



الجمهورية الجزائرية الديمقراطية
REPUBLIQUE ALGERIENNE DEMOCRATIQUE ET POPULAIRE



وزارة التعليم العالي والبحث العلمي
MINISTERE DE L'ENSEIGNEMENT SUPERIEUR ET DE LA RECHERCHE SCIENTIFIQUE
جامعة باجي مختار - عنابة

UNIVERSITE BADJI MOKHTAR-ANNABA

FACULTE DES SCIENCES

DEPARTEMENT DE BIOLOGIE

LABORATOIRE DE BIOSURVEILLANCE ENVIRONNEMENTALE

THÈSE

Présentée en vue de l'obtention du diplôme de Doctorat

DOMAINE : Science de la Nature et de la Vie

FILIERE : SCIENCES BIOLOGIQUES

SPECIALITE: PHYSIOPATHOLOGIE

THEME

**Analyse de l'Expression des Facteurs
Immuno- Hormonaux dans le Cancer du Sein**

Présentée par : Mme BOULMAIZ Amel

Membres de Jury :

Mr FRIH Hacene	(Prof)	Président de jury	UBMA
Mme BERREDJEM Hajira	(Prof)	Directrice de thèse	UBMA
Mr REZGOUNE Mohamed Larbi	(Prof)	Examinateur	Univ. Constantine
Mme HARBI Karima	(MCA)	Examinatrice	UBMA

Année : 2024/2025

الحمد لله الذي وفقني وأعانني على إتمام هذا العمل، والذي منحني القوة والصبر في
مسيرتي العلمية. إليه أرفع شكري وحمدي

REMERCIEMENTS

Un immense Merci à Madame le Professeur BERREDJEM Hajira, ma directrice de thèse, pour son soutien inestimable tout au long de mon parcours. Je lui suis profondément reconnaissante de m'avoir offert l'opportunité de réaliser cette thèse, mais également pour le temps, la confiance et l'encadrement précieux qu'elle m'a accordés au fil des années. Merci d'avoir veillé sur moi, pour tout ce que vous m'avez appris, pour vos conseils avisés, votre disponibilité et votre exigence bienveillante. Plus que tout, je vous remercie de m'avoir transmis votre passion pour la recherche et de m'avoir encouragée à aller toujours plus loin. Je suis honorée de faire partie de votre équipe et vous suis infiniment reconnaissante d'avoir cru en moi.

Je tiens à remercier les membres du jury qui m'ont fait l'honneur d'évaluer mon travail de thèse.

Je remercie tout particulièrement le Professeur FRIH Hacène, président du jury, un enseignant d'exception qui m'a transmis son savoir avec rigueur et bienveillance. Son expertise et son engagement académique sont une source d'inspiration inestimable. Merci également pour nous avoir donné la chance d'accéder à la formation doctorale élaborée par vos soins.

Je remercie également le Professeur HARBI Karima et le Professeur REZGOUNE Mohamed Larbi d'avoir accepté d'être les rapporteurs de cette thèse. Leur rigueur scientifique et leurs précieuses analyses ont enrichi ce travail.

*Un remerciement particulier au **Professeur OUALI Kheireddine**, directeur du laboratoire de biosurveillance environnementale de m'avoir accueilli dans son laboratoire. Votre rigueur scientifique et votre engagement sont une source d'inspiration pour moi et toute l'équipe.*

*J'exprime ma profonde gratitude à **Dr CHEIKCHOUK Khadidja**, membre invitée du jury, une personne admirable avec qui j'ai eu le privilège de travailler en collaboration, et dont le travail est toujours accompli avec le sourire et un professionnalisme exemplaire.*

*Je tiens également à remercier le **Professeur DJEDI Hanene**, Directrice du Centre de lutte contre le cancer d'Annaba, pour son soutien et son engagement en faveur de la recherche et des patients.*

*Enfin, une mention spéciale au **Dr CHENIKI Rachid** et **Madame Zina**, l'ingénieur de laboratoire de l'hôpital **El Bouni**, dont l'aide précieuse m'a été d'un grand soutien tout au long de ces années de thèse.*

DEDICACE

À la mémoire de mon père, pour sa confiance et ses conseils avisés.

À la mémoire de ma mère, pour son soutien inébranlable.

*Bien que partis trop tôt, je sais combien vous seriez fiers et heureux
de me voir accomplir ce chemin.*

*Merci à mon Mari et mes enfants, Eyed, Adem et Yasser, qui
illuminent mes journées par leurs espiègleries et leur bonne humeur.*

*Un grand merci à ma sœur Amina et mon frère Tayeb, qui sont
toujours à mes côtés dans les moments difficiles.*

Votre soutien et votre présence me sont précieux.

In God we trust, all others bring data.

(William Edwards Deming, 1900-1993)

Résumé

Ce travail de recherche s'inscrit dans le cadre de l'intégration de l'intelligence artificielle (IA) en oncologie. L'un des défis actuels consiste à identifier de nouveaux biomarqueurs moléculaires pour soutenir le développement de thérapies ciblées et personnalisées. Les méthodes de détection conventionnelles, bien qu'efficaces, restent souvent coûteuses et peuvent manquer de précision diagnostique et pronostique. L'objectif principal de ce travail est d'évaluer le potentiel des modèles d'IA dans la classification et la stratification des patientes atteintes de cancer du sein. Deux approches complémentaires ont été mises en œuvre. La première, une étude prospective, visait à prédire l'expression de HER2 à l'aide de modèles supervisés d'apprentissage automatique entraînés sur des données cliniques et biologiques de 109 patientes prises en charge dans le centre de lutte contre le cancer d'Annaba. Quatre modèles ont été développés et évalués. Les résultats ont montré des performances prometteuses, soulignant l'intérêt des outils algorithmiques pour améliorer la précision du diagnostic et limiter les erreurs des méthodes classiques. La deuxième approche a reposé sur une analyse non supervisée menée sur une cohorte de 531 patientes issues de la base TCGA. Cette étude multiomique a intégré des données de méthylation de l'ADN, d'expression génique et des variables cliniques afin de stratifier les patientes et d'identifier des sous-groupes aux profils moléculaires et cliniques distincts. L'analyse de clustering a permis d'identifier deux sous-groupes majeurs : le premier comprenait des patientes jeunes avec un profil à implication immunitaire, tandis que le second regroupait des patientes plus âgées, de phénotype hormonodépendant. Ces résultats ont mis en évidence une hétérogénéité moléculaire et clinique pouvant influencer la réponse thérapeutique. Une analyse de survie a également été réalisée à l'aide de modèles classiques et d'apprentissage profond. Le modèle de régression de Cox, ainsi que des réseaux de neurones artificiels (ANN) et convolutifs (CNN), ont été appliqués. Ces approches ont affiné la stratification pronostique et identifié des signatures associées à l'évolution de la maladie. L'ensemble des résultats renforce la pertinence de l'IA en oncologie, ouvrant la voie à une médecine plus personnalisée.

Mots-clés : Cancer du sein, Intelligence artificielle, Biomarqueurs moléculaires, Analyse multiomique, Modèles prédictifs.

Abstract

This research contributes to the integration of artificial intelligence (AI) in oncology. One of the current challenges lies in identifying novel molecular biomarkers to support the development of targeted and personalized therapies. Conventional detection methods, although reliable, often remain costly and may lack diagnostic and prognostic accuracy. The primary objective of this work is to assess the potential of AI models in classifying and stratifying breast cancer patients. Two complementary approaches were implemented. The first approach, based on a prospective study, aimed to predict HER2 expression using supervised machine learning models trained on clinical and biological data from 109 patients treated at a cancer center. Four models were developed and evaluated. The results showed promising performance, highlighting the potential of algorithmic tools to improve diagnostic accuracy and minimize errors inherent to conventional techniques.

The second approach involved an unsupervised analysis conducted on a cohort of 531 patients from the TCGA database. This multi-omic investigation combined DNA methylation profiles, gene expression data, and clinical variables to stratify patients and identify subgroups with distinct molecular and clinical characteristics. The clustering analysis identified two major subgroups. The first included younger patients with an immuno-involved profile, while the second involved older patients with a hormone-dependent phenotype. These findings revealed underlying molecular and clinical heterogeneity with potential therapeutic implications.

Survival analysis was also performed using both classical and deep learning models. A Cox proportional hazards model and advanced neural network architectures, including artificial neural networks (ANN) and convolutional neural networks (CNN), were applied. These models enhanced prognostic stratification and helped identify signatures linked to disease progression. Both the Cox-based and CNN-based approaches supported the relevance of AI in oncology, providing valuable perspectives for personalized medicine.

Keywords: Breast Cancer, Artificial Intelligence, Molecular Biomarkers, Multi-omics Analysis, Predictive Models.

ملخص

يهدف هذا العمل البحثي إلى دمج الذكاء الاصطناعي (AI) في مجال علم الأورام. من بين التحديات الحالية، تبرز أهمية تحديد المؤشرات الحيوية الجزيئية الجديدة لدعم تطوير علاجات موجهة وشخصية. رغم أن طرق الكشف التقليدية فعالة، إلا أنها غالبًا ما تكون مكلفة وقد تفتقر إلى الدقة في التشخيص والتنبؤ. الهدف الأساسي من هذا البحث هو تقييم قدرة نماذج الذكاء الاصطناعي على تصنيف المريضات المصابات بسرطان الثدي وتحديد الفئات الفرعية بينهن. تم اعتماد مقاربتين تكمليتين. ركزت الأولى، وهي دراسة استباقية، على التنبؤ بتعبير مستقبل HER2 باستخدام نماذج تعلم آلي خاضعة للإشراف، تم تدريبها على بيانات سريرية وبيولوجية من 109 مريضات في مركز لمكافحة السرطان. أظهرت النتائج أداءً واعدًا، مما يؤكد فائدة النماذج الحسابية في تحسين دقة التشخيص والحد من الأخطاء المرتبطة بالطرق التقليدية.

أما المقاربة الثانية، فاعتمدت على تحليل غير خاضع للإشراف أُجري على مجموعة من 531 مريضة من قاعدة بيانات TCGA. جمعت الدراسة متعددة الأوميكس بين بيانات مثيلة الحمض النووي، وتعبير الجينات، والمتغيرات السريرية، بهدف تصنيف المريضات واكتشاف مجموعات فرعية ذات خصائص جزيئية وسريرية متميزة. أظهر التحليل التجمي وجود مجموعتين رئيسيتين: الأولى تضم مريضات شابات بملامح مناعية، والثانية تشمل مريضات أكبر سنًا بنمط يعتمد على الهرمونات. تكشف هذه النتائج عن تباين بيولوجي وسريري قد يؤثر على الاستجابة للعلاج.

كما تم إجراء تحليل للبقاء على قيد الحياة باستخدام نماذج كلاسيكية ونماذج تعلم عميق. تم تطبيق نموذج كوكس للمخاطر، إلى جانب الشبكات العصبية الاصطناعية (ANN) والتلافيفية (CNN). أسهمت هذه النماذج في تحسين التنبؤ بمسار المرض وتحديد بصمات حيوية ترتبط بالتطور السريري. تؤكد النتائج مجددًا أهمية الذكاء الاصطناعي في تطوير طب شخصي أكثر دقة في مجال الأورام.

الكلمات المفتاحية: سرطان الثدي، الذكاء الاصطناعي، المؤشرات الحيوية الجزيئية، تحليل متعدد الأوميكس، النماذج التنبؤية.

Table des Matières

Liste des figures

Liste des tableaux

Liste des abréviations

Résumé

Abstract

ملخص

INTRODUCTION 1

REVUE DE LA BIBLIOGRAPHIE

1. LE CANCER DU SEIN	3
1.1. Anatomie du sein et de la glande mammaire	3
1.2. Physiologie du sein	4
1.2.1. Rôle de l'œstrogène et de la progestérone	7
1.3. Physiopathologie du cancer du sein	8
1.3.1. Bases moléculaires.....	9
1.3.2. Processus de la carcinogenèse mammaire	9
2. EPIDEMIOLOGIE	12
2.1. Dans le monde	12
2.2. En Algérie	12
2.3. Facteurs de risque	13
2.3.1. Les facteurs de risque non modifiables	13
2.3.2. Les facteurs de risque modifiables	14
3. CLASSIFICATION ET DIAGNOSTIC DU CANCER DU SEIN	14
3.1. Classification histologique	15
3.2. Classification moléculaire	17
3.3. Classification selon le stade TNM	19
3.4. Classification concordante TNM/Stade	21
4. DIAGNOSTIC, PRONOSTIC ET TRAITEMENT DU CANCER DU SEIN	22
4.1. Signes et symptômes cliniques	22
4.2. Diagnostic	22
4.3. Évaluation du Pronostic	23
4.4. Traitement	24
5. LES MARQUEURS TUMORAUX	24
5.1. L'œstrogène	25
5.2. La progestérone	28
5.3. Human Epidermal Growth Factor Receptor 2 (HER2)	29
5.4. Indexe de l'activité cellulaire proliférative: Ki-67	33
6. FACTEURS IMMUNITAIRES	34
6.1. Microenvironnement tumorale	34
6.2. PDCD1 (PD-1) et CD274 (PD-L1)	36
6.3. TLR4 et TLR9	37

6.4. CTLA4	38
6.5. NF-κB1	39
6.6. Interactions entre facteurs immunitaires et hormonaux	40
7. EPIGENETIQUE ET CANCER	42
7.1. Définition de l'épigénétique	42
7.2. Méthylation de l'ADN	43
7.3. Méthylation de l'ADN et cancers	45
7.3.1. Hypométhylation globale	47
7.3.2. Hyperméthylation des gènes suppresseurs de tumeurs	48
8. INTELLIGENCE ARTIFICIELLE ET IMPORTANCE DANS LES PATHOLOGIES	48
8.1. Systèmes de Support à la Décision Clinique (SSDC)	48
8.2. L'IA dans le domaine sanitaire, la médecine de précision et la gestion des ressources	49
8.3. Avantages de l'Intelligence Artificielle (IA)	51
9. OBJECTIFS	52
9.1. Objectif principal	52
9.2. Objectifs secondaires	52

MATERIEL ET METHODES

PREMIERE PARTIE : ETUDE PROSPECTIVE : MODELISATION PREDICTIVE SUPERVISEE DU STATUT HER2

1. DESCRIPTION DE LA POPULATION D'ETUDE	55
1.1. Les critères d'inclusion	55
1.2. Les critères d'exclusion	55
2. COLLECTE DES DONNEES	56
3. IMMUNOHISTOCHIMIE DES BIOMARQUEURS (ER, PR, HER2 et Ki-67)	56
3.1. Préparation des échantillons	57
3.2. Technique Immuno-histochimique (IHC)	57
4. TRAITEMENT DES DONNEES PAR METHODE SUPERVISEE	58
4.1. Prétraitement des Données	58
4.1.1. Nettoyage des données	59
4.1.2. Sélection des caractéristiques	59
4.1.3. Transformation des données	59
4.1.4. Imputation des données manquantes	59
4.1.5. Partitionnement des données (Train/Test)	59
4.1.6. Validation croisée	59
4.1.7. Préparation des données pour l'analyse ML	59
4.1.8. Évaluation des performances	59
4.2. Description des algorithmes utilisés pour la prédiction du statut HER2	60
4.2.1. Régression Logistique	61
4.2.2. Random Forest	61
4.2.3. LightGBM	62
4.2.4. CatBoost	63
4.3. Évaluation des performances des modèles	64

**DEUXIEME PARTIE : ETUDE RETROSPECTIVE : ANALYSE MULTIOMIQUE (TCGA)
PAR STRATIFICATION MOLECULAIRE ET ENRICHISSEMENT DES VOIES KEGG**

1. POPULATION D'ETUDE ET SOURCE DE DONNEES	67
1.1. Les critères d'inclusion	67
1.2. Les critères d'exclusion	67
1.3. Variables clés	68
2. COLLECTE ET PRETRAITEMENT DES DONNEES	68
2.1. Extraction des données	68
2.2. Imputation des valeurs manquantes	68
2.3. Normalisation et correction des effets de Lot	68
3. CLUSTERING ET METHODES D'ANALYSE	68
3.1. Analyse descriptive	68
3.2. Clustering non supervisé	69
3.3. Analyses statistiques et visualisation	70
4. METHODOLOGIE DE L'ANALYSE KEGG	70
4.1. Analyse des voies KEGG	70
4.2. Préparation des données	70
4.3. Enrichissement fonctionnel avec KEGG	71

**TROISIEME PARTIE : MODÉLISATION DE LA SURVIE : BENCHMARK COMPARATIF
ENTRE COX (APPRENTISSAGE HYBRIDE), ANN ET CNN**

1. PRÉTRAITEMENT DES DONNÉES POUR LA MODÉLISATION DE LA SURVIE	72
2. MODÉLISATION GLOBALE PAR COX HAZARD MODEL	73
2.1. Indicateurs statistiques calculés	73
2.2. Analyse par clusters à partir de la méthylation et de l'expression des gènes	73
3. MODÉLISATION PAR APPRENTISSAGE PROFOND (DEEP LEARNING)	74
3.1. Réseau de Neurone Artificiel (ANN)	74
3.2. Réseau de Neurones Convolutif (CNN)	75
4. ÉVALUATION DES MODELES	76

RESULTATS ET DISCUSSION

**PREMIERE PARTIE : ETUDE PROSPECTIVE. MODELISATION PREDICTIVE
SUPERVISEE DU STATUT HER2**

1. DESCRIPTION CLINICO-PATHOLOGIQUE DE LA POPULATION ETUDIEE	77
2. ASSOCIATION DES CARACTÉRISTIQUES CLINIQUES ET BIOLOGIQUES SELON LE STATUT HER2	80
3. PERFORMANCES COMPAREES DES DIFFERENTS MODELES DE MACHINE LEARNING POUR PREDIRE LE STATUT HER2	83

3.1. Modèle de régression logistique (LR)	83
3.2. Modèles Random Forest (RF) et LightGBM	84
3.3. Modèle CatBoost (CB)	85
3.4. Matrices de confusion des modèles LR, RF, LGBM et CB	85
<u>DEUXIEME PARTIE : ETUDE RETROSPECTIVE : ANALYSE MULTIOMIQUE (TCGA)</u>	
PAR STRATIFICATION MOLECULAIRE ET ENRICHISSEMENT DES VOIES KEGG	
1. CARACTERISTIQUES DEMOGRAPHIQUES ET CLINICO-PATHOLOGIQUES DES PATIENTES	92
1.1. Répartition selon l'âge	93
1.2. Répartition selon le stade pathologique et le stade TNM	93
1.3. Répartition selon le stade ménopausique et la survie	94
2. DISTRIBUTION DES GENES METHYLES ET EXPRIMES (IMMUNITAIRES ET HORMONAUX) DANS LA POPULATION TOTALE	95
2.1. Gènes méthylés	95
2.1.1. Gènes immunitaires	95
3.1.2. Gènes hormonaux	97
2.2. Expression des gènes méthylés	98
2.2.1. Expression des gènes immunitaires	98
2.2.2. Expression des gènes hormonaux	100
3. METHODE DE CLUSTERING PAR INTELLIGENCE ARTIFICIELLE : K-MEANS ET METHODE AGGLOMERATIVE	101
3.1. Clustering par K-Means	102
3.2. Clustering Hiérarchique Agglomératif	103
3.3. Similitudes et différences entre le Cluster 0 et le Cluster 1	105
3.4. Justification du choix de l'algorithme de clustering K-means	105
4. DISTRIBUTION DES PARAMETRES DEMOGRAPHIQUES ET CLINIQUES PAR CLUSTERS (K-MEANS)	106
4.1. Distribution de l'âge en fonction des clusters	106
4.2. Distribution des stades TNM en fonction des clusters	107
4.3. Distribution du stade pathologique par cluster	109
4.4. Distribution des sous-types moléculaires PAM 50 par cluster	110
4.5. Distribution du statut ménopausique par cluster	111
4.6. Distribution de la survie par cluster	111
5. DISTRIBUTION DES BIOMARQUEURS IMMUNITAIRES ET HORMONAUX PAR CLUSTER	112
5.1. Analyse des gènes méthylés	113
5.1.1. Gènes Immunitaires	113
5.1.2. Gènes hormonaux	115
5.2. Analyse de l'expression des gènes méthylés	117
5.2.1. Gènes Immunitaires	117
5.2.2. Gènes hormonaux	119
5.3. Moyenne des biomarqueurs immunitaires et hormonaux en fonction des deux Clusters	121
6. ETUDE DES CORRELATIONS	122
6.1. Corrélation entre les gènes méthylés et leurs expressions	122

