

الجمهورية الجزائرية الديمقراطية الشعبية

République Algérienne Démocratique et Populaire

وزارة التعليم العالي والبحث العلمي

Ministère de l'Enseignement Supérieur et de la Recherche Scientifique

جامعة 8 ماي 1945 قالمة

Université 8 Mai 1945 Guelma

Faculté des Sciences de la Nature et de la Vie, Sciences de la Terre et de l'Univers



Mémoire en vue de l'obtention du diplôme de Master

Domaine : Sciences de la Nature et de la Vie

Filière : Sciences Biologiques

Spécialité/Option : Biochimie Appliquée

Département : Biologie

**Thème : Etude bio-informatique de l'expression des gènes
différentiellement exprimés chez les patients atteints du
cancer colorectal**

Présenté par :

M^{elle} BAHLOUL Nor El Houda.

M^{elle} BOUACHA Rania.

Devant le jury composé de :

Président : Dr. TOUATI Hassen	M.C.B	Université de Guelma
Examinatrice : Dr. BRAIK Asma	M.C.B	Université de Guelma
Encadreur : Dr. OUMEDDOUR A.	M.C.A	Université de Guelma

Juin 2025

Remercîments

*Au terme de ce travail, je tiens à exprimer mes remerciements et ma profonde gratitude, avant tout à **Dieu** le tout puissant et miséricordieux qui m'a donné le courage et la force pour mener à bout ce modeste travail.*

*Je remercie chaleureusement mon encadreur Monsieur « **OUMEDDOUR Abdealkader** » pour les enseignements clairs et précis que nous avons reçus de vous tout au long de nos études. Votre gentillesse, vos qualités humaines, et votre compétence sont incomparables. Veuillez trouver ici, Professeur, l'expression de nos sincères remerciements.*

*Je remercie Monsieur « **TOUATI Hassen** », qui nous fait l'honneur de présider le jury et de juger notre travail.*

*Je voudrais également remercier Madame « **BRAIK Asma** », pour avoir accepté de faire partie du jury et d'évaluer notre travail.*

*Un grand merci à vous, nos **enseignants en biochimie appliquée**. Puisse Dieu vous accorder sa protection et une longue vie en bonne santé. Mes remerciements vont enfin à toute personne qui a contribué de près ou de loin à l'élaboration de ce travail.*

MERCI infiniment

اهداء

الحمد لله الذي ما نجحنا وما علونا وما

تفوقنا إلا برضاه الحمد لله الذي ما اجتزنا دربا ولا تخطينا جهدا إلا بفضلته وإليه ينسب الفضل والكمال والإكمال.

أهدي ثمرة عملي إلى من قال فيهما الله عز وجل ﴿وَاخْفِضْ لَهُمَا جَنَاحَ الذُّلِّ مِنَ الرَّحْمَةِ وَقُلْ رَبِّ ارْحَمْهُمَا كَمَا رَبَّيَانِي صَغِيرًا﴾ سورة الإسراء، الآية 24

إلى العزيز الذي حمّلت اسمه فخراً، وإلى من كلّله الله بالهيبة والوقار، إلى من حصّد الأشواك عن دربي، وزرع لي الراحة بدلاً منها، إلى أبي عياد، لم يحزن ضهر أبي ما كان يحمله، لكن ليحملي، من أجلي انحدب، وكنت أحجّب عن نفسي مطالبتها، فكان يكشف عما أشتهي الحُجبا،

فشكراً لكونك أبي.

وإلى من علّمتني الأخلاق قبل أن أتعلّمها إلى الجسر الصاعد بي إلى الجنّة، إلى اليد الخفية التي أزالّت عن طريقي العقبات، ومن ظلّت دعواتها تحمل اسمي ليلاً ونهاراً،

أمي وسيلة، محبوبتي ومُلهمتي.

إلى من وهبني الله نعمة وجودهم، إلى مصدر قوتي، وأرضي الصلبة، وجدار قلبي المتين، إخوتي: سالم، حسام، خميسي، عادل وأخواتي: رزيقة، صليحة، وإلى ما إن ضاقت بي الدنيا، وسعت بخطاهم، وإن سقطت، كانوا أول من رفعوني بكلماتهم.

إلى من رافقني بالقلب قبل الدرب، صديقتي لينة، وأحبّتي هديل، أحلام، وزميلتي رانيا.

ها أنا اليوم طويّت صفحة من التعب، وسجّلت في تاريخي فخراً لا يُنسى، لم أعد أتساءل عن ملامح الوصول، فقد رأيته في عيوني. تلاشت غيوم التعب، وابتمت الأفق بعد عتمة الانتظار، ها هي الخطى التي كانت تتعثّر أحياناً، قد وجدت مستقرّها في قمة الإنجاز، وبين طيات الطريق، تنفّست سلاماً، وفرحاً، وامتناناً.

وآخر دعوانا أن الحمد لله رب العالمين

نور

اهداء

الحمد لله الذي بنعمته تتم الصالحات، وبفضله وكرمه تحققت أمنيّتي بإنهاء هذه المرحلة من دراستي.
إلى من كانت دعواتها سندي في كل حين، إلى نبض قلبي، إلى أمي الغالية، أقدم لك ثمرة تعبتي وفخري.
إلى أبي العزيز، الداعم الأول لي، الذي غمرني بحنانه وثقته، كل الشكر والعرفان.
إلى أخي أيمن، شكراً لوجودك الدائم، ولروحك المشجعة التي منحني القوة في كل لحظة.
إلى أخواتي خولة، فطيمة، وأسماء، أنتنّ النور الذي يضيء دربي، شكراً لكنّ من القلب.
إلى صديقتي الوفية بشرى، رفيقة اللحظات الصعبة والجميلة، أدامك الله سنداً لي.
إلى زميلتي نور الهدى، التي تشرفت بمرافقتها في إنجاز هذه المذكرة، شكراً لتعاونك وروحك العالية.
وإلى كل من ساندني ووقف بجانبني، من قريب أو بعيد، أقول لكم: جزاكم الله خير الجزاء.

رانيا

Listes des abréviations

Liste des abréviations

ADK : Adénocarcinome	GEO : Gene Expression Omnibus.
ADN : Acide Désoxyribonucléique.	GEPIA2 : Gene Expression Profiling Interactive Analysis2.
Akt : Protein Kinase B (PKB/Akt).	GO : Gene Ontology.
APC : Adenomatous Polyposis Coli.	GPL : Gene Platform.
ARNm : Acide ribonucléique messenger.	GSE : Gene Sample Expression.
ATP : adenosine Triphosphate	IGF-1 : Insulin-like Growth Factor 1
AURKA : Aurora Kinase A.	KLF6 : Kruppel-Like Factor 6.
BCL-2 – B-cell Lymphoma 2	MAPK : Mitogen-Activated Protein Kinase.
BP : Biological Process.	MCODE : molecular complex detection.
BUB1 : BUB1 mitotic checkpoint serine/threonine kinase.	MELK : Maternal Embryonic Leucine Zipper Kinase.
CC : Cellular Component.	MED1 : Mediator Subunit 1.
CCR : Cancer Colorectal.	MICI : Maladie Inflammatoire Chronique de l'Intestin.
CDCA5 : Cell Division Cycle Associated.	MLH1 . : MutL Homolog
CDC20 : Cell Division Cycle20.	MSH2 : MutS Homolog 2.
CDK1 : Cyclin-Dependent Kinase 1.	MSH6 : MutS Homolog 6.
DAVID : Database for Annotation Visualization and Integrated Discovery.	MYC : MYC Proto-Oncogene, BHLH Transcription Factor.
DEG : Differentially Expressed Genes.	NEK2 : NIMA-Related Kinase 2.
E2F3 : E2F Transcription Factor 3.	OTX2 : orthodenticle Homeobox 2.
E2F1 : E2F Transcription Factor 1.	PAF : polypose adénomateuse familiale
FAK : Focal Adhesion Kinase.	PHF8 : PHD Finger Protein 8.
FC : Fold Change.	PI3K : Phosphoinositide 3-Kinase.
Fn : <i>Fusobacterium nucleatum</i> .	

Listes des abréviations

PMS2: *Postmeiotic Segregation Increased*

2

PTTG1 : *Pituitary Tumor-Transforming
Gene1*

SP1: *Specificity Protein 1*

TP53: *Tumor Protein p53.*

UBE2C: *Ubiquitin-Conjugating Enzyme
E2 C*

YBX1: *Y-Box Binding Protein 1*

TPX2: *Targeting Protein for Xenopus
kinesin-like protein 2.*

Résumé

Le cancer colorectal (CCR) est la troisième cause de décès par cancer dans le monde et un problème de santé majeur. Le développement de nouvelles thérapies ciblées susceptibles d'améliorer la survie des patients suscite un intérêt croissant. Cette étude visait à identifier de nouveaux biomarqueurs impliqués dans la progression du CCR et susceptibles d'être utilisés comme thérapies ciblées. Des gènes différentiellement exprimés (*DEG*) ont été sélectionnés parmi trois profils d'expression génique, GSE52060, GSE83889 et GSE106582, à l'aide de l'outil *GEO2R* et du logiciel de diagramme de Venn. L'ontologie génétique et les voies KEGG ont ensuite été réalisées à l'aide du logiciel DAVID. Le réseau PPI a ensuite été construit à l'aide de *STRING* et visualisé à l'aide du logiciel *Cytoscape*, et les gènes clés ont été extraits à l'aide du plug-in *cytoHubba*. L'analyse de survie a été réalisée à l'aide du traceur de *Kaplan-Meier*, tandis que l'expression des gènes clés dans le CCR a été vérifiée à l'aide de l'outil *GEPIA2*. Enfin, les facteurs de transcription des gènes pivots ont été déterminés à l'aide de la base de données *NetworkAnalyst*. Au total, 440 *DEG* ont été identifiés dans les trois ensembles de données, dont 277 gènes surexprimés, enrichis dans le cycle cellulaire, et 163 gènes sous-exprimés, principalement enrichis dans l'organisation de la matrice extracellulaire et la voie de signalisation du cycle cellulaire. Dix gènes ont été identifiés : PTTG1, BUB1, AURKA, CDCA5, MELK, CDC20, TPX2, CDK1, UBE2C et NEK2. La surexpression de cinq gènes clés, dont AURKA, PTTG1, CDC20, CDCA5 et UBE2C, était associée à une faible survie dans le cancer colorectal (CCR) et était régulée par plusieurs facteurs de transcription impliqués dans la division cellulaire et la régulation positive de l'activité de l'ubiquitine protéine ligase. Cette étude a identifié de nouveaux biomarqueurs moléculaires fiables pouvant servir de cibles thérapeutiques potentielles pour le traitement du CCR.

Mots clés: Cancer colorectal, *Gene Ontology*, DAVID, gènes clés

Abstract

Colorectal cancer (CRC) is the third most common cause of cancer death worldwide and a major health problem. The development of new targeted therapies that may improve patient survival remains an area of growing interest. This study aimed to identify new biomarkers involved in CRC progression that could be used as potential targeted therapies. Differentially expressed genes (DEGs) were selected from three gene expression profiles, GSE52060, GSE83889 and GSE106582, using the *GEO2R* tool and Venn diagram software. Gene Ontology and *KEGG* pathways were then performed using DAVID software. Next, the *PPI network* was constructed using *STRING* and visualized using *Cytoscape* software, and *hub* genes were extracted using the *cytoHubba* plug-in. Survival analysis was performed using the *Kaplan–Meier Plotter*, while the expression of hub genes in CCR was verified using the GEPIA2 tool. Finally, transcription factors of hub genes were determined using the *NetworkAnalyst* database. A total of 440 DEGs were identified in the three datasets, including 277 upregulated genes that were enriched in the cell cycle, and 163 downregulated genes that were mainly enriched in the extracellular matrix organization and cell cycle signaling pathway. Ten genes were identified: PTTG1, BUB1, AURKA, CDCA5, MELK, CDC20, TPX2, CDK1, UBE2C and NEK2. The overexpression of five hub genes including AURKA, PTTG1, CDC20, CDCA5, UBE2C was associated with poor survival in CRC and was regulated by several transcription factors involved in cell division a positive regulation of ubiquitin protein ligase activity. This study identified new reliable molecular biomarkers that can serve as potential therapeutic targets for CRC treatment.

Keywords: Colorectal cancer, Gene Ontology, DAVID, Hub genes.

ملخص

سرطان القولون والمستقيم (CRC) هو ثالث أكثر أسباب الوفاة بالسرطان شيوعاً في العالم ومشكلة صحية كبيرة. لا يزال تطوير علاجات موجهة جديدة قد تُحسِّن من بقاء المرضى على قيد الحياة مجالاً ذا اهتمام متزايد. هدفت هذه الدراسة إلى تحديد مؤشرات حيوية جديدة متورطة في تطوّر سرطان القولون والمستقيم والتي يمكن استخدامها كعلاجات موجهة محتملة. تم اختيار الجينات المعبر عنها تفاضلياً (*DEGs*) من ثلاثة ملفات تعريف للتعبير الجيني، GSE83889 و GSE52060 و GSE106582، باستخدام أداة GEO2R وبرنامج رسم مخططات فين. بعد ذلك، تم إجراء تحليل مصطلحات علم الجينات (*Gene Ontology*) ومسارات KEGG باستخدام برنامج DAVID. تم إنشاء شبكة التفاعلات البروتينية (*PPI*) باستخدام أداة STRING وتم تصورها باستخدام برنامج Cytoscape، كما تم استخراج الجينات المحورية باستخدام إضافة cytoHubba. تم إجراء تحليل البقاء باستخدام أداة Kaplan–Meier Plotter، في حين تم التحقق من التعبير الجينات المحورية في سرطان القولون والمستقيم باستخدام أداة GEPIA2. وأخيراً، تم تحديد عوامل النسخ للجينات المحورية باستخدام قاعدة بيانات NetworkAnalyst تم تحديد ما مجموعه 440 جيناً معبراً عنه تفاضلياً في المجموعات الثلاث، بما في ذلك 277 جيناً مرتفع التعبير تم إثراؤها في دورة الخلية، و163 جيناً منخفض التعبير تم إثراؤها بشكل رئيسي في تنظيم مصفوفة خارج الخلية ومسار إشارات دورة الخلية. تم تحديد عشرة جينات وهي: PTTG1 و BUB1 و AURKA و CDCA5 و MELK و CDC20 و TPX2 و CDK1 و UBE2C و NEK2. كان فرط التعبير عن خمسة جينات محورية، بما في ذلك AURKA و PTTG1 و CDC20 و CDCA5 و UBE2C، مرتبطاً بسوء البقاء في سرطان القولون والمستقيم وتم تنظيمها بواسطة عدة عوامل نسخ متورطة في انقسام الخلايا والتنظيم الإيجابي لنشاط إنزيم يوبكوتين ليغاز. حددت هذه الدراسة مؤشرات حيوية جزيئية جديدة موثوقة يمكن أن تكون أهدافاً علاجية محتملة لعلاج سرطان القولون والمستقيم.

