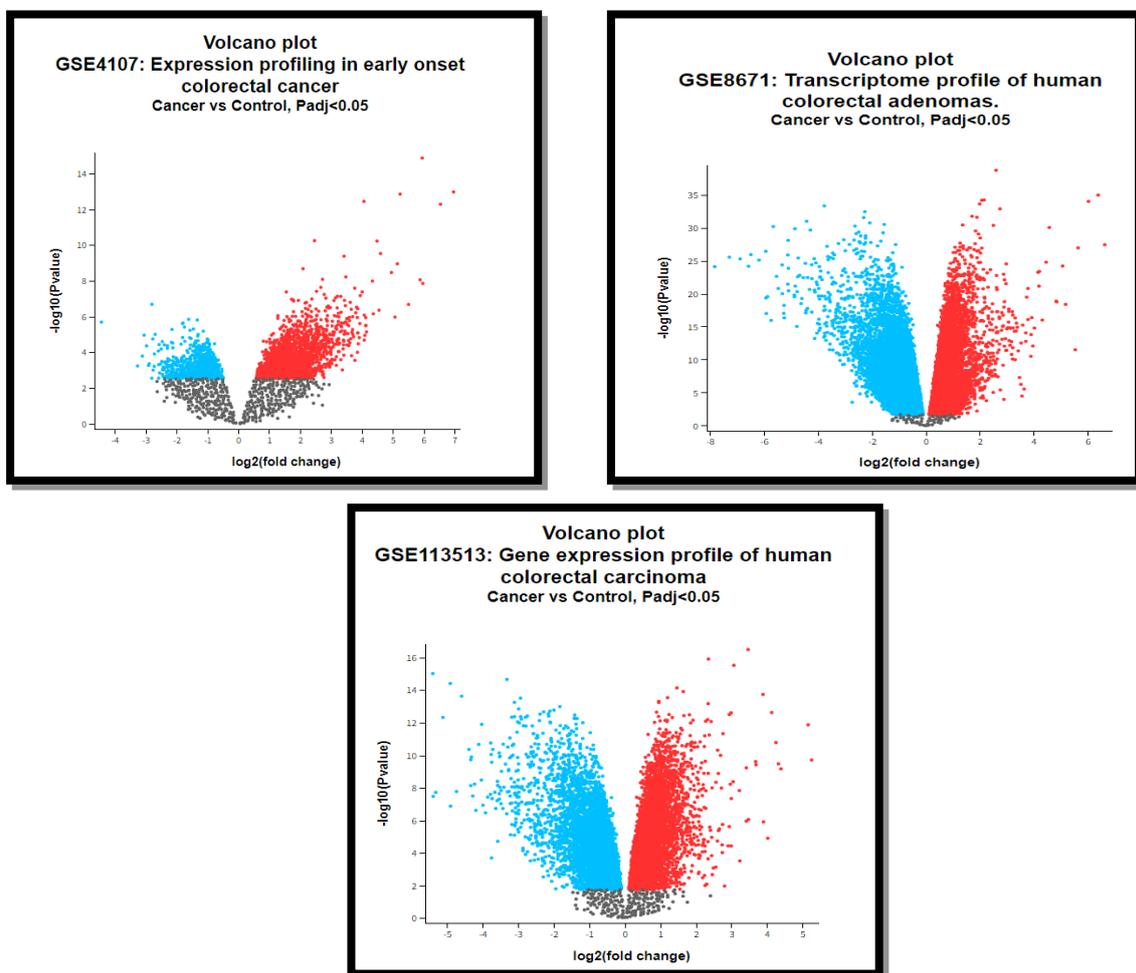


Figure 07: Les tracés volcaniques (*Volcano plots*) des trois ensembles de données GEO, le bleu représente les gènes sous-exprimés, le rouge représente les gènes surexprimés ($P < 0:05$ et $\log |\text{FC}| > 1$).

Nous avons utilisé la base de données *Venn Diagram* pour déterminer les gènes à expression différentiels commun (DEG) entre les trois profils. Comme le montre la figure 08, un effectif de 222 DEG communs ont été identifiés parmi les trois ensembles de données GEO, dont 47 gènes surexprimés et 175 gènes sous-exprimés.

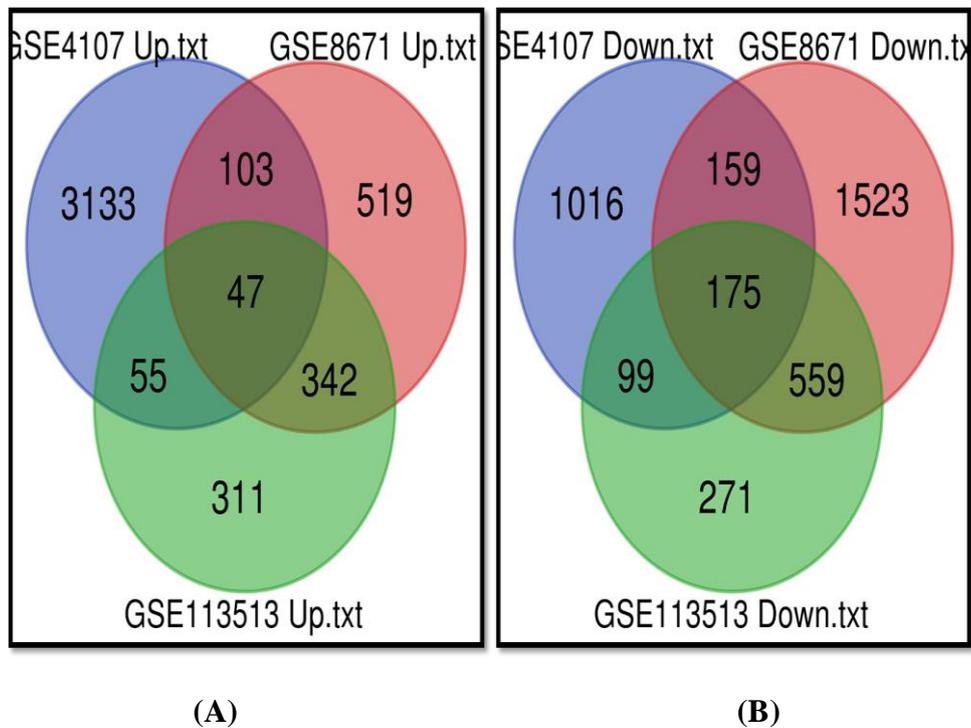


Figure 08 : DEG communs aux trois ensembles de données GEO. **(A)** 47 gènes surexprimés. **(B)** 175 gènes sous-exprimés.

La liste des DEGs communs surexprimés et sous-exprimés sont montrés dans le tableau 04.

Tableau 04 : Liste des DEGs communs dans le cancer colorectal.