6.2. Corrélation entre les expressions géniques des marqueurs immunitaires et hormonaux	124
7. ANALYSE DE LA MÉTHYLATION ET L'EXPRESSION DES GÈNES IMMUNITAIRES ET HORMONAUX SELON LE STATUT MÉNOPAUSIQUE	127
7.1. Analyse différentielle de la méthylation des gènes et leur expression selon le statut ménopausique	127
7.1.1. Gènes Méthylés	127
7.1.2. Gènes Exprimés	129
7.2. Analyse des corrélations croisées des gènes méthylés et de leur expression selon le statut ménopausique	130
7.3. Distribution des gènes méthylés et leur expression en fonction du statut ménopausique et selon les clusters	131
7.3.1. Distribution des gènes immunitaires méthylés dans le cluster 0	131
7.3.2. Distribution des gènes hormonaux méthylés dans le cluster 0	133
7.3.3. Distribution des gènes immunitaires méthylés exprimés dans le cluster 0	134
7.3.4. Distribution des gènes hormonaux méthylés exprimés dans le cluster 0	135
7.3.5. Distribution des gènes immunitaires méthylés dans le cluster 1	136
7.3.6. Distribution des gènes hormonaux méthylés dans le cluster 1	138
7.3.7. Distribution des gènes immunitaires méthylés exprimés dans le cluster 1	139
7.3.8. Distribution des gènes hormonaux méthylés exprimés dans le cluster 1	140
8. ANALYSE DE LA MÉTHYLATION ET L'EXPRESSION DES GÈNES IMMUNITAIRES ET HORMONAUX SELON LA SURVIE	141
8.1. Distribution des gènes méthylés et leur expression dans la population totale	141
8.1.1. Gènes méthylés	141
8.1.2. Gènes exprimés	143
8.2. Survie globale en fonction des profils épigénétiques et transcriptomiques	144
8.2.1. Gènes immunitaires : méthylation et expression génique	144
8.2.2. Gènes hormonaux : méthylation et expression génique	148
8.2.3. Impact de la méthylation et de l'expression génique sur la survie des patientes	149
8.3. Analyse de la survie par cluster	151
9. PRONOSTIC VITAL : COURBE ROC	154
9.1. Analyse ROC dans la population totale	155
9.1.1. Courbe ROC des gènes méthylés (Biomarqueurs Epigénétiques)	155
9.2. Analyse ROC de la combinaison génique dans le groupe des gènes méthylés et celui des gènes exprimés dans la population totale	157
9.2.1. Cas des gènes méthylés (Biomarqueurs Epigénétiques)	157
9.2.2. Cas des gènes exprimés (Biomarqueurs Transcriptionnels)	158
9.3. Analyse ROC par cluster	160
9.3.1. Cas du cluster 0	160
9.3.2. Cas du cluster 1	163
10. ANALYSE PAR CLUSTER DES VOIES KEGG RÉVÉLANT DES SIGNATURES MOLÉCULAIRES TUMORALES	166
10.1. Cas du cluster 0	166
10.2. Cas du cluster 1	169

TROISIEME PARTIE : MODÉLISATION DE LA SURVIE : BENCHMARK COMPARATIF
ENTRE COX (APPRENTISSAGE HYBRIDE), ANN ET CNN

1- MODELISATION DE LA SURVIE PAR LE MODELE COX – HAZARD	172
1.1. Application du modèle Cox-Hazard sur les clusters 0 et 1	174
2. MODELES DEEP LEARNING	179
2.1. Réseaux de Neurones Artificiels (ANN)	179
2.1.1. Performance Globale	179
2.1.2. Convergence du Modèle	180
2.2. Réseaux de Neurones Convolutifs (CNN)	182
2.2.1. Performance Globale	182
2.2.2. Convergence du Modèle	183
DISCUSSION	186
CONCLUSION	195
REFERENCES BIBLIOGRAPHIQUES	197
ANNEXE	

Liste des Figures

Figure 1. (A) Représentation anatomique du sein ; (B) Ebauches d'acini au niveau du lobule	3
Figure 2. Ganglions lymphatiques de la région mammaire	4
Figure 3. Sein féminin de l'enfance à la lactation, avec la structure des canaux correspondants et les coupes transversales des tissus	5
Figure 4. Régulation hormonale de la lactation	6
Figure 5. Voie de la synthèse du lait maternel	7
Figure 6. Processus de cancérisation: (A) Initiation ; (B) Progression ; (C) Métastase	10
Figure 7. Principaux effecteurs et phases de cancérisation du sein	11
Figure 8. Nombre estimé de nouveaux cas en 2022 dans le monde chez les femmes	12
Figure 9. Taux d'incidence et de mortalité standardisés par âge estimés en 2022 en Algérie chez les femmes	13
Figure 10. (A) Carcinome canalaire et (B) carcinome lobulaire invasifs (infiltrants) du sein	14
Figure 11. Catégorisation moléculaire du cancer du sein : classification moléculaire, thérapie et pronostic	18
Figure 12. Hypothèse du développement tumoral selon les sous-types moléculaires du cancers du sein	19
Figure 13. Voies de signalisation génomiques d'Er α	25
Figure 14. Les différentes sources de l'œstrogène (avant et après la ménopause)	26
Figure 15. Rôle de l'aromatase dans la synthèse des œstrogènes	27
Figure 16. Mécanismes d'induction de la prolifération cellulaire par la progestérone (A) au niveau des cellules mammaires (B) cas de la transformation tumorale	29
Figure 17. Réponse immunitaire contre le cancer du sein exprimant HER2 et effets immunomodulateurs des traitements par trastuzumab, ciblant HER2	30
Figure 18. Représentation schématique de la voie de signalisation HER2	31
Figure 19. Mécanisme sous-jacent au rôle des TLRs dans l'immunité tumorale	32
Figure 20. Régulation du cycle cellulaire par Ki-67	33
Figure 21. Le microenvironnement tumoral complexe du cancer du sein	34
Figure 22. Interaction entre les différentes cellules qui composent le système immunitaire et le microenvironnement tumoral	35
Figure 23. Augmentation de la prolifération des lymphocytes T et de la production de cytokines lors du blocage de PD-1 et PD-L1	36
Figure 24. Représentation schématique des voies de signalisation des récepteurs Toll-like (TLR) ...	37

Figure 25. Inhibiteur de point de contrôle immunitaire	38
Figure 26. La voie NF- κ B régule le microenvironnement tumoral	39
Figure 27. Composition de l'infiltrat immunitaire dans le cancer du sein en fonction du statut des récepteurs hormonaux	41
Figure 28. Régulations épigénétiques dans le cancer	43
Figure 29. Méthylation et ADN méthyltransférases (DNMTs)	44
Figure 30. Lien entre l'épigénétique et le cancer	46
Figure 31. Rôle de l'intelligence artificielle dans l'analyse traditionnelle des données de santé et dans la médecine de précision	49
Figure 32. Modélisation et analyse de données hétérogènes intégrant des informations sur les soins de santé, la génomique, la métabolomique, le phénotype et le mode de vie propres au patient, ainsi que des données d'annotation accessibles au public, y compris les gènes, les variants, les maladies, les médicaments et les biomarqueurs	50
Figure 33. Méthodologie et évaluation des performances des modèles prédictifs de classification du statut HER2 dans le cancer du sein	60
Figure 34. Fonction sigmoïde	61
Figure 35. Random forest algorithm dans le machine learning	62
Figure 36. LightGBM Histogramme algorithm	63
Figure 37. CatBoost algorithm	64
Figure 38. Réseau de Neurone Artificiel	75
Figure 39. Conventional Neural Network	76
Figure 40. Variation de la distribution par âge dépendante (a) du statut HER2 et (b) du sous-type hormonal	78
Figure 41. Importance des variables clinico-pathologiques dans la prédiction du statut HER2 par RF	82
Figure 42. Indicateurs de performance (a) Acc., Recall, Prec. et F1-Score; (b) Spécificité et Sensibilité des modèles proposés	84
Figure 43. Matrices de confusion des modèles LR, RF, LGBM et CB	86
Figure 44. Aire Sous la Courbe (AUC) et Perte pour chaque Modèle (a) Courbe ROC avec scores AUC ; (b) Perte pour les modèles implémentés	87
Figure 45. Score/valeur de validation croisée en fonction du nombre de caractéristiques à chaque itération	89
Figure 46. Distribution de l'âge au diagnostic pathologique initial des patientes	93
Figure 47. Distribution en fonction des sous type moléculaire PAM 50 (a) et des stades T (b), N (c) et M (d) dans la population étudiée	94

Figure 48. Distribution en fonction du stade ménopausique (a) et de la survie (b) dans la population étudiée	95
Figure 49. Heatmaps des profils de méthylation d'ADN des gènes immunitaires (a, c) et hormonaux (b,d) selon : (a, b) toutes les patientes et (c, d) les groupes de sous-types moléculaires.	96
Figure 50. Distribution des gènes méthylés, immunitaires et hormonaux, selon les sous-types moléculaires PAM 50 du cancer du sein	97
Figure 51. Heatmaps des profils d'expression des gènes immunitaires (a, c) et hormonaux (b,d) selon : (a, b) toutes les patientes et (c, d) les groupes de sous-types moléculaires	99
Figure 52. Distribution de l'expression des gènes méthylés, immunitaires et hormonaux, selon les sous-types moléculaires PAM 50 du cancer du sein	101
Figure 53. Segmentation des patientes en deux clusters basées sur la méthylation de l'ADN, l'expression génique des biomarqueurs immunitaires et hormonaux et les caractéristiques clinico-pathologiques	102
Figure 54. Diagramme de Venn représentant les similitudes et les différences entre le Cluster 0 et le Cluster 1	105
Figure 55. Distribution de l'âge par cluster (K-Means)	106
Figure 56. Distribution des patientes selon les clusters (K-Means) des stades T (a), N (b) et M (c).	108
Figure 57. Distribution du stade pathologique par cluster (K-Means) chez les patientes	109
Figure 58. Distribution des sous-types moléculaires PAM 50 par cluster (K-Means)	110
Figure 59. Distribution du statut ménopausique par Clusters (K-Means)	111
Figure 60. Représentation de la survie globale par Kaplan Meyer en fonction des clusters	112
Figure 61. Distribution des gènes immunitaire méthylés selon les deux clusters	113
Figure 62. Distribution des gènes hormonaux méthylés selon les deux clusters	116
Figure 63. Distribution de l'expression des gènes immunitaires selon les deux clusters	118
Figure 64. Distribution de l'expression des gènes hormonaux selon les deux clusters	120
Figure 65. Comparaison des moyennes de la méthylation des biomarqueurs et des niveaux d'expression génique entre les deux clusters (K-means)	121
Figure 66. Corrélation entre les niveaux de méthylation de l'ADN méthylé et son expression selon les deux Clusters (K-means)	124
Figure 67. Matrice de corrélation entre les expressions géniques des marqueurs immunitaires et hormonaux	125
Figure 68. Distribution des gènes immunitaires méthylés selon le statut ménopausique dans le cluster 0	132
Figure 69. Distribution des gènes hormonaux méthylés selon le statut ménopausique dans le cluster 0	134

Figure 70. Distribution des gènes immunitaires exprimés selon le statut ménopausique dans le cluster 0	135
Figure 71. Distribution des gènes hormonaux exprimés selon le statut ménopausique dans le cluster 0	136
Figure 72. Distribution des gènes immunitaires méthylés selon le statut ménopausique dans le cluster 1	137
Figure 73. Distribution des gènes hormonaux méthylés selon le statut ménopausique dans le cluster 1	138
Figure 74. Distribution des gènes immunitaires exprimés selon le statut ménopausique dans le cluster 1	140
Figure 75. Distribution des gènes hormonaux exprimés selon le statut ménopausique dans le cluster 1	141
Figure 76. Distribution des gènes méthylés, immunitaires et hormonaux, selon la survie dans la population totale	142
Figure 77. Distribution de l'expression des gènes méthylés, immunitaires et hormonaux, selon la survie dans la population totale	143
Figure 78. Courbes de survie de Kaplan-Meier en fonction des gènes immunitaires : méthylation et expression génique	146
Figure 79. Courbes de survie de Kaplan-Meier en fonction des gènes hormonaux : méthylation et expression génique	149
Figure 80. Evaluation de la valeur prédictive des biomarqueurs épigénétiques par la courbe ROC.	156
Figure 81. Evaluation de la valeur prédictive des biomarqueurs transcriptionnels par la courbe ROC	157
Figure 82. Evaluation de la valeur prédictive de la combinaison des biomarqueurs épigénétiques par la courbe ROC	158
Figure 83. Evaluation de la valeur prédictive de la combinaison des biomarqueurs ranscriptionnels par la courbe ROC	160
Figure 84. Evaluation de la valeur prédictive des biomarqueurs épigénétiques du cluster 0 par la courbe ROC	161
Figure 85. Evaluation de la valeur prédictive des biomarqueurs transcriptionnels du cluster 0 par la courbe ROC	163
Figure 86. Evaluation de la valeur prédictive des biomarqueurs épigénétiques du Cluster 1 par la courbe ROC	164
Figure 87. Evaluation de la valeur prédictive des biomarqueurs transcriptionnels du Cluster 1 par la courbe ROC	166
Figure 88. Analyse des voies KEGG des gènes différentiellement exprimés dans le cluster 0.	167
Figure 89. Analyse des voies KEGG des gènes différentiellement exprimés dans le cluster 1.	170
Figure 90. Courbes de perte et d'accuracy du modèle ANN	180
Figure 91. Matrice de confusion du modèle ANN	181

Figure 92. Courbe AUC-ROC du modèle ANN	182
Figure 93. Courbes de perte et d'accuracy du modèle CNN	183
Figure 94. Matrice de confusion du modèle CNN	184
Figure 95. Courbe AUC-ROC du modèle CNN	185

Liste des Tableaux

Tableau 1. Données minimales du compte rendu histo-pathologique en cas de cancer du sein	15
Tableau 2. Classification OMS des cancers du sein	16
Tableau 3. Caractéristiques des sous-types moléculaires du cancer du sein	17
Tableau 4. Classification TNM des cancers du sein	20
Tableau 5. Classification des stades des cancers du sein en fonction de la classification TNM	21
Tableau 6. Distribution des patients selon le statut HER2	77
Tableau 7. Analyse de régression logistique (LR) évaluant l'association entre les caractéristiques cliniques et biologiques des patientes en fonction du statut HER2	79
Tableau 8. Caractéristiques clinico-pathologiques des patientes	92
Tableau 9. Clustering par méthode K-Means	103
Tableau 10. Clustering Hiérarchique Agglomératif	104
Tableau 11. Corrélation entre les gènes méthylés et leurs expressions dans chaque cluster	122
Tableau 12. Valeurs de p ajustées pour les corrélations entre l'expression des gènes immunitaires et hormonaux	126
Tableau 13. Résultats des tests de normalité (Shapiro-Wilk) et du test de Kruskal-Wallis pour la méthylation des gènes selon le statut ménopausique	128
Tableau 14. Résultats des tests de normalité (Shapiro-Wilk) et du test de Kruskal-Wallis pour la méthylation des gènes hormonaux selon le statut de ménopausique	128
Tableau 15. Tests de normalité et test de Kruskal-Wallis pour l'expression des gènes immunitaires selon le statut ménopausique.....	129
Tableau 16. Tests de normalité et test de Kruskal-Wallis pour l'expression des gènes hormonaux selon le statut ménopausique	130
Tableau 17. Analyse conjointe de la méthylation de l'ADN et de l'expression génique en fonction du statut ménopausique	131
Tableau 18. Variations de l'expression génique en fonction des groupes de survie.	150
Tableau 19. Comparaison de la méthylation des gènes entre les clusters et les statuts de survie	152
Tableau 20. Comparaison de l'expression génique entre les clusters et les statuts de survie.	153
Tableau 21. Analyse de survie par le modèle de Cox-Hazard dans la population totale	172

Tableau 22. Analyse de survie par le modèle de Cox-Hazard pour le Cluster 0	174
Tableau 23. Analyse de survie par le modèle de Cox-Hazard pour le Cluster 1	175
Tableau 24. Analyse de survie par le modèle de Cox-Hazard dans le Cluster 0	177
Tableau 25. Analyse de survie par le modèle de Cox-Hazard dans le Cluster 1	178
Tableau 26. Performance du modèle ANN par classe de survie	181
Tableau 27. Performance du modèle ANN par classe de survie	184

Liste des Abréviations

A

450k	Illumina HumanMethylation450 BeadChip	Plateforme d'analyse de la méthylation de l'ADN à 450 000 sites
AIC	Akaike Information Criterion	Critère d'information d'Akaike
AKT1	AKT Serine/Threonine Kinase 1	Gène de la voie PI3K impliqué dans la survie cellulaire
ANN	Artificial Neural Network	Réseau de neurones artificiels
ANOVA	Analysis of Variance	Analyse de la variance
AUC-ROC	Area Under the Receiver Operating Characteristic Curve	Aire sous la courbe ROC

B

BMI / IMC	Body Mass Index	Indice de Masse Corporelle
BRCA	Breast Cancer	Cancer du sein

C

CB	CatBoost	Algorithme d'apprentissage automatique
CD80	Cluster of Differentiation 80	Protéine immunorégulatrice
CD86	Cluster of Differentiation 86	Protéine costimulatrice exprimée sur les cellules présentatrices d'antigènes
CD274 / PDL1	Programmed Death-Ligand 1	Ligand immunitaire codé par le gène CD274
CI 95% / IC95%	95% Confidence Interval	Intervalle de Confiance à 95 %
CNN	Convolutional Neural Network	Réseau de neurones convolutionnels
Coef.	Coefficient	Paramètre estimé dans un modèle statistique
Concordance (c-index)	Concordance Index	Mesure de performance pour les modèles de survie
CTLA4	Cytotoxic T-Lymphocyte Associated Protein 4	Récepteur immunitaire inhibiteur
CYP19A1	Cytochrome P450 Family 19 Subfamily A Member 1	Gène de l'aromatase, enzyme clé dans la synthèse des œstrogènes

D

DNA / ADN	Deoxyribonucleic Acid	Acide Désoxyribonucléique
DNMT	DNA Methyltransferase	Enzyme impliquée dans la méthylation de l'ADN
DSS	Disease-Specific Survival	Survie spécifique à la maladie

E

EGFR	Epidermal Growth Factor Receptor	Récepteur de facteur de croissance épidermique
ER / ESR1	Estrogen Receptor 1	Récepteur des œstrogènes
ERBB2 / HER2	Erb-B2 Receptor Tyrosine Kinase 2	Gène codant pour HER2
Exp(Coef)	Hazard Ratio (HR)	Risque relatif estimé dans un modèle de Cox

F

F1-score	F1 Score	Mesure harmonique entre la précision et le rappel
FISH	Fluorescence In Situ Hybridization	Technique de marquage génétique
FP	False Positive	Faux positif
FN	False Negative	Faux négatif
F-value	F Statistic	Statistique utilisée dans l'ANOVA

H

HDAC	Histone Deacetylase	Enzyme impliquée dans la régulation de la chromatine
HR	Hazard Ratio	Risque relatif

I

IA	Artificial Intelligence	Intelligence Artificielle
IHC	Immunohistochemistry	Immunohistochimie
IQR	Interquartile Range	Écart interquartile

J

J(f)	Regularization Term	Terme de régularisation — évite le surapprentissage
-------------	---------------------	---

K

KEGG	Kyoto Encyclopedia of Genes and Genomes	Base de données de voies biologiques
Ki67	Ki67	Marqueur de prolifération cellulaire
K-Means	K-Means Clustering	Algorithme de partitionnement

L

L2	L2 Regularization	Régularisation par norme quadratique
LGBM	Light Gradient Boosting Machine	Algorithme d'apprentissage supervisé
Log2FC	Log2 Fold Change	Logarithme base 2 du Fold Change
Log-vraisemblance	Log-likelihood	Mesure de l'ajustement d'un modèle
LR	Logistic Regression	Régression logistique

M

M	Metastasis	Métastase (classification TNM)
MAP kinases	Mitogen-Activated Protein Kinases	Protéines de signalisation cellulaire
MaxPooling	Max Pooling	Technique de sous-échantillonnage dans les CNN
ML	Machine Learning	Apprentissage automatique

N

N	Node	Nombre de ganglions atteints (classification TNM)
NF-κB1 / NFKB1	Nuclear Factor Kappa B Subunit 1	Facteur de transcription inflammatoire
NGHT	Nottingham Grading Histologic Tumor	Grade histologique tumoral

O

OR	Odds Ratio	Rapport de cotes
OS	Overall Survival	Survie globale

P

PAM50	Prediction Analysis of Microarray 50	Classification des sous-types moléculaires du cancer du sein
PCA	Principal Component Analysis	Réduction de dimension
PCR	Polymerase Chain Reaction	Réaction de polymérisation en chaîne
PD1 / PDCD1	Programmed Cell Death 1	Récepteur immunitaire PD-1
PGR / PR	Progesterone Receptor	Récepteur de la progestérone
PI3K	Phosphoinositide 3-Kinase	
PI3K/AKT/mTOR	PI3K/AKT/mTOR Pathway	Voie de signalisation impliquée dans la croissance cellulaire
PIK3CA	PI3K Catalytic Subunit Alpha	Gène de la voie PI3K
pCR	Pathological Complete Response	Réponse pathologique complète
Precision	Precision	Proportion de vrais positifs parmi les positifs prédits

R

Recall	Recall	Sensibilité ; proportion de vrais positifs parmi les cas positifs
ReLU	Rectified Linear Unit	Fonction d'activation en réseaux neuronaux
RF	Random Forest	Forêt aléatoire
RCP	Multidisciplinary Team Meeting	Réunion de Concertation Pluridisciplinaire
ROC	Receiver Operating Characteristic	Courbe ROC

S

SADM	Clinical Decision Support System	Système d'Aide à la Décision Médicale
Se	Standard Error	Erreur standard
Specificity	Specificity	Proportion de vrais négatifs parmi les cas négatifs
Spearman	Spearman Coefficient	Corrélation non paramétrique
SSDC	Statistically Determined Subtype of Cancer	Sous-type Statistiquement Déterminé du Cancer

T

T	Tumor Size	Taille tumorale (classification TNM)
TCGA	The Cancer Genome Atlas	Base de données omiques sur le cancer

TLR	Toll-Like Receptor	Récepteurs de l'immunité innée
TLR4 / TLR9	Toll-Like Receptors 4 / 9	Récepteurs immunitaires spécifiques
TN	True Negative	Vrai négatif
TNBC	Triple Negative Breast Cancer	Cancer du sein triple négatif
TNM	Tumor Node Metastasis	Classification anatomopathologique
TP	True Positive	Vrai positif
Transcript	Transcriptomics	Étude de l'expression des gènes
Tregs	Regulatory T Cells	Lymphocytes T régulateurs
T-SNE	t-distributed Stochastic Neighbor Embedding	Technique de visualisation de données

U

UCSC Xena	UCSC Xena Platform	Portail pour l'analyse de données génomiques
UMI	Unique Molecular Identifier	Séquence d'identification unique pour l'analyse de l'ARN

W

W_i	Sample Weight	Poids de l'échantillon i — Utilisé dans la fonction de perte pondérée
-------------------------	---------------	--

Z

z	Z Statistic	Test de Wald dans les modèles de survie
-----------------------	-------------	--



Introduction

Le cancer du sein demeure l'une des principales causes de morbidité et de mortalité chez les femmes dans le monde, représentant environ 2,3 millions de nouveaux cas en 2020, selon l'Organisation Mondiale de la Santé (**Lei et al., 2021**). Cette maladie complexe touche des femmes de tous âges et milieux socio-économiques, avec une prévalence comparable dans les pays développés et en développement (**Barzaman et al., 2020**). Ces données témoignent de la nécessité urgente de mieux comprendre les mécanismes biologiques et environnementaux sous-jacents pour améliorer les stratégies de prévention, de diagnostic et de traitement. Le diagnostic et le traitement du cancer du sein sont rendus complexes par la diversité des sous-types moléculaires et cliniques. Parmi ces sous-types, les cancers HER2-positifs représentent environ 20 % des cas, se distinguant par leur agressivité et leur réponse aux thérapies ciblées. Toutefois, la détection précise du statut HER2 repose sur des méthodes comme l'immunohistochimie (IHC) ou la fluorescence *in situ* hybridization (FISH), qui, bien que fiables, peuvent être coûteuses et parfois inconclusives (**Horimoto et al., 2022**). Cette difficulté met en lumière une lacune critique dans les approches actuelles : le besoin de méthodes diagnostiques et pronostiques complémentaires, plus accessibles et automatisées, notamment dans les contextes cliniques à ressources limitées.