الكلمات المفتاحية: سرطان القولون والمستقيم، علم دلالة الجينات، DAVID، الجينات المحورية

Tableau 01 : Proposition des descripteurs T, N et M pour la huitième édition de la classification TNM pour le cancer colorectal.	07
Tableau 02 : Répartition par stade selon la 8ème classification TNM pour le CCR	08
Tableau 03 : Liste des <i>DEG</i> communs aux trois <i>datasets</i>	16
Tableau 04 : Analyse <i>GO</i> des <i>DEG</i> communs dans le cancer colorectal.....	17
Tableau 05 : Analyse <i>KEGG</i> des <i>DEG</i> courants dans le cancer colorectal.....	19
Tableau 06 : Facteurs de transcription régulant les gènes clés identifiés.....	23

Figure 01 : Analyse de l'incidence des cancers à l'échelle mondiale	03
Figure 02 : Évolution de l'incidence du cancer colorectal chez les hommes et les femmes en Algérie en fonction de l'âge	04
Figure 03 : Analyse de la mortalité des cancers à l'échelle mondiale	04
Figure 04 : Les tracés volcaniques des trois <i>Datasets</i> , le bleu représente les gènes sous-exprimés, le rouge représente les gènes surexprimés	14
Figure 05 : Les <i>DEGs</i> communs aux trois <i>datasets</i> . (A) les 277 gènes surexprimés $P < 0.05$ et $\log FC > 1$ et (B) les 163 gènes sous-exprimés	15
Figure 06 : Le réseau <i>PPI</i> , les gènes surexprimés sont montrés en rouge, les gènes sous-exprimés sont montrés en bleu.	20
Figure 07 : Les dix gènes centraux classés par les algorithmes topologiques <i>MCC</i> et <i>degree de Cytoscape</i>	20
Figure 08 : Analyse de la survie des cinq gènes exprimés dans le cancer colorectal. Les tissus cancéreux sont montrés en rouge, les tissus normaux sont montrés en gris.	21
Figure 09 : La vérification de l'expression des gènes clés dans les tissus de cancer colorectal par l'outil <i>GEPIA2</i> , (<i>LUSC : lung squamous cell carcinoma</i>).	22

Table des matières

Introduction	1
1. Cancer Colorectal	3
1.1 Définition.....	3
1.2 Epidémiologie des cancers colorectaux.....	3
1.2.1 Taux d'incidence	3
1.2.2 Taux de Mortalité	4
1.3 Facteurs de risque.....	5
1.3.1 Facteurs génétiques	5
1.3.2 Facteurs pronostiques et mode de vie	5
1.3.3 Facteurs endocriniens et métaboliques	6
1.3.4 Facteurs infectieux.....	7
1.4 Le mécanisme de cancérogenèse	7
1.4.1 Les étapes de cancérogenèse	7
1.5 La classification	8
1.6 Diagnostic	9
1.6.1 Symptomatologie.....	9
1.6.2 Examen clinique	10
1.6.3 Bilan biologique	10
1.6.4 Tests de dépistage.....	10
1.6.5 Explorations endoscopiques.....	11
1.6.6 Imagerie médicale	11
2. Matériels et méthodes	11
2.1 L'objectif de travail.....	11
2.2 Méthodes d'analyse.....	11
2.2.1 Source de données et informations sur les puces à ADN.....	11
2.2.2 Traitement des données et identification des DEG.....	11
2.2.3 Étude de l'ontologie des gènes et de l'enrichissement des voies	12
2.2.4 Utilisation de PPI pour l'analyse des réseaux et des modules.....	12
2.2.5 Sélection des gènes clés.....	12
2.2.6 Analyse de survie globale et validation de l'expression des gènes clés	12
2.2.7 Construction du réseau des facteurs de transcription associés au gènes clés	13
3. Résultats	15
3.1 Analyse des données	15
3.1.1 Étude de l'ontologie génique et de l'enrichissement des voies	17

3.1.2 Étude de Kyoto Encyclopedia of Genes and Genomes (KEGG)	19
3.1.3 Analyse de l'interaction protéine-protéine	20
3.1.4 Sélection des gènes centraux.....	21
3.1.5 Analyse de survie globale des gènes clés.....	21
3.1.6 Validation de l'expression des gènes clés	22
3.1.7 Analyse de la régulation transcriptionnelle des gènes sélectionnés	23
4. Discussion	25
5. Conclusion.....	31

Introduction

Introduction

Le cancer colorectal est l'un des types de cancer les plus répandus dans le monde. En 2020, il représentait 10 % de l'ensemble des nouveaux cas de cancer et 9,4 % des décès liés au cancer, ce qui le classe respectivement au troisième rang en termes d'incidence et au deuxième rang en termes de mortalité selon l'Organisation mondiale de la santé et le Global Cancer Observatory [1] [2].

En 2022, on a enregistré près de 150 000 nouveaux cas et environ 53 000 décès aux États-Unis [3]. Le CRC se caractérise par une accumulation de mutations génétiques qui entraînent une prolifération anormale des cellules épithéliales du côlon ou du rectum. Il est défini comme un groupe de tumeurs malignes se développant dans la muqueuse du côlon ou du rectum. Ces tumeurs prennent souvent naissance à partir de polypes bénins, qui peuvent devenir malins avec le temps [4].

Parmi les principaux facteurs de risque associés à ce cancer, on peut citer l'âge avancé, les antécédents familiaux — en effet, plus de 11 % des patients atteints d'un cancer colorectal ont au moins un parent au premier degré atteint de la même maladie. De plus, des études sur les jumeaux suggèrent qu'environ 35 % des cancers colorectaux sont d'origine héréditaire. Ainsi, environ un tiers des cas seraient attribuables à des facteurs génétiques [5], ainsi que l'obésité, un mode de vie sédentaire, le tabagisme et la consommation excessive d'alcool [6]. Le dépistage précoce du cancer colorectal joue un rôle crucial dans l'amélioration des taux de guérison et de survie. De plus, la coloscopie constitue l'une des méthodes diagnostiques les plus fiables pour détecter les polypes avant leur progression vers des tumeurs malignes [7].

À une époque de progrès rapides en biomédecine et de compréhension accrue de la base génétique des maladies, l'accès aux technologies avancées est devenu de plus en plus accessible. Cela peut permettre d'identifier des anomalies moléculaires spécifiques à un patient, qui peuvent ensuite être ciblées pour le traitement. Connus sous le nom de tests de biomarqueurs pour les thérapies ciblées au niveau moléculaire, ces tests complexes ont le potentiel de permettre la sélection du traitement le plus bénéfique en fonction des fondements moléculaires de la maladie du patient. De tels tests sont essentiels pour libérer le potentiel de la médecine de précision[8].

Le but de cette étude est d'identifier les gènes centraux liés à la pathogenèse du cancer colorectal. Pour ce faire, nous avons analysé trois jeux de données d'expression génique (GSE52060, GSE83889 et GSE106582), à partir desquels des gènes différenciellement

exprimés (DEGs) communs ont été identifiés. L'analyse bioinformatique a permis de détecter les altérations génétiques au niveau du génome. Afin de déterminer les fonctions potentielles de ces biomarqueurs, des analyses d'enrichissement ont été réalisées, notamment l'Ontology des Gènes (GO), la base de données Kyoto Encyclopedia of Genes and Genomes (KEGG), et l'analyse d'enrichissement par ensembles de gènes (GSEA). Un réseau d'interactions protéine-protéine (PPI) des DEGs a ensuite été construit à l'aide de la base de données STRING, et le plug-in Cytoscape-cytoHubba a été utilisé pour analyser les modules géniques et identifier cinq gènes centraux importants. Pour approfondir l'étude des rôles de ces gènes centraux sélectionnés dans la pathogenèse du cancer colorectal, des analyses d'enrichissement GO et KEGG ont été menées. De plus les facteurs de transcriptions qui régulaient la transcription des gènes de centraux ont également été étudiés.

Enfin, une analyse de survie de Kaplan-Meier a été réalisée afin d'évaluer la valeur pronostique.

Ce mémoire comporte deux parties :

Une partie théorique sur le cancer colorectal : Cette section aborde la définition, épidémiologie, facteurs de risque, classification, cancérogénèse, diagnostic.

Une partie bioinformatique : se focalise sur l'analyse des données des patients atteints de cancer colorectal à l'aide de plusieurs logiciels et outils bioinformatique.

1. Cancer Colorectal

1.1 Définition

Le cancer colorectal est une tumeur maligne qui affecte les cellules tapissant l'intérieur du côlon ou du rectum, faisant partie du gros intestin et de l'appareil digestif. Il se développe généralement à partir de polypes adénomateux, qui sont des excroissances bénignes de la paroi interne du côlon ou du rectum. Ces polypes peuvent se transformer lentement en tumeurs malignes sur une période de plusieurs années, souvent entre 5 et 10 ans [9].

1.2 Epidémiologie des cancers colorectaux

1.2.1 Taux d'incidence

A. À l'échelle Mondiale

Le cancer colorectal est le troisième cancer le plus fréquent dans le monde après ceux du sein et du poumon en termes d'incidence. En 2020, environ 1,93 million de nouveaux cas de cancer colorectal ont été recensés, principalement dus à des habitudes de vie malsaines (alimentation, la sédentarité et l'obésité), impliquées dans 55% des cas [10].

En 2022, plus de 1,9 million de nouveaux cas ont été estimée. Les pays développés enregistrent des taux 3 à 4 fois supérieurs à ceux des pays en développement. Les régions les plus touchées sont l'Europe (Danemark et Norvège en tête), l'Australie / Nouvelle-Zélande et l'Amérique du Nord. Tandis que, l'Afrique, ainsi que l'Asie du Sud et centrale, présentent des taux relativement faibles[11].

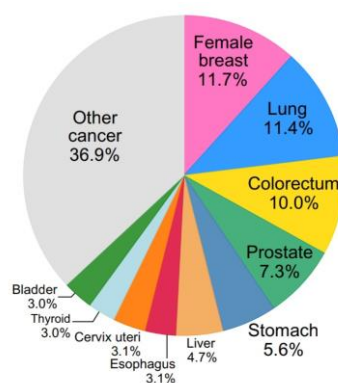


Figure 01 : Analyse de l'incidence des cancers à l'échelle mondiale [10].

B. À l'échelle Algérienne

En Algérie, l'incidence du cancer colorectal reste faible par rapport aux pays développés, bien qu'elle soit en augmentation constante [12]. En 2020, le CCR est fréquent et occupe la première place parmi les cancers digestifs dans la wilaya de Tlemcen [13].

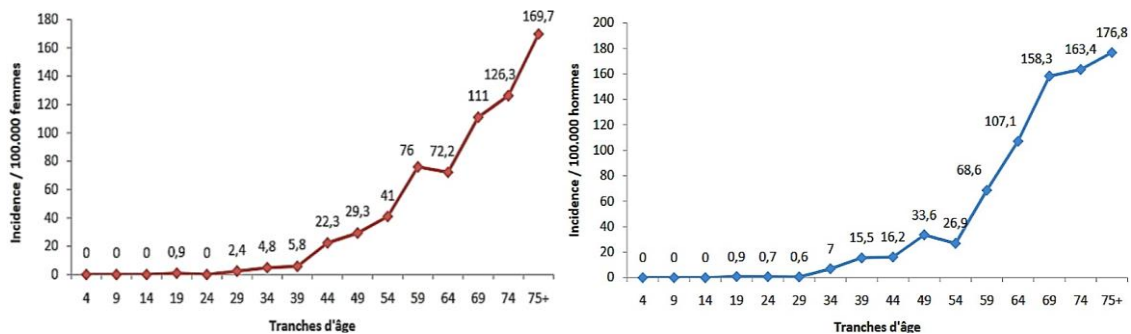


Figure 02 : Évolution de l'incidence du cancer colorectal chez les hommes et les femmes en Algérie en fonction de l'âge [14].

1.2.2 Taux de Mortalité

Le cancer colorectal était la deuxième cause de décès par cancer dans le monde, après le cancer du poumon, avec environ 935173 décès en 2020 [10]. Selon les estimations de GLOBOCAN, fournies par l'Agence Internationale de Recherche sur le Cancer, le nombre de décès diminué à 904000 en 2022, indiquant une légère baisse de la mortalité. Cette réduction pourrait être attribuée à l'amélioration des stratégies de dépistage précoce et des traitements [11].

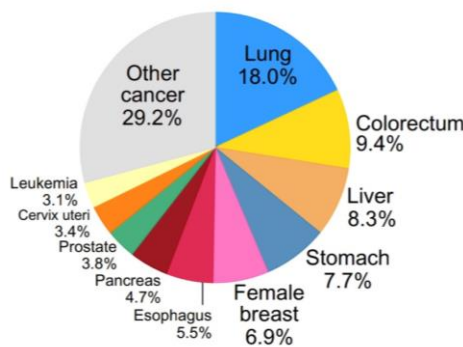


Figure 03 : Analyse de la mortalité des cancers [10].

1.3 Facteurs de risque

1.3.1 Facteurs génétiques

1.3.1.1 Syndrome de lynche

Le syndrome de lynche est à l'origine de 3 à 5% des cancers colorectaux [15]. Il résulte d'une mutation constitutionnelle transmise de manière autosomique dominante affectant principalement les gènes MLH1 (25%), MSH2 (35%), MSH6 (2%) et PMS2 [16].