DEG communes	symbol du gene
Sur-exprimés 47	CD44, GDPD5, DUSP4, DCUN1D5, PITX1, MCM10, PTP4A3, FXYD5, WDR74, LRP8, EVA1A, AJUBA, DUOXA2, ATP11A, ARNTL2, STC1, EIF5A2, TSPAN5, RRP12, IL33, KLK10, ACSL6, TESC, ZAK, RCL1, NKD1, PHLDA1, SULT1C2, NEBL, AURKA, EXOSC3, MEGF6, MTFR2, CLPB, PLEKHA8P1, CXCL8, CADPS, ZC3HAV1L, LRRC34, FAM92A1, TMEM158, CYP4X1, CAD, KIAA1549, APCDD1, SERPINE1, SSX2IP.
Sous-exprimés 175	PLCE1, XDH, PTPRR, HSD17B2, STYK1, ZG16, TMCC3, DHRS11, HSD11B2, TRPM6, CES2, PYY, MAOA, PDZD3, B3GALT5, ZZEF1, TSPAN7, HHLA2, AHCYL2, IL6R, PKIB, CDHR5, PBLD, SLC51B, AQP8, BRINP3, AOC1, LYPD8, CD177, HRCT1, SLC25A34, ADH1C, SGK2, SLC17A4, S100B, UGT2A3, GCNT2, LRRC19, CLIC5, CDKN2B, LAMA1, CWH43, MEP1A, MALL, PLOD2, CA12, GUCA2A, SLC22A5, PTPRC, SLC23A1, SPPL2A, CA4, APPL2, 44621, NAAA, KRT20, HSD3B2, SEMA6D, SLC26A2, VSTM2A, ISX, GDPD2, KIAA1211, GCNT3, ACACB, SLC46A3, MS4A12, SOWAHA, VSIG2, RIMKLA, HIST1H2BD, ABCC13, CEACAM1, C2orf88, HIST1H1C, CDA, SLC25A23, TEX11, SLC4A4, MEP1B, SI, TMEM37, PDE6A, GSTA1, MTM1, LDHD, ITLN1, TSPAN3, PHLPP2, TUBAL3, DEFB1, ARL14, RMDN2, PTGDR, SLC2A13, SRI, TFCP2L1, TNFSF10, MIER3, SCIN, SLC26A3, CDKN2B-AS1, TMEM171, SMPD1, SLC16A9, CLDN23, EDN3, HIGD1A, FGFR2, ITPKA, DDX60, MS4A8, PCK1, CPM, GALNT12, SMPDL3A, WDR78, ITM2C, TINAG, USP2, CLMN, HPGD, SCGB2A1, MOGAT2, TLR3, DHRS9, PNLIPRP2, CCDC68, ENTPD5, CNTN3, ABCG2, GUCA2B, CHP2, RORA, CAPN13, CLCA4, MYO1A, SLC36A1, NXPE4, SORBS1, IL1R2, GDPD3, AKAP9, SLC30A10, NR5A2, FMO5, PAQR5, CCL28, CA2, SLC41A2, ANPEP, CEACAM7, PADI2, MGAT4A, CAPN5, CMBL, TMEM220, CR2, NHSL1, AMN, GCG, GBA3, UGP2, CLDN8, EGLN3, ANK3, HEPACAM2, PLAC8, CA1, RETSAT, FXYD3, GPAT3, GRAMD3, HAPLN1, CDHR2 .

3.1.2. Analyse de l'ontologie des gènes et de l'enrichissement des voies

L'analyse d'enrichissement des voies de GO a été effectuée sur les DEG communs (les gènes surexprimés et sous-exprimés). Le critère de sélection pour l'analyse était une valeur $P < 0,05$. L'analyse comprenait *Biological process* (BP), *Cellular component* (CC), *Molecular function* (MF).

Dans les BP, les gènes surexprimés étaient principalement enrichis dans les voies suivantes : *protein maturation, positive regulation of cellular biosynthetic process, rRNA processing*. Les gènes sous-exprimés étaient enrichis dans les voies suivantes : *one-carbon metabolic process, epithelial cell differentiation, bicarbonate transport, transmembrane transport, sodium ion transport, cellular response to cAMP, prostaglandin metabolic process, intermicrovillar adhesion, positive regulation of extrinsic apoptotic signaling pathway, chloride transport, ion transport, excretion, positive regulation of Fc-gamma receptor signaling pathway, involved in phagocytosis, xenobiotic metabolic process, N-acylethanolamine metabolic process*.

Dans les CC, les gènes sur-exprimés étaient plus enrichis dans les voies suivantes : *nucleolus*. Les gènes sous-exprimés étaient plus enrichis dans les voies : *extracellular exosome, apical plasma membrane, integral component of membrane, brush border membrane, brush border, plasma membrane, microvillus membrane, extracellular region, anchored component of membrane, extracellular space, integral component of plasma membrane, spanning component of plasma membrane, endoplasmic reticulum, basolateral plasma membrane, meprin A complex, secretory granule membrane*.

Dans les MF, les gènes sur-exprimés étaient en grande partie enrichis dans les voies suivantes : *enzyme binding, protein binding*. Les gènes sous-exprimés étaient surtout enrichis dans les voies suivants: *carbonate dehydratase activity, hydro-lyase activity, zinc ion binding, acetylglucosaminyltransferase activity, receptor binding, phosphoric diester hydrolase activity, oxidoreductase activity, guanylate cyclase activator activity, flavin adenine dinucleotide binding, calcium ion binding, identical protein binding, N-acetyllactosaminide beta-1,6-N-acetylglucosaminyltransferase activity, protein homodimerization activity, steroid delta-isomerase activity, transmembrane transporter activity, transporter activity, arylesterase activity, primary amine oxidase activity*.

L'analyse GO des DEG commun est présentée dans le tableau 05.

Tableau 05 : Analyse du GO des DEG communs dans le cancer colorectal.

Catégorie	Terme	Nm	%	P-Value
Les gènes surexprimés				
GOTERM_BP_DIRECT	<i>protein maturation</i>	3	6.7	7.9E-4
GOTERM_BP_DIRECT	<i>positive regulation of cellular biosynthetic process</i>	2	4.4	6.3E-3
GOTERM_BP_DIRECT	<i>rRNA processing</i>	3	6.7	2.9E-2
GOTERM_CC_DIRECT	<i>Nucleolus</i>	8	17.8	2.4E-2
GOTERM_MF_DIRECT	<i>enzyme binding</i>	5	11.1	9.5E-3
GOTERM_MF_DIRECT	<i>protein binding</i>	34	75.6	3.9E-2
Les gènes sous-exprimés				
GOTERM_BP_DIRECT	<i>one-carbon metabolic process</i>	5	3.0	2.0E-4
GOTERM_BP_DIRECT	<i>epithelial cell differentiation</i>	6	3.6	9.3E-4
GOTERM_BP_DIRECT	<i>bicarbonate transport</i>	4	2.4	1.7E-3
GOTERM_BP_DIRECT	<i>transmembrane transport</i>	8	4.7	2.9E-3
GOTERM_BP_DIRECT	<i>sodium ion transport</i>	5	3.0	6.3E-3
GOTERM_BP_DIRECT	<i>cellular response to cAMP</i>	4	2.4	1.2E-2
GOTERM_BP_DIRECT	<i>prostaglandin metabolic process</i>	3	1.8	1.3E-2
GOTERM_BP_DIRECT	<i>intermicrovillar adhesion</i>	2	1.2	1.6E-2
GOTERM_BP_DIRECT	<i>positive regulation of extrinsic apoptotic signaling pathway</i>	3	1.8	2.0E-2
GOTERM_BP_DIRECT	<i>chloride transport</i>	4	2.4	2.1E-2
GOTERM_BP_DIRECT	<i>ion transport</i>	5	3.0	2.4E-2
GOTERM_BP_DIRECT	<i>Excretion</i>	3	1.8	2.6E-2
GOTERM_BP_DIRECT	<i>positive regulation of Fc-gamma receptor signaling pathway involved in phagocytosis</i>	2	1.2	3.2E-2
GOTERM_BP_DIRECT	<i>xenobiotic metabolic process</i>	4	2.4	3.5E-2