De nos jours, on pense que les mécanismes qui régissent l'apparition du cancer sont la conséquence non seulement de défauts génétiques mais aussi de modifications épigénétiques. Ainsi, l'épigénétique est devenue un domaine de recherche très intéressant et de plus en plus exploré afin de trouver de nouveaux moyens de prévention et de traitement des néoplasies comme c'est le cas pour le cancer du sein (**Basse et Arock, 2015**). La méthylation de l'ADN émerge comme un mécanisme clé dans la tumorigenèse. Ce processus épigénétique, impliquant des modifications chimiques qui répriment l'expression génique, est fréquemment altéré dans le cancer. L'hyperméthylation des gènes suppresseurs de tumeurs ou l'hypométhylation des oncogènes contribue à la prolifération incontrôlée des cellules cancéreuses (**Kulis et Esteller, 2010**). Ces altérations peuvent apparaître tôt dans le processus tumoral, offrant une opportunité pour le diagnostic précoce et l'identification de cibles thérapeutiques potentielles (**Ma et al., 2023**). Malgré ces avancées, peu d'études ont exploré de manière exhaustive les interactions entre méthylation de l'ADN, expression génique et marqueurs immunitaires dans le contexte du cancer du sein. Les marqueurs immunitaires tels que PDCD1, CD274 (PD-L1), TLR4, TLR9, CTLA4 et NF- κ B1 jouent des rôles essentiels dans les interactions entre la tumeur et le système immunitaire. Leur implication dans la modulation des signaux immunitaires, la mort cellulaire programmée et la croissance

tumorale en fait des cibles prometteuses pour les thérapies immunomodulatrices (Li et al., 2017 ; Karapetyan et al., 2020 ; Navarrete-Bernal et al., 2020 ; Ralser et al., 2021 ; Zhang et al., 2021). Par ailleurs, les récepteurs hormonaux tels que l'œstrogène, la progestérone, et HER2 influencent directement la prolifération et la progression tumorales (Chumsri et al., 2011 ; Andrahennadi et al., 2021). Toutefois, les mécanismes d'interaction entre ces marqueurs et les profils de méthylation restent encore mal compris. Bien que des progrès significatifs aient été réalisés dans la caractérisation moléculaire du cancer du sein, l'intégration des données multiomiques de méthylation, expression génique et paramètres cliniques reste une approche sous-exploitée. Ces lacunes soulèvent des questions de recherche majeures :

1. Comment les altérations épigénétiques, notamment la méthylation de l'ADN, influencent-elles l'expression génique des biomarqueurs immunitaires et hormonaux dans le cancer du sein ?
2. Quelle est la corrélation entre les altérations multiomiques (méthylation, expression génique) et les courbes de survie des patientes atteintes de cancer du sein ?
3. Les signatures multiomiques permettent-elles d'identifier des cibles thérapeutiques pertinentes ou de prédire la réponse aux traitements de manière plus personnalisée ?
4. Quels modèles d'intelligence artificielle offrent la meilleure performance pour prédire la survie à partir des données omiques intégrées ?

À travers cette étude, nous cherchons à explorer par des méthodes d'intelligence artificielle les profils de méthylation de gènes immunitaire et hormonaux clés, impliqués dans le cancer du sein en parallèle avec leur expression et leurs implications cliniques. Tirant parti des données de la cohorte TCGA, cette recherche vise à fournir une vision multiomique du cancer du sein, identifier des cibles moléculaires potentielles et améliorer la personnalisation des traitements pour les patientes.

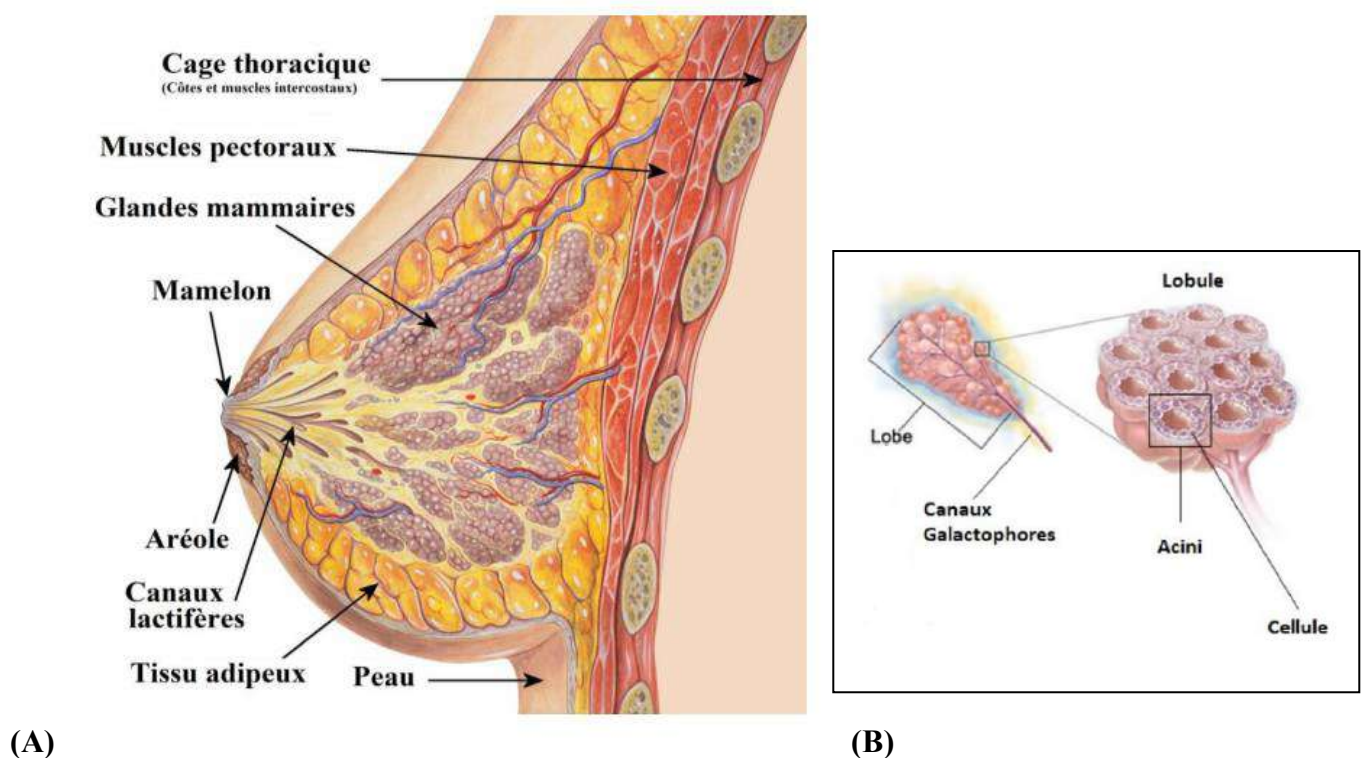


Revue de la **B**ibliographie

1. LE CANCER DU SEIN

1.1. Anatomie du sein et de la glande mammaire

Le sein est une glande endocrine sudoripare complexe située sur la partie antérieure du thorax. Il est composé de trois types principaux de tissus : glandulaire, conjonctif et adipeux. La glande mammaire est composée de 15 à 20 lobes disposés en cercles et séparés par du tissu adipeux et conjonctif, qui assure la structure et le soutien du sein. Chaque lobe est formé de 20 à 40 unités appelées lobules et de canaux qui fusionnent pour former un canal lobulaire unique de deux millimètres de diamètre permettant l'excrétion du lait. Ces lobules sont constitués de 10 à 100 alvéoles appelées également acinis glandulaires, où se déroule la production lactée (**Gremmo-Féger, 2015 ; Tortora et Derrickson, 2016**). Chaque lobe est relié à un système de conduits galactophores, qui se dilatent sous le mamelon en de petites cavités appelées sinus galactophores, servant de réservoirs temporaires pour le lait. Ces conduits se terminent par des pores lactifères situés au sommet du mamelon (**Figure 1**). L'aréole mammaire est une zone pigmentée entourant le mamelon ; elle est dotée de glandes sébacées modifiées, qui aident à maintenir l'élasticité et la souplesse de la peau dans cette région sensible (**Tortora et Derrickson, 2016**).



Le sein est parcouru par de nombreux vaisseaux sanguins et lymphatiques. Le réseau lymphatique est composé de trois principales voies de drainage : les voies axillaires, transpectoraux et mammaires internes. Les ganglions lymphatiques régionaux sont subdivisés en niveaux selon leur localisation par rapport au muscle pectoral : le niveau I (axillaire inférieur), le niveau II (axillaire moyen) et le niveau III (axillaire apical) (**Figure 2**) (**Virginia et al., 2017**).

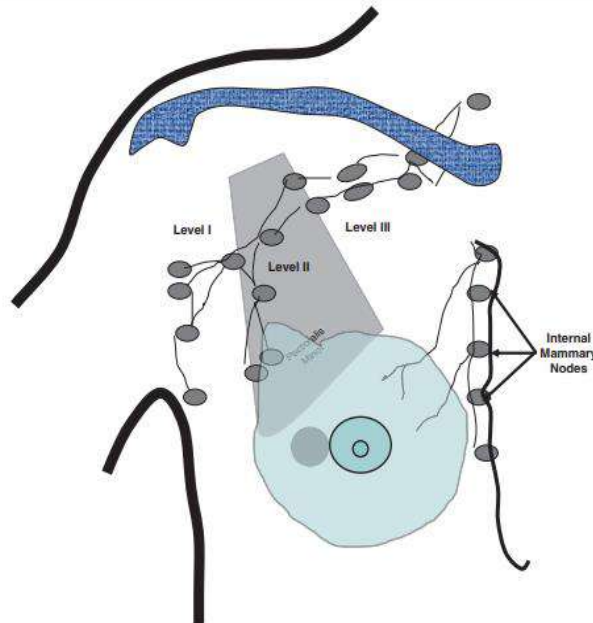


Figure 2. Ganglions lymphatiques de la région mammaire (**Jatoi et Kaufmann, 2010**).

La plupart des ganglions axillaires sont indiqués comme étant de niveau I, niveau II et niveau III, sur la base de leur relation avec le muscle pectoral mineur. A noter également les ganglions mammaires internes situés à côté du bord du sternum.

1.2. Physiologie du sein

La physiologie du sein est étroitement liée à son rôle dans la lactation et à la régulation hormonale. Les changements mammaires sont produits par des conditions physiologiques communes telles que l'âge, le cycle menstruel, la grossesse, l'allaitement, la lactation, la ménopause, les changements de poids corporel, certaines substances contenant des hormones et le type d'alimentation (**Bazira et al., 2022**). Le sein ne fait pas réellement partie de l'appareil reproducteur femelle, toutefois, les glandes mammaires sont considérées comme d'importantes glandes accessoires. A la naissance, la structure de la glande mammaire est inachevée; le tissu mammaire ne subit aucune croissance visible et reste quiescent jusqu'à la puberté. Lors de la puberté, leur développement va se différencier selon le sexe : chez l'homme, les glandes

mammaires présenteront un aspect atrophié, tandis que chez la femme ces glandes se développent en même temps que les ovaires (**Figure 3**). En réalité, sous l'influence des hormones stéroïdiennes ovariennes qui augmentent de manière cyclique, on observe une interaction physiologique entre le développement de ces glandes et le cycle génital. Durant chaque cycle menstruel, les seins présentent des variations de la concentration plasmatique des hormones. Ces variations demeurent faibles jusqu'à la première grossesse avec remodelage des structures permettant la lactation. En dehors des grossesses, le sein connaît une phase d'involution, qui prend fin à la ménopause (**Hassiotou et Geddes, 2012**).

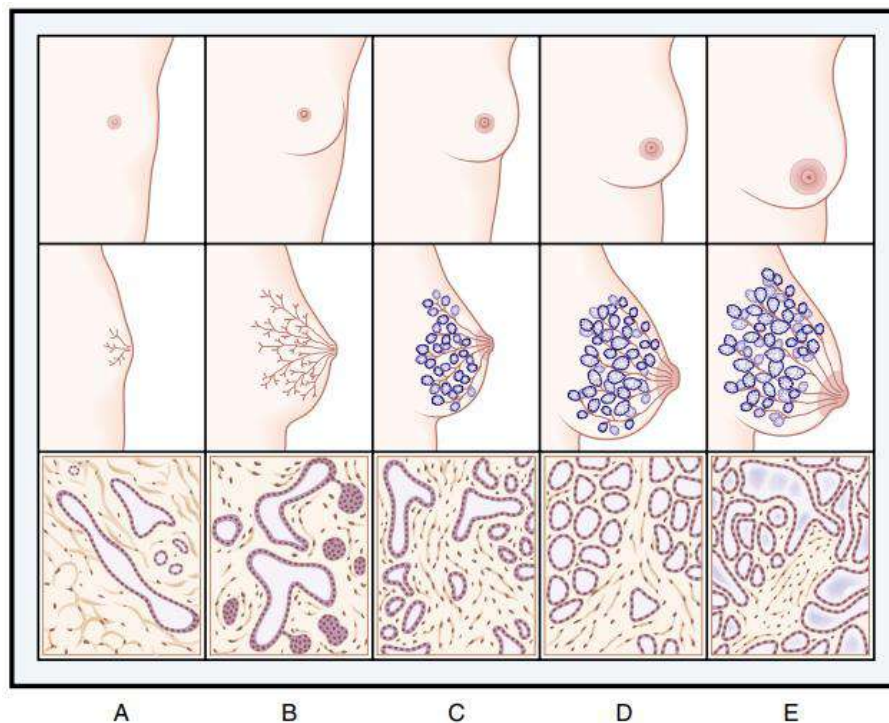


Figure 3. Sein féminin de l'enfance à la lactation, avec la structure des canaux correspondants et les coupes transversales des tissus (**Adetola et al., 2010**).

(A-C) Développement progressif du système ductale et lobulaire-alvéolaire périphérique bien différencié. **(D)** Poussée ductale et intensification du développement lobulaire-alvéolaire périphérique pendant la grossesse. Les cellules luminales glandulaires commencent à synthétiser activement les graisses et les protéines du lait vers le terme ; seules de petites quantités sont libérées dans la lumière. **(E)** Avec le retrait post-partum des stéroïdes sexuels lutéaux et placentaires et du lactogène placentaire, la prolactine est capable d'induire une activité sécrétoire complète des cellules alvéolaires et la libération de lait dans l'alvéole et les petits canaux.

Le sein est une glande endocrine qui joue un rôle crucial dans la production de lait, principalement sous l'influence d'hormones telles que la prolactine (**Figure 4 et 5**). Celle-ci, sécrétée par l'hypophyse, stimule le développement des acini glandulaires, permettant ainsi la

lactation. Pendant la grossesse, les niveaux d'œstrogènes et de progestérone augmentent, entraînant une préparation des tissus mammaires à la lactation. Après l'accouchement, la chute des niveaux hormonaux et la stimulation des mamelons par la succion du nourrisson déclenchent la libération de prolactine et d'ocytocine, ce qui favorise la production et l'éjection du lait (**Javed et al., 2013**).

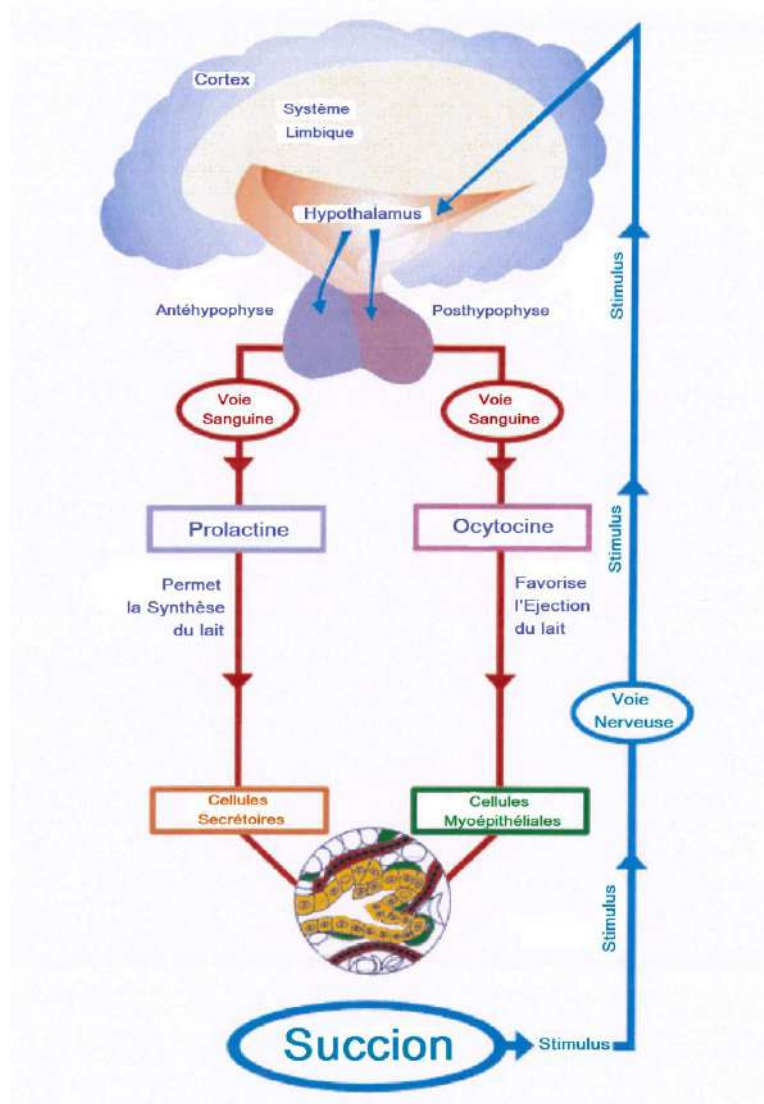


Figure 4. Régulation hormonale de la lactation (**Damis et al., 2012**).

Les ganglions et vaisseaux lymphatiques qui traversent le sein aident à surveiller et à contrôler l'invasion microbienne, tout en étant impliqués dans la circulation des cellules immunitaires. Cependant, les états pathologiques, comme le cancer, entraînent des dérangements significatifs qui affectent non seulement la fonction glandulaire mais aussi l'intégrité immunitaire et lymphatique (**Martaindale, 2018**).