1.3.1.2 La polypose Adénomateuse Familial

La polypose adénomateuse familiale (PAF) est un syndrome héréditaire autosomique dominante [17], liée à une mutation du gène APC qui a été localisé au niveau du bras long du chromosome 5 [18]. caractérisé par le développement précoce de de centaines à des milliers de polypes dans le côlon et le rectum [19].

1.3.1.3 Héritabilité

Chez un homme ou une femme ayant un antécédent familial du CCR le risque de développer cette pathologie varie en fonction du nombre de parents affectés, du degré de parenté et de l'âge auquel le cancer a été diagnostiqué chez le membre de la famille concerné. Ce risque est plus élevé lorsque plusieurs parents au premier degré sont touchés ou lorsque le diagnostic est posé à un âge précoce [20].

1.3.1.4 Maladies inflammatoires chronique de l'intestin(MICI)

Le cancer colorectal associé aux MICI est considéré comme un modèle de cancérogenèse induite par l'inflammation. L'inflammation chronique provoque des dommages oxydatifs à l'ADN, activant des gènes promoteurs de tumeurs et inactivant des gènes suppresseurs. Contrairement aux polypes sporadiques, de larges zones de muqueuse inflammée dans les MICI subissent une transformation néoplasique diffuse, favorisée par des altérations génétiques et épigénétiques précoces sous l'effet de l'inflammation persistante [21].

1.3.2 Facteurs pronostiques et mode de vie

1.3.2.1 Antécédents personnels de CCR

Une personne ayant déjà eu un CCR présente un risque accru de récurrence, particulièrement durant les cinq premières années suivant la résection complète de la tumeur [22].

1.3.2.2 Tabagisme

Le tabagisme est impliqué dans le développement du cancer colorectal, en particulier chez les patients atteints de MICI. L'association entre les années-paquets fumées et le risque de CRC reste significative même après ajustement de l'inflammation. L'arrêt du tabac est ainsi recommandé pour réduire le risque de CRC[23].

1.3.2.3 Consommation de viandes rouges et de transformée

L'augmentation de la consommation de 120g/jour de viande rouge accroît le risque de CCR de 1,28(IC à 95% : 1,18-1,39), tandis qu'une consommation de 30g/jour de viande transformée ce risque de 1,09(IC à 95% :1,05-1,13) [24]. Le fer héminique contenu dans ces viandes favorise la production de radicaux libres et la formation de composés N-nitrosés, substances aux propriétés carcinogènes [25].

1.3.2.4 L'avancée en âge

Le cancer colorectal touche majoritairement les personnes âgées [26]. Mais il reste rare chez les jeunes adultes [27].l'Age moyen de survenue d'un CCR est 69,5 ans chez l'homme et 72,8 ans chez la femme [28].

1.3.2.5 L'alcoolisme

Le risque de CCR est proportionnel à la durée et à la quantité d'alcool consommé [29].Une consommation d'alcool comprise entre 30g /jour (soit plus de 2 verres) et 45g/jour (3verres) est associée à un risque relatif de 1,16 (IC à 95% : 0,99-1,36) [30].

1.3.2.6 Obésité et apport calorique

Une méta-analyse de 56 études a révélé qu'une augmentation de 5kg/m² de l'indice de masse corporelle (IMC) accroît le risque de CCR de 18%. Une obésité définie par un IMC≥30kg /m² est associée à un risque accru de 41% par rapport aux individus ayant un IMC<23kg/m² [31]. Cette corrélation est plus marquée pour le cancer de côlon que pour celui du rectum et plus prononcée chez les hommes que chez les femmes[32]. Le tissu adipeux sécrète des cytokines pro-inflammatoires favorisant un microenvironnement tumoral propice à la carcinogenèse [33].

1.3.3 Facteurs endocriniens et métaboliques

1.3.3.1 L'acromégalie

Les patients atteints d'acromégalie présentent un risque multiplié par 2 à 3 de développer des adénomes et un CCR. Cette association est attribuée à une hypersécrétion d'hormone de croissance et de facteur de croissance insulino-mimétique1 (IGF-1), stimulant la prolifération cellulaire [22].

1.3.4 Facteurs infectieux

1.3.4.1 Infection par *Fusobacterium nucleatum*

Près de 18% des cancers mondiaux sont liés à des agents infectieux [34]. *Fusobacterium nucleatum* (Fn) est impliqué dans la carcinogenèse colorectale en modulant la réponse immunitaire de la muqueuse intestinale et en favorisant l'instabilité génétique [35]. Des concentrations élevées de Fn ont été détectées dans l'ADN tumoral et dans les métastases coliques [36].

1.4 Le mécanisme de cancérogenèse

La carcinogenèse est un processus en plusieurs étapes impliquant des événements moléculaires et cellulaires aux niveaux génétique et épigénétique, entraîné par l'interaction de facteurs environnementaux, génétiques et métaboliques. Quelle que soit la séquence spécifique des altérations génétiques ou épigénétiques impliquées, ce processus est généralement classé en trois étapes principales : initiation, promotion, Progression.

1.4.1 Les étapes de cancérogenèse

a) Initiation

L'initiation du cancer débute lorsque des mutations ou des altérations épigénétiques s'accumulent, conférant à une cellule un avantage de croissance ou de survie. Ces mutations peuvent être héréditaires ou acquises. Si les mécanismes de réparation ou l'apoptose échouent, un clone de cellules initiées apparaît, marquant le début de cette phase par des dommages irréversibles à l'ADN [37].

b) Promotion

La promotion est la phase au cours de laquelle des agents promoteurs favorisent la prolifération des cellules déjà initiées sans provoquer de mutations supplémentaires. Cette étape est réversible et dépend de l'exposition continue à l'agent promoteur [38].

c) Progression

Les cellules initiées prolifèrent, provoquant une augmentation rapide de la taille de la tumeur. Avec cette croissance, les cellules peuvent accumuler d'autres mutations, ce qui entraîne une hétérogénéité croissante de la population cellulaire[39].

1.5 La classification

La classification TNM, établie par l'Union Internationale Contre le Cancer (UICC), est essentielle pour évaluer le pronostic d'une maladie cancéreuse et orienter les décisions thérapeutiques. Elle repose sur trois critères : T (taille et extension de la tumeur primitive), N (présence et nombre de ganglions lymphatiques régionaux envahis) et M (présence [M1] ou absence [M0] de métastases). Le foie représente l'organe cible principal des métastases dans 75 % des cas [40].

Tableau 01 : Proposition des descripteurs T, N et M pour la huitième édition de la classification TNM pour le cancer colorectal [41].

T : évaluation de l'extension de la tumeur	
Tx	Tumeur primitive non mesurable/évaluable
T0	Absence de tumeur primitive ou non localisable
Tis	Tumeur localisée à l'épithélium ou lamina propria, tumeur in situ
T1	Envahissement de la sous-muqueuse
T2	Envahissement de la musculuse
T3	Envahissement de la sous-séreuse ou des tissus péri-colorectaux non péritonisés
T4	Atteinte de la séreuse ou des structures adjacentes 4a : perforation du péritoine viscéral 4b : adhésion ou invasion d'un organe voisin
N : évaluation de l'extension cancéreuse aux ganglions lymphatiques	
Nx	Ganglions non évalués/évaluables
N0	Pas de métastase ganglionnaire régionale
N1	1 à 3 ganglions métastatiques régionaux N1a : 1 ganglion régional atteint N1b : 2 ou 3 ganglions atteints N1c : nodules tumoraux satellites dans la sous séreuse.
N2	4 ou + ganglions métastatiques régionaux N2a : 4 à 6 ganglions atteints N2b : 7 ou + ganglions atteints

M : présence de métastases à distance	
M0	Absence de métastases à distance
M1	Métastases à distance identifiées M1a : Métastases distantes localisées à 1 seul organe (péritoine exclus) M1b : Métastases distantes dans plusieurs organes (péritoine exclus) M1c : Métastases péritonéales avec ou sans atteintes d'autres organes

Tableau 02 : Répartition par stade selon la 8^{ème} classification TNM pour le CCR [41].

Stade	T	N	M
0	Tis	N0	M0
I	T1-2	N0	M0
IIA	T3	N0	M0
IIB	T4a	N0	M0
IIC	T4b	N0	M0
IIIA	T1-2	N1/N1c	M0
IIIB	T1	N2a	M0
	T3-4a	N1/N1c	M0
	T2-3	N2a	M0
IIIC	T1-2	N2b	M0
	T4a	N2a	M0
	T3-4a	N2b	M0
	T4b	N1-2	M0
IVA	Any T	Any N	M1a
IVB	Any T	Any N	M1b
IVC	Any T	Any N	M1c

1.6 Diagnostic

1.6.1 Symptomatologie

Le cancer colorectal peut évoluer de façon asymptomatique pendant une longue période, puis se manifester par les signes cliniques suivants :

- Douleurs abdominales ou pelviennes persistantes.
- Troubles récents du transit intestinal (alternance entre constipation et diarrhée).
- Anémie d'origine inexpliquée.
- Méléna (présence de sang noir dans les selles).
- Amaigrissement, anorexie, et fatigue chronique.
- Apparition d'une complication aiguë révélatrice (occlusion, perforation ou hémorragie) [42].

1.6.2 Examen clinique

- Examen de l'abdomen : La recherche d'une masse tumorale abdominale palpable, dure, irrégulière, mal limité, plus ou moins fixée, en générale peu sensible en regard du cadre colique et d'ascite [43].
- Le toucher rectal : Est le temps capital de l'examen clinique dans le diagnostic de cancer du rectum, une exécution correcte permet d'avoir accès aux tumeurs du bas et du moyen rectum. Il évalue la distance qui sépare le pôle inférieur de la tumeur de la marge anale, mais aussi du bord supérieur du sphincter anal pour un examinateur entraîné en faisant contracter le sphincter de l'anus [44].

1.6.3 Bilan biologique

- Numération formule sanguine (NFS) pour détecter une anémie.
- Bilan biochimique hépatique et rénal.
- Dosage du marqueur tumoral CEA (antigène carcino-embryonnaire) : est le marqueur le plus fréquemment analysé lorsqu'une tumeur du tractus gastro-intestinal est suspectée. Il s'agit d'une glycoprotéine produite par les cellules du côlon, jouant un rôle important dans l'adhésion et la reconnaissance cellulaire. Une élévation de son taux sérique peut être liée au processus de cancérogenèse [45].

1.6.4 Tests de dépistage

- Test de sang occulte dans les selles :
Ce test repose sur la détection de sang invisible dans les selles en utilisant la propriété peroxydase de l'hémoglobine. Une réaction chimique avec du peroxyde d'hydrogène provoque un changement de couleur en présence de sang [46].
- Test immunologique fécal :
Utilise des anticorps monoclonaux ou polyclonaux pour détecter l'hémoglobine humaine dans les selles. Il n'est pas affecté par les aliments ou médicaments et cible spécifiquement le sang provenant du côlon et du rectum [47].
- Test ADN fécal :
Détection des mutations ou changements génétiques dans l'ADN des cellules présentes dans les selles. Principalement utilisé pour identifier des polypes cancéreux ou des cellules tumorales [48].

1.6.5 Explorations endoscopiques

➤ Coloscopie optique :

Un tube flexible muni d'une caméra (coloscope) est inséré via le rectum pour examiner tout le côlon. Des biopsies peuvent être prélevées ou des polypes retirés pendant la procédure [49].

➤ Sigmoidoscopie flexible :

Utilise un endoscope court pour observer la partie inférieure du côlon (sigmoïde et rectum). Moins complet que la coloscopie mais moins complexe [50].

1.6.6 Imagerie médicale

➤ La colonographie par tomodensitométrie (CTC) est une méthode d'imagerie non invasive utilisant un scanner du côlon, après une préparation intestinale et une insufflation d'air ou de dioxyde de carbone pour une meilleure visualisation [51].

➤ Imagerie par résonance magnétique (IRM) : Utilise des champs magnétiques et des ondes radio pour produire des images détaillées des organes et tissus. Souvent utilisé pour évaluer l'étendue du cancer dans les tissus environnants [52].

➤ La coloscopie par capsule (CCE) est un outil sûr et efficace pour détecter le cancer colorectal (CCR) et les polypes dans un contexte de dépistage. La CCE utilise une capsule jetable qui traverse le côlon par péristaltisme pour enregistrer des séquences vidéo en couleur des deux extrémités. Les deux caméras de l'appareil permettent une couverture à 344 degrés de la muqueuse intestinale [51].

Méthodes

2. Méthodes

2.1 L'objectif de travail

L'objectif de notre travail est d'analyser l'expression des gènes différentiellement exprimés liés au cancer colorectal à l'aide des outils bioinformatiques afin de déchiffrer les mécanismes moléculaires associés à cette pathologie.

2.2 Méthodes d'analyse

2.2.1 Source de données et informations sur les puces à ADN

Gene Expression Omnibus(GEO) (<http://www.ncbi.nlm.nih.gov/geo>) est une base de données publique contenant un grand nombre d'ensembles de données de séquençage et de micro réseaux à haut débit soumis par des instituts de recherche du monde entier.

Nous avons exploité des données de puces à ADN obtenues à partir de ce référentiel public. Pour analyser le cancer colorectal, nous avons sélectionné trois profils d'expression génique (GSE52060, GSE83889, GSE106582) chacun contenant des échantillons tumoraux et normaux. La matrice GSE52060 comprend 23 échantillons de tissu tumoral colorectal (CCR) et 23 échantillons de tissu normal, basés sur la plateforme GPL6947 (*Illumina Human HT-12 v3.0 expression beadchip*). De même, la matrice GSE83889 contient 101 échantillons de tissus tumoraux et 35 échantillons de tissus normaux, tandis que la matrice GSE106582 inclut 77 échantillons tumoraux et 117 échantillons normaux, toutes deux analysées sur la plateforme GPL10558 (*Illumina Human HT-12 v4.0 expression Beadchip*). Ces ensembles de données permettent une analyse comparative de l'expression génique entre les tissus cancéreux et Normaux, fournissant ainsi une base solide pour l'identification des gènes différentiellement exprimés dans le cancer colorectal.