GOTERM_BP_DIRECT	<i>N-acylethanolamine metabolic process</i>	2	1.2	4.0E-2
GOTERM_CC_DIRECT	<i>extracellular exosome</i>	50	29.6	3.2E-11
GOTERM_CC_DIRECT	<i>apical plasma membrane</i>	19	11.2	1.3E-9
GOTERM_CC_DIRECT	<i>integral component of membrane</i>	79	46.7	8.7E-9
GOTERM_CC_DIRECT	<i>brush border membrane</i>	9	5.3	2.9E-8
GOTERM_CC_DIRECT	<i>brush border</i>	7	4.1	1.1E-5
GOTERM_CC_DIRECT	<i>plasma membrane</i>	61	36.1	4.4E-4
GOTERM_CC_DIRECT	<i>microvillus membrane</i>	4	2.4	1.3E-3
GOTERM_CC_DIRECT	<i>extracellular region</i>	31	18.3	1.8E-3
GOTERM_CC_DIRECT	<i>anchored component of membrane</i>	6	3.6	2.1E-3
GOTERM_CC_DIRECT	<i>extracellular space</i>	28	16.6	3.1E-3
GOTERM_CC_DIRECT	<i>integral component of plasma membrane</i>	23	13.6	3.4E-3
GOTERM_CC_DIRECT	<i>spanning component of plasma membrane</i>	3	1.8	5.6E-3
GOTERM_CC_DIRECT	<i>endoplasmic reticulum</i>	18	10.7	9.2E-3
GOTERM_CC_DIRECT	<i>basolateral plasma membrane</i>	7	4.1	1.3E-2
GOTERM_CC_DIRECT	<i>mepirin A complex</i>	2	1.2	1.6E-2
GOTERM_CC_DIRECT	<i>secretory granule membrane</i>	4	2.4	4.9E-2
GOTERM_MF_DIRECT	<i>carbonate dehydratase activity</i>	4	2.4	2.8E-4
GOTERM_MF_DIRECT	<i>hydro-lyase activity</i>	4	2.4	6.3E-4
GOTERM_MF_DIRECT	<i>zinc ion binding</i>	18	10.7	7.0E-4
GOTERM_MF_DIRECT	<i>acetylglucosaminyltransferase activity</i>	4	2.4	3.6E-3
GOTERM_MF_DIRECT	<i>receptor binding</i>	10	5.9	5.4E-3
GOTERM_MF_DIRECT	<i>phosphoric diester hydrolase activity</i>	3	1.8	1.0E-2
GOTERM_MF_DIRECT	<i>oxidoreductase activity</i>	7	4.1	1.1E-2
GOTERM_MF_DIRECT	<i>guanylate cyclase activator activity</i>	2	1.2	1.6E-2
GOTERM_MF_DIRECT	<i>flavin adenine dinucleotide binding</i>	4	2.4	1.7E-2
GOTERM_MF_DIRECT	<i>calcium ion binding</i>	13	7.7	1.9E-2
GOTERM_MF_DIRECT	<i>identical protein binding</i>	23	13.6	1.9E-2
GOTERM_MF_DIRECT	<i>N-acetyllactosaminide beta-1,6-N-acetylglucosaminyltransferase activity</i>	2	1.2	2.4E-2
GOTERM_MF_DIRECT	<i>protein homodimerization activity</i>	12	7.1	3.0E-2
GOTERM_MF_DIRECT	<i>transmembrane transporter activity</i>	5	3.0	4.3E-2
GOTERM_MF_DIRECT	<i>steroid delta-isomerase activity</i>	2	1.2	3.2E-2
GOTERM_MF_DIRECT	<i>transporter activity</i>	3	1.8	4.4E-2
GOTERM_MF_DIRECT	<i>arylesterase activity</i>	2	1.2	4.8E-2
GOTERM_MF_DIRECT	<i>primary amine oxidase activity</i>	2	1.2	4.8E-2

3.1.3. Analyse de *Kyoto Encyclopedia of Genes and Genomes (KEGG)*

Le critère de sélection de l'analyse KEGG était la valeur $P < 0,05$. Les résultats montrent treize voies de signalisation identifiées. Les voies de signalisation des DEGs communs étaient principalement enrichies dans : *Hippo signaling pathway*, pour les gènes surexprimés et *Metabolic pathways*, *Nitrogen metabolism*, *Proximal tubule bicarbonate reclamation*, *Bile secretion*, *Steroid hormone biosynthesis*, *Drug metabolism - cytochrome P450*, *Drug metabolism - other enzymes*, *Pyruvate metabolism*, *Pancreatic secretion*, *Retinol metabolism*, pour les gènes sous-exprimés.

L'analyse de la voie KEGG des DEG commun est présentée dans le tableau 06.

Tableau 06 : Analyse KEGG des DEG courants dans le cancer colorectal.

Catégorie	Terme	Nm	%	P-Value
Les gènes surexprimés				
KEGG_PATHWAY	<i>Hippo signaling pathway</i>	3	6.7	3.3E-2
Les gènes sous-exprimés				
KEGG_PATHWAY	<i>Metabolic pathways</i>	42	24.9	6.1E-9
KEGG_PATHWAY	<i>Nitrogen metabolism</i>	4	2.4	7.9E-4
KEGG_PATHWAY	<i>Proximal tubule bicarbonate reclamation</i>	4	2.4	2.0E-3
KEGG_PATHWAY	<i>Bile secretion</i>	6	3.6	3.0E-3
KEGG_PATHWAY	<i>Steroid hormone biosynthesis</i>	5	3.0	4.5E-3
KEGG_PATHWAY	<i>Drug metabolism - cytochrome P450</i>	5	3.0	8.1E-3
KEGG_PATHWAY	<i>Drug metabolism - other enzymes</i>	5	3.0	1.2E-2
KEGG_PATHWAY	<i>Pyruvate metabolism</i>	4	2.4	1.5E-2
KEGG_PATHWAY	<i>Pancreatic secretion</i>	5	3.0	2.6E-2
KEGG_PATHWAY	<i>Retinol metabolism</i>	4	2.4	3.9E-2

3.1.4. Analyse de l'interaction protéine-protéine

Pour mieux évaluer les interactions entre les DEG communs identifiés, nous avons construit un réseau PPI à l'aide du site web STRING. Le réseau PPI de DEG communs est illustré dans la figure 09. Un total de 192 nœuds (*nodes*) et 221 bords (*edges*) ont été sélectionnés pour tracer un réseau PPI composé de 41 gènes surexprimés et 175 gènes sous-exprimés. Ensuite, par l'utilisation le plug-in MCODE du Cytoscape (seuil de degré = 2, seuil de score de nœud = 0,2, k-core = 2 et profondeur max. = 100), nous avons identifié un ensemble de 10 gènes (ZG16, CLCA4, MS4A12, GUCA2B, GUCA2A, SLC26A3, AQP8, PTPRC, CD44, CXCL8).