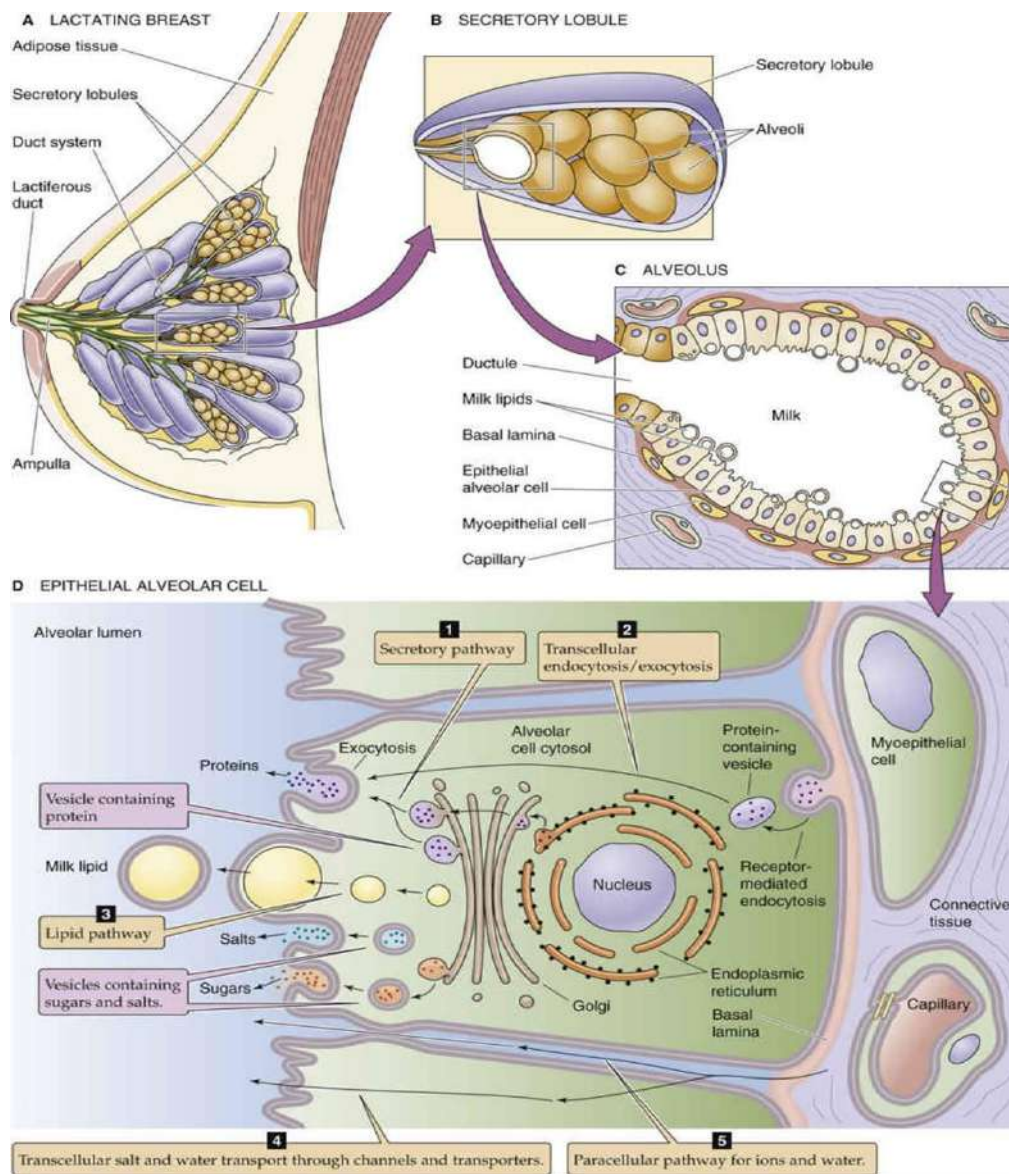


Figure 5. Voie de la synthèse du lait maternel (Baron et Boulpaep, 2016).

1.2.1. Rôle de l'œstrogène et de la progestérone

Les hormones œstrogéniques et progestatives contrôlent la croissance de la glande mammaire, sous l'action de l'hormone hypothalamique : Gonadotropin-Releasing-Hormone (GnRH ou LH-RH). La GnRH stimule l'hypophyse entraînant la sécrétion, de manière pulsatile, la Folliculo-Stimulating Hormone (FSH) et la Luteinizing Hormone (LH), qui à leur tour vont conduire à l'ovulation et la stimulation de la sécrétion des hormones sexuelles féminines œstrogéniques (estradiol, estrone, estriol) et progestative (progestérone). Ces hormones ont des actions bien spécifiques sur les glandes mammaires.

* **L'œstrogène** : sous l'action des œstrogènes, et à partir de la puberté la glande mammaire et plus particulièrement les canaux galactophores vont subir un développement de leur volume. Lors de la ménopause, on observe la fin du développement mammaire, voire une régression suite à une diminution importante d'œstrogène. L'absence de l'œstrogène conduit à la déshydratation du tissu conjonctif du sein qui perd en élasticité.

* **La progestérone** : cette hormone va permettre une différenciation alvéo-acineuse et va exercer ses effets, principalement après l'ovulation et au cours de la grossesse. Toutefois, cette hormone ne fonctionne qu'en synergie avec les œstrogènes. En effet, lorsque la concentration de l'œstrogène diminue, la progestérone maintient le bon fonctionnement des tissus mammaires hypertrophiés (Houdebine, 1986 ; Badowska-Kozakiewicz et al., 2015).

1.3. Physiopathologie du cancer du sein

Le processus de cancérogenèse chez la femme est multifactoriel. Il implique divers gènes, que ce soit dans des cas héréditaires (anomalies génétiques constitutionnelles) ou sporadiques (anomalies génétiques somatiques). Le cancer du sein est souvent caractérisé par des modifications moléculaires, aussi bien les amplifications d'oncogènes, ou les mutations ou les délétions de gènes suppresseurs de tumeurs. Ces gènes codent pour des protéines impliquées dans de diverses cascades de transduction, comme les ligands et les récepteurs aux facteurs de croissance, des facteurs de transcription ou encore des protéines régulant le cycle cellulaire. Actuellement, les 3 événements les plus observés dans le cancer du sein lors de l'inactivation de gènes suppresseurs de tumeurs sont : les mutations, les pertes d'hétérozygotie (ou perte d'allèles), et les altérations épigénétiques (Rountree et al., 2001 ; Widschwendter et Jones, 2002 ; Thompson and Easton, 2004).

L'hétérogénéité du cancer du sein, à l'instar de toute tumeur solide, signifie qu'il y a plusieurs voies conduisant à la carcinogenèse mammaire, chacune comportant des étapes distinctes. Le cancer solide ne se présente pas comme une condition statique, mais comme une progression pathologique où les cellules cancéreuses modifiées ne prennent le dessus sur les autres que si leurs changements leur donnent une capacité sélective avantageuse (Hanahan et Weinberg, 2000).

1.3.1. Bases moléculaires

Le cancer du sein résulte de modifications complexes au niveau moléculaire et cellulaire qui transforment des cellules mammaires normales en cellules cancéreuses (**Zhang, 2023**). Ainsi, les cellules épithéliales du parenchyme mammaire peuvent se transformer de manière anormale, entraînant une prolifération incontrôlée. Les tumeurs mammaires, souvent localisées dans les unités lobulaires du parenchyme, peuvent altérer la fonction normale de la glande (**Khan et al., 2023**). Les tumeurs peuvent également induire des modifications de la vascularisation et du drainage lymphatique, ce qui peut perturber l'homéostasie et favoriser la propagation de cellules malignes à d'autres parties du corps (**Aranda-Gutierrez et al., 2023**).

Ces transformations sont généralement déclenchées par une interaction entre des facteurs génétiques, hormonaux et environnementaux. Parmi les gènes les plus impliqués, nous citons *BRCA1* et *BRCA2*, essentiels à la réparation de l'ADN, et dont les mutations augmentent significativement le risque de développer un cancer du sein et de l'ovaire. Un autre gène, *HER2* (*ERBB2*), joue également un rôle important ; son amplification ou sa surexpression entraîne une prolifération cellulaire excessive, associée à des tumeurs plus agressives (**Dumbrava et al., 2019 ; Morsberger et al., 2022 ; Seo et al., 2023**). Sur le plan moléculaire, plusieurs altérations contribuent à la progression du cancer du sein, notamment la dérégulation des voies de signalisation impliquant des facteurs de croissance comme l'EGF et le PDGF, favorisant une croissance et une survie cellulaires anormales. L'inactivation des gènes suppresseurs de tumeurs comme le gène *TP53*, qui régulent le cycle cellulaire et l'apoptose, peuvent être mutés, empêchant l'élimination des cellules endommagées et facilitant la progression tumorale. Ces altérations génétiques et moléculaires aboutissent à une prolifération cellulaire incontrôlée, à une résistance à l'apoptose et, potentiellement, à la propagation du cancer à d'autres tissus (**Gajria et Chandrapaty, 2011**).

1.3.2. Processus de la carcinogenèse mammaire

Le cancer du sein se développe principalement à partir des cellules épithéliales du sein ou des cellules souches mammaires, à la base d'un amalgame de changements génétiques et épigénétiques, ainsi que d'interactions aberrantes dans le microenvironnement. Ce processus se déroule en plusieurs étapes comprenant l'initiation, la promotion et la progression (**Figure 6**).

La phase d'invasion, suite à la métastase, met en jeu des molécules d'adhésion telles que les cadhérines et les intégrines, ainsi que des protéases capables de dégrader la matrice extracellulaire (**Mego et al., 2010**).

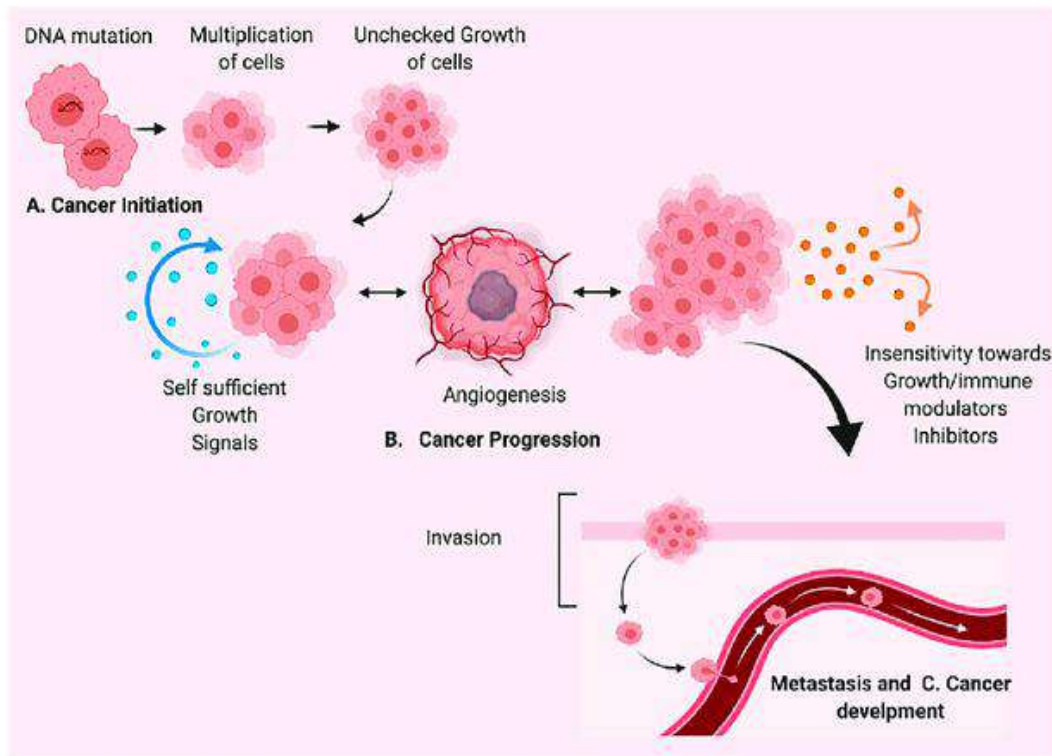


Figure 6. Processus de cancérisation: (A) Initiation ; (B) Progression ; (C) Métastase (Chaudhry et al., 2022).

***La phase d'initiation** : se caractérise par la transformation de la cellule normale en cellule cancéreuse, résultant de multiples altérations génomiques échappant aux processus de réparation de l'ADN. L'activation des oncogènes et/ou l'inhibition des gènes suppresseurs de tumeurs sont des éléments clés de cette étape. Plusieurs facteurs hormonaux peuvent favoriser l'initiation du cancer du sein (Rocheffort, 2008).

* **La promotion** : est la deuxième étape, où une cellule mutée par un initiateur est susceptible aux effets des promoteurs. La phase de promotion se manifeste par une prolifération anarchique générant une tumeur à partir de la cellule cancéreuse.

* **La phase de progression** : implique la croissance tumorale, sa dissémination *via* la circulation sanguine pour former des métastases, et l'acquisition de propriétés spécifiques par la cellule cancéreuse. Ces propriétés comprennent une autosuffisance en facteurs de croissance, une insensibilité aux signaux extérieurs, une invasion des tissus, une réplification illimitée, une induction de l'angiogenèse et un échappement à l'apoptose (Feng et al., 2018 ; Pavitra et al., 2023).

Au premier stade, le cancer est généralement asymptomatique et présente un faible potentiel de propagation. Ainsi, l'hyperplasie canalaire ou lobulaire atypique est une lésion précancéreuse caractérisée par des couches de cellules anormales dans le canal ou le lobule. Elle est considérée comme le précurseur du carcinome canalaire ou lobulaire *in situ*, qui est une lésion non invasive. Au cours de ce processus, le contrôle de la prolifération, la survie, la différenciation et la migration deviennent déréglementés. Une fois que les cellules sont envahies, le risque de développer des métastases augmente considérablement. Les ganglions lymphatiques sont le site primaire du cancer métastatique du sein. A ce stade, les cellules doivent envahir la membrane basale, entrer dans le système vasculaire (intravasation), survivre en l'absence d'adhérence, quitter le système vasculaire (extravasation) et établir une nouvelle tumeur (**Figure 7**).

Au fil du temps, les cancers *in situ* peuvent évoluer vers une forme invasive, envahissant le tissu mammaire environnant, puis se propager aux ganglions lymphatiques voisins (métastases régionales) ou à d'autres organes du corps, à savoir, le poumon, le cerveau, le foie et les os (métastases à distance). Les métastases généralisées sont la principale cause de décès chez les femmes atteintes de cancer du sein (**Vargo-Gogola and Rosen, 2007 ; Jin and Mu, 2015**).

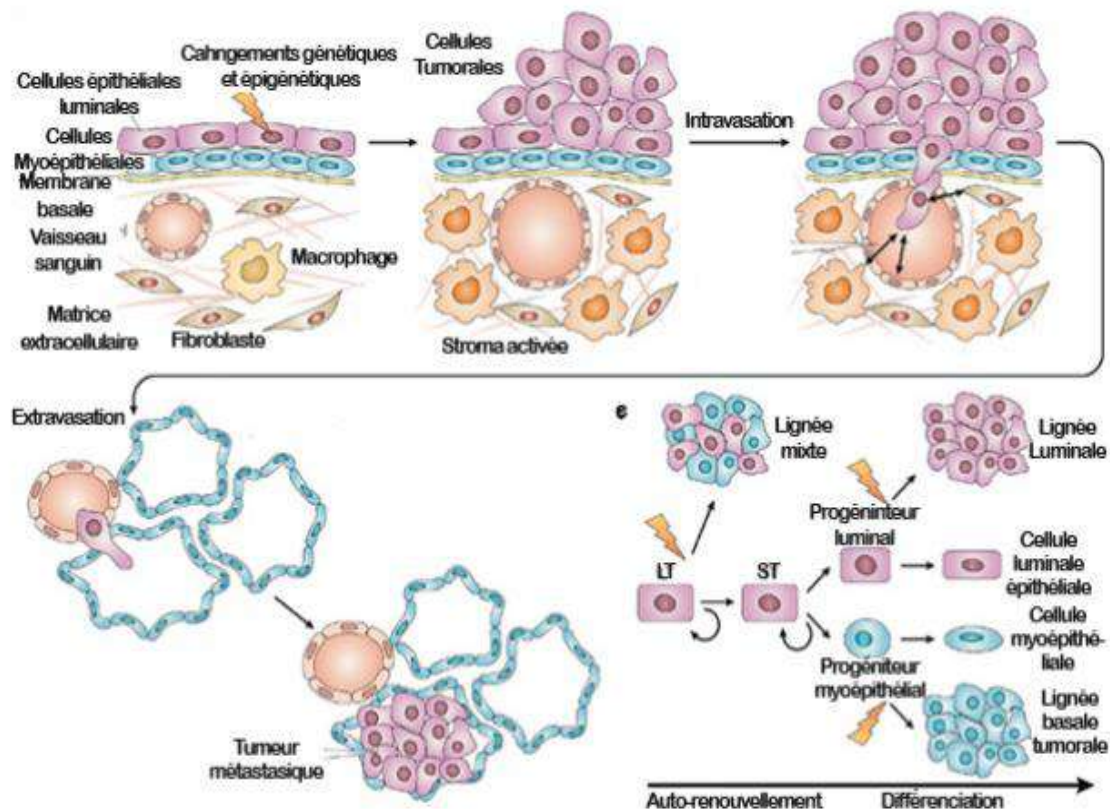


Figure 7. Principaux effecteurs et phases de cancérisation du sein (Vargo-Gogola et Rosen, 2007).

2. EPIDEMIOLOGIE

2.1. Dans le monde

Sur le plan international, le cancer du sein est le cancer le plus fréquent chez la femme. Il est la première cause de décès par cancer dans le monde, avec plus de 666 103 (15.4%) cas par an. Il demeure un défi majeur de santé mondiale, marqué par une incidence estimée à 2 261 419 (24,5%) nouveaux cas (**Figure 8**) (**Globocan, 2022**). La répartition mondiale du cancer du sein est très inégale. Les taux les plus élevés sont retrouvés dans les pays industrialisés (Amérique du Nord, Europe sauf le Japon). Les taux les plus bas sont enregistrés en Afrique et en Asie. L'augmentation de l'incidence du cancer du sein chez la femme. est due à des paramètres épidémiologiques, marqués par l'évolution démographique, l'augmentation de l'espérance de vie, les transformations de l'environnement et les changements des habitudes alimentaires (**Hamdi-Cherif et al., 2015**).

Une tendance à la baisse de la mortalité au fil des années, a été observée et découle en grande partie des efforts significatifs en matière de détection précoce et d'amélioration des traitements adaptés ([Global Cancer Observatory \(iarc.fr\)](http://Global Cancer Observatory (iarc.fr))).

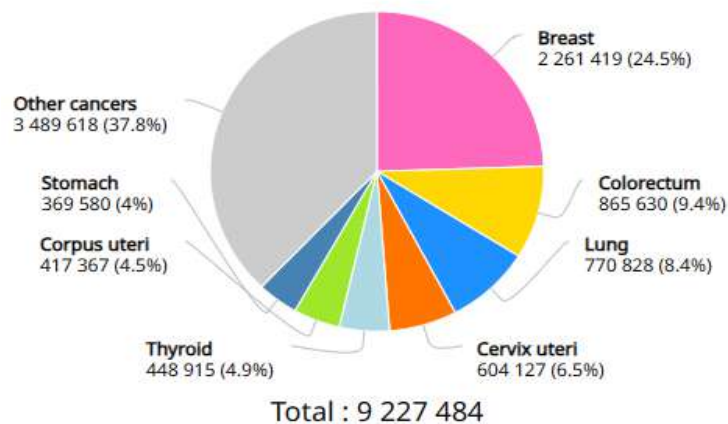


Figure 8. Nombre estimé de nouveaux cas en 2022 dans le monde chez les femmes (**Globocan, 2022**).

2.2. En Algérie

Le cancer du sein représente un problème majeur de santé publique en Algérie, avec une incidence estimée à 14 601 (43,1%) nouveaux cas chaque année et un taux de mortalité de 4 893 (28,8%) (**Figure 9**) (**Globocan, 2022**). L'Algérie présente un taux d'incidence élevé du cancer du sein avec 55,78 nouveaux cas pour 100 000 habitants, se plaçant en première position pour l'incidence et la mortalité liée au cancer du sein, avec un taux de 18,55 pour 100 000 habitants

(Boundedjar et al., 2022). Cependant, le pays ne dispose pas encore d'un programme organisé de dépistage de masse, en raison de la jeunesse de la population et d'un manque de sensibilisation parmi les femmes. Une enquête a révélé que seulement 28,2 % des femmes connaissaient l'auto-examen des seins et seulement 16,5 % l'avaient pratiqué (Lounis et al., 2023).