2.2.2 Traitement des données et identification des DEG

GEO2R (www.ncbi.nlm.nih.gov/geo/ge2r) est un outil d'analyse en ligne. Pour identifier les DEG des trois profils d'expression d'ARNm obtenus à partir de la base de données *GEO*.

Nous avons utilisé *GEO2R* avec $P < 0,05$ et $|\log_2\text{-fold change (FC)}| > 1$, utilisés comme critères de sélection. Nous l'avons utilisé pour comparer les profils d'expression génique entre différents groupes afin de déterminer les DEG entre le groupe pathologique et le groupe témoin. Les fichiers de données au format TXT ont ensuite été traités à l'aide de l'outil *Venn Diagram* (<https://bioinformatics.psb.ugent.be/webtools/Venn/>) afin d'identifier les DEG communs aux

trois profils. Les gènes Présentant un $\log FC > 0$ ont été classés comme surexprimés (*upregulated genes*), tandis que ceux avec un $\log FC < 0$ ont été considérés comme sous-exprimés (*downregulated genes*).

2.2.3 Étude 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 des données transcriptomique ou génomique à haut débit. Cette analyse repose sur l'utilisation d'un outil bioinformatique en ligne, le logiciel *DAVID* (Version 6.8, <https://david.ncifcrf.gov/>), permettant d'effectuer l'analyse d'enrichissement des voies *GO* (*Biological process* :BP, *Cellular component* :CC, et *Molecular function* :MF) ainsi que du *kyoto Encyclopedia of Genes and Genomes* (KEGG). Le critère de sélection appliqué pour cette analyse était un seuil de $P < 0,05$.

2.2.4 Utilisation de PPI pour l'analyse des réseaux et des modules

Pour explorer les interactions entre les DEGs, nous avons utilisé une base de donnée 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.

2.2.5 Sélection des gènes clés

Nous avons employé le plug-in *CytoHubba* du logiciel *Cytoscape* pour identifier les gènes pertinents. Ensuite, nous avons sélectionné les dix principaux gènes centraux, classés selon les algorithmes topologiques *Maximal Clique Centrality* (MCC) et *Degree*.

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

Nous avons utilisé l'outil *Kaplan Meier Plotter* (<http://kmplot.com/analysis/index.php?p=service>) pour évaluer l'impact de dix gènes centraux sur la survie globale (OS) des patients atteints de CCR. Pour estimer le taux de survie globale des patients atteints de cancer colorectal, les échantillons ont été divisés en groupes à expression élevée et faible en fonction de la médiane d'expression. 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 présentés sur le graphique.

Ensuite, l'outil *Gene Expression Profiling Interactive Analysis* *GEPIA2*

(<http://gepia2.cancer-pku.cn/>) a été utilisé pour analyser les données d'expression de séquençage d'ARN des patients atteints de CCR en les comparant à des échantillons témoins (avec une valeur de coupure de $|\text{Log2FC}| = 1$ et un seuil de $p = 0,05$).

2.2.7 Construction du réseau des facteurs de transcription associés au gènes clés

Network Analyst (<http://www.networkanalyst.ca/faces/home.xhtml>) a été utilisé pour explorer les interactions TF-gène (*transcription factor – genes*) pour les gènes soumis et évaluer l'effet du facteur de transcription sur l'expression et les voies fonctionnelles du gène clé. Dans cette étude, les facteurs de transcription des gènes clés ont été prédits à partir de cette base de données et un réseau de régulation transcriptionnelle a été construit et visualisé à l'aide du *Network Analyst*.

Résultats

3. Résultats

3.1 Analyse des données

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 $|\log_2\text{-fold change (FC)}| > 1$. L'ensemble de données GSE52060 a révélé 807 *DEG*, dont 363 gènes surexprimés et 444 gènes sous-exprimés. L'ensemble GSE83889 a identifié 526 *DEG*, comprenant 355 gènes surexprimés et 171 gènes sous-exprimés. Enfin, l'ensemble GSE106582 a mis en évidence 27 *DEG*, avec 11 gènes surexprimés et 16 gènes sous-exprimés. Les tracés volcaniques (*Volcano Plot*) des trois ensembles de données *GEO* sont présentés dans la figure 04.

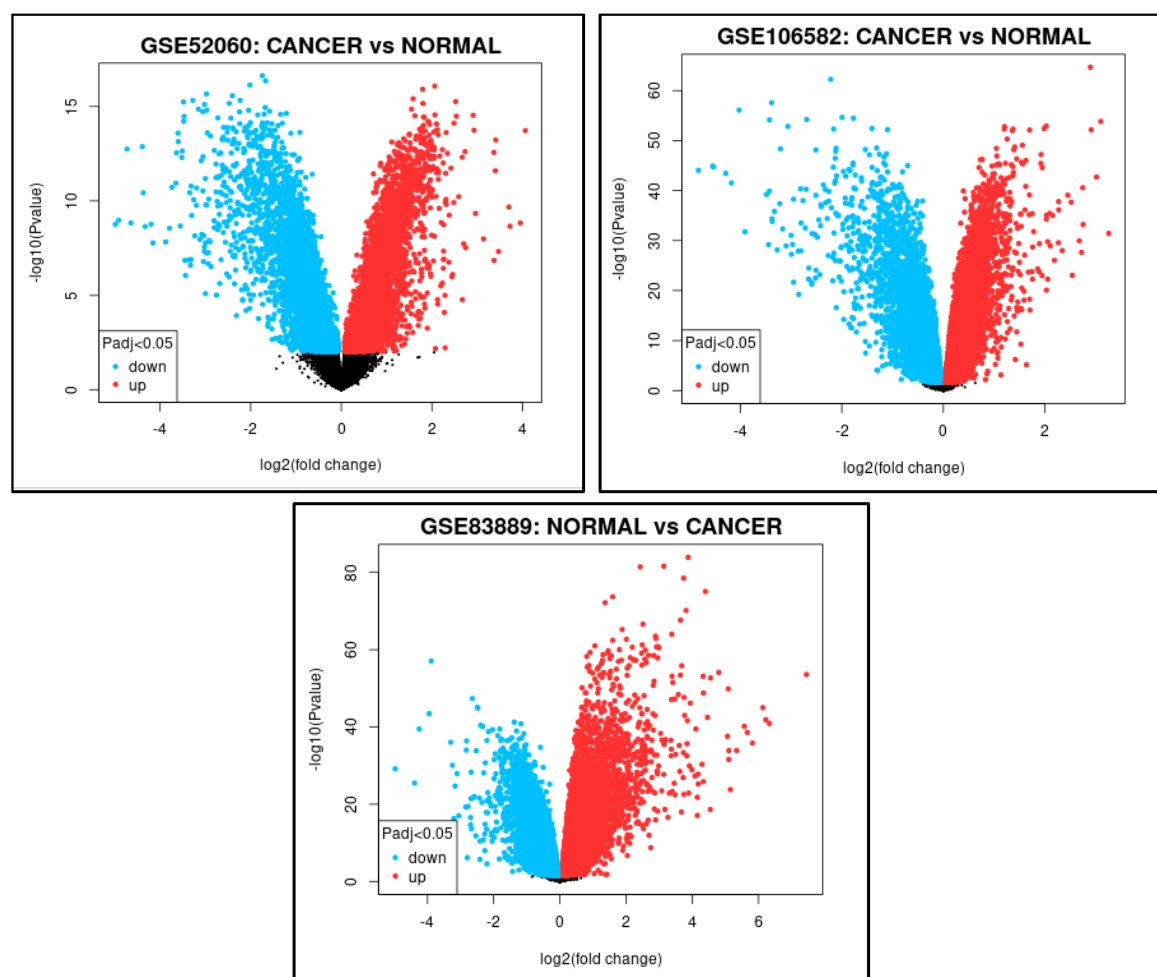


Figure 04 : Les tracés volcaniques (*Volcano plots*) des trois *Datasets*, le bleu représente les gènes sous-exprimés, le rouge représente les gènes surexprimés ($P < 0 :05$ et $\log |\text{FC}| > 1$).

À l'aide de la base de données *Venn Diagram*, nous avons identifié les *DEGs* communs aux trois profils. La figure 05 montre que 440 *DEG* ont été trouvés dans les trois *Dataset*, incluant 277 gènes surexprimés et 163 gènes sous-exprimés.

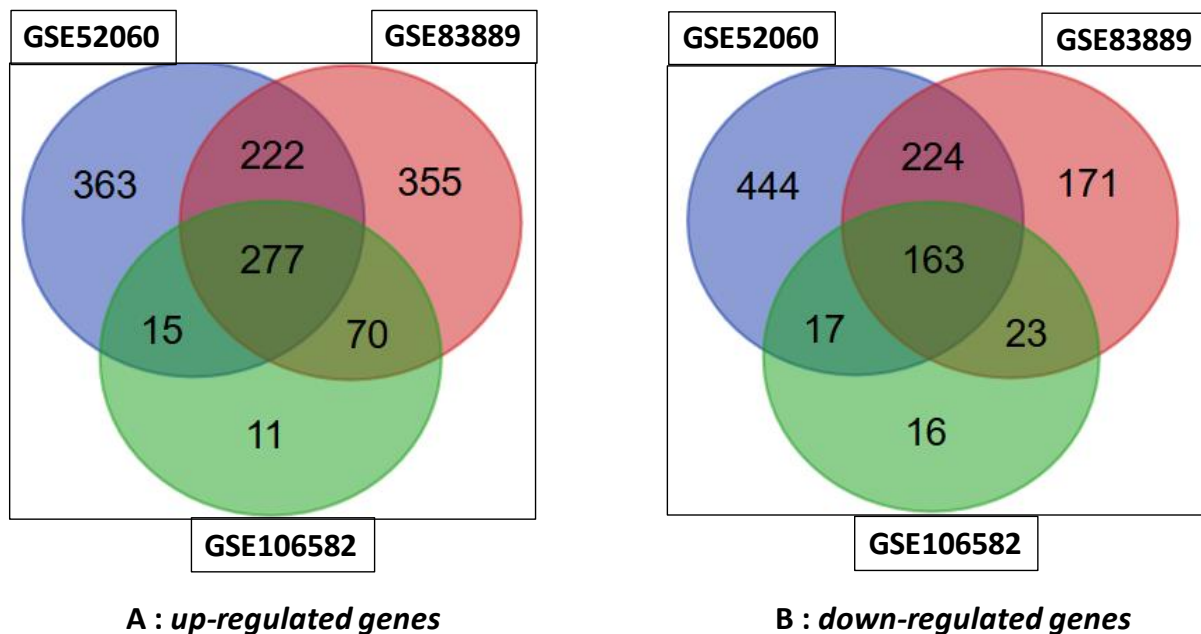


Figure 05 : Les *DEGs* communs aux trois *datasets*. (A) les 277 gènes surexprimés $P < 0.05$ et $\log |FC| > 1$ et (B) les 163 gènes sous-exprimés $P < 0.05$ et $\log |FC| > 1$.

La liste des *DEGs* communs surexprimés et sous-exprimés dans les trois *datasets* sont montrés dans le tableau 3.

Tableau 03 : Liste des *DEGs* communs aux trois *datasets*.