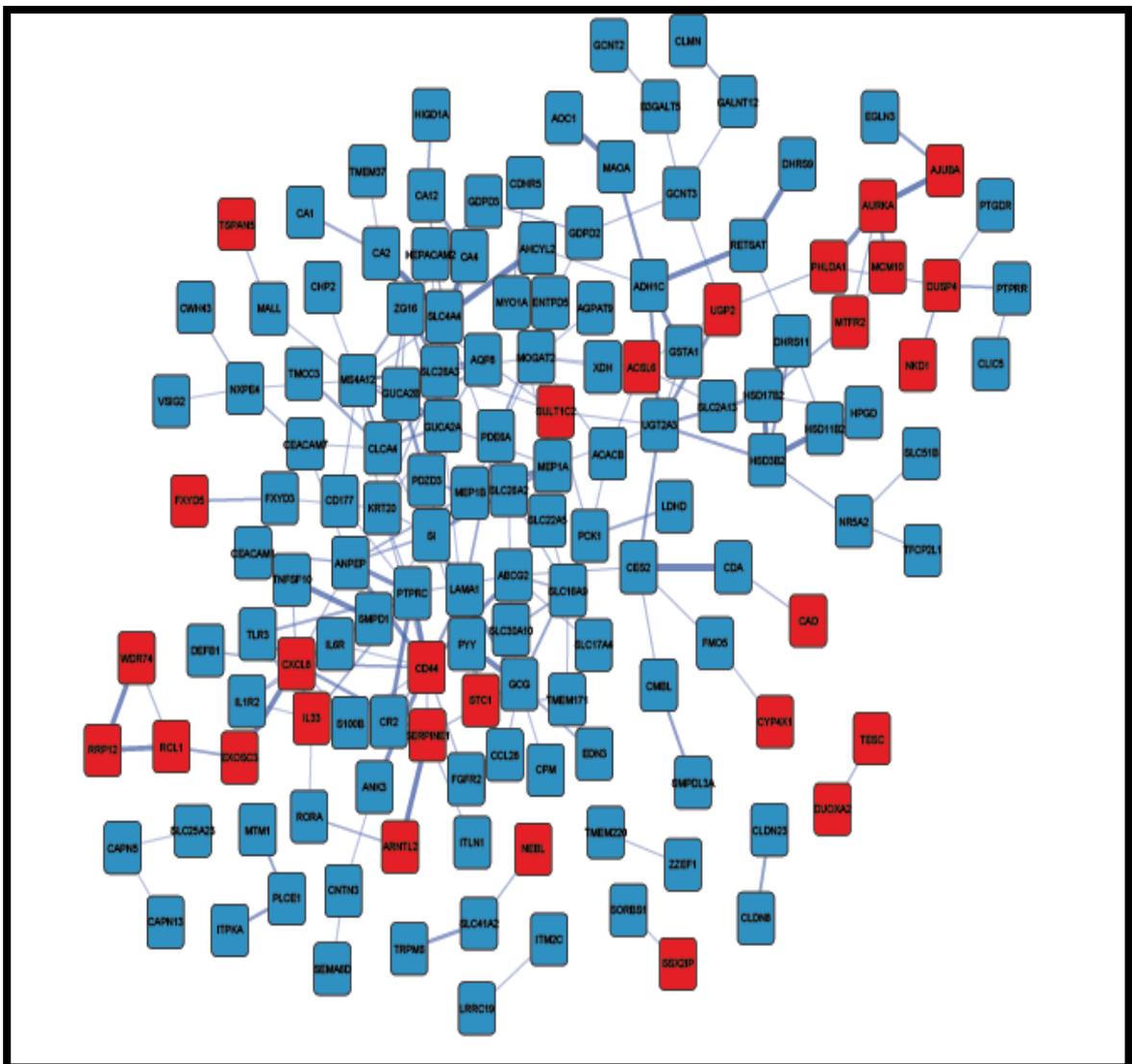


Figure 09 : Le réseau PPI, les gènes surexprimés en rouge, les gènes sous-exprimés en bleu.

3.1.5. Sélection des gènes clés

A partir cytoHubba, Nous avons trouvé deux unités fonctionnels du réseau PPI. L'une contenait 7 gènes : ZG16, CLCA4, MS4A12, GUCA2B et GUCA2A, SLC26A3, AQP8. L'autre contenait 3 gènes : PTPRC, CD44, CXCL8. Ces dix gènes sont considérés comme les gènes clés associés au cancer colorectal (figure 10).

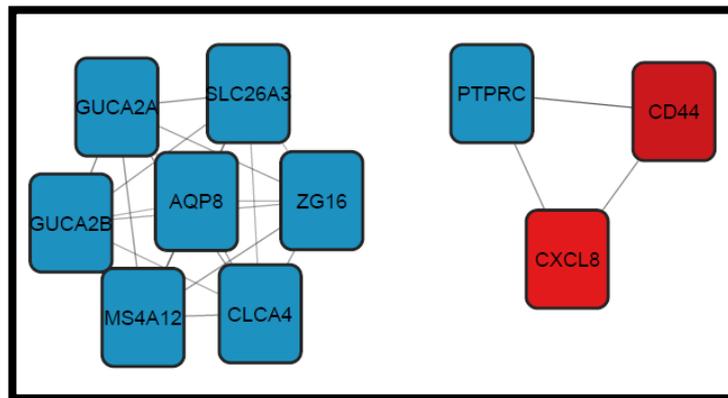


Figure 10 : Les dix principaux gènes clés classés par les algorithmes topologiques *Maximal Clique Centrality (MCC)* et *Degree*.

3.1.6. Analyse de survie globale et validation de l'expression des gènes clés

Les résultats obtenus sur la base de données GEPIA2 ont montré que les gènes ZG16, CXCL8, AQP8 (figure 11), étaient significativement exprimés de manière différentielle entre les tissus tumoraux et les tissus normaux.

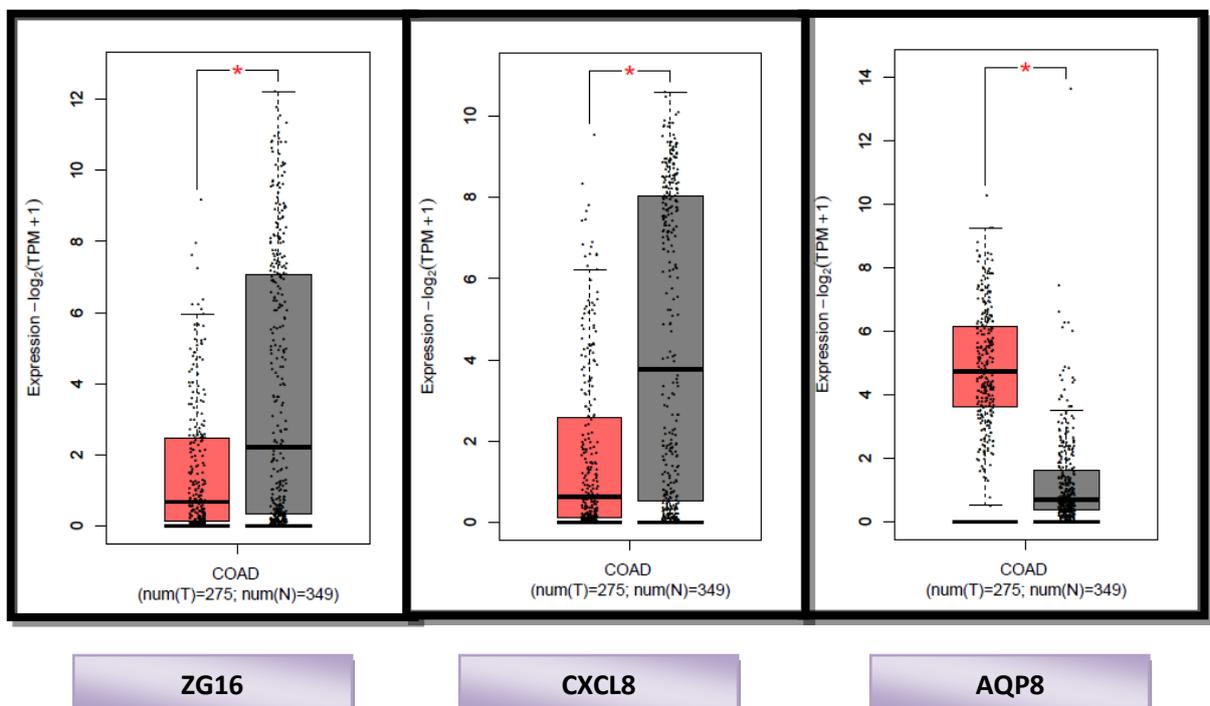


Figure 11 : Les trois gènes exprimés de manière significative entre les tissus cancéreux et les tissus normaux. Les tissus cancéreux en rouge, les tissus normaux en gris. (COAD : *Colon adenocarcinoma*).

L'analyse de la survie totale des dix gènes (Figure 12) montre que les niveaux d'expression de ces gènes (AQP8, CXCL8, ZG16) sont associés à la mauvaise survie des patients atteints du cancer colorectal.