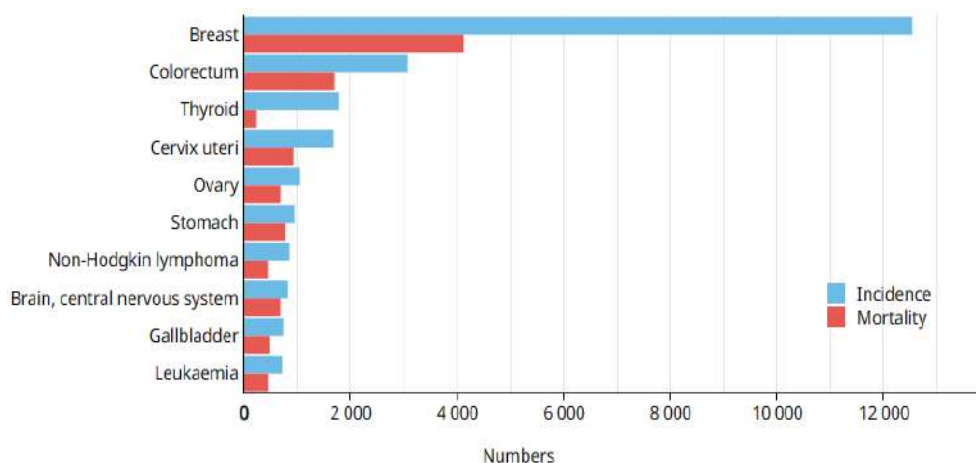


Figure 9. Taux d'incidence et de mortalité standardisés par âge estimés en 2022 en Algérie chez les femmes (Globocan, 2022).

2.3. Facteurs de risque

Le cancer du sein est une maladie multifactorielle, influencée par divers facteurs de risque. Il existe un manque de connaissances et d'informations des patientes et de certains professionnels de la santé concernant ces facteurs de risque. L'âge est un facteur de risque important; la majorité des cancers se développe après 50 ans. Il reste rare chez les femmes de moins de 30 ans, sauf en cas de prédispositions génétiques.

Ces facteurs sont divisés en deux : facteurs non modifiables et facteurs modifiables (Manouchehri et al., 2022 ; Rehman et al., 2022).

2.3.1. Les facteurs de risque non modifiables

Ils comprennent les antécédents familiaux de cancer du sein, impliquant des gènes comme *BRCA1* et *BRCA2*, ainsi que des gènes à faible pénétrance comme *TP53*, *PTEN*, *CDH1*, *STK11*, *ATM*, *PALB2*, *BRIP1* et *CHEK2* (Heisey et Carroll, 2016 ; Foulkes, 2021 ; Wang, 2021). L'âge précoce des premières menstruations et la ménopause tardive augmentent également le risque de cancer du sein (Surakasula, 2014 ; Khalis, 2018). Une densité mammaire élevée et des maladies bénignes du sein sont aussi associées à un risque accru de cancérisation (Astley et al., 2018 ; Johansson et al., 2021).

2.3.2. Les facteurs de risque modifiables

Ils incluent les traitements hormonaux substitutifs de la ménopause, la parité, l'âge à la première grossesse, l'allaitement, l'alimentation, l'obésité, l'indice de Masse Corporelle (IMC), l'activité physique, les perturbateurs endocriniens, et la perturbation du rythme circadien (Gompel, 2019 ; Zouré, 2016 ; Nelson, 2022 ; D'cunha, 2023). Ces facteurs soulignent l'importance de stratégies de prévention personnalisées pour réduire le risque de cancer du sein (Frikha et Chlif, 2021).

3. CLASSIFICATION DU CANCER DU SEIN

Vu le caractère hétérogène du cancer du sein, il existe plusieurs classifications dans le but d'adopter la meilleure décision thérapeutique qui repose sur les caractéristiques histologiques et moléculaires (Kamina, 1984). Par conséquent, on peut distinguer différentes formes de cancers du sein en fonction des cellules à partir desquelles il se développe. La plupart des cancers du sein (95%) sont des adénocarcinomes, ce qui signifie qu'ils se forment suite aux transformations subies par les cellules épithéliales. Si les cellules cancéreuses sont limitées aux canaux (Figure 10A) ou aux lobules (Figure 10B) du sein, le carcinome est dit *in situ*. En revanche, si les cellules cancéreuses traversent la membrane "basale" des canaux ou des lobules et envahissent les tissus avoisinants, le carcinome est dit infiltrant. La propagation des cellules cancéreuses peut entraîner des métastases soit par voie lymphatique dans les ganglions situés sous l'aisselle, soit par voie veineuse. Contrairement aux carcinomes canaux, les carcinomes lobulaires présentent un meilleur pronostic (Galant et al., 2010).

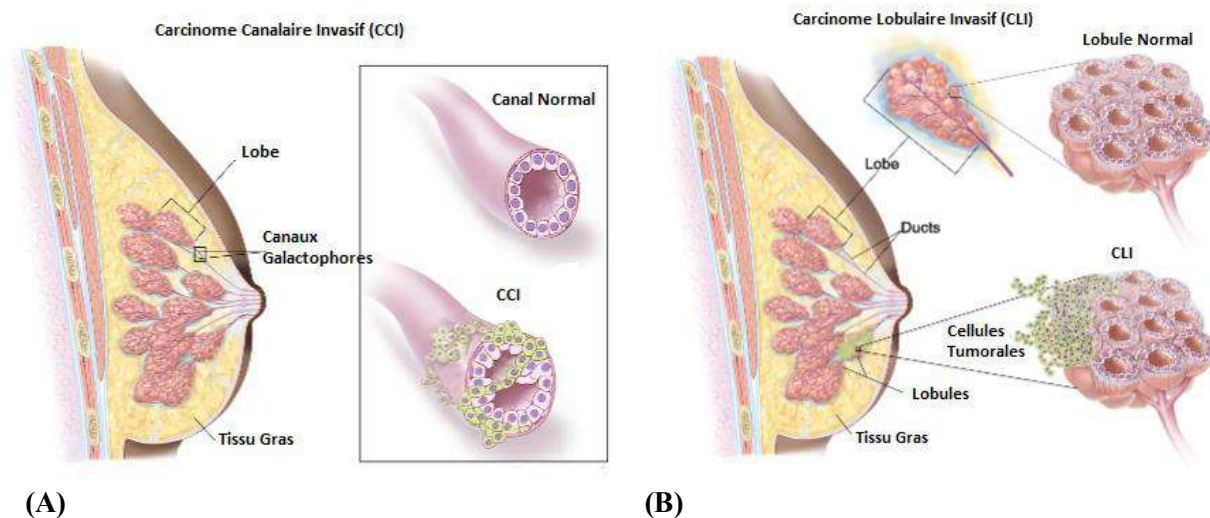


Figure 10. (A) Carcinome canalaire et (B) carcinome lobulaire invasifs (infiltrants) du sein

<https://www.delawarebreastcare.com/fr/breast-cancer-2>

3.1. Classification histologique

Le cancer du sein peut être classé en deux catégories principales : *in situ* et invasif. Les formes *in situ* incluent le carcinome canalaire *in situ* (CCIS) et le carcinome lobulaire *in situ* (CLIS), où la prolifération néoplasique est confinée au système canalo-lobulaire sans invasion du stroma. Les carcinomes invasifs incluent le carcinome invasif de type spécial (NST) et d'autres sous-types avec des caractéristiques spécifiques (Aydiner et al, 2019).

La première étape du pathologiste consiste en un diagnostic histo-pathologique correct. Il est tout d'abord essentiel d'avoir un ensemble de caractères fondamentaux (Tableau 1). Le pathologiste doit fournir dans son compte rendu : le diagnostic lésionnel, les limites d'exérèse et l'envahissement ganglionnaire. Il devra préciser les paramètres à valeur pronostique ou importants pour la prise en charge thérapeutique.

Tableau 1. Données minimales du compte rendu histo-pathologique en cas de cancer du sein (Penault-Llorca, 2014).

Renseignements généraux
Identification de la patiente
Renseignements cliniques
Type de prélèvement
Côté du prélèvement, quadrant
Description macroscopique (préciser si encrage des limites, radio pièce, mesures)
Examen extemporané
Description microscopique (peut être courte)
Conclusions (+/- rappel du contexte clinique)
• Type histo-pathologique
• Taille
• Grade SBR, mSBR
• Emboles
• Ganglions
• Sentinelle
• Curage nombre de ganglions envahis/nombre de ganglions prélevés
• Limites d'exérèse
• Statut des récepteurs hormonaux
• Statut HER2

L'histopathologie des cancers du sein est basée sur la diversité des traits morphologiques présents dans les tumeurs. La classification histopathologique, qui est adoptée dans le monde entier, comprend 20 grands types de tumeurs et 18 sous types mineurs (Tableau 2). La plupart des noms donnés dépendent de la région mammaire où les cellules cancéreuses ont débuté leur prolifération. Le carcinome canalaire est le type le plus fréquent.

Tableau 2. Classification OMS des cancers du sein (Tavassoli et al., 2004).

Carcinome non infiltrant ou in situ
Carcinome intracanauxaire ou canalaire <i>in situ</i> (CCIS)
Carcinome lobulaire <i>in situ</i>
Carcinome infiltrant ou invasif
Carcinome infiltrant de type non spécifique (canauxaire TNS)
Carcinome de type mixte
Carcinome pléomorphe
Carcinome avec cellules géantes ostéoclastiques
Carcinome avec aspects choriocarcinomateux
Carcinome avec aspects mélanocytaires
Carcinome lobulaire infiltrant
Carcinome tubuleux
Carcinome cribiforme infiltrant
Carcinome médullaire Carcinome produisant de la mucine
Carcinome mucineux
Cystadénocarcinome et carcinome à cellules cylindriques sécrétantes
Carcinome à cellules en bague à chaton
Tumeurs neuroendocrines du sein
Carcinome neuroendocrine de type solide
Carcinoïde atypique
Carcinome à petites cellules
Carcinome neuroendocrine à grandes cellules
Carcinome papillaire infiltrant
Carcinome micropapillaire infiltrant
Carcinome apocrine
Carcinome métaplasique
Carcinome métaplasique de type épithélial pur
Carcinome épidermoïde
Adénocarcinome avec métaplasie à cellules fusiformes
Carcinome adénosquameux
Carcinome mucoépidermoïde
Carcinome métaplasique mixte à composante épithéliale et conjonctive
Carcinome à cellules riches en lipides
Carcinome sécrétant
Carcinome oncocylique
Carcinome adénoïde kystique
Carcinome à cellules acineuses
Carcinome à cellules claires (riches en glycogène)
Carcinome sébacé
Carcinome inflammatoire
Maladie de Paget du mamelon

Le grade histologique est un élément important qui permet une nette stratification pronostique et la distinction de trois catégories de cancer du sein. Le grade de Scarff, Bloom et Richardson (SBR) modifié par Elston et Ellis (Nottingham) est celui qui est actuellement admis. Il est basé sur 3 critères (<https://www.carenity.com/infos-maladie/cancer-du-sein-155>) :

- le degré de différenciation tubulo-glandulaire de la tumeur ;
- l'irrégularité de la taille des noyaux ;
- l'index mitotique (nombre de mitoses) ;

On évalue alors :

- l'aspect **histologique** : bien (1), moyennement (2) ou peu différencié (3) ;
- l'aspect du **noyau** : faible (1), moyenne (2) ou forte (3) **anisonucléose** (tous les noyaux des cellules d'un tissu ne sont pas de la même taille) ;
- le nombre de **mitoses** sur 10 champs : absence (1), peu < 12 (2) ou beaucoup > 12 (3).

Ainsi, le total des scores donne le grade SBR :

- **SBR I** : scores 3, 4 ou 5 pour les tumeurs **peu agressives** ;
- **SBR II** : scores 6 ou 7 pour les tumeurs **intermédiaires** ;
- **SBR III** : scores 8 ou 9 pour les tumeurs **agressives** ;

3.2. Classification moléculaire

La première classification moléculaire des cancers du sein a été établie par Sorlie et Perou (2001). Grâce aux avancées en génomique, il a été possible de classer les tumeurs en sous-types basés sur l'expression génique, définis par la signature PAM50: basal-like, HER2-enriched, luminal, et normal-like (Perou et al., 2000). La classification actuelle inclut quatre sous-types moléculaires basés sur les biomarqueurs (ER, PR, HER2 et Ki-67) (Tableau 3 ; Figure 11).

Tableau 3. Caractéristiques des sous-types moléculaires du cancer du sein (Herschowitz et al., 2007).

Sous-type	Incidence	IHC	Pronostic
Luminal A	50–60%	RE+ et/ou RP+ Her2 négatif Ki67 bas	Bon
Luminal B	15%	RE+ et/ou RP+ Her2 négatif Ki67 élevé	Intermédiaire
Basal like	10–20%	RE+ et/ou RP+ Her2 positif Her2 négatif	Péjoratif
Her2 positif	15–20%	RE négatif RP négatif Her 2 surexprimé ou amplifié	Péjoratif
Claudin low	12–14%	RE négatif RP négatif Her2 négatif	Péjoratif

1/ **Luminal A** : RH+/HER2- avec forte expression des ER/PR, faible prolifération (Ki-67). Représentant 50 à 60% des cas, ces tumeurs ont un bon pronostic et répondent bien à l'hormonothérapie.

2/ **Luminal B** : RH+/HER2- ou RH+/HER2+ avec faible expression des ER/PR, prolifération élevée (Ki-67). Environ 10% des cas, plus proliférantes mais chimio-sensibles.

3/ **HER2 positive** : HER2+/RH- ou HER2+/RH+ avec un grade SBR élevé. Environ 15% des cas, agressives mais répondant bien aux traitements anti-HER2.

4/ **Basal-like ou triple-négatif** : RH-/HER2- avec un grade SBR élevé. Environ 15% des cas, agressives et souvent traitées par chimiothérapie en raison de l'absence de récepteurs hormonaux (Sotiriou et al., 2006 ; Goldhirsch et al., 2013).

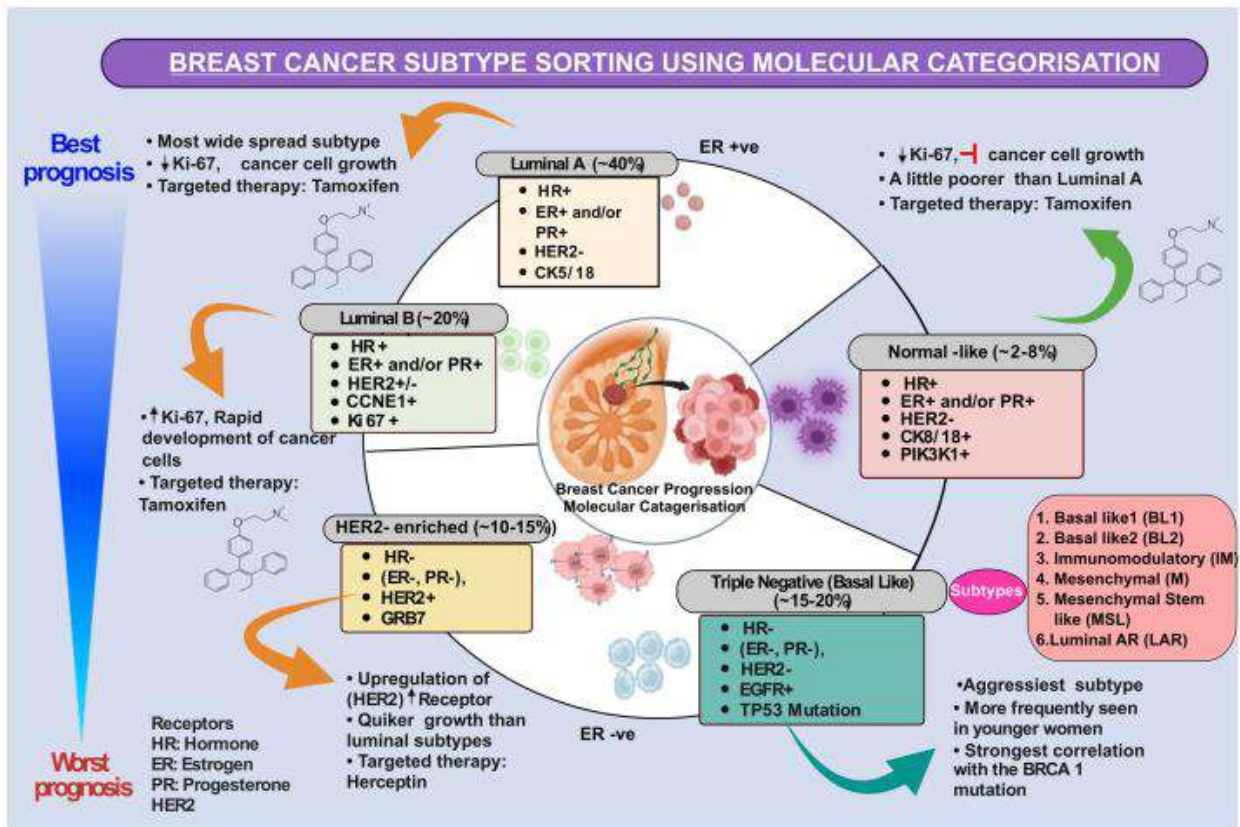


Figure 11. Catégorisation moléculaire du cancer du sein : classification moléculaire, thérapie et pronostic (Pal et al., 2024).

L'expression hormonale a un impact négatif sur la prolifération cellulaire et le grade de la tumeur. Le sous-type Luminal A répond mieux au traitement endocrinien, tandis que le TNBC est avancé, nucléaire et mitotiquement actif et présente un mauvais pronostic (ER+et/ou PR+ (récepteur d'œstrogène positif et/ou récepteur de progestérone positif), HER2- (récepteur du facteur de croissance épidermique humain 2 négatif), HR+ (récepteur hormonal positif)).

Le sous-type Claudin-low a été ajouté à la classification initiale en 2007 (Herschkowitz et al., 2007). Ces tumeurs sont caractérisées par l'absence d'expression des marqueurs luminaux, la faible ou absence d'expression des protéines de jonction et d'adhésion cellulaire (E-Cadhérine, Claudins 3, 4, 7) et par la surexpression des gènes de la réponse immunitaire. Elles sont typiquement triples négatives, ce en quoi elles se rapprochent du sous-groupe basal-like. L'identification du sous-type claudin-low permet de générer des hypothèses sur le développement tumoral (Figure 12) (Prat et Perou, 2009).

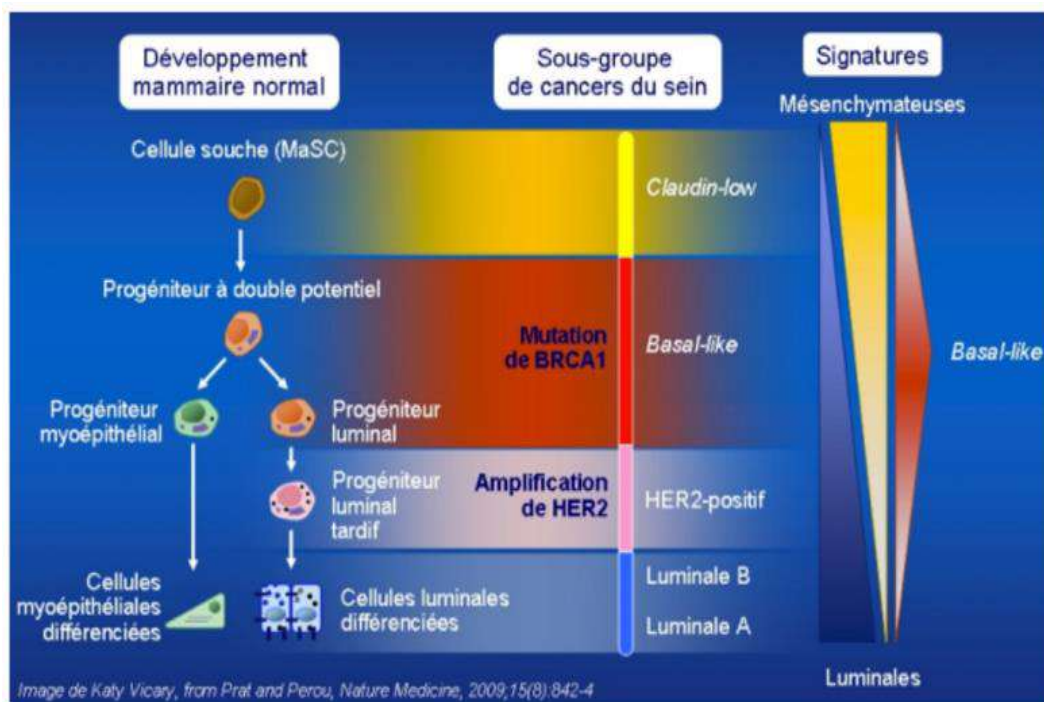


Figure 12. Hypothèse du développement tumoral selon les sous-types moléculaires du cancers du sein (Prat et Perou, 2009).

3.3. Classification selon le stade TNM

La classification standard TNM (Tableau 4), utilisée au niveau international, a été établie en 1959 par l'*American Joint Committee on Cancer* (AJCC). Il s'agit d'une méthode de classification des cancers basée sur trois critères principaux : la taille de la tumeur (T), l'implication des ganglions lymphatiques (N), et la présence de métastases (M). Chaque composante est évaluée individuellement, ce qui permet de déterminer le stade global du cancer (Edge et al., 2010).