	Liste des gènes
Gènes surexprimés	PLCE1 KCNMA1 SLC22A18AS FGL2 OTOP2 GSTA1 HSD17B2 HMGCS2 FAM134B C15orf48 ZG16 TPSG1 ITLN1 UGT2B17 AGR3 TSPAN3 LRRN2 DHRS11 MYOM1 AKR1B10 HSD11B2 AKAP7 ACVRL1 TUBAL3 PAPSS2 DEFB1 ARL14 PIGZ TRPM6 A1CF ADAMDEC1 PDCD4 CES2 GHR PYY MAOA SRI SLCO2A1 REP15 UGT1A10 DNASE1L3 BCHE SCIN CKB SLC26A3 PLCL2 PDZD3 HSD17B11 BDKRB2 ZZEF1 TSPAN7 HHLA2 PAQR8 AKR1B15 PIGR AHCYL2 MYOT RCAN2 MFAP5 FCGBP CLEC3B CHGA CAPN9 MCOLN2 DSC2 MT1M TP53INP2 PKIB CDHR5 LRRC31 RDH5 C10orf99 FHL1 ABCB1 KLF4 SULT1A2 TMEM171 MT1G CLCA1 PDE9A ABCA8 PPARGC1A ANKRD20A1 LIMA1 MUC12 MUC2 SLC16A9 GSN GLDN NEDD4L CLDN23 PLS1 INSL5 SLC51B EDN3 SPON1 SULT1A1 AQP8 SGK1 HIGD1A ITPKA AOC1 HSPA2 LYPD8 MS4A8 TNFRSF17 CD177 C1orf115 SCGN PLAC9 BTNL3 MYH11 LGALS4 SLC25A34 KBTBD11 PCK1 ADH1C CIDEC SMPDL3A SLC9A2 SLC17A4 BMP2 UGT2A3 CYP4F12 LRRC19 SST ITM2C RUNDC3B SCNN1B DENND2A CLIC5 EPB41L3 CDKN2B HPGD LAMA1 ZBTB7C CWH43 MT1X GGT6 MUC4 MEP1A MALL KLF9 SCGB2A1 HOXD1 LINC00997 PLA2G10 GPT PDK4 CA12 TMPRSS2 ECI2 EPHX2 GUCA2A SEMA6A ATP2A3 PTGER4 TMEM100 CA7 CPNE8 LGALS2 NR3C2 TPSAB1 DPT MMP28 MFAP4 GUCA2B CHP2 SLC22A23 OGN SCARA5 CAPN13 CLCA4 MT1E UGDH MYO1A CA4 SEPP1 CTSG NXPE4 SCG2 ETFDH APPL2 MYLK LYVE1 NAAA TMIGD1 NAT2 KRT20 SPDEF NEU4 CXCL12 CTSE ENDOD1 PDE5A NR5A2 CPA3 OSBPL1A RBM47 PID1 FMO5 ABHD3 SYTL2 LRRC26 VIP CA2 BCAS1 FUCA1 SMIM14 SLC26A2 ANPEP ADH1A KCNIP4 CEACAM7 ITM2A ARHGAP44 ADTRP CAPN5 SRPX COL17A1 ST6GALNAC6 TSPAN1 CMBL MT1F MATN2 BEST2 B3GNT6 ADH1B PLPP1 GCG CHST5 GCNT3 GBA3 UGP2 RNF125 FRZB CLDN8 MS4A12 FAM107A MXD1 PGM5 GPX3 VSIG2 TCEA3 FAM46C MT1H HEPACAM2 MT1A VIPR1 PLAC8 CEACAM1 NXPE1 C2orf88 GPA33 LPAR1 MAMDC2 BEST4 CA1 STMN2 RETSAT FABP2 SULT1B1 SDCBP2 FAM3D SLC4A4 SFRP1 FXYP3 SELENBP1 SI CFD GPAT3 JCHAIN SLC51A PARM1 KLRB1 ST6GALNAC1 SLC44A4
Gènes sous-exprimés	TKT PLAU KLK6 GDF15 TPX2 TOP1MT MMP7 GPSM2 TRIB3 SERPINB5 FOXQ1 AQP9 SOX4 FASN CDC25B RPL29 COL1A1 RPP40 ANLN BOP1 MRGBP CDK1 TOMM34 CLDN2 VSNL1 PHLDA1 GTF3A COL7A1 SAPCD2 CST1 NEBL CEL AHCY CXCL8 AURKA RIPK2 RPS21 PAQR4 IFITM2 PPM1H SOX9 PAFAH1B3 CELSR3 HOXB8 FAM3B DUSP14 SNAR-A1 COL5A2 REG1B THY1 ASCL2 SCD ETV4 CCDC86 SPP1 CA9 OLR1 NELFCD H19 TGIF2 WDR4 PYCR1 MELK CSE1L RNF43 CDC20 CENPN DKC1 GART TUBB3 FAM89A UBD MYC TIMP1 BUB1 BGN TNFRSF12A TRIP13 MMP1 S100A11 CBX2 PTTG1 CDCA5 UBE2T CKS2 TEAD4 ECT2 MMP11 CDH3 CCNO TMEM158 IFITM3 MMP3 KRT80 LY6G6D FAP DDIT4 TCN1 SLC5A6 SLC7A5 SNHG15 CXCL2 UBE2C SLC35D3 FXYP5 SRPX2 HES6 PSAT1 PUS7 LRP8 CENPW NFE2L3 CEP55 CEMIP NKD2 MSX1 DBNDD1 C19orf48 TOP2A TACSTD2 WDR72 COMP GZMB CDC45 THBS2 DIO2 MTHFD1L CXCL1 TMEM97 REG1A DPEP1 HSPH1 COL4A1 PTTG3P SULF1 RPS7 EVA1A LY6E HILPDA CCNF IMPDH1 MIF CLDN1 ENC1 GTF2IRD1 STC2 COL10A1 SNTB1 COL11A1 CKMT2 CTHRC1 IFITM1 HMGAI1 TESC SULT2B1 REG3A TGFB1 CDKN3 NME1 COL1A2 NEK2 GINS2 S100A9.

3.1.1 Étude de l'ontologie génique 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 été une valeur $P < 0,05$. L'analyse a compris *Biological process* (BP), *Cellular component* (CC), *Molecular function* (MF).

Dans les BP, on observe que les gènes surexprimés ont été principalement enrichis dans les voies suivantes : *detoxification of copper ion*, *xenobiotic metabolic process*, *negative regulation of growth*, *cellular response to zinc ion*, *cellular response to copper ion*, *cellular response to cadmium ion*. En revanche, les gènes sous-exprimés ont été principalement enrichis dans des voies suivantes : *extracellular matrix organization*, *cell division*, *mitotic cell cycle*, *collagen fibril organization*, *collagen catabolic process*, *apoptotic process*.

Dans les CC, les résultats de l'analyse ont montré que les gènes surexprimés ont été principalement impliqués dans les voies suivantes : *extracellular exosome*, *plasma membrane*, *apical plasma membrane*, *collagen-containing extracellular matrix*, *extracellular region*, *extracellular space*. Pour les gènes sous-exprimés ont été principalement enrichis dans les voies suivantes : *extracellular space*, *collagen-containing extracellular matrix*, *collagen trimer*, *extracellular region*, *nucleoplasm*, *endoplasmic reticulum lumen*.

Dans les MF, les gènes surexprimés ont été principalement enrichis dans les voies : *carbonate dehydratase activity*, *extracellular matrix structural constituent*, *estradiol 17-beta-dehydrogenase [NAD(P)+] activity*, *all-trans-retinol dehydrogenase (NAD+) activity*, *oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor*, *chloride channel activity*. Les gènes sous-exprimés ont été enrichis dans les voies suivantes : *extracellular matrix structural constituent*, *extracellular matrix structural constituent conferring tensile strength*, *protein binding*, *serine-type endopeptidase activity*, *molecular function activator activity*, *extracellular matrix binding*.

L'analyse GO des DEG communs est présentée dans le tableau suivant :

Tableau 04 : Analyse GO des DEG communs dans le cancer colorectal.

Catégorie	Terme	N	%	Valeur de P
Les gènes surexprimés				
GOTERM_BP_DIRECT	<i>detoxification of copper ion</i>	08	2,9	5,9E-10
GOTERM_BP_DIRECT	<i>xenobiotic metabolic process</i>	14	5,1	6,9E-9
GOTERM_BP_DIRECT	<i>negative regulation of growth</i>	07	2,6	7,4E-8
GOTERM_BP_DIRECT	<i>cellular response to cadmium ion</i>	07	2,6	3,8E-7
GOTERM_BP_DIRECT	<i>cellular response to zinc ion</i>	07	2,6	5,0E-7
GOTERM_BP_DIRECT	<i>cellular response to copper ion</i>	07	2,6	6,5E-7
GOTERM_CC_DIRECT	<i>extracellular exosome</i>	72	26,5	2,0E-13
GOTERM_CC_DIRECT	<i>plasma membrane</i>	23	8,5	1,9E-8
GOTERM_CC_DIRECT	<i>apical plasma membrane</i>	114	41,9	2,0E-8
GOTERM_CC_DIRECT	<i>collagen-containing extracellular matrix</i>	21	7,7	1,4E-7

GOTERM_CC_DIRECT	<i>extracellular space</i>	49	18,0	2,0E-6
GOTERM_CC_DIRECT	<i>extracellular region</i>	55	20,2	7,5E-6
GOTERM_MF_DIRECT	<i>carbonate dehydratase activity</i>	05	1,8	6,1E-5
GOTERM_MF_DIRECT	<i>extracellular matrix structural constituent</i>	08	2,9	1,1E-3
GOTERM_MF_DIRECT	<i>estradiol 17-β-dehydrogenase [NAD(P)⁺] activity</i>	04	1,5	2,2E-3
GOTERM_MF_DIRECT	<i>all-trans-retinol dehydrogenase (NAD⁺) activity</i>	04	1,5	2,2E-3
GOTERM_MF_DIRECT	<i>oxidoreductase activity, acting on the CH-OH group of donors, NAD or NADP as acceptor</i>	05	1,8	2,8E-3
GOTERM_MF_DIRECT	<i>Chloride channel activity</i>	06	2,2	3,1E-3
les gènes sous-exprimées				
GOTERM_BP_DIRECT	<i>extracellular matrix organization</i>	13	08	7,8E-9
GOTERM_BP_DIRECT	<i>cell division</i>	15	9,3	3,6E-6
GOTERM_BP_DIRECT	<i>mitotic cell cycle</i>	08	4,9	1,0E-4
GOTERM_BP_DIRECT	<i>collagen fibril organization</i>	06	3,7	1,8E-4
GOTERM_BP_DIRECT	<i>collagen catabolic process</i>	05	3,1	3,1E-4
GOTERM_BP_DIRECT	<i>Apoptotic process</i>	05	3,1	4,1E-4
GOTERM_CC_DIRECT	<i>extracellular space</i>	43	26,5	6,7E-11
GOTERM_CC_DIRECT	<i>collagen trimer</i>	10	6,2	8,0E-8
GOTERM_CC_DIRECT	<i>collagen-containing extracellular matrix</i>	16	9,9	2,6E-7
GOTERM_CC_DIRECT	<i>Nucleoplasm</i>	57	35,2	2,2E-6
GOTERM_CC_DIRECT	<i>extracellular region</i>	38	23,5	7,0E-6
GOTERM_CC_DIRECT	<i>endoplasmic reticulum lumen</i>	11	6,8	1,6E-4
GOTERM_MF_DIRECT	<i>extracellular matrix structural constituent</i>	11	6,8	4,7E-8
GOTERM_MF_DIRECT	<i>extracellular matrix structural constituent conferring tensile strength</i>	07	4,3	1,5E-6
GOTERM_MF_DIRECT	<i>protein binding</i>	133	82,1	1,5E-5
GOTERM_MF_DIRECT	<i>serine-type endopeptidase activity</i>	08	4,9	9,1E-4
GOTERM_MF_DIRECT	<i>molecular function activator activity</i>	04	2,5	1,3E-3
GOTERM_MF_DIRECT	<i>extracellular matrix binding</i>	09	5,6	1,6E-3

3.1.2 Étude de Kyoto Encyclopedia of Genes and Genomes (KEGG)

Le critère de sélection de l'analyse KEGG était la valeur $P < 0,05$. Les résultats ont identifié douze voies de signalisation. Les voies de signalisation associées aux DEG étaient principalement enrichies dans les voies suivantes : *Metabolic pathways*, *Mineral absorption*, *Pentose and glucuronate interconversions*, *Drug metabolism - cytochrome P450*, *Bile secretion*, *Retinol metabolism* pour les gènes surexprimés et *Cell cycle*, *Cytoskeleton in muscle cells*, *Protein digestion and absorption*, *ECM-receptor interaction*, *IL-17 signaling pathway*, *Amoebiasis* pour les gènes sous-exprimés.

L'analyse de la voie KEGG des DEG commun est présentée dans les tableaux :

Tableau 05 : Analyse KEGG des DEG courants dans le cancer colorectal.

Catégorie	Terme	N	%	Valeur de P
Les gènes surexprimés				
KEGG_PATHWAY	<i>Metabolic pathways</i>	51	18,8	3,1E-6
KEGG_PATHWAY	<i>Mineral absorption</i>	09	3,3	8,3E-6
KEGG_PATHWAY	<i>Pentose and glucuronate interconversions</i>	07	2,6	2,9E-5
KEGG_PATHWAY	<i>Drug metabolism - cytochrome P450</i>	06	3,3	3,2E-5
KEGG_PATHWAY	<i>Bile secretion</i>	09	3,3	1,4E-4
KEGG_PATHWAY	<i>Retinol metabolism</i>	08	2,9	1,5E-4
les gènes sous-exprimées				
KEGG_PATHWAY	<i>Cell cycle</i>	09	5,6	2,7E-4
KEGG_PATHWAY	<i>Cytoskeleton in muscle cells</i>	10	6,2	7,8E-4
KEGG_PATHWAY	<i>Protein digestion and absorption</i>	07	4,3	8,6E-4
KEGG_PATHWAY	<i>ECM-receptor interaction</i>	06	3,7	2,5E-3
KEGG_PATHWAY	<i>IL-17 signaling pathway</i>	06	3,7	3,4E-3
KEGG_PATHWAY	<i>Amoebiasis</i>	06	3,7	4,8E-3

3.1.3 Analyse de l'interaction protéine-protéine

Pour mieux évaluer les interactions entre les DEGs communs identifiés, nous avons effectué un réseau PPI à l'aide du site Web *STRING*. Le réseau PPI de DEGs communs est illustré dans la Figure 06. Le réseau PPI tracé est composé de 11 gènes surexprimés et de 28 gènes sous-exprimés. Ensuite, par l'utilisation le plug-in MCODE du *Cytoscape*, nous avons identifié un ensemble de gènes centraux (PTTG1, BUB1, AURKA, CDCA5, MELK, CDC20, TPX2, CDK1, UBE2C, NEK2).

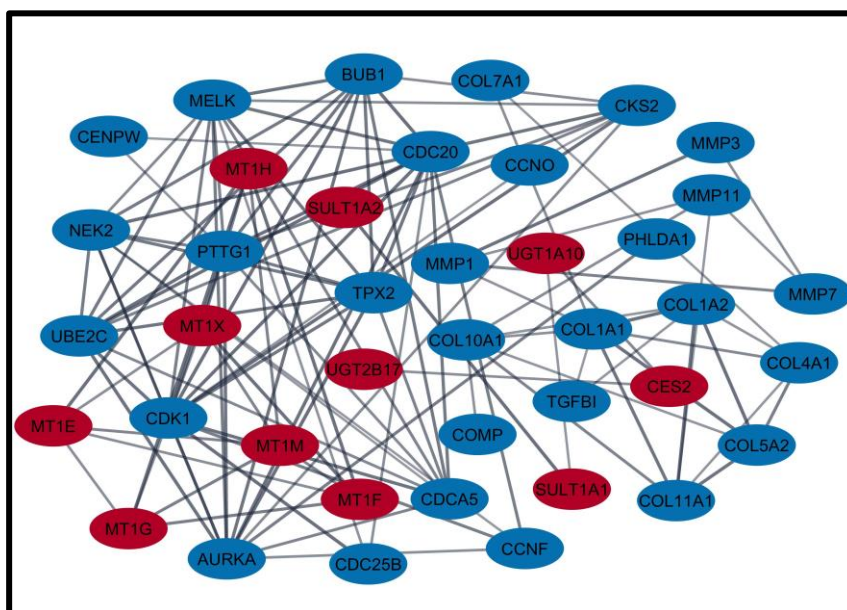


Figure 06 : Le réseau PPI, les gènes surexprimés sont montrés en rouge, les gènes sous-exprimés sont montrés en bleu.

3.1.4 Sélection des gènes centraux

Nous avons trouvé une unité fonctionnelle du réseau PPI comportant dix gènes (PTTG1, BUB1, AURKA, CDCA5, MELK, CDC20, TPX2, CDK1, UBE2C, NEK2). Ces dix gènes sont considérés comme les gènes centraux associés à la pathologie du cancer colorectal (Figure 07).