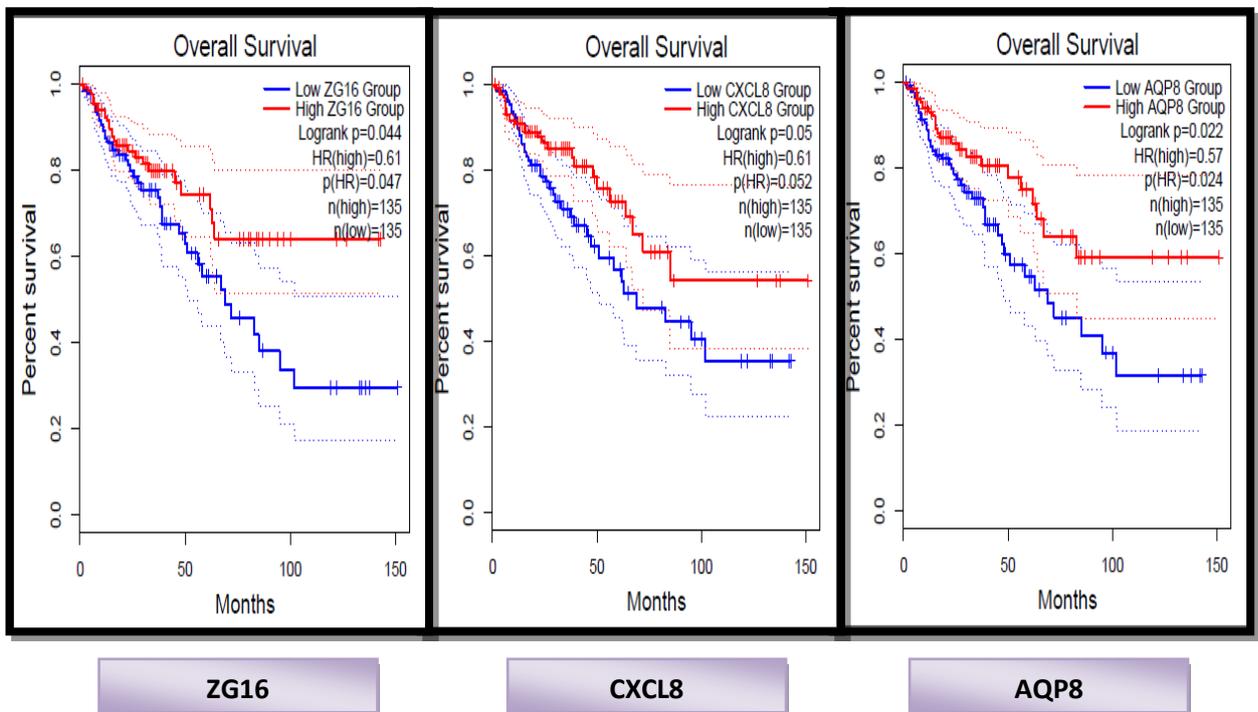


Figure 12 : La vérification de la survie globale des trois gènes clés.

3.2. Discussion

Le CCR est l'une des tumeurs malignes courantes du tube digestif dans le monde. Les facteurs alimentaires et environnementaux, ainsi que les mutations génétiques sont les principales causes du CCR [34].

Les principales stratégies de traitement du cancer colorectal sont la chirurgie, la radiothérapie, la chimiothérapie adjuvant et la pharmacothérapie à ciblage moléculaire. Malgré l'existence de nombreux traitements du cancer colorectal, le taux de survie à 5 ans du cancer colorectal reste inférieur à 40 % en raison des récives et des métastases. Par conséquent, l'identification de nouveaux bio-marqueurs pour comprendre les changements moléculaires à l'origine du cancer colorectal est essentielle pour mettre en place des stratégies efficaces de diagnostic précoce et de nouvelles thérapies ciblées [35].

Dans ce travail, nous avons analysé et identifié l'expression différentielle des gènes clés associés au cancer colorectal par l'utilisation des outils et méthodes bioinformatiques. D'abord, 222 DEG communs ont été identifiés parmi les trois ensembles de données GEO, dont 47 gènes surexprimés et 175 gènes sous-exprimés. L'analyse GO sur les DEG communs selon les BP a montré que les gènes surexprimés étaient plus enrichis en *protein maturation*, *positive regulation of cellular biosynthetic process*, *rRNA processing*. Ces résultats sont cohérents avec le concept que diverses protéines jouent un rôle central dans la progression et la prolifération des tumeurs, tandis que certaines protéines agissent comme dérégulation du cancer et comme suppresseur de tumeur [36]. En revanche les gènes sous-exprimés étaient plus enrichis en *one-carbon metabolic process*, *epithelial cell differentiation*, *bicarbonate transport*, *transmembrane transport*, *sodium ion transport*, *cellular response to cAMP*, *prostaglandin metabolic process*, *intermicrovillar adhesion*, *positive regulation of extrinsic apoptotic signaling pathway*, *chloride transport*, *ion transport*, *excretion*, *positive regulation of Fc-gamma receptor signaling pathway*, *involved in phagocytosis*, *xenobiotic metabolic process*, *N-acylethanolamine metabolic process*. Cette découverte est conforme à une étude qui indique que le métabolisme à un seul carbone (*one-carbon metabolic process*) fournit aux cellules les éléments constitutifs, ainsi que le pouvoir réducteur, nécessaires pour maintenir des taux de prolifération élevés, et est donc essentiel pour soutenir le cancer [37]. Quant au transfert du bicarbonate (*bicarbonate transport*), il existe une étude indiquant que le transport du bicarbonate joue un rôle important dans le diagnostic et le traitement de nombreux cancers [35]. Le transport

des ions sodium (*sodium ion transport*) dans les tissus colorectaux normaux, les canaux ioniques et les transporteurs maintiennent l'équilibre eau-électrolyte et l'homéostasie acide/base. Cependant, le dysfonctionnement de ces canaux et transporteurs ioniques conduit au développement et à la progression du cancer colorectal [38]. La signalisation de *cAMP* peut avoir à la fois des rôles suppresseurs et promoteurs de tumeurs selon les types de tumeurs et le contexte [39]. Les effets de la prostaglandine (*prostaglandin metabolic process*) dans le cancer dépendent de plusieurs facteurs, tels que le tissu cible endommagé, la concentration plasmatique des prostaglandines et de leurs sous-types, la présence de mutations génétiques et les différentes voies de signalisation intracellulaires impliquées, feront pencher la balance soit vers la régression tumorale, soit expansion du cancer [40]. Aussi une autre étude indiquant que le métabolisme xénobiotique (*xenobiotic metabolic process*) joue un rôle important dans la progression du cancer du côlon [41].

L'analyse KEGG a montré que les DEG communs étaient principalement enrichies dans : *Hippo signaling pathway*, *Metabolic pathways*, *Nitrogen metabolism*, *Proximal tubule bicarbonate reclamation*, *Bile secretion*, *Steroid hormone biosynthesis*, *Drug metabolism - cytochrome P450*, *Drug metabolism - other enzymes*, *Pyruvate metabolism*, *Pancreatic secretion*, *Retinol metabolism*. La voie Hippo (*Hippo signaling pathway*), est étroitement liée à la différenciation tumorale, aux métastases ganglionnaires et au stade TNM, ce qui suggère que cette voie de signal pourrait servir d'indicateur pronostique pour les patients atteints de CCR [42]. Les voies métaboliques (*Metabolic pathways*) sont activement reprogrammées au cours du développement du CCR, entraînant une augmentation de la glycolyse, de la glutaminolyse et de la synthèse des acides gras. Cette reprogrammation métabolique est désormais reconnue comme une caractéristique du cancer [43]. Pour les acides biliaires (*Bile secretion*), il existe de nombreux rapports scientifiques décrivant les acides biliaires, en particulier les acides biliaires secondaires, comme de puissants cancérigènes ou promoteurs de cancers du côlon [44]. Les cytochromes P450 sont des enzymes clés dans la formation et le traitement du cancer. Ils interviennent dans l'activation métabolique de nombreux pré-cancérogènes et participent à l'inactivation et à l'activation des médicaments anticancéreux [45]. L'acide rétinol (*Retinol metabolism*) est un métabolite de la vitamine A et est essentiel à la croissance cellulaire normale et le métabolisme aberrant de l'acide rétinol est impliqué dans la tumorigenèse [46]. Les enzymes métabolisant les médicaments (*Drug metabolism*) dans les tumeurs sont capables

de bio-transformer une variété de xénobiotiques, y compris les antinéoplasiques, entraînant leur activation ou leur détoxification [47].