Tableau 4. Classification TNM des cancers du sein (Edge et al., 2010).

<p>Tumeur primaire T</p> <p>Tx : La tumeur primitive ne peut pas être évaluée T0 : La tumeur primitive n'est pas palpable</p> <ul style="list-style-type: none"> • Tis : Carcinome in situ • Tis (DCIS) : Cancer intra-canalariare in situ • Tis (LCIS) : Cancer lobulaire in situ • Tis (Paget) : Maladie de Paget du mamelon sans tumeur sous-jacente <p>T1 : Tumeur ≤ 2 cm dans sa plus grande dimension T1mic : Micro-invasion ≤ 1 mm dans son plus grand diamètre</p> <ul style="list-style-type: none"> • T1a : 1 mm < tumeur ≤ 5 mm dans sa plus grande dimension • T1b : 5 mm < tumeur ≤ 1 cm dans sa plus grande dimension • T1c : 1 cm < tumeur ≤ 2 cm dans sa plus grande dimension <p>T2 : 2 cm < Tumeur ≤ 5 cm dans sa plus grande dimension T3 : Tumeur > 5 cm dans sa plus grande dimension T4 : Tumeur, quelle que soit sa taille, avec une extension directe à la paroi thoracique (a) ou à la peau (b)</p> <ul style="list-style-type: none"> • T4a : Extension à la paroi thoracique en excluant le muscle pectoral • T4b : Œdème (y compris peau d'orange) ou ulcération de la peau du sein, ou nodules de perméation situés sur la peau du même sein • T4c : T4a + T4b • T4d : carcinome inflammatoire
<p>Ganglions lymphatiques régionaux pN</p> <p>Nx : l'envahissement des ganglions lymphatiques régionaux ne peut pas être évalué (par exemple déjà enlevés chirurgicalement ou non disponibles pour l'analyse anatomopathologique du fait de l'absence d'évidement)</p> <p>N0 : Absence de signe d'envahissement ganglionnaire régional histologique et absence d'examen complémentaire à la recherche de cellules tumorales isolées</p> <ul style="list-style-type: none"> • N0(i-) : absence d'envahissement ganglionnaire régional histologique, étude immunohistochimique négative (IHC) • N0(i+) : absence d'envahissement ganglionnaire régional histologique, IHC positive, avec des amas cellulaires $\leq 0,2$ mm (considéré comme sans métastase ganglionnaire) • N0(mol-) : absence d'envahissement ganglionnaire régional histologique, biologie moléculaire négative (RT-PCR : reverse transcriptase polymerase chain reaction) • N0(mol+) : absence d'envahissement ganglionnaire régional histologique, biologie moléculaire positive (RT-PCR) <p>N1mi : micrométastases > 0,2 mm et ≤ 2 mm N1 : envahissement de 1 à 3 ganglions axillaires et/ou envahissement des ganglions de la CMI détecté sur ganglion sentinelle sans signe clinique</p> <ul style="list-style-type: none"> • N1a : envahissement de 1 à 3 ganglions axillaires • N1b : envahissement des ganglions de la CMI détecté sur ganglion sentinelle sans signe clinique • N1c : envahissement de 1 à 3 ganglions axillaires et envahissement des ganglions de la CMI détecté sur ganglion sentinelle sans signe clinique (pN1a + pN1b) <p>N2 : Envahissement de 4 à 9 ganglions axillaires ou envahissement des ganglions mammaires internes homolatéraux suspects, en l'absence d'envahissement ganglionnaire axillaire</p> <ul style="list-style-type: none"> • N2a : envahissement de 4 à 9 ganglions axillaires avec au moins un amas cellulaire > 2 mm • N2b : envahissement des ganglions mammaires internes homolatéraux suspects, en l'absence d'envahissement ganglionnaire axillaire

<p>N3 : Envahissement d'au moins 10 ganglions axillaires ou envahissement des ganglions sous-claviculaires (niveau III axillaire) ou envahissement des ganglions mammaires internes homolatéraux suspects avec envahissement ganglionnaire axillaire ou envahissement de plus de 3 ganglions axillaires et envahissement des ganglions de la CMI détecté sur ganglion sentinelle sans signe clinique ou envahissement des ganglions sus-claviculaires homolatéraux</p> <ul style="list-style-type: none"> • N3a : Envahissement d'au moins 10 ganglions axillaires (avec au moins un amas cellulaire > 2 mm) ou envahissement des ganglions sous-claviculaires • N3b : envahissement des ganglions mammaires internes homolatéraux suspects avec envahissement ganglionnaire axillaire ou envahissement de plus de 3 ganglions axillaires et envahissement des ganglions de la CMI détecté sur ganglion sentinelle sans signe clinique • N3c : envahissement des ganglions sus-claviculaires homolatéraux
Métastases à distance (M)
<p>Mx : Renseignements insuffisants pour classer les métastases à distance M0 : Absence de métastases à distance M1 : Présence de métastase(s) à distance</p>

3.4. Classification concordante TNM/Stade

Cette classification permet d'évaluer avec précision la gravité de cette maladie et de recommander le traitement le plus approprié. Elle permet de catégoriser les tumeurs en différentes étapes en fonction de leur taille, de leur propagation aux ganglions lymphatiques et de la présence de métastases. Chaque stade reflète un degré de progression du cancer, ce qui guide les médecins dans la prise en charge des patientes pour déterminer le traitement le plus approprié et prédire le pronostic (Tableau 5) (Edge et al., 2010).

Tableau 5. Classification des stades des cancers du sein en fonction de la classification TNM (Lacroix-Triki et Penault-Llorca, 2017).

Stade	Tumeur (T)	Ganglion(N)	Métastase (M)	Description
0	Tis (in situ)	N0	M0	Cancer non invasif
I	T1	N1	M1	Tumeur <2 cm, pas de propagation à l'extérieur du sein
II	T0/1	N1	M0	Tumeur de 2 à 5 cm et/ou atteinte ganglionnaire satellite mineur
	T2	N0/1		
	T3	N0		
III	T0/1/2	N2	M0	Atteinte local importante et/ou atteinte ganglionnaire satellite majeure
	T4	N0/1/2		
	Tout T	N3		
IV	Tout T	Tout N	M1	Tumeur avancée localement et métastase à distance

4. DIAGNOSTIC, PRONOSTIC ET TRAITEMENT DU CANCER DU SEIN

4.1. Signes et symptômes cliniques

Le signe le plus commun et facile à reconnaître lors d'un cancer du sein est la modification de la forme et de l'aspect du sein. Généralement, la présence d'une boule (ou masse) non douloureuse, à contours pas très bien définis, est le premier signe de découverte souvent de façon fortuite, par autopalpation ou lors d'un examen gynécologique. D'autres signes peuvent être observés tels qu'un changement de taille, de forme ou d'apparence du sein, une modification de la peau (lésions, couleur rougeâtre, apparition de fossettes, grossissement de veine, aspect de peau d'orange). Un autre indicateur souvent constaté est une masse solide indolore au niveau de l'aisselle, là où sont localisés les ganglions axillaires (**Alleman et al., 2010**).

4.2. Diagnostic

Le développement du cancer du sein est un processus complexe, et sa prévention demeure un défi mondial. Le diagnostic précoce est essentiel pour améliorer les taux de survie (**Ifeanyi Obeagu et al., 2023**). Lorsqu'une anomalie est détectée lors des examens de dépistage ou en présence de symptômes, des investigations sont menées pour confirmer ou exclure le cancer.

Le diagnostic du cancer du sein repose sur plusieurs méthodes :

1. **Examen Clinique** : réalisé par un oncologue ou un sénologue, il comprend une inspection visuelle pour repérer des changements de forme, couleur ou texture, ainsi qu'une palpation pour détecter des masses ou ganglions anormaux (**Henderson et Ferguson, 2018 ; Ngan et al., 2020**).
2. **Imagerie** : des techniques telles que la mammographie, l'échographie mammaire et l'IRM mammaire sont essentielles pour détecter les anomalies. La mammographie utilise des rayons X (**Schaller et al., 2015**), tandis que l'échographie utilise des ultrasons (**Candelaria, 2013**). L'IRM permet de différencier les anomalies bénignes de celles malignes (**D'Olne et al., 2022**).
3. **Biopsie** : en cas de doute, une biopsie mammaire est réalisée pour prélever des tissus à des fins d'analyse, souvent guidée par échographie. La biopsie est essentielle pour établir un diagnostic précis (**Vazquez-Sequeiros et Boixeda de Miquel, 2005**).

4.3. Évaluation du Pronostic

L'évaluation du pronostic du cancer du sein est cruciale pour prédire la probabilité de survie à long terme et de récurrence de la maladie (**Figure 11**). Plusieurs facteurs sont pris en compte :

1. **Âge du patient** : l'âge peut influencer le pronostic, les jeunes et les plus âgées ayant des perspectives différentes.
2. **Surexpression de HER2** : les cancers surexprimant HER2 nécessitent souvent un traitement ciblé. Le pronostic dépend de la réponse au traitement anti-HER2.
3. **Ki-67** : c'est un marqueur de prolifération cellulaire; un taux élevé est associé à un risque accru de récurrence.
4. **Récepteurs hormonaux** : la présence des récepteurs d'œstrogène et de progestérone influence le pronostic et le traitement. Les cancers hormonaux positifs ont généralement un meilleur pronostic.
5. **Stade du cancer** : évalué selon la classification TNM ; les stades 0 et I ont un pronostic plus favorable.
6. **Grade tumoral** : est une mesure de l'anomalie des cellules cancéreuses. Un grade élevé indique une malignité accrue et un pronostic moins favorable.
7. **État des ganglions lymphatiques** : plus il y a de ganglions atteints, moins le pronostic est favorable.
8. **Statut BRCA** : les mutations *BRCA1* et *BRCA2* sont liées à un risque accru de cancer du sein et influencent le pronostic et les options de traitement.

Le pronostic du cancer du sein est souvent exprimé en termes de taux de survie à 5 ans, mais il est essentiel de noter que de nombreuses personnes survivent bien au-delà de cette période, avec un suivi médical régulier et un mode de vie sain ayant un impact significatif sur les résultats (**Stankov et al., 2012 ; Phung et al., 2019 ; Susini et al., 2022**).

Toutefois, prédire de manière individuelle le pronostic et la réponse au traitement sur la base de la classification anatomoclinique serait insuffisant. De nouvelles techniques de biologie moléculaire ayant permis le développement de plusieurs signatures génomiques devraient permettre d'évaluer la biologie tumorale individuelle ainsi que les particularités de l'hôte. À terme, l'enjeu est de combiner les données anatomocliniques et moléculaires afin de prédire au mieux le pronostic et la réponse au traitement (**Mailliez, 2014**).

4.4. Traitement

Le traitement du cancer du sein est basé sur le stade de la maladie, le type de cancer et l'état de santé général du patient. Les tests génomiques, tels que MammaPrint et Blueprint, évaluent le risque de récurrence en analysant l'activité de gènes clés, aidant ainsi à personnaliser le traitement (*Breast Cancer Treatment* - NCI) (Cardoso, 2016 ; Audeh, 2019 ; Haan, 2022).

Les options thérapeutiques comprennent :

1. **Chirurgie** : ablation de la tumeur, pouvant être conservatrice (lumpectomie) ou mastectomie totale (Admoun et Mayrovitz, 2021 ; Wang, 2023).
2. **Chimiothérapie** : médicaments administrés oralement ou par voie intraveineuse pour réduire ou tuer les cellules cancéreuses (Asaoka, 2020 ; Wang et Mao, 2020).
3. **Radiothérapie** : utilisation de rayons haute énergie pour détruire les cellules cancéreuses résiduelles après chirurgie (Hausmann, 2020).
4. **Thérapie hormonale** : blocage des hormones nécessaires à la croissance des cellules cancéreuses (Rosso, 2023).
5. **Thérapie ciblée** : en particulier pour les cancers HER2-positifs, où des médicaments comme le trastuzumab (Herceptin) sont utilisés pour cibler spécifiquement les cellules exprimant HER2 (Jallah et Dweh, 2023).
6. **Immunothérapie** : l'immunothérapie est une option majeure pour les TNBC avancés PD-L1+, avec des réponses prometteuses (Schneble et al., 2015 ; Debien et al., 2023).

Le diagnostic précoce, une évaluation précise du pronostic et un traitement ciblé, notamment pour les cancers HER2 positifs, sont essentiels pour améliorer les résultats et la survie des patientes atteintes de cancer du sein.

5. LES MARQUEURS TUMORAUX

Les marqueurs tumoraux (hormone, protéine, antigène ...) sont des éléments clés du dépistage du cancer. Il s'agit de substances, révélatrices de la présence d'une tumeur, et naturellement présentes dans notre organisme. Cette substance peut aussi provenir des cellules tumorales elles-mêmes. L'augmentation ou la présence de ces marqueurs n'est pas systématiquement révélatrice de cancer. Les valeurs de leur concentration peuvent indiquer le stade du cancer, sa progression et son agressivité. Il peut également être utile pour évaluer la réussite ou l'échec de la réponse au traitement ou apprécier le risque de récurrences (Nagpal et al., 2016).

Les principaux marqueurs tumoraux impliqués dans le cancer du sein sont les facteurs hormonaux : l'œstrogène, la progestérone, HER2 et l'aromatase. Un autre marqueur lié au cycle cellulaire, Ki-67, est également impliqué dans cette pathologie. Ces marqueurs affectent la croissance tumorale, la signalisation cellulaire et la réponse immunitaire.

5.1. L'œstrogène

Le cancer du sein est hormono-dépendant dans la majorité des cas ; les cellules cancéreuses sont sensibles aux hormones, plus particulièrement aux œstrogènes sécrétés par les ovaires et le tissu graisseux mammaire (Travis et Key, 2003). Les œstrogènes sont des hormones stéroïdiennes qui dérivent du cholestérol. Chez les femmes, l'hormone la plus prédominante et la plus puissante est le 17 β -estradiol, également connu sous le nom d'E2 qui exerce ses effets en se liant à ses récepteurs ER α et Er β (Figure 13) (Björnström et Sjöberg, 2005). Ces deux isoformes sont faiblement exprimés dans le tissu mammaire normal. Toutefois, les cellules tumorales mammaires expriment plus d'ER α que d'ER β et 70% des cancers mammaires sont ER α positifs (Renoir et al., 2013 ; Hamilton et al., 2015).

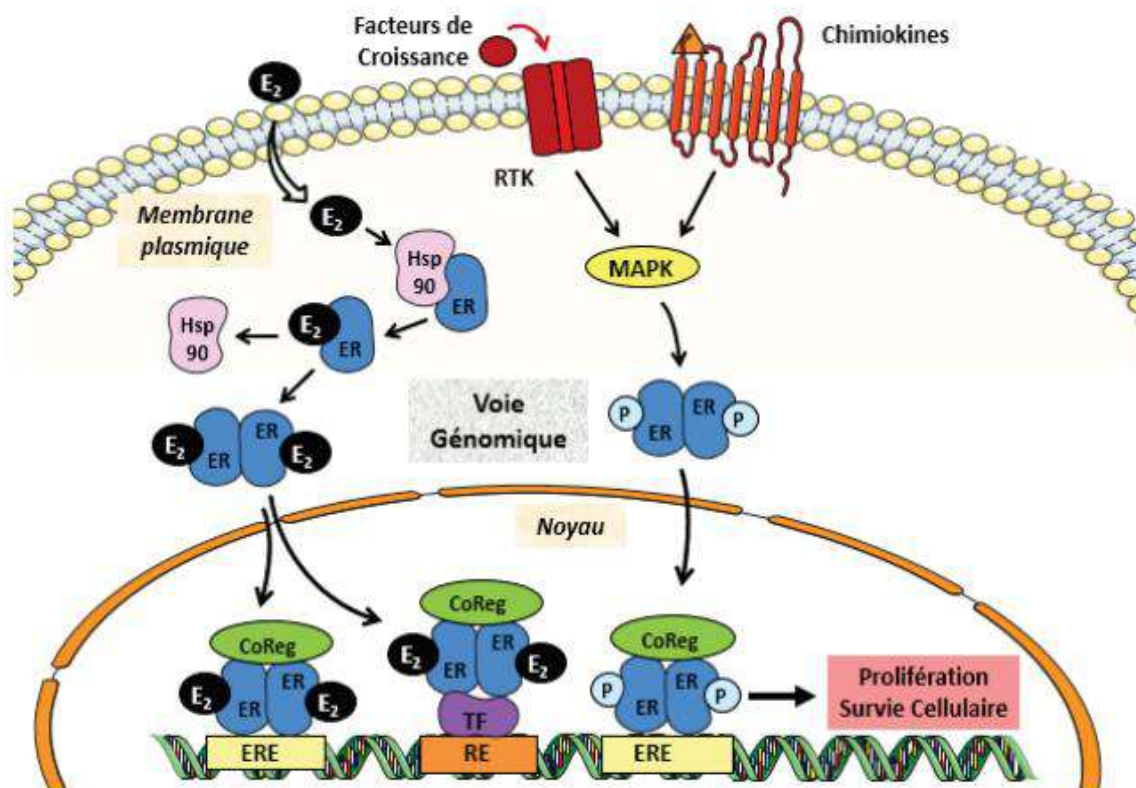


Figure 13. Voies de signalisation génomiques d'ER α (Le Romancer et al., 2011).

ER α se détache de Hsp90 et se dimérise, suite à la fixation d'E2. Le complexe formé par le ligand et le récepteur se déplace vers le noyau et s'attache, soit directement, soit indirectement par l'intermédiaire de facteurs de transcription, aux éléments de réponse spécifiques sur l'ADN. Il initie donc la transcription des gènes cibles liés à la survie cellulaire, en requérant les corégulateurs appropriés. Toutefois, l'ER α peut être activé sans E2 : des facteurs de croissance peuvent déclencher sa phosphorylation, qui est suivie de sa translocation vers le noyau, stimulant ainsi les voies de prolifération et de survie cellulaire.

E2: 17- β -œstradiol; **Hsp 90:** Heat shock protein 90; **ER:** Récepteur oestrogénique; **CoReg:** Co-régulateurs ; **ERE:** Eléments de Réponse à l'œstrogène ; **FT:** Facteurs de Transcription ; **RE:** éléments de réponse ; **RTK:** Récepteur Tyrosine Kinase ; **MAPK:** Mitogen-Activated Protein Kinase.

L'œstrogène est connu pour favoriser la croissance et la prolifération des cellules tumorales, en se liant à des récepteurs spécifiques présents sur ces cellules. Environ deux tiers des cancers du sein expriment des récepteurs à œstrogènes, ce qui signifie qu'ils sont sensibles à l'action stimulante de cette hormone. Lorsqu'il se lie à ces récepteurs, l'œstrogène active des voies de signalisation qui entraînent une augmentation de la prolifération cellulaire et une résistance à l'apoptose. Des études ont montré que même les tumeurs sans récepteurs à œstrogènes peuvent croître plus rapidement en présence d'œstrogènes. Cela est dû à l'effet indirect de l'œstrogène sur le microenvironnement tumoral, en particulier en améliorant la vascularisation (Russo et Russo, 2006 ; Péqueux et al., 2012 ; Rothenberger et al., 2018).

Des taux élevés d'œstrogène dans le corps, surtout à la ménopause, augmenteraient considérablement le risque de survenu d'un cancer du sein (Figure 14).

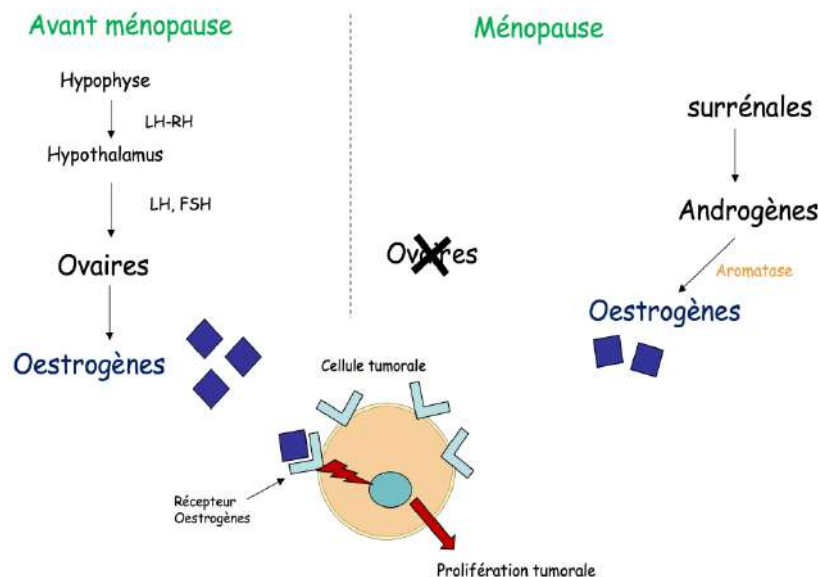


Figure 14. Les différentes sources de l'œstrogène (avant et après la ménopause) ([Inhibiteurs de l'Aromatase](#)).