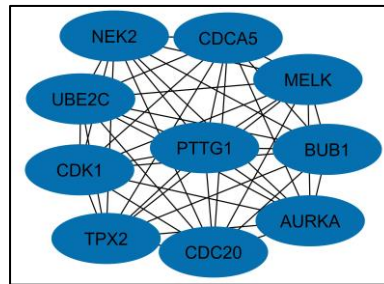


Figure 07 : Les dix gènes centraux classés par les algorithmes topologiques *MCC* et *degree* de Cytoscape.

3.1.5 Analyse de survie globale des gènes clés

Les résultats d'analyse de la survie globale des gènes clés ont indiqué que les niveaux d'expression de cinq gènes (AURKA, PTTG1, CDC20, CDCA5, UBE2C) sont associés à une mauvaise survie chez les patients atteints de cancer colorectal (Figure 08).

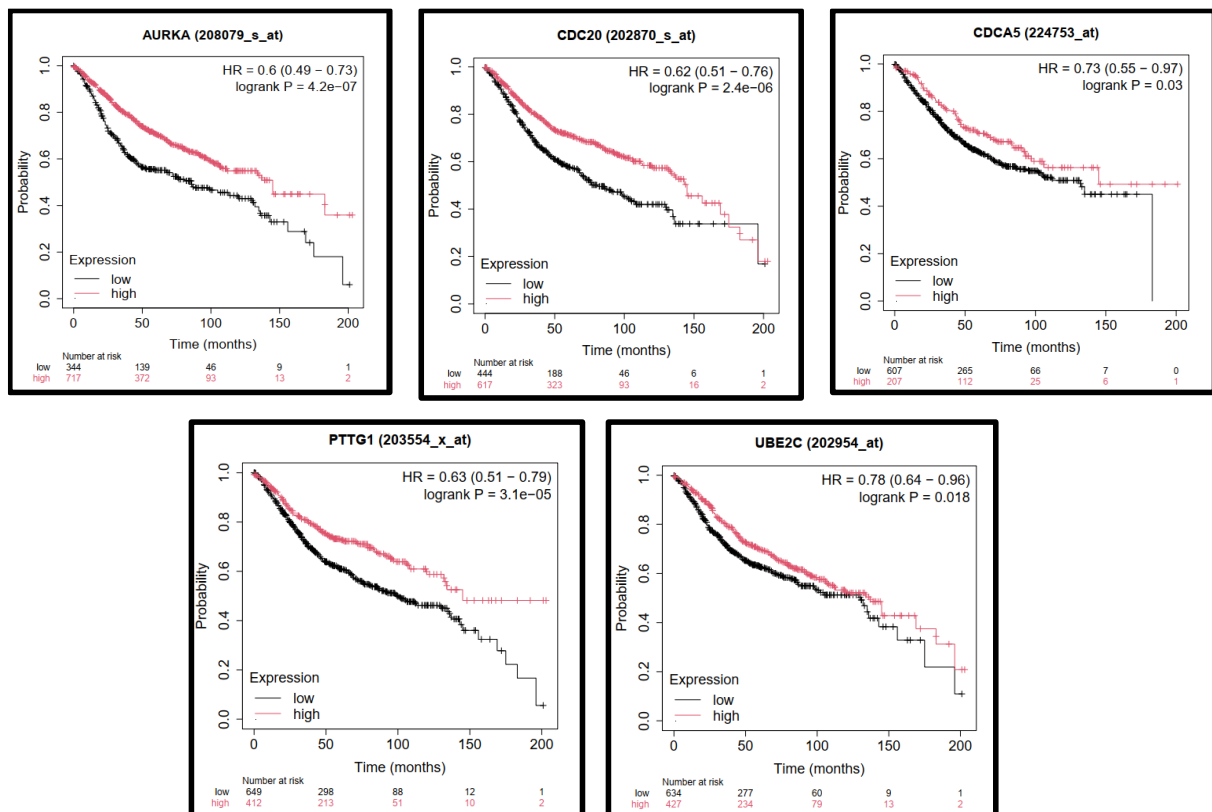


Figure 08 : Analyse de la survie des cinq gènes exprimés dans le cancer colorectal. Les tissus cancéreux sont montrés en rouge, les tissus normaux sont montrés en gris.

3.1.6 Validation de l'expression des gènes clés

Les résultats obtenus à partir de la base de données GEPIA2 ont révélé une différence d'expression significative pour les gènes clés : AURKA, PTTG1, CDC20, CDCA5 et UBE2C entre les tissus tumoraux et les tissus normaux (Figure 09).

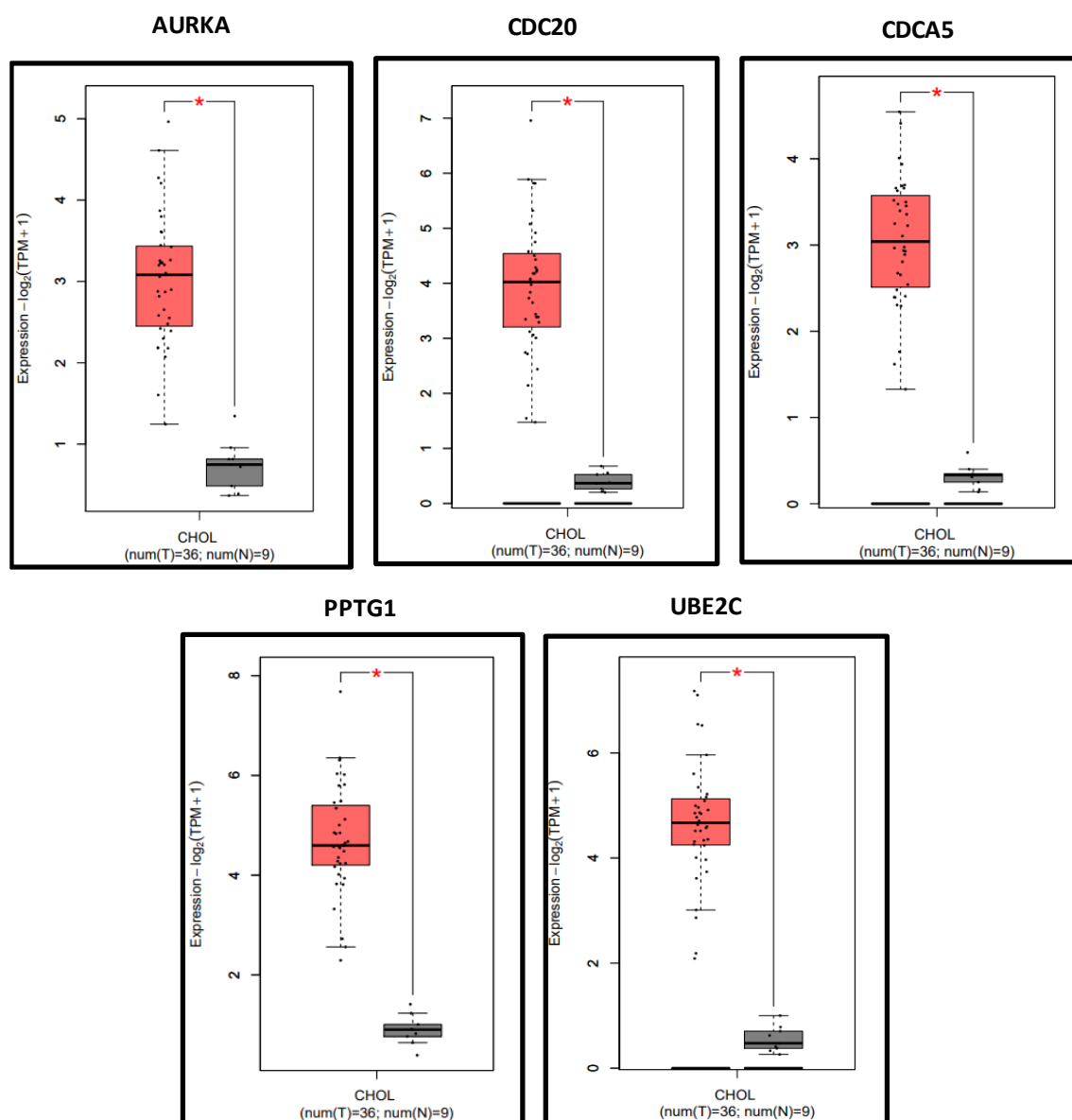


Figure 09 : La vérification de l'expression des gènes clés dans les tissus de cancer colorectal par l'outil GEPIA2, (LUSC : lung squamous cell carcinoma).

3.1.7 Analyse de la régulation transcriptionnelle des gènes sélectionnés

Un réseau de régulation gène–facteur de transcription (TF) a été construit pour les gènes sélectionnés, et les résultats de cette analyse ont été convertis en tableau afin de faciliter leur visualisation et interprétation. Ce réseau a révélé que plusieurs facteurs de transcription régulent plus d'un gène clé, ce qui suggère des interactions étroites entre ces TFs et les gènes clés (DEGs). Plus précisément, AURKA est régulé par quatre TFs (E2F1, E2F3, OTX2, MED1), PTTG1 par trois TFs (TP53, SP1, KLF6), UBE2C par deux TFs (MED1, MYC), et CDC20 par deux TFs (PHF8, YBX1).

Les résultats détaillés de cette analyse sont présentés dans le tableau ci-dessous.

Tableau 06 : Facteurs de transcription régulant les gènes clés identifiés.

Gènes	Facteur de transcription
AURKA	E2F1, E2F3, OTX2, MED1
PTTG1	TP53, SP1, KLF6
UBE2C	MED1, MYC
CDC20	PHF8, YBX1

Discussion

4. Discussion

Le cancer colorectal (CCR) représente une pathologie complexe caractérisée par une grande diversité génétique et moléculaire, ce qui complique considérablement son étude et son traitement. Cette hétérogénéité résulte principalement de l'accumulation de mutations dans des gènes cruciaux impliqués dans des processus biologiques fondamentaux tels que la prolifération cellulaire, la réparation de l'ADN et la signalisation cellulaire [53]. Dans notre étude, l'analyse fonctionnelle des gènes surexprimés dans le cancer colorectal a révélé un enrichissement significatif dans plusieurs voies biologiques clés : *negative regulation of growth*, *collagen-containing extracellular matrix*, *carbonate dehydratase activity*.

Des études ultérieures ont montré que la régulation négative de la croissance (*negative regulation of growth*) est un mécanisme essentiel empêchant la prolifération cellulaire incontrôlée, notamment dans le cancer colorectal. Des gènes suppresseurs comme p53 et TGF- β interviennent en bloquant le cycle cellulaire, en cas d'anomalies [54]. Le TGF- β joue un rôle crucial en inhibant la croissance des cellules anormales au début de la tumorigenèse, mais peut être inactivé par des mutations [55]. De même, la perte du gène APC, inhibiteur de la voie Wnt/ β -caténine, entraîne une prolifération anormale [56]. L'inactivation combinée de p53, APC et TGF- β favorise la progression vers des formes plus agressives du cancer [57].

La matrice extracellulaire (MEC) contenant du collagène joue un rôle crucial dans le maintien de l'homéostasie tissulaire et la régulation des interactions cellulaires dans des conditions normales. En effet, Il a été suggéré que le collagène fournit une architecture structurale essentielle pour les cellules épithéliales et mésenchymateuses, facilitant ainsi leur adhérence, leur migration et leur différenciation. Cependant, dans le contexte du cancer colorectal, cette matrice peut subir des modifications profondes qui favorisent la progression tumorale [58]. Selon Kalluri (2016), les altérations de la MEC, notamment la surproduction de collagène et sa réorganisation par des enzymes comme les métalloprotéinases matricielles (MMPs), créent un microenvironnement propice à l'invasion et à la dissémination des cellules cancéreuses [59]. De plus, Erler et *al.*, (2009) ont prouvé que la rigidité accrue de la MEC, induite par une accumulation anormale de collagène, stimule la signalisation mécanique via des intégrines, ce qui peut activer des voies oncogéniques telles que la voie FAK/PI3K/Akt, favorisant ainsi la prolifération et la survie des cellules tumorales [60].

Notre étude a montré que les gènes sous-exprimés étaient majoritairement associés à des processus fondamentaux tels que : *cell division*, *mitotic cell cycle*, *apoptotic process*, *nucleoplasm*, *protein binding* et *molecular function activator activity*. Une étude ultérieure a affirmé que le processus de division cellulaire (*cell division*) est essentiel pour la prolifération cellulaire normale, mais des anomalies dans sa régulation, souvent causées par des mutations dans des gènes, comme KRAS et PIK3CA, favorisent une prolifération incontrôlée dans le cancer colorectal [56]. De plus, Lamond et *al.*, (1998) ont souligné que le nucléoplasme constitue un composant principal du noyau cellulaire, où se produisent de nombreuses réactions biochimiques essentielles à la régulation de l'expression génique [61]. Les altérations de ces voies peuvent perturber les processus cellulaires normaux, entraînant une division cellulaire incontrôlée. En outre, Markowitz et *al.*, (2009) ont montré que le cycle cellulaire est perturbé par des mutations dans des gènes clés, tels que APC et TP53, entraînant une instabilité génomique caractéristique de ce type de cancer [57].

Jones et *al.*, (2008), ont montré l'interaction des protéines (*protein binding*) intervient dans les interactions moléculaires nécessaires à la signalisation cellulaire et à la stabilité des complexes protéiques [62]. Lorsque ces interactions sont perturbées, cela peut conduire à une perte de contrôle des mécanismes de réparation de l'ADN, favorisant ainsi la formation tumorale [54]. Les fonctions moléculaires liées à l'activité d'activation (*molecular function activator activity*) jouent un rôle crucial dans la régulation des processus biologiques normaux, notamment en initiant et modulant les voies de signalisation cellulaire. À l'état normal, ces fonctions permettent de maintenir l'homéostasie tissulaire et de contrôler des processus essentiels comme la prolifération cellulaire et la différenciation [63]. Cependant, dans le cancer colorectal, une dérégulation de ces activités peut entraîner une activation excessive de voies oncogéniques. Selon Xia P. et *al.*, (2015), l'activation anormale des kinases activatrices, telles que MAPK et PI3K, est fréquemment observée dans le cancer colorectal, favorisant ainsi la survie et la prolifération des cellules cancéreuses [64]. De plus, Wang et Lin. (2008) ont confirmé que l'activation des facteurs de transcription, comme AP-1, peut contribuer à la progression tumorale en augmentant l'expression des gènes pro-inflammatoires et pro-angiogéniques [65]. Il a été confirmé que le processus apoptotique (*apoptotic process*), un mécanisme crucial pour éliminer les cellules endommagées, est inhibé dans le cancer colorectal. Cela permet la survie des cellules tumorales malgré les dommages cellulaires, souvent à cause d'une surexpression de protéines anti-apoptotiques comme BCL-2 [66].