Ensuite, Dans cette étude aussi nous avons également effectué une analyse de module sur le réseau PPI construit et sélectionné les dix gènes les plus importants (ZG16, CLCA4, MS4A12, GUCA2B, GUCA2A, SLC26A3, AQP8, PTPRC, CD44, CXCL8). Grâce à la base de données GEPIA2 nous obtenons que les gènes ZG16, CXCL8, AQP8, étaient significativement exprimés de manière différentielle entre les tissus tumoraux et les tissus normaux. L'analyse de l'OS A montré que les expressions des trois gènes clés (AQP8, CXCL8, ZG16) peuvent indiquer que les chances de survie des patients atteints de cancer colorectal sont faibles.

C-X-C Motif Chemokine Ligand 8, également connue sous le nom de CXCL8, est une chimio kiné CXC pro-inflammatoire. De plus en plus de preuves démontrent que la signalisation CXCL8/CXCR1/2 joue un rôle régulateur important dans le microenvironnement tumoral et est importante pour la progression tumorale et les métastases [48]. Qian Liu et al. ont trouvé in vivo, la réparixine ou l'anticorps CXCR1 peut épuiser sélectivement les CSC et les cellules tumorales via la signalisation FASL/FAS et peut inhiber la croissance tumorale et les métastases dans une xénogreffe tumorale. Alors que la réparixine et le paclitaxel ont montré un effet synergique pour arrêter le cycle cellulaire et inhiber la tumorsphere in vitro [49]. Ning et al. ont constaté que, que ce soit in vivo ou in vitro, la surexpression de CXCL8 favorise la croissance tumorale, les métastases et l'angiogenèse, ce qui signifie que CXCL8 peut être une cible thérapeutique importante pour le cancer colorectal [35]. D'un point de vue personnel, les gènes CXCL8 peuvent jouer un rôle majeur dans le développement du cancer du côlon et du rectum et peuvent également être l'une des raisons de sa guérison.

Zymogen Granule Protein 16 (ZG16) est une protéine soluble qui n'est exprimée que dans les cellules épithéliales de l'intestin grêle, du rectum et du côlon [35]. Hui Meng et al, ont démontré une régulation négative séquentielle de ZG16 au cours de la progression du CCR et ont identifié une association significative de ZG16 avec les phénotypes moléculaires et cliniques du CCR. Compte tenu de sa grande perte de tissu dans le CCR et de sa grande similitude avec la chaîne de Jacalin, ils ont émis l'hypothèse que la perte de ZG16 pourrait perturber le système de surveillance bien régulé, faciliter l'invasion bactérienne dans le

hôte et provoquer des changements inflammatoires locaux, qui peuvent être directement liés à une augmentation des risques de développement du cancer [50].

Aquaporin 8 (AQP8), sont composées d'une série de petits transporteurs membranaires. Ils peuvent être divisés en deux catégories en fonction de leur perméabilité : les AQP1, 2, 4, 5 et 8 sont utilisés comme transporteurs sélectifs pour l'eau ; Les AQP3, 7, 9 et 10 peuvent transporter de l'eau, du glycérol et d'autres petits solutés [35]. Hui Zhanget et al. Ils ont prouvé que la surexpression d'AQP8 limitait les capacités de prolifération, de migration et d'invasion des cellules CRC in vitro. In vivo, la régulation à la hausse de l'AQP8 a également supprimé la croissance des cellules CRC. Les analyses mécanistes ont montré que l'AQP8 était une cible directe de miR-92a. L'expression d'AQP8 a été modulée négativement par miR-92a. L'analyse des sauvetages a indiqué que miR-92a facilitait la croissance et l'invasion des cellules CCR en modulant l'expression d'AQP8 [51]. Wu et al. Découvert que l'AQP8 joue un rôle important dans la croissance et la métastase des cellules cancéreuses colorectales ; la surexpression d'AQP8 a inhibé la croissance et l'invasion des cellules colorectales SW480 et HT-29 [35].

Chloride Channel Accessory 4 (CLCA4), est un membre de la famille des protéines des canaux chlorure activés par le calcium et est exprimé dans les cellules épithéliales intestinales et dans le sein, l'utérus, la prostate, l'épididyme, les testicules et les tissus cérébraux. CLCA4 est également connu comme un suppresseur de tumeur, qui peut favoriser le développement de nombreux types de tumeurs malignes. Des études ont montré que CLCA4 peut inhiber la différenciation tumorale dans le cancer du sein. Après l'inactivation de CLCA4, il a été constaté que les tumeurs induisaient la différenciation et la métastase des cellules tumorales par la transition épithéliale-mésenchymateuse. Les expériences in vitro dans le cancer colorectal montrent que la surexpression de CLCA4 peut inhiber la migration et l'invasion des cellules cancéreuses en supprimant la transition épithéliale-mésenchymateuse (EMT) via la signalisation PI3K/ATK [35].

Le CD44, une glycoprotéine transmembranaire non kinase, est surexprimé dans plusieurs types de cellules, y compris les cellules souches cancéreuses, et présente fréquemment des variantes épissées alternatives qui joueraient un rôle dans le développement et la progression du cancer [52]. Lei Du et al. Ils ont constaté que le CD44 est un marqueur robuste et revêt une importance fonctionnelle pour l'initiation du cancer colorectal. à travers que les cellules CD44+ présentaient une croissance groupée et qu'elles ne

colocalisaient pas avec les cellules CD133+ dans le cancer colorectal. Aussi que 100 cellules CD44+ de la tumeur d'un patient ont initié une tumeur xénogreffe *in vivo*. Une seule cellule CD44 + d'une tumeur pourrait former une sphère *in vitro* qui possède des propriétés caractéristiques de cellule souche et était capable de générer une tumeur de xénogreffe ressemblant aux propriétés de la tumeur primaire [53].

Le CD45 (PTPRC), est un récepteur exprimé exclusivement sur toutes les cellules nucléées du système hématopoïétique. Il se caractérise par l'expression de plusieurs iso-formes, spécifiques à un certain type cellulaire et à l'état de développement ou d'activation de la cellule, et l'un des acteurs clés de l'initiation de la signalisation des récepteurs des cellules T [54].

On pensait à l'origine que SLC26A3 était un facteur de transcription et un suppresseur potentiel de tumeur du côlon basé sur des bibliothèques de protéines provenant de cellules cancéreuses normales du côlon et du côlon. Cependant, des études ultérieures ont montré que SLC26A3 code pour une protéine membranaire qui se localise spécifiquement sur la membrane tubulaire de la bordure en brosse (BB) des cellules épithéliales différenciées du côlon. Alors que le mécanisme d'action de cette protéine dans le côlon normal n'est pas clair, la mutation de SLC26A3 est connue pour conduire à la diarrhée chlorée congénitale, démontrant que SLC26A3 est un transporteur qui joue un rôle dans l'absorption colique du chlorure. Certaines études sur les mécanismes *in vivo* et *in vitro* de SLC26A3 dans la barrière épithéliale colique indiquent que SLC26A3 se lie directement aux protéines des jonctions serrées (TJ) et affecte leur expression cellulaire. L'inactivation ou la surexpression de SLC26A3 conduit à la protéine TJ et à la perméabilité épithéliale, et le traitement au TNF- α régule à la baisse les sous-exprimés dans l'adénome et affecte l'intégrité de la barrière épithéliale intestinale en activant la voie de signalisation NF- κ B. La surexpression de SLC26A3 peut également inverser partiellement les dommages induits par le TNF- α en stabilisant les protéines TJ. Cette étude suggère que SLC26A3 joue un rôle important dans la protection de la barrière épithéliale et pourrait fournir des cibles thérapeutiques pour le traitement de l'environnement intestinal à l'avenir [38].

MS4A12 est une membrane couvrant 4 domaines. En tant que protéine cellulaire, MS4A12 participe à la composition de la membrane cellulaire, à la différenciation cellulaire, à la prolifération et à la régulation du cycle cellulaire. Les membres de la famille MS4A font probablement partie du complexe de surface cellulaire, qui a différentes fonctions de

transduction du signal. MS4A12 peut favoriser la prolifération et l'invasion des cellules cancéreuses colorectales en influençant le récepteur du facteur de croissance épidermique.