Une tumeur mieux vascularisée reçoit plus de nutriments et d'oxygène, ce qui favorise sa croissance. L'œstrogène peut également moduler la réponse immunitaire en affectant la polarisation des lymphocytes T et la production de cytokines. Ces effets complexes soulignent l'importance d'une compréhension approfondie de l'influence de l'œstrogène sur les interactions entre le système immunitaire et les cellules cancéreuses. Les traitements anti-hormonaux, tels que les inhibiteurs des récepteurs d'œstrogènes ou les médicaments qui réduisent la production d'œstrogènes (comme les inhibiteurs de l'aromatase), peuvent être efficaces pour traiter les cancers hormono-dépendants.

- ❖ **L'Aromatase** : c'est une enzyme clé dans la biosynthèse des œstrogènes, et son expression est associée à certains cancers hormono-dépendants, notamment le cancer du sein. Codée par le gène *CYP19A1*, elle catalyse la conversion des androgènes, tels que la testostérone et l'androstènedione, en œstrogènes comme l'estradiol et l'estrone (**Figure 15**). Cette conversion se produit dans divers tissus, notamment les gonades, le tissu adipeux et le cerveau. Son activité est modulée par des facteurs tels que l'âge, l'obésité et les niveaux d'insuline, et son expression est souvent accrue dans les tissus tumoraux, en particulier dans le cancer du sein et de l'endomètre (**Warsy et al., 2017**).

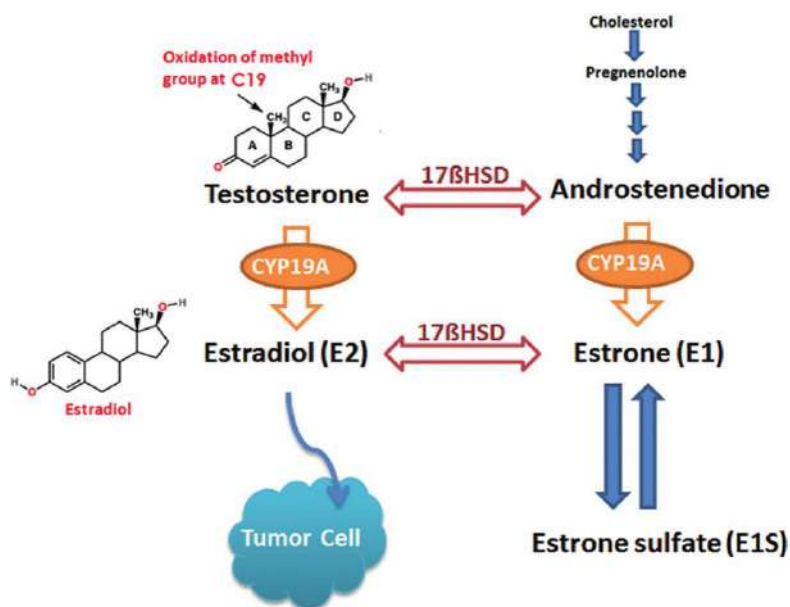


Figure 15. Rôle de l'aromatase dans la synthèse des œstrogènes (**Warsy et al., 2017**).

Dans le cancer du sein, l'aromatase joue un rôle crucial en contribuant à des niveaux élevés d'œstrogènes, lesquels stimulent la prolifération des tumeurs hormono-dépendantes (RE+). Les inhibiteurs de l'aromatase (IA), comme l'anastrozole et le létrozole, sont utilisés pour bloquer la production d'œstrogènes chez les femmes ménopausées atteintes de cancer du sein, inhibant ainsi la croissance des cellules tumorales. Des recherches récentes ont également mis en évidence un rôle plus complexe de l'aromatase dans la modulation de la réponse immunitaire dans le microenvironnement tumoral. Elle pourrait moduler la production de cytokines pro-inflammatoires, créant ainsi une inflammation chronique favorable à la croissance tumorale (Mukhopadhyay et al., 2015 ; Caciolla et al., 2020). De plus, il est suggéré que l'activation des récepteurs Toll-like (TLR) pourrait interagir avec les voies de signalisation de l'aromatase, influençant son expression et son activité dans le contexte tumoral (Shi et al., 2019).

5.2. La Progesterone

La progestérone peut avoir des effets mitigés sur le cancer, en fonction du contexte hormonal et des caractéristiques spécifiques de la tumeur. Tout comme l'œstrogène, la progestérone peut influencer la croissance des cellules tumorales. Dans certains cas, elle peut favoriser la progression tumorale, tandis que dans d'autres, elle peut avoir des effets protecteurs. Les mécanismes par lesquels la progestérone affecte le cancer du sein incluent son interaction avec les récepteurs hormonaux, qui modulent l'expression de gènes impliqués dans la prolifération cellulaire et l'apoptose (Figure 16). Les effets de la progestérone sur le système immunitaire sont moins bien compris que ceux de l'œstrogène. Cependant, des études suggèrent qu'elle peut moduler la réponse immunitaire et l'inflammation au sein du microenvironnement tumoral. Elle pourrait influencer la polarisation des lymphocytes T et la production de cytokines, affectant ainsi l'équilibre entre les réponses immunitaires pro-tumorales et anti-tumorales. Il existe une interaction significative entre les récepteurs d'œstrogènes et de progestérone dans le cancer du sein. L'activation des récepteurs d'œstrogènes peut influencer l'expression des récepteurs de progestérone et *vice versa*. Cette interaction est cruciale pour comprendre comment ces hormones influencent les traitements hormonaux. Par exemple, certains traitements ciblant les récepteurs d'œstrogènes peuvent également affecter l'activité des récepteurs de progestérone, ce qui a des implications pour l'efficacité thérapeutique. Enfin, bien que les interactions de la progestérone avec le système immunitaire soient encore moins bien comprises, des études montrent qu'elle peut également moduler la réponse immunitaire et l'inflammation dans le Microenvironnement tumoral (Chakravorty et al., 2023).

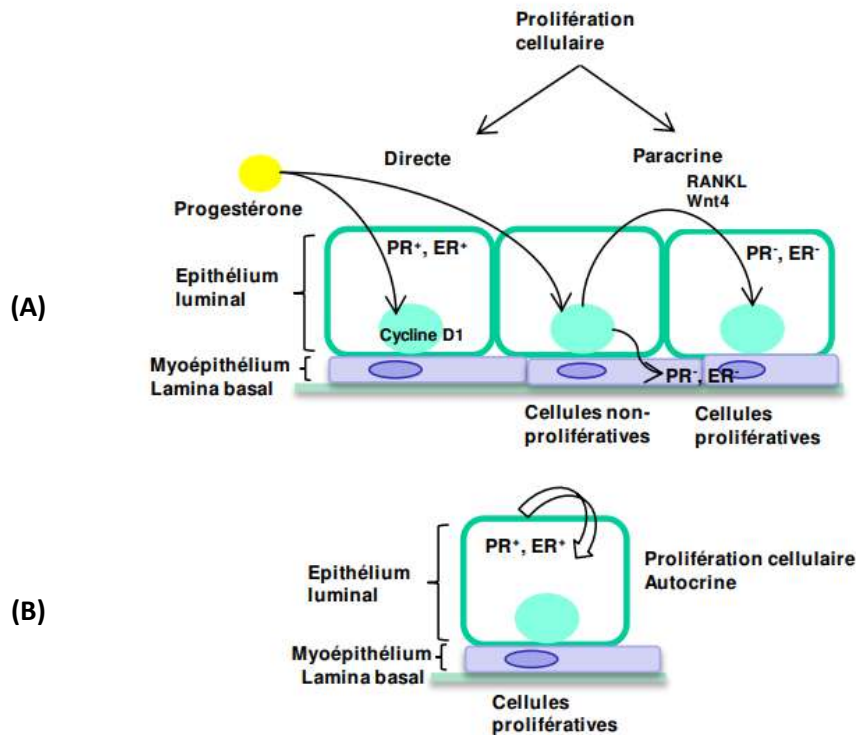


Figure 16. Mécanismes d'induction de la prolifération cellulaire par la progestérone (A) au niveau des cellules mammaires (B) cas de la transformation tumorale (Briskin, 2013 ; Knutson et Lange, 2014).

(A) La multiplication des cellules PR+/ER+ induite par la progestérone est réalisée par un mécanisme intrinsèque direct (la cycline D1). La progestérone déclenche la libération de facteurs paracrine (RANKL, Wnt4...) qui favorisent la multiplication des cellules voisines PR-/ER- à travers un processus paracrine. (B) Les cellules PR+/ER+ se transforment en cellule cancéreuse proliférative sous l'effet d'un mécanisme autocrine.

5.3. Human Epidermal Growth Factor Receptor 2 (HER2)

La découverte de HER2 remonte aux années 1980, lorsque des chercheurs ont identifié ce récepteur au sein du gène ERBB2 (Erb-B2 Receptor Tyrosine Kinase 2). À l'époque, des études démontraient sa similitude avec le récepteur du facteur de croissance épidermique humain, d'où son appellation. Cependant, c'est en 1987 que sa surexpression a été associée au cancer du sein, mettant ainsi en lumière son rôle potentiel dans cette maladie. Cette découverte a ouvert la voie à de nombreuses recherches visant à comprendre le rôle de l'HER2 dans la cancérogenèse et à développer des traitements ciblant ce récepteur (Raghav et Moasser, 2023).

L'amplification du gène HER2 est située sur le chromosome 17 en 17q21.1. Ce gène appartient à la famille des récepteurs ERBB, un groupe de quatre glycoprotéines transmembranaires avec une activité tyrosine kinase. En raison de son rôle dans la pathogenèse du cancer du sein, le gène

HER2 est considéré comme la cible privilégiée pour le traitement de cette maladie. L'hyperactivation du gène HER2 joue un rôle essentiel dans le développement du cancer du sein invasif en induisant une prolifération incontrôlée des cellules mammaires (Iqbal et Iqbal, 2014). Cette découverte a conduit au développement d'une thérapeutique ciblée par l'anticorps monoclonal trastuzumab (Herceptin®) (Figure 17). La réponse thérapeutique à ce traitement nécessite une évaluation rigoureuse du statut du gène dans les tumeurs ; seules les patientes présentant une forte expression de la protéine ou une amplification du gène sont éligibles pour ce traitement (Rubin et al., 2023).

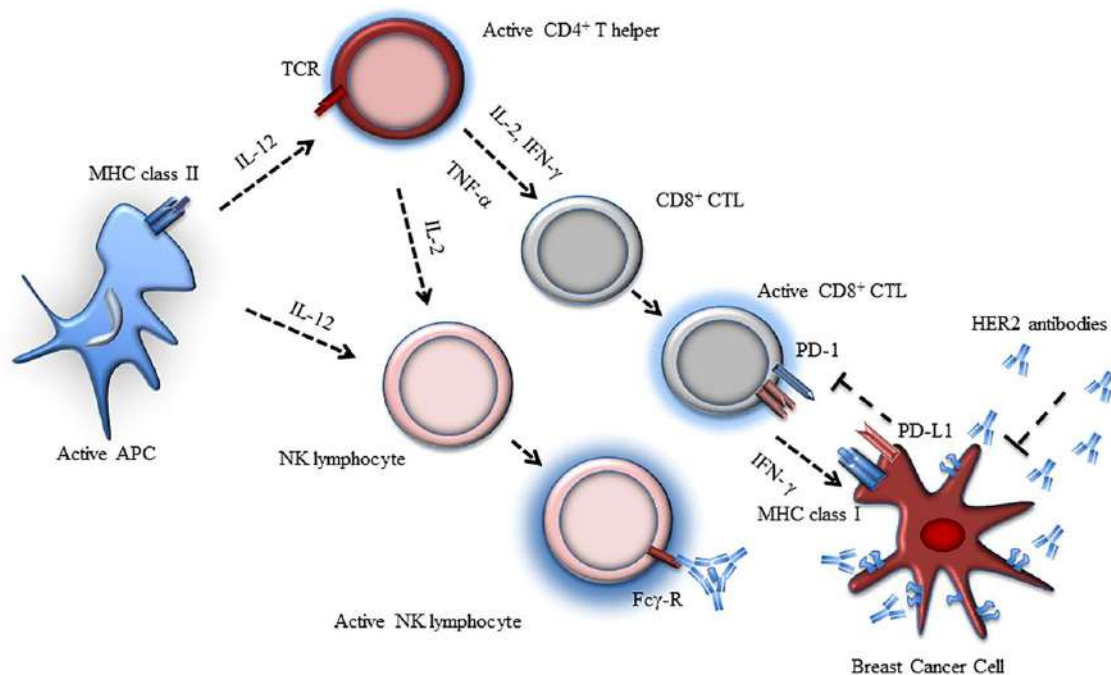


Figure 17. Réponse immunitaire contre le cancer du sein exprimant HER2 et effets immunomodulateurs des traitements par trastuzumab, ciblant HER2 (Iqbal et Iqbal, 2014).

Au niveau cellulaire, HER2 est principalement localisé à la surface cellulaire, où il joue un rôle clé dans les voies de signalisation liées à la croissance et à la différenciation cellulaire. Sa localisation membranaire agit dans son interaction avec d'autres protéines de signalisation et pour sa fonction régulatrice dans le contexte du cancer du sein. La localisation de l'HER2 dans l'organisme varie selon les tissus. Il est principalement exprimé dans les tissus épithéliaux, tels que la peau, le tractus gastro-intestinal, les voies respiratoires et les glandes mammaires. Sa surproduction dans certains tissus, comme les glandes mammaires, est associée à des implications pathologiques, telles que le cancer du sein (Jeong et al., 2017).

Les rôles de l'HER2 dans la signalisation cellulaire sont multiples et essentiels pour la régulation de la croissance, la survie et la différenciation cellulaire. En tant que récepteur de type tyrosine kinase, HER2 active des voies de signalisation intracellulaire, telles que la voie des MAP kinases et la voie PI3-kinase/AKT, qui sont impliquées dans la régulation de la croissance et de la prolifération cellulaires. De plus, l'HER2 interagit avec d'autres récepteurs de la famille de l'EGFR pour former des homodimères ou des hétérodimères, amplifiant ainsi les signaux de croissance cellulaire (**Figure 18**). Les interactions complexes et ces voies de signalisation font de l'HER2 une cible thérapeutique importante dans le traitement des cancers surexprimant ce récepteur (Hsu et Hung, 2016).

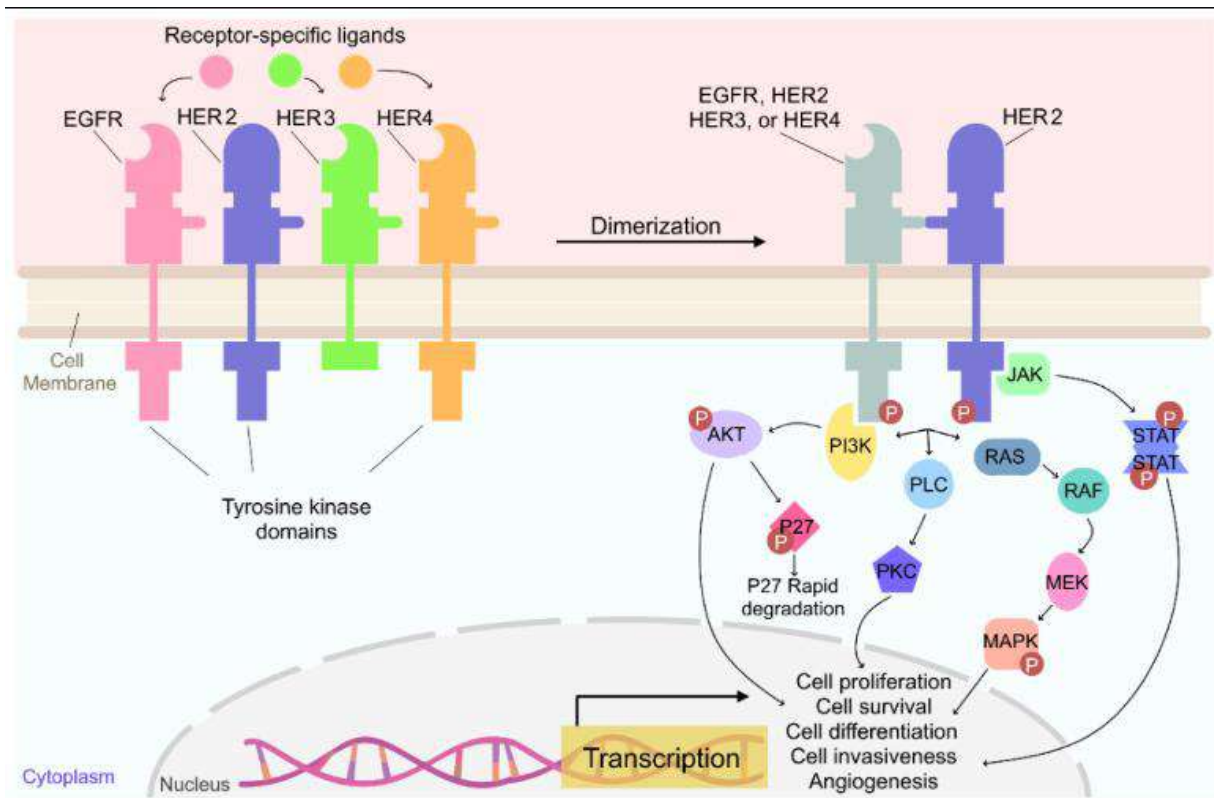


Figure 18. Représentation schématique de la voie de signalisation HER2 (Najjar et al., 2022).

Contrairement aux autres récepteurs de la famille EGFR, HER2 ne se lie à aucun ligand connu. Au lieu de cela, HER2 est activé après hétérodimérisation avec d'autres récepteurs activés de la famille EGFR ou par hétérodimérisation avec des récepteurs HER2 activés. La dimérisation des récepteurs entraîne la phosphorylation des résidus de tyrosine et la transduction du signal qui en résulte. PI3K/AKT, RAS/MEK/MAPK, JAK/STAT et PKC sont les voies de signalisation les plus courantes par lesquelles plusieurs cascades en aval sont activées, ce qui favorise de nombreux effets, notamment la prolifération cellulaire, la survie, la différenciation, l'angiogenèse et l'invasion. En outre, l'activation de PI3K/AKT entraîne également la dégradation de l'inhibiteur du cycle cellulaire $p27^{Kip1}$ et favorise ainsi la progression du cycle cellulaire.

Pour établir le diagnostic initial des cancers du sein infiltrants, il est essentiel de déterminer le statut HER2, dans le but de définir la classe moléculaire. Actuellement, la recherche de la surexpression des récepteurs HER2 dans les cellules mammaires est systématique lors de l'évaluation anatomopathologique. HER2 constitue un facteur pronostic de rechute et de survie globale défavorable (Méry et al., 2021).

Les récepteurs Toll-like (TLR), tels que TLR4 et TLR9, jouent un rôle crucial dans la modulation de la réponse immunitaire et peuvent influencer la signalisation de HER2. L'activation des TLR peut induire une production de cytokines pro-inflammatoires, renforçant l'inflammation dans le microenvironnement tumoral, ce qui peut à son tour moduler l'expression et l'activité de HER2. En raison de la capacité spécifique à reconnaître les DAMPs *via* les TLRs, la cascade de signalisation est déclenchée (Figure 19), initiant ainsi les cellules immunitaires innées et adaptatives, telles que les cellules dendritiques (DCs), les macrophages, les NK et les lymphocytes T (Huang et al., 2022).

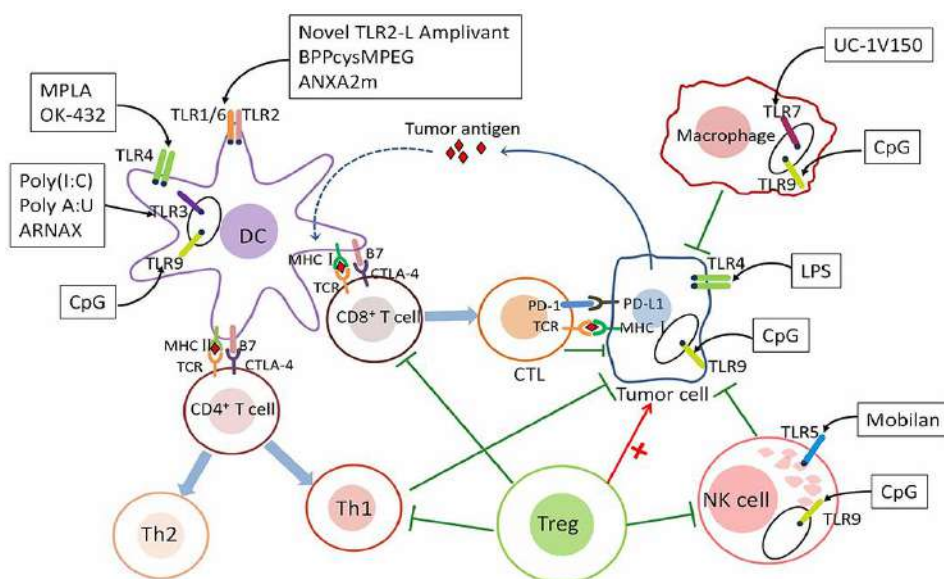


Figure 19. Mécanisme sous-jacent au rôle des TLRs dans l'immunité tumorale (Huang et al., 2022).