L'analyse KEGG a montré que les DEG communs étaient principalement enrichis dans : *cell cycle*, *metabolic pathways*, *drug metabolism – cytochrome P450*. Malumbres et al. (2009) ont démontré que le cycle cellulaire joue un rôle fondamental dans la régulation des processus cellulaires essentiels, notamment la prolifération, la réparation de l'ADN et la mort cellulaire programmée. Cette voie est contrôlée par des protéines spécifiques appelées cyclines et kinases dépendantes des cyclines (CDKs), qui agissent comme des régulateurs clés pour assurer une progression ordonnée du cycle cellulaire [67]. Les altérations dans cette voie de signalisation peuvent entraîner une prolifération cellulaire incontrôlée et une résistance à l'apoptose, facilitant ainsi la progression tumorale [56]. Par exemple, des mutations affectant les gènes suppresseurs de tumeurs, comme TP53 ou APC, sont fréquemment observées dans le cancer colorectal, ce qui contribue à la dérégulation de la division cellulaire [57].

Les voies métaboliques (*metabolic pathways*) jouent un rôle central dans la régulation des fonctions biologiques de l'organisme. En conditions physiologiques normales, les cellules s'appuient sur des processus tels que le cycle de Krebs (cycle TCA) et la phosphorylation oxydative pour la production d'ATP, ainsi que sur des voies impliquées dans la biosynthèse des protéines, des lipides et des acides nucléiques. Cependant, dans les cellules cancéreuses, ces voies subissent des réajustements métaboliques profonds afin de soutenir une prolifération rapide et d'assurer leur survie dans un microenvironnement souvent hostile. DeBerardinis et Chandel. (2016) ont mis en évidence que les cellules tumorales réorientent leur métabolisme, notamment celui du glucose et de la glutamine, afin de soutenir les processus biosynthétiques et de maintenir l'homéostasie redox intracellulaire [68]. Ce phénomène s'inscrit dans le cadre des propriétés fondamentales du cancer décrit par Hanahan et Weinberg. (2011) parmi lesquelles figure la reprogrammation du métabolisme énergétique, désormais considérée comme une caractéristique distinctive des cellules tumorales [54]. Ces altérations métaboliques ne se limitent pas à la promotion de la croissance tumorale, mais influencent également la réponse aux traitements, notamment en modulant l'activité des enzymes impliquées dans le métabolisme des médicaments, telles que les enzymes du cytochrome P450 [69]. En plus, Il a été prouvé que les variations génétiques affectant les gènes du cytochrome P450 peuvent altérer l'efficacité de la chimiothérapie et accroître la toxicité médicamenteuse chez certains patients [70]. De plus, une partie de ces polymorphismes est associée à une résistance accrue des tumeurs aux traitements chimiothérapeutiques [71]. Enfin, la variabilité génétique du cytochrome P450 influence considérablement les réponses thérapeutiques ainsi que le profil de toxicité chez les patients cancéreux [72].

L'étude des interactions PPI a révélée 10 gènes centraux y compris NEK2, CDCA5, UBE2C, PTTG1, MELK, BUB1, AURKA, CDC20, TPX2, et CDK1. La base des données GEPIA2 a montré que cinq gènes AURKA, PTTG1, CDCA5, UBE2C et CDC20 étaient significativement exprimés de manière différentielle entre les tissus tumoraux et les tissus normaux. L'expression de ces cinq gènes clés peut indiquer que les chances de survie des patients atteints de cancer colorectal sont faibles.

AURKA (STK15/BTAK) est un membre de la famille des sérine/thréonine kinases impliquée dans l'entrée en mitose, la formation du fuseau bipolaire, le contrôle de la maturation du centrosome et la ségrégation chromosomique pendant la mitose [73]. Le rôle des Aurora kinases dans la mitose et la tumorigénèse est bien documenté [74]. AURKA est nécessaire à l'entrée en mitose, à l'alignement des chromosomes et à la cytokinèse, et son fonctionnement anormal peut entraîner une division cellulaire aberrante et une aneuploïdie, ce qui augmente à son tour l'instabilité génomique et contribue à la carcinogenèse [75].

CDCA5, ou protéine associée au cycle cellulaire 5, est essentielle à la cohésion et à la ségrégation des chromatides sœurs [76]. CDCA5 maintient le complexe cohésine, préservant la cohésion des chromatides sœurs et assurant une ségrégation chromosomique précise pendant la mitose et la méiose. Elle est également nécessaire à la réparation de l'ADN [77]. De plus, CDCA5 affecte l'activité des facteurs de transcription et des protéines associées au cycle cellulaire, favorisant ainsi la croissance des cellules cancéreuses et leur participation à l'apoptose[78]. Elle participe également à diverses voies de signalisation, telles que la voie ERK et la voie PI3k/AKT/mTOR[79]. Ainsi, CDCA5 pourrait être une cible potentielle pour le traitement du cancer colorectal en inhibant sa fonction [80].

Le gène CDC20, également désigné par "*Cell Division Cycle Protein 20*", appartient à la famille des protéines cyclines impliquées dans la régulation du cycle cellulaire. Il joue un rôle essentiel dans l'activation du complexe anaphase-promoting (APC) et la formation du complexe ubiquitine-ligase E3 APC, favorisant ainsi la progression de la mitose par la dégradation de régulateurs clés du cycle cellulaire [81]. CDC20 agit comme un oncogène par diverses voies, notamment la promotion de la transition épithélio-mésenchymateuse (EMT), la suppression de l'apoptose, l'inhibition de l'infiltration immunologique et les voies de signalisation. CDC20 contribue significativement à la progression du cancer colorectal, et le développement d'inhibiteurs spécifiques de CDC20 pourrait représenter une avancée précieuse dans le traitement et l'intervention contre le cancer[82] .

Le gène PTTG1 (*Pituitary Tumor-Transforming Gene 1*), aussi appelé "*securin*", code pour une protéine clé dans la régulation du cycle cellulaire, la séparation des chromosomes et l'inhibition de l'apoptose [83]. PTTG1 intervient dans plusieurs voies de signalisation, notamment les voies TGF- β et PI3K/AKT, favorisant ainsi la transition épithélio-mésenchymateuse (EMT), l'invasion et la métastase des cellules cancéreuses [84]. Sur le plan diagnostique, le dosage de PTTG1 peut servir de biomarqueur pour prédire l'agressivité et l'évolution du cancer colorectal, en particulier aux stades précoces [85].

L'UBE2C (*Ubiquitin-conjugating enzyme E2 C*), également connu sous le nom d'UBCH10, code pour une enzyme clé dans le système ubiquitine-protéasome, qui régule la dégradation des protéines du cycle cellulaire. Il facilite la mitose en ciblant des protéines spécifiques pour leur dégradation[86]. Il module également l'activité de la kinase Aurora A, favorisant ainsi une division cellulaire non contrôlée. De plus, l'UBE2C interfère avec l'apoptose en influençant l'expression de protéines pro- et anti-apoptotiques, contribuant à la résistance au traitement. Ces rôles sont associés à des anomalies telles que l'amplification du chromosome 20q [87]. En plus, il est considéré comme un biomarqueur prometteur pour la prédiction de la survie et la distinction des tissus cancéreux des tissus normaux [88]. Son inhibition représente une stratégie potentielle, en particulier contre les tumeurs résistantes, en bloquant la progression du cycle cellulaire et en induisant l'apoptose [89].

En résumé, la présente étude a identifié un ensemble de cinq gènes clés, dont AURKA, CDCA5, UBE2C, PTTG1 et CDC20, qui étaient partiellement différents de ceux identifiés lors de précédentes études sur les biomarqueurs du cancer colorectal. Tous ces gènes étaient significativement associés à une faible survie des patients atteints du cancer colorectal, et leurs voies étaient principalement enrichies dans cette maladie. Ces résultats offrent des informations clés sur la progression du cancer colorectal et le développement de nouvelles thérapies anticancéreuses potentielles.

Conclusion

5. 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 440 gènes communs entre les trois profils d'expression des patients, dont 277 gènes surexprimés et 163 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 (AURKA, BUB1, UBE2C, PTTG1, CDC20, CDCA5, TPX2, MELK, CDK1, NEK2).
- Cinq gènes clés (AURKA, CDCA5, UBE2C, PTTG1, CDC20) ont été confirmés comme étant exprimés de manière différentielle entre les tissus cancéreux et les tissus normaux. L'expression de ces cinq gènes est associée à la mauvaise survie des patients atteints de cancer colorectal.
- Facteurs de transcription identifiés tel que : E2F1, E2F3, OTX2, SP1, MED1, TP53, MYC, PHF8, et YBX1.

Nos résultats indiquent que les cinq 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.

Glossaire

Adénocarcinome : est une tumeur maligne dérivée des cellules glandulaires de l'épithélium, caractérisée par sa capacité à produire des substances de type glandulaire telles que le mucus ou des enzymes.

Cytoscape : est un logiciel open-source utilisé pour visualiser des réseaux moléculaires complexes et intégrer ces réseaux à des données d'expression génomique et d'autres données biologiques.

DAVID : est une plateforme bioinformatique qui fournit des outils pour l'annotation fonctionnelle et l'enrichissement des gènes afin d'interpréter les données génomiques à un niveau biologique systémique.

Fold Change : est un indicateur qui mesure le rapport d'expression d'un gène entre deux conditions biologiques différentes, indiquant une augmentation ou une diminution relative de l'expression

Gene Ontology : est une ressource bioinformatique qui décrit les fonctions des gènes à travers trois catégories principales : les processus biologiques, les fonctions moléculaires et les composants cellulaires.

GEO : est une base de données publique du NCBI qui archive et partage des données sur l'expression des gènes obtenues par diverses technologies comme les microarrays et le séquençage à haut débit.

GEPIA2 : est un outil en ligne interactif pour l'analyse du profil d'expression des gènes dans les bases de données TCGA et GTEx, permettant la comparaison entre tissus normaux et tumoraux

KEGG : est une ressource bioinformatique qui relie les informations génétiques aux fonctions biologiques à travers des cartes de voies métaboliques et de réseaux cellulaires.

Kaplan-Meier Plotter : est un outil en ligne permettant d'évaluer l'impact de l'expression des gènes sur la survie des patients atteints de cancer, en utilisant des courbes de survie de Kaplan-Meier

NCBI : est une institution américaine qui fournit des bases de données biologiques majeures, y compris *PubMed*, *GenBank*, et *GEO*, pour soutenir la recherche biomédicale.

P- value : est une mesure statistique utilisée pour déterminer la signification des résultats expérimentaux. Une valeur p faible indique que l'observation est peu probable due au hasard.

Venn Diagram: est un graphique qui montre les relations logiques et les chevauchements entre différents ensembles de données, souvent utilisé en biologie pour visualiser des gènes partagés.

volcano plot :est un graphique de dispersion qui affiche la relation entre la significativité statistique (valeur p) et l'ampleur du changement (*fold change*) pour des gènes ou des protéines

Références bibliographiques

Références bibliographiques :

1. Siegel, R.L., et al., *Cancer statistics*, 2023. CA: a cancer journal for clinicians, 2023. **73**(1): p. 17-48.
2. Aiman, W., et al., *BRAF inhibitors in BRAF-mutated colorectal cancer: A systematic review*. Journal of Clinical Medicine, 2023. **13**(1): p. 113.
3. Tabernero, J., J. Ros, and E. Élez. *The Evolving Treatment Landscape in BRAF-V600E-Mutated Metastatic Colorectal Cancer*. in *American Society of Clinical Oncology educational book. American Society of Clinical Oncology. Annual Meeting*, 2022.
4. ER, F., *A genetic model for colorectal tumorigenesis*. Cell, 1990. **61**.
5. Xu, Y. and B. Pasche, *TGF- β signaling alterations and susceptibility to colorectal cancer*. Human molecular genetics, 2007. **16**(R1): p. R14-R20.
6. Chan, A.T., et al., *Aspirin dose and duration of use and risk of colorectal cancer in men*. Gastroenterology, 2008. **134**(1): p. 21-28.
7. Zauber, A.G., et al., *Colonoscopic polypectomy and long-term prevention of colorectal-cancer deaths*. New England Journal of Medicine, 2012. **366**(8): p. 687-696.
8. Moses, H.L., J.K. Phillips, and L.A. Graig, *Biomarker tests for molecularly targeted therapies: key to unlocking precision medicine*. 2016.
9. Markowitz, S.D. and M.M. Bertagnolli, *Molecular basis of colorectal cancer molecular origins of cancer*. The New England Journal of Medicine, 2009. **361**(25): p. 2449.
10. Sung, H., et al., *Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA: a cancer journal for clinicians, 2021. **71**(3): p. 209-249.
11. Bray, F., et al., *Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries*. CA: a cancer journal for clinicians, 2024. **74**(3): p. 229-263.
12. Abderrahmane, R., et al., *TP53 Arg 72Pro and MDM2 SNP309 polymorphisms and colorectal cancer risk: a west Algerian population study*. Pathology & oncology research, 2015. **21**: p. 629-635.
13. Samira, B.A. and M. Kaouel, *Survie sur 05 ans des Cancers Colorectaux dans la Wilaya de Tlemcen*. Algerian Journal of Health Sciences: p. 28.
14. MAZOUZI, C., R. BENYAHIA, and N. FERRAH, *Epidemiology of Colorectal Cancer in Algeria*.
15. FREBOURG, T., et al., *Le cancer colorectal héréditaire non polyposique: définition, génétique, diagnostic et surveillance médicale*. Gastroentérologie clinique et biologique, 2003. **27**(8-9): p. 708-714.
16. Jiricny, J., *The multifaceted mismatch-repair system*. Nature reviews Molecular cell biology, 2006. **7**(5): p. 335-346.
17. Björk, J., *Epidemiology of familial adenomatous polyposis in Sweden: changes over time and differences in phenotype between males and females*. Scandinavian journal of gastroenterology, 1999. **34**(12): p. 1230-1235.
18. Aoki, K. and M.M. Taketo, *Adenomatous polyposis coli (APC): a multi-functional tumor suppressor gene*. Journal of cell science, 2007. **120**(19): p. 3327-3335.
19. Half, E., D. Bercovich, and P. Rozen, *Familial adenomatous polyposis*. Orphanet journal of rare diseases, 2009. **4**: p. 1-23.
20. Liang, P.S., T.Y. Chen, and E. Giovannucci, *Cigarette smoking and colorectal cancer incidence and mortality: Systematic review and meta-analysis*. International journal of cancer, 2009. **124**(10): p. 2406-2415.
21. Shah, S.C. and S.H. Itzkowitz, *Colorectal cancer in inflammatory bowel disease: mechanisms and management*. Gastroenterology, 2022. **162**(3): p. 715-730. e3.
22. Chanson, P. and S. Salenave, *Acromegaly*. Orphanet journal of rare diseases, 2008. **3**: p. 1-17.
23. Marafini, I. and G. Monteleone, *Smoking and colorectal cancer in inflammatory bowel disease: Quantity matters?* United European Gastroenterology Journal, 2023. **11**(7): p. 589.