D'autres études ont montré que le facteur de transcription spécifique intestinal CDX2 médié par l'interférence ARN est un transe-activateur de l'expression génique favorisant la croissance dans le cancer colorectal, suggérant que MS4A12 pourrait être une cible thérapeutique potentielle pour le cancer colorectal. Il a découvert que MS4A12 participait à la différenciation des cellules cancéreuses du côlon. L'analyse de survie a montré que les patients avec une faible expression de MS4A12 avaient une faible survie, suggérant que MS4A12 pourrait être un marqueur moléculaire pour le diagnostic et le pronostic. De plus, jing Han et al. Ils ont constaté que MS4A12 était faiblement exprimé chez les patients atteints de cancer colorectal primaire par analyse bioinformatique. Pendant ce temps, l'analyse de la survie des patients atteints de cancer colorectal primaire a montré que les patients avec une faible expression de MS4A12 avaient une survie plus mauvaise que ceux avec une expression élevée de MS4A12, suggérant que MS4A12 était impliqué dans l'apparition et le développement du cancer colorectal primaire et pouvait inhiber la progression du cancer [55].

Il existe peu d'études sur le facteur 2A d'activation de *guanylate cyclase-activating factor* 2A (GUCA2A), et les mécanismes ne sont toujours pas clairs. Et les fonctions barrières. GUCY2C est un récepteur transmembranaire exprimé sur les cellules épithéliales intestinales et joue un rôle important dans la coordination du mécanisme de homéostasie. Des études récentes ont montré qu'il existe un lien entre le silence de GUCY2C et le dysfonctionnement intestinal, y compris la tumorigenèse. Par conséquent, nous avons émis l'hypothèse que GUCA2A affecte le développement du cancer colorectal en régulant GUCY2C [35].

Conclusion

Conclusion

Le but de cette étude était d'identifier les gènes clé impliqués dans le développement et la progression du cancer colorectal par l'utilisation des méthodes bioinformatiques.

L'étude nous a permis d'arriver aux résultats suivants :

- ✓ Identification de 222 gènes communs entre les trois profils d'expression des patients, dont 47 gènes surexprimés et 175 gènes sous-exprimés.
- ✓ Les analyses d'enrichissement fonctionnel (GO) et (KEGG) ont montré que les gènes étaient enrichis dans de nombreux processus biologiques, cellulaires et moléculaires associés au cancer colorectal.
- ✓ 10 gènes ont été trouvés : (ZG16-AQP8-CXCL8-CD44-PTPRC-GUCA2B-MS4A12-SLC26A3-GUCA2A-CLCA4).
- ✓ trois gènes clés (ZG16-AQP8-CXCL8) ont été confirmés comme étant exprimés de manière différentielle entre les tissus cancéreux et les tissus normaux.
- ✓ L'expression de ces trois gènes est associée à la mauvaise survie des patients atteints de cancer colorectal.

Nos résultats indiquent que les trois gènes clés peuvent être des bio-marqueurs potentiels pour le diagnostic du cancer colorectal, de plus, ils peuvent jouer un rôle thérapeutique efficace dans l'avenir du traitement du cancer colorectal.

Références bibliographiques

Références Bibliographiques:

1. Shelton, B.K. Introduction to colorectal cancer. in Seminars in oncology nursing. 2002. Elsevier.
2. Xi, Y. and P. Xu, Global colorectal cancer burden in 2020 and projections to 2040. *Translational Oncology*, 2021. 14(10): p. 101174.
3. Abes, F., Z. Adjaouti, and M. Belaidi, Etude rétrospective et prospective du cancer colorectal en général et de la thérapie ciblée impliquée dans son traitement en particulier, 2015, Université Mouloud Mammeri.
4. Boussios, S., et al., The developing story of predictive biomarkers in colorectal cancer. *Journal of personalized medicine*, 2019. 9(1): p. 12.
5. Chan, S.K., et al., Meta-analysis of colorectal cancer gene expression profiling studies identifies consistently reported candidate biomarkers. *Cancer epidemiology and prevention biomarkers*, 2008. 17(3): p. 543-552.
6. Scott, E.V.C.a.S.M., Physiology and function of the colon. *Advanced Nutrition and Dietetics in Gastroenterology*, 2014. First Edition: p. 28-32.
7. Society, A.C., Colorectal cancer facts & figures 2020–2022. Published online, 2020: p. 48.
8. David Bême, J.C., Cancer colorectal : la survie dépendrait de la localisation de la tumeur. <https://www.doctissimo.fr/>, 2016.
9. Gallot, D., Anatomie chirurgicale du côlon. *EMC Tech Chir Appar Dig*, 2006. 1: p. 1-8.
10. Nigam, Y., J. Knight, and N. Williams, Gastrointestinal tract 5: the anatomy and functions of the large intestine. *Nursing Times*, 2019. 115(10): p. 50-53.
11. Arbizu, R. and S. Nurko, Colon: structure and function. 2016.

12. Aubrey-Jones, D., The Colon. TeachMeAnatomy.Disponible sur: <<https://teachmeanatomy.info/abdomen/gi-tract/colon/>>. 2018.
13. Slawomir Marecik, J.P., and Leela M. Prasad, Rectal Anatomy: Clinical Perspective. Rectal Cancer, 2018: p. 1-23.
14. Professeur Aimery de GRAMONT, P.T.A., Professeur Florence HUGUET, Professeur Yann PARC, Professeur Julien TAIEB, Le cancer colorectal en questions. Publications de la fondation A.R.C.A.D, 2018.
15. LAFQUIR, J., Cancer du rectum: acquisitions recentes et perspectives, 2009.
16. Futura-Sciences, Rectum : qu'est-ce que c'est ? Disponible sur : <www.futura-sciences.com> (consulté le 09.03.2022). , 2021.
17. Lee, S., Le côlon et le rectum. Société canadienne du cancer, Disponible sur : <<https://cancer.ca/fr/cancer-information/cancer-types/colorectal/what-is-colorectal-cancer/the-colon-and-rectum>>, 2014.
18. Janney, A., Powrie, F. & Mann, E.H., Fact Sheet on Colorectal Cancer. Cancer Association of South Africa (CANSAs), 2020: p. 1-18.
19. Simon, K., Colorectal cancer development and advances in screening. Clinical interventions in aging, 2016. 11: p. 967.
20. (INCa), I.I.n.d.c., Les traitements des cancers du côlon. collection Guides patients Cancer info, 2020.
21. Shussman, N. and S.D. Wexner, Colorectal polyps and polyposis syndromes. Gastroenterology report, 2014. 2(1): p. 1-15.
22. villedieu, A., Le cancer colorectal Association pour le Dépistage des Maladies Cancéreuses en Essonne : < <http://www.admc91.org> >, 2008.
23. Al-Shamsi, H.O., et al., Cancer in the Arab World, 2022, Springer Nature.

24. Hermann Brenner, M.K., Christian Peter Pox, Colorectal cancer. *The Lancet*, 2014: p. 1490-1502.
25. Rodney Hull , F.Z.F., Meryl Oyomno, Zodwa Dlamini, Colorectal Cancer Genetics, Incidence and Risk Factors: In Search for Targeted Therapies. *Cancer Management and Research*, 2020: p. 9869-9882.
26. Singleterry, J., The costs of cancer: addressing patient costs. American Cancer Society Cancer Action Network, 2017.
27. ANSWERS, A., Colorectal Cancer. the American Society of Clinical Oncology., 2021: p. 01-48.
28. Ducreux, P.M., le cancer colorectal. Fondation ARC pour la recherche sur le cancer, 2014: p. 01-46.
29. Foundation, N., Colon Cancer. the NCCN Guidelines for Patients 2021: p. 1-74.
30. Dondaine, L., Etude du rôle d'un donneur de monoxyde d'azote (Glycéril Trinitrate) dans la mort cellulaire immunogène induite par des chimiothérapies (FOX) et/ou radiothérapie dans le cancer colique. memoires pour l'obtention du Diplôme de l'Ecole Pratique des Hautes Etudes., 2017.
31. Parades, V.d., Cancer colon et rectum, stade TNM. SNFCP Colo-Protologie, 2015: p. 1-2.
32. Alvin C. Silva, M.A.K.H., MD. Jonathan A. Leighton, MD Jacques P. Heppell, MD, CT Colonography with Intravenous Contrast Material: Varied Appearances of Colorectal Carcinoma. *RadioGraphics*, 2005: p. 1321-1334.
33. Akkoca, A.N., et al., TNM and Modified Dukes staging along with the demographic characteristics of patients with colorectal carcinoma. *International journal of clinical and experimental medicine*, 2014. 7(9): p. 2828.
34. Hammad, A., M. Elshaer, and X. Tang, Identification of potential biomarkers with colorectal cancer based on bioinformatics analysis and machine learning. *Mathematical Biosciences and Engineering*, 2021. 18(6): p. 8997-9015.