Références bibliographiques

24. Larsson, S.C. and A. Wolk, *Meat consumption and risk of colorectal cancer: a meta-analysis of prospective studies*. International journal of cancer, 2006. **119**(11): p. 2657-2664.
25. Cross, A.J., et al., *A large prospective study of meat consumption and colorectal cancer risk: an investigation of potential mechanisms underlying this association*. Cancer research, 2010. **70**(6): p. 2406-2414.
26. Gordon, P.H. and S. Nivatvongs, *Neoplasms of the colon, rectum, and anus* 2007: CRC Press.
27. Rougier, P., *Cancers du côlon et du rectum: mieux les dépister et mieux les traiter*. La Revue du praticien, 2004. **54**: p. 133-142.
28. Viguier, J., et al., *Cancer du côlon*. Encycl méd Chir, Gastro-entérologie, 2003: p. 9-068.
29. LoConte, N.K., et al., *Alcohol and cancer: a statement of the American Society of Clinical Oncology*. Journal of Clinical Oncology, 2018. **36**(1): p. 83-93.
30. Cho, E., et al., *Alcohol intake and colorectal cancer: a pooled analysis of 8 cohort studies*. Annals of internal medicine, 2004. **140**(8): p. 603-613.
31. Ning, Y., L. Wang, and E. Giovannucci, *A quantitative analysis of body mass index and colorectal cancer: findings from 56 observational studies*. Obesity reviews, 2010. **11**(1): p. 19-30.
32. Matsuo, K., et al., *Association between body mass index and the colorectal cancer risk in Japan: pooled analysis of population-based cohort studies in Japan*. Annals of oncology, 2012. **23**(2): p. 479-490.
33. Soltani, G., et al., *Obesity, diabetes and the risk of colorectal adenoma and cancer*. BMC endocrine disorders, 2019. **19**: p. 1-10.
34. Ray, K., *Fusobacterium nucleatum found in colon cancer tissue—could an infection cause colorectal cancer?* Nature reviews Gastroenterology & hepatology, 2011. **8**(12): p. 662-662.
35. Sears, C.L. and D.M. Pardoll, *Perspective: alpha-bugs, their microbial partners, and the link to colon cancer*. Journal of Infectious Diseases, 2011. **203**(3): p. 306-311.
36. Kostic, A.D., et al., *Genomic analysis identifies association of Fusobacterium with colorectal carcinoma*. Genome research, 2012. **22**(2): p. 292-298.
37. Koya, A.I. and S.A. Ibrahim, *Carcinogenesis*, 2024.
38. Pitot, H.C. and Y.P. Dragan, *Facts and theories concerning the mechanisms of carcinogenesis*. The FASEB journal, 1991. **5**(9): p. 2280-2286.
39. Barrett, J.C., *Mechanisms of multistep carcinogenesis and carcinogen risk assessment*. Environmental health perspectives, 1993. **100**: p. 9-20.
40. Boutegrabet, W., *Histopathologie spectrale du cancer colique: développements d'outils chimiométriques pour le traitement automatisé des images tissulaires en spectroscopie moyen-infrarouge*, 2022, Université de Strasbourg.
41. Lopez, J.-B.R., *Mécanismes moléculaires et cellulaires de la formation des TSIP: médiateurs de la dissémination des cancers colorectaux*, 2022, Université Paris-Saclay.
42. Caravati-Jouveanceaux, A., *Les patients guéris d'un cancer colorectal retrouvent-ils à long terme un niveau de qualité de vie et de réinsertion comparable à celui de témoins de population?*, 2012, Université de Franche-Comté.
43. Stracci, F., M. Zorzi, and G. Grazzini, *Colorectal cancer screening: tests, strategies, and perspectives*. Frontiers in public health, 2014. **2**: p. 210.
44. Seitz, J., G. Houvenaeghel, and L. Dahan, *Tumeurs du colon et du rectum (148)*. 2007.
45. Świdarska, M., et al., *The diagnostics of colorectal cancer*. Contemporary Oncology/Współczesna Onkologia, 2014. **18**(1): p. 1-6.
46. Levin, B., et al., *Screening and surveillance for the early detection of colorectal cancer and adenomatous polyps, 2008: a joint guideline from the American Cancer Society, the US Multi-Society Task Force on Colorectal Cancer, and the American College of Radiology*. Gastroenterology, 2008. **134**(5): p. 1570-1595.
47. Lee, J.K., et al., *Accuracy of fecal immunochemical tests for colorectal cancer: systematic review and meta-analysis*. Annals of internal medicine, 2014. **160**(3): p. 171-181.
48. Imperiale, T.F., et al., *Multitarget stool DNA testing for colorectal-cancer screening*. New England Journal of Medicine, 2014. **370**(14): p. 1287-1297.

Références bibliographiques

49. Rex, D.K., et al., *Quality indicators for colonoscopy*. Official journal of the American College of Gastroenterology| ACG, 2006. **101**(4): p. 873-885.
50. Atkin, W.S., et al., *Once-only flexible sigmoidoscopy screening in prevention of colorectal cancer: a multicentre randomised controlled trial*. The Lancet, 2010. **375**(9726): p. 1624-1633.
51. Gude, S.S., et al., *Colorectal cancer diagnostic methods: the present and future*. Cureus, 2023. **15**(4).
52. Taylor, F.G., et al., *Preoperative magnetic resonance imaging assessment of circumferential resection margin predicts disease-free survival and local recurrence: 5-year follow-up results of the MERCURY study*. Journal of Clinical Oncology, 2014. **32**(1): p. 34-43.
53. Holubekova, V., et al., *Differential gene expression of immunity and inflammation genes in colorectal cancer using targeted RNA sequencing*. Frontiers in Oncology, 2023. **13**: p. 1206482.
54. Hanahan, D. and R.A. Weinberg, *Hallmarks of cancer: the next generation*. Cell, 2011. **144**(5): p. 646-674.
55. Neuzillet, C., et al., *Targeting the TGF β pathway for cancer therapy*. Pharmacology & therapeutics, 2015. **147**: p. 22-31.
56. Fearon, E.R. and B. Vogelstein, *A genetic model for colorectal tumorigenesis*. Cell, 1990. **61**(5): p. 759-767.
57. Markowitz, S.D. and M.M. Bertagnolli, *Molecular basis of colorectal cancer*. New England Journal of Medicine, 2009. **361**(25): p. 2449-2460.
58. Ricard-Blum, S., *The collagen family*. Cold Spring Harbor perspectives in biology, 2011. **3**(1): p. a004978.
59. Kalluri, R., *The biology and function of fibroblasts in cancer*. Nature Reviews Cancer, 2016. **16**(9): p. 582-598.
60. Erler, J.T., et al., *Hypoxia-induced lysyl oxidase is a critical mediator of bone marrow cell recruitment to form the premetastatic niche*. Cancer cell, 2009. **15**(1): p. 35-44.
61. Lamond, A.I. and W.C. Earnshaw, *Structure and function in the nucleus*. Science, 1998. **280**(5363): p. 547-553.
62. Jones, S., et al., *Core signaling pathways in human pancreatic cancers revealed by global genomic analyses*. Science, 2008. **321**(5897): p. 1801-1806.
63. Robinson, M.J. and M.H. Cobb, *Mitogen-activated protein kinase pathways*. Current opinion in cell biology, 1997. **9**(2): p. 180-186.
64. Xia, P. and X.-Y. Xu, *PI3K/Akt/mTOR signaling pathway in cancer stem cells: from basic research to clinical application*. American journal of cancer research, 2015. **5**(5): p. 1602.
65. Wang, X. and Y. Lin, *Tumor necrosis factor and cancer, buddies or foes? 1*. Acta Pharmacologica Sinica, 2008. **29**(11): p. 1275-1288.
66. Watson, A., *Apoptosis and colorectal cancer*. Gut, 2004. **53**(11): p. 1701-1709.
67. Malumbres, M. and M. Barbacid, *Cell cycle, CDKs and cancer: a changing paradigm*. Nature Reviews Cancer, 2009. **9**(3): p. 153-166.
68. DeBerardinis, R.J. and N.S. Chandel, *Fundamentals of cancer metabolism*. Science advances, 2016. **2**(5): p. e1600200.
69. Zanger, U.M. and M. Schwab, *Cytochrome P450 enzymes in drug metabolism: regulation of gene expression, enzyme activities, and impact of genetic variation*. Pharmacology & therapeutics, 2013. **138**(1): p. 103-141.
70. Bethke, L., et al., *Polymorphisms in the cytochrome p 450 genes CYP1A2, CYP1B1, CYP3A4, CYP3A5, CYP11A1, CYP17A1, CYP19A1 and colorectal cancer risk*. BMC cancer, 2007. **7**: p. 1-7.
71. De Chaisemartin, L. and M.-A. Lorient, *Pharmacogénétique des médicaments anticancéreux*. Pathologie biologie, 2005. **53**(2): p. 116-124.
72. Ingelman-Sundberg, M., *Genetic polymorphisms of cytochrome P450 2D6 (CYP2D6): clinical consequences, evolutionary aspects and functional diversity*. The pharmacogenomics journal, 2005. **5**(1): p. 6-13.
73. Katsha, A., et al., *Aurora kinase A in gastrointestinal cancers: time to target*. Molecular cancer, 2015. **14**: p. 1-13.

Références bibliographiques

74. Dotan, E., et al., *Relationship of increased aurora kinase A gene copy number, prognosis and response to chemotherapy in patients with metastatic colorectal cancer*. British journal of cancer, 2012. **106**(4): p. 748-755.
75. Goktas, S., et al., *Prognostic role of Aurora-A expression in metastatic colorectal cancer patients*. J buon, 2014. **19**(19): p. 686-91.
76. Zhang, N. and D. Pati, *Sororin is a master regulator of sister chromatid cohesion and separation*. Cell cycle, 2012. **11**(11): p. 2073-2083.
77. Zhang, N. and D. Pati, *Handcuff for sisters: a new model for sister chromatid cohesion*. Cell cycle, 2009. **8**(3): p. 399-402.
78. Zhang, N., et al., *Interaction of Sororin protein with polo-like kinase 1 mediates resolution of chromosomal arm cohesion*. Journal of Biological Chemistry, 2011. **286**(48): p. 41826-41837.
79. Fu, G., et al., *CDCA5 functions as a tumor promoter in bladder cancer by dysregulating mitochondria-mediated apoptosis, cell cycle regulation and PI3k/AKT/mTOR pathway activation*. Journal of Cancer, 2020. **11**(9): p. 2408.
80. Bao, X., et al., *Integrated multi-omics analyses identify CDCA5 as a novel biomarker associated with alternative splicing, tumor microenvironment, and cell proliferation in colon cancer via pan-cancer analysis*. Journal of Cancer, 2024. **15**(3): p. 825.
81. He, W. and J. Meng, *CDC20: A novel therapeutic target in cancer*. American journal of translational research, 2023. **15**(2): p. 678.
82. Xian, F., et al., *The potential role of CDC20 in tumorigenesis, cancer progression and therapy: A narrative review*. Medicine, 2023. **102**(36): p. e35038.
83. Ren, Q. and B. Jin, *The clinical value and biological function of PTTG1 in colorectal cancer*. Biomedicine & Pharmacotherapy, 2017. **89**: p. 108-115.
84. Lin, P.-C., et al., *Identification of phosphorylated cyclin-dependent kinase 1 associated with colorectal cancer survival using label-free quantitative analyses*. PLoS One, 2016. **11**(7): p. e0158844.
85. Cao, F., Y.-Y. Chen, and H.-C. Wang, *GLI1 and PTTG1 expression in colorectal carcinoma patients undergoing radical surgery and their correlation with lymph node metastasis*. World Journal of Gastrointestinal Surgery, 2024. **16**(5): p. 1328.
86. Wagner, S.A., et al., *A proteome-wide, quantitative survey of in vivo ubiquitylation sites reveals widespread regulatory roles*. Molecular & Cellular Proteomics, 2011. **10**(10): p. M111. 013284.
87. Liu, P.-F., et al., *UBE2C is a potential biomarker for tumorigenesis and prognosis in tongue squamous cell carcinoma*. Diagnostics, 2020. **10**(9): p. 674.
88. Jalali, P., et al., *UBE2C: A pan-cancer diagnostic and prognostic biomarker revealed through bioinformatics analysis*. Cancer Reports, 2024. **7**(4): p. e2032.
89. Wang, Z., et al., *Roles of F-box proteins in cancer*. Nature Reviews Cancer, 2014. **14**(4): p. 233-247.