35. Xu, H., et al., Identification and verification of core genes in colorectal cancer. *BioMed Research International*, 2020. 2020.
36. Singh, A., Indispensable Role of Protein in Cancer. *Journal of Practical Biochemistry and Biophysics*, 2016. 1(1).
37. Rosenzweig, A., J. Blenis, and A.P. Gomes, Beyond the Warburg effect: how do cancer cells regulate one-carbon metabolism? *Frontiers in cell and developmental biology*, 2018: p. 90.
38. Zhang, M., et al., Physiological and pathophysiological role of ion channels and transporters in the colorectum and colorectal cancer. *Journal of Cellular and Molecular Medicine*, 2020. 24(17): p. 9486-9494.
39. Zhang, H., et al., Complex roles of cAMP–PKA–CREB signaling in cancer. *Experimental Hematology & Oncology*, 2020. 9(1): p. 1-13.
40. Jara-Gutiérrez, Á. and V. Baladrón, The role of prostaglandins in different types of cancer. *Cells*, 2021. 10(6): p. 1487.
41. Wen, L. and Z. Han, Identification and validation of xenobiotic metabolism-associated prognostic signature based on five genes to evaluate immune microenvironment in colon cancer. *Journal of Gastrointestinal Oncology*, 2021. 12(6): p. 2788.
42. Liang, K., et al., Expression of hippo pathway in colorectal cancer. *Saudi journal of gastroenterology: official journal of the Saudi Gastroenterology Association*, 2014. 20(3): p. 188.
43. Brown, R.E., S.P. Short, and C.S. Williams, Colorectal cancer and metabolism. *Current colorectal cancer reports*, 2018. 14(6): p. 226-241.
44. Nguyen, T.T., et al., Role of bile acids in colon carcinogenesis. *World journal of clinical cases*, 2018. 6(13): p. 577.
45. Rodriguez-Antona, C. and M. Ingelman-Sundberg, Cytochrome P450 pharmacogenetics and cancer. *Oncogene*, 2006. 25(11): p. 1679-1691.

46. Brown, G.T., et al., The expression and prognostic significance of retinoic acid metabolising enzymes in colorectal cancer. *PloS one*, 2014. 9(3): p. e90776.
47. Michael, M. and M.M. Doherty, Tumoral drug metabolism: overview and its implications for cancer therapy. *Journal of Clinical Oncology*, 2005. 23(1): p. 205-229.
48. Xiao, Y.-C., et al., CXCL8, overexpressed in colorectal cancer, enhances the resistance of colorectal cancer cells to anoikis. *Cancer letters*, 2015. 361(1): p. 22-32.
49. Liu, Q., et al., The CXCL8-CXCR1/2 pathways in cancer. *Cytokine & growth factor reviews*, 2016. 31: p. 61-71.
50. Meng, H., et al., Loss of ZG16 is associated with molecular and clinicopathological phenotypes of colorectal cancer. *Bmc Cancer*, 2018. 18(1): p. 1-11.
51. Zhang, H., et al., Identification of the AQP8-miR-92a network associated with the aggressive traits of colorectal cancer. *Biochemical and Biophysical Research Communications*, 2020. 527(1): p. 218-225.
52. Chen, C., et al., The biology and role of CD44 in cancer progression: therapeutic implications. *Journal of hematology & oncology*, 2018. 11(1): p. 1-23.
53. Du, L., et al., CD44 is of functional importance for colorectal cancer stem cells. *Clinical cancer research*, 2008. 14(21): p. 6751-6760.
54. Rheinländer, A., B. Schraven, and U. Bommhardt, CD45 in human physiology and clinical medicine. *Immunology Letters*, 2018. 196: p. 22-32.
55. Han, J., et al., CLCA4 and MS4A12 as the significant gene biomarkers of primary colorectal cancer. *Bioscience reports*, 2020. 40(8).

Glossaire :

Adénocarcinome : est une tumeur maligne développée aux dépens d'un épithélium glandulaire. Le terme est à distinguer de celui d'adénome qui désigne une tumeur développée aux dépens d'un épithélium glandulaire, mais bénigne.

APC (*Adenomatous Polyposis Coli*) : est une molécule de régulation de la concentration en β -caténine qui agit avec la cadhérine E, laquelle intervient dans l'adhérence cellulaire.

BRAF (*V-raf murine Sarcoma Viral oncogene Homolog B*): est un gène humain qui code pour une protéine appelée B-Raf. Le gène est également appelé proto-oncogène B-Raf et *V-raf murine Sarcoma Viral oncogene Homolog B*.

Cytoscape : est un logiciel libre utilisé en bioinformatique pour la visualisation de réseaux d'interaction moléculaire. Plus généralement, il permet l'analyse et la visualisation des réseaux complexes.

DAVID (*Database for Annotation Visualization and Integrated Discovery*): fournit un ensemble complet d'outils d'annotation fonctionnelle permettant aux chercheurs de comprendre la signification biologique derrière de grandes listes de gènes.

FC (*Fold Change*) : est une mesure décrivant à quel point une quantité change entre une mesure originale et une mesure ultérieure. Il est défini comme le rapport entre les deux grandeurs ; pour les quantités A et B, le pli de B par rapport à A est B/A.

GEO (*Gene Expression Omnibus*) : est un référentiel public international qui archive et distribue librement les puces à ADN, le séquençage de nouvelle génération et d'autres formes d'ensembles de données génomiques fonctionnelles à haut débit.

GO (*Gene Ontology*) : est une initiative bioinformatique majeure visant à unifier la représentation des attributs des gènes et des produits génétiques dans toutes les espèces.

HR (*hazard ratio*) : est le rapport des taux de risque correspondant aux conditions décrites par deux niveaux d'une variable explicative.

KEGG (*Kyoto Encyclopedia of Genes and Genomes*) : est une collection de bases de données traitant des génomes, des voies biologiques, des maladies, des médicaments et des substances chimiques.

KRAS (*Kirsten Ras*) : est un gène qui fournit des instructions pour fabriquer une protéine appelée K-Ras, une partie de la voie RAS/MAPK. La protéine relaie les signaux provenant de l'extérieur de la cellule vers le noyau de la cellule.

NCBI (*National Center for Biotechnology Information*) : une série de bases de données pertinentes pour la biotechnologie et la biomédecine et constitue une ressource importante pour les outils et services bioinformatiques. Les principales bases de données comprennent GenBank pour les séquences d'ADN et PubMed, une base de données bibliographique pour la littérature biomédicale.

OS (*Overall Survival*) : Il s'agit du pourcentage de personnes dans une étude ou un groupe de traitement encore en vie pendant une période donnée après le diagnostic. C'est une méthode de description du pronostic dans certaines conditions de la maladie.

P (Valeur de probabilité) : est utilisé dans le contexte du test d'hypothèse nulle afin de quantifier la signification statistique d'un résultat.

Venn diagram : est un style de diagramme largement utilisé qui montre la relation logique entre les ensembles, popularisé par John Venn dans les années 1880.

Volcano plot : est un type de nuage de points utilisé pour identifier rapidement les changements dans de grands ensembles de données composés de données répliquées.