Certaines études suggèrent que l'activation des TLRs peut exacerber la signalisation de HER2, contribuant à la progression tumorale (González-Reyes et al., 2010 ; Oblak et Jerala, 2011), tandis que dans d'autres contextes, ces interactions pourraient renforcer la réponse immunitaire anti-tumorale (Zheng et Ma, 2022). Ainsi, les TLRs, en interagissant avec HER2, peuvent à la fois promouvoir la progression du cancer ou, dans certains cas, favoriser une meilleure

reconnaissance immunitaire des cellules tumorales, rendant cette interaction un axe de recherche prometteur pour les thérapies ciblées (Hu et al., 2024).

5.4. Indexe de l'activité cellulaire proliférative: Ki-67

C'est un antigène nucléaire exprimé par les cellules en prolifération, dans toutes les phases du cycle cellulaire (G1, S, G2 et M), à l'exception de la phase G0 (Figure 20). Il est retrouvé au niveau nucléolaire et intervenant dans la synthèse de l'ARN. Il permet d'évaluer le taux de prolifération des cellules tumorales (Romero et al., 2014 ; Nielsen et al., 2021). L'évaluation du Ki-67 est essentiellement utilisée comme un indicateur pronostique pour orienter la décision d'une thérapie adjuvante, mais aussi en tant qu'élément prédictif de la réponse au traitement néo-adjuvant dans le cas du cancer du sein RE+/HER2-. Dans certains types de cancer du sein tels que RE-/HER2+ et RE-/HER2-, une valeur élevée de Ki-67 en postnéo-adjuvant est associé à un pronostic défavorable (Penault-Llorca et al., 2017).

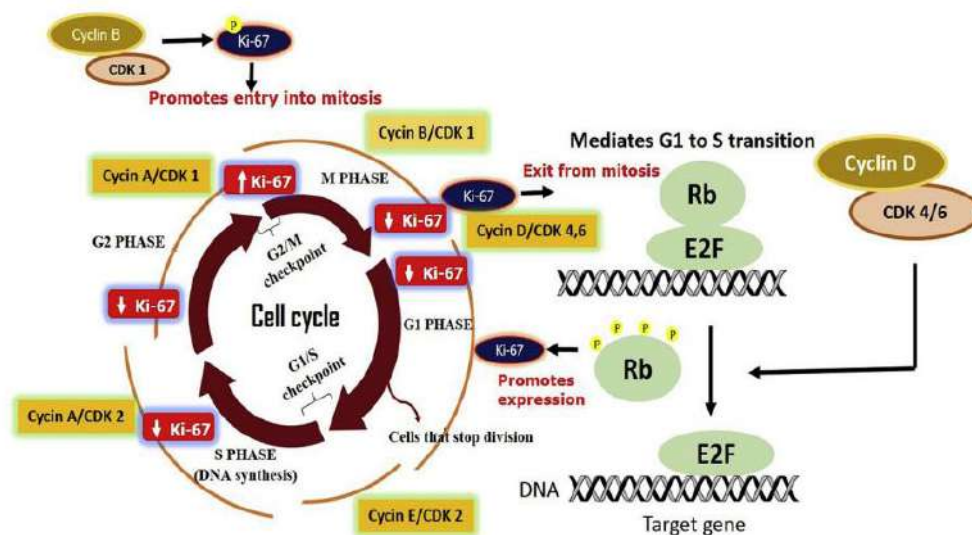


Figure 20. Régulation du cycle cellulaire par Ki-67 (Menon et al., 2019).

Ki-67 est présent dans toutes les phases du cycle cellulaire, sauf pendant la phase de repos G0. Des niveaux variables de Ki-67 varient de la phase G1 à la phase M, les niveaux les plus élevés étant observés au début de la phase M. Ils diminuent ensuite lentement et atteignent leur niveau le plus bas pendant l'anaphase et la télophase de la phase M. La régulation est assurée par différentes molécules effectrices. La phosphorylation de Ki-67 par CDK-1 pendant la phase M entraîne le passage à la mitose. Le complexe CDK 4/6 déclenche la phosphorylation de la famille des protéines du rétinoblastome (Rb) au cours de la phase G1 et libère les facteurs de transcription associés à E2F. Comme la protéine Ki-67 est liée aux protéines E2F dans des conditions cancéreuses, lors de la surexpression de l'E2F, l'ARNm de Ki-67 s'accumule.

6. FACTEURS IMMUNITAIRES

6.1. Microenvironnement Tumoral

Le microenvironnement tumoral constitue un écosystème dynamique composé de cellules stromales et immunitaires qui, sous l'influence des cellules cancéreuses, régissent les voies de signalisation biochimique et mécanique *via* la synthèse et le remodelage de la matrice extracellulaire (MEC), la formation de réseaux vasculaires (néoangiogenèse) et la promotion de la croissance tumorale (**Figure 21**). Dans le carcinome mammaire, la signalisation médiée par les récepteurs hormonaux joue un rôle central dans la régulation de la prolifération et de l'invasivité tumorales, non seulement de manière autonome, mais aussi par des interactions paracrines entre les cellules cancéreuses et le stroma (**Boyle, 2023**). Un dialogue dynamique entre les composants stromaux et la tumeur est indispensable à la progression et à la croissance de la tumeur (**Deshmukh et al., 2017 ; Charan et al., 2020**).

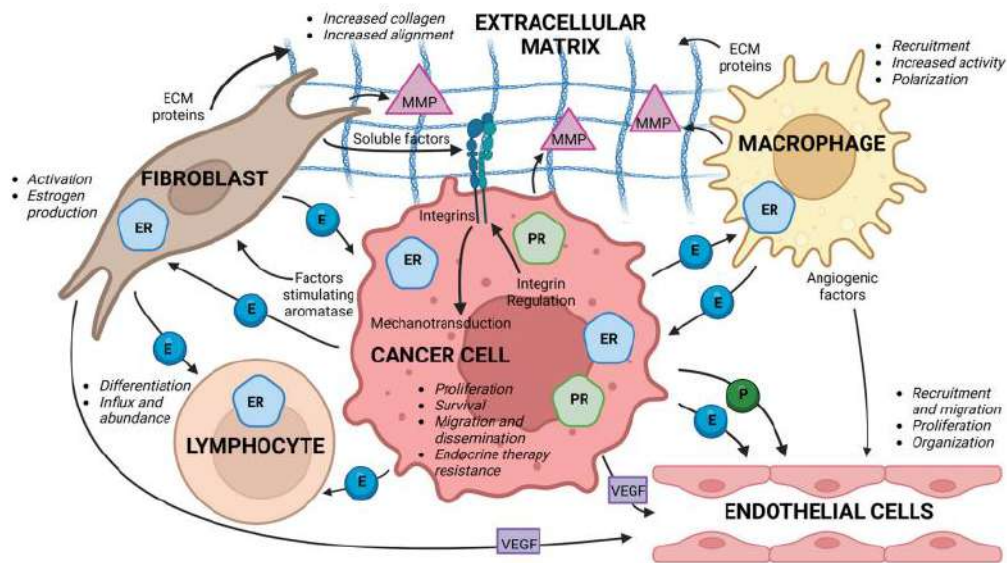


Figure 21. Le microenvironnement tumoral complexe du cancer du sein (**Boyle, 2023**).

Le microenvironnement tumoral du cancer du sein comprend différentes cellules, notamment des fibroblastes, des adipocytes, des macrophages et des cellules dendritiques, et sécrète différents facteurs de croissance et cytokines qui régulent la croissance et le développement des cellules tumorales (**Figure 22**). Comprendre l'impact du microenvironnement sur le cancer du sein est essentiel pour développer des approches thérapeutiques innovantes et améliorer les résultats cliniques (**Li et al., 2021**).

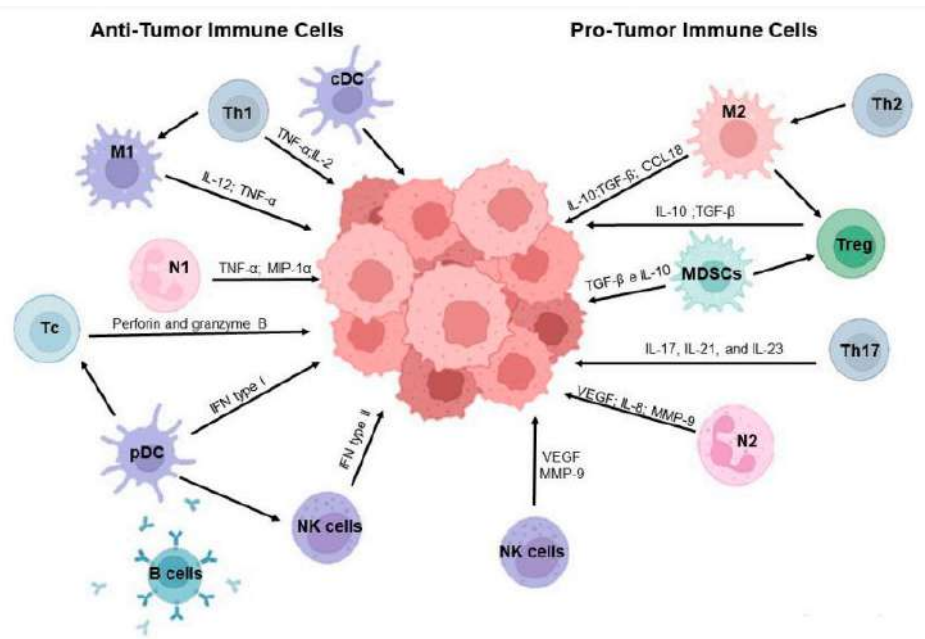


Figure 22. Interaction entre les différentes cellules qui composent le système immunitaire et le microenvironnement tumoral (Moura et al., 2025).

cDC - cellules dendritiques conventionnelles ; *IL* - interleukine ; *M1-M1* et *M2-M2* macrophages ; *MDSC* - cellules suppressives dérivées des myéloïdes ; *MIP-1-alpha* - protéine inflammatoire-1- α macrophagiques ; *MMP-9* - métalloprotéinases-9 de la matrice ; *N1 - N1* et *N2 - N2* – neutrophiles ; *NK* – Natural killer ; *pDC*-cellules dendritiques plasmacytoïdes ; *Tc*- cellules T cytotoxiques ; *Th1-H* helper cells 1 ; *TNF- α* - tumor necrosis factor alpha ; *Treg*-regulatory T cells ; *VEGF*-vascular endothelial growth factor.

L'influence des cellules immunitaires sur le développement et la survie des tumeurs mammaires varie selon le sous-type tumoral et la composition des populations cellulaires inflammatoires au sein du microenvironnement tumoral (MET). Le carcinome mammaire triple négatif (TNBC) se distingue par son fort caractère immunogène, contrairement aux autres sous-types. Les cellules tumorales interagissent avec diverses cellules du système immunitaire, modulant ainsi la prolifération et la progression tumorale ainsi que l'efficacité des traitements (Denkert et al., 2018).

Au sein du MET, les cellules immunitaires exercent une double fonction, jouant aussi bien un rôle anti-tumoral que pro-tumoral, en favorisant l'immunosuppression et la croissance cancéreuse (Figure 22). Parmi ces populations, on retrouve en abondance : Macrophages, Neutrophiles, Cellules NK (Natural Killer), Cellules dendritiques, Cellules suppressives dérivées de la lignée myéloïde (MDSC), Cellules lymphoïdes innées (ILC) et Lymphocytes. Chacune de ces populations contribue de manière distincte à la dynamique tumorale et à la réponse aux thérapies (Li et al., 2022 ; Moura et al., 2025).

Les macrophages associés aux tumeurs (TAMs) constituent une composante majeure du microenvironnement tumoral (TME) dans le cancer du sein, où ils exercent une influence déterminante sur la progression tumorale. Représentant une fraction importante du TME, ces cellules jouent des rôles multiples dans la carcinogenèse, l'immunosuppression et la résistance aux chimiothérapies, impactant ainsi le potentiel métastatique de la tumeur ainsi que le pronostic et l'évolution clinique des patientes. Leur recrutement au sein du TME est médié par des signaux chimiokines, après quoi ils sécrètent divers facteurs pro-tumoraux comme le VEGF (vascular endothelial growth factor), favorisant ainsi l'angiogenèse et la survie des cellules cancéreuses (Moura et al., 2025).

6.2. PDCD1 (PD-1) et CD274 (PD-L1)

Parmi les facteurs de la progression tumorale, les protéines PD-1 et PD-L1, codées respectivement par les gènes *PDCD1* et *CD274*, jouent un rôle crucial dans la régulation de la réponse immunitaire. Leur surexpression par les lymphocytes T peut favoriser l'échappement tumoral limitant ainsi l'activation excessive de ces lymphocytes. La protéine PD-L1, qui est exprimée par les cellules tumorales peut induire l'épuisement des lymphocytes T. Ainsi, les cellules tumorales peuvent proliférer sans être attaquées par le système immunitaire. Ces deux protéines peuvent faciliter l'échappement tumoral et représentent des cibles potentielles pour l'immunothérapie anti-cancéreuse. L'inhibition de l'interaction entre PD-1 et PD-L1 représente une approche prometteuse en immunothérapie. Des anticorps monoclonaux anti-PD-1 et anti-PD-L1 ont été développés pour bloquer cette interaction, réactivant ainsi la réponse immunitaire contre les cellules tumorales (Figure 23).

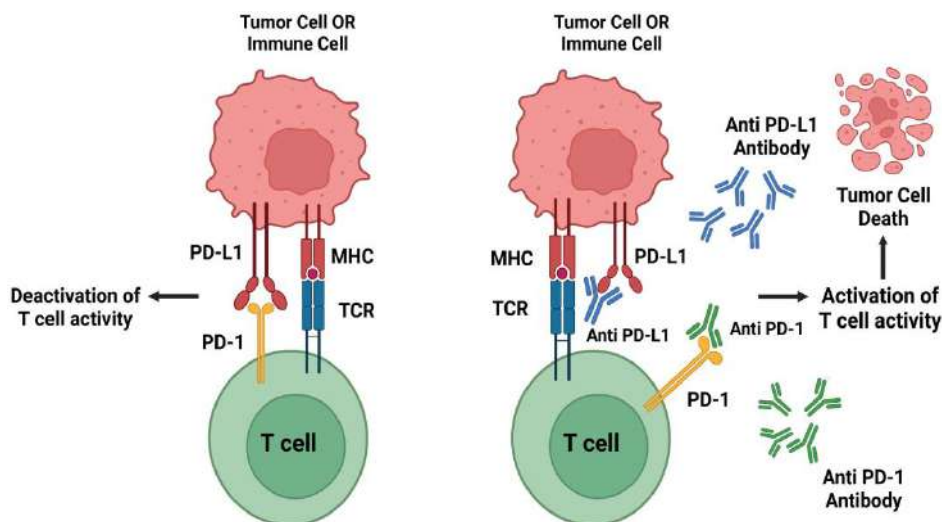


Figure 23. Augmentation de la prolifération des lymphocytes T et de la production de cytokines lors du blocage de PD-1 et PD-L1 (Parvez et al., 2023).

6.3. TLR4 et TLR9

Les récepteurs TLR4 et TLR9 appartiennent à la famille des *Toll-like receptors*, impliqués dans la détection des motifs moléculaires associés aux pathogènes. Leur activation joue un rôle clé dans l'initiation de la réponse immunitaire innée, tout en modulant la réponse immunitaire adaptative contre les cellules cancéreuses (**Figure 24**). Lorsqu'ils sont activés, ces récepteurs stimulent à la fois la production de cytokines pro-inflammatoires et la présentation antigénique, contribuant ainsi à renforcer la lutte contre les tumeurs (**Cook et al., 2004 ; Delneste et al., 2007**). Le récepteur TLR4 est principalement reconnu pour sa capacité à détecter le lipopolysaccharide (LPS) des bactéries Gram-négatives, tandis que TLR9 perçoit l'ADN bactérien ou viral riche en motifs CpG. Une fois activés, ces récepteurs recrutent des molécules adaptatrices telles que MyD88 et TRIF, déclenchant des cascades de signalisation intracellulaire qui aboutissent à la production de cytokines pro-inflammatoires et à la stimulation de la présentation antigénique.

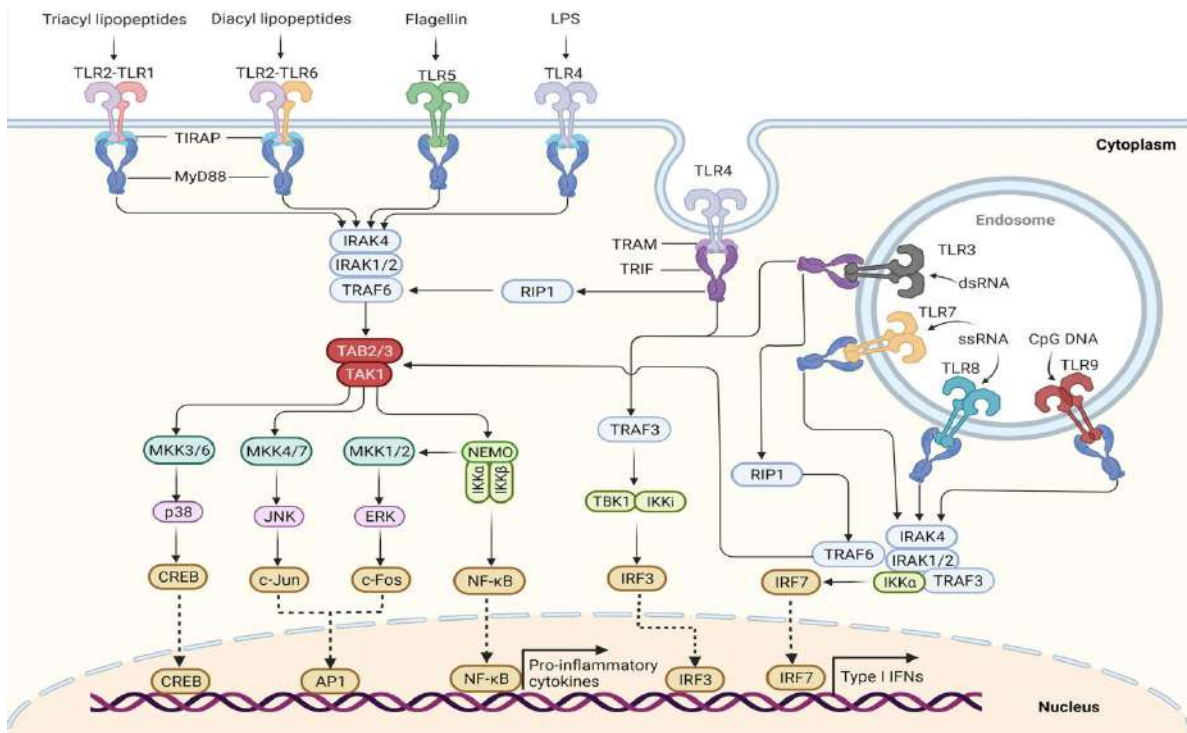


Figure 24. Représentation schématique des voies de signalisation des récepteurs Toll-like (TLR) (Stierschneider et Wiesner, 2023).

Les **PAMPs** interagissent avec les domaines extracellulaires des **TLR**, initiant une signalisation via la dimérisation des récepteurs et le recrutement de protéines adaptatrices (**MyD88**, **TIRAP**, **TRIF**, **TRAM**). Cette activation déclenche des cascades moléculaires impliquant des kinases (**IRAK**, **TAK-1**, **IKK**) et des facteurs de transcription (**NF-κB**, **AP-1**, **IRF**) régulant les réponses immunitaires et inflammatoires.

L'activation des TLRs permet non seulement de déclencher une réponse immunitaire innée rapide, mais elle ouvre également la voie à une réponse adaptative plus ciblée. Cela se traduit par :

- **Production de cytokines** : les cytokines produites modulent l'activité des cellules immunitaires, favorisant l'inflammation et l'activation des lymphocytes T et B.
- **Présentation antigénique** : les cellules dendritiques, en particulier, s'appuient sur les signaux des TLR pour améliorer leur capacité à présenter des antigènes aux lymphocytes T, renforçant ainsi la réponse adaptative (Yamamoto et al., 2004 ; Allison et al., 2015).

Dans le contexte du cancer, l'activation des récepteurs TLR4 et TLR9 aide à surmonter l'immunosuppression tumorale en stimulant les réponses immunitaires innées et adaptatives (Chuang et al., 2020).

6.4. CTLA4

C'est une protéine inhibitrice, exprimée par les lymphocytes T, qui régule la réponse immunitaire en se liant aux molécules B7-1/B7-2 (CD80/CD86) sur les cellules présentatrices d'antigènes. Cette interaction inhibe l'activation des cellules T, limitant leur capacité à attaquer les cellules tumorales et favorisant ainsi la tolérance immunitaire. Lorsque des inhibiteurs de points de contrôle immunitaire, comme les anticorps anti-CTLA-4, bloquent cette interaction, les cellules T sont activées et peuvent mieux reconnaître et détruire les cellules cancéreuses (Figure 25), renforçant ainsi la réponse antitumorale (National Cancer Institute, n.d.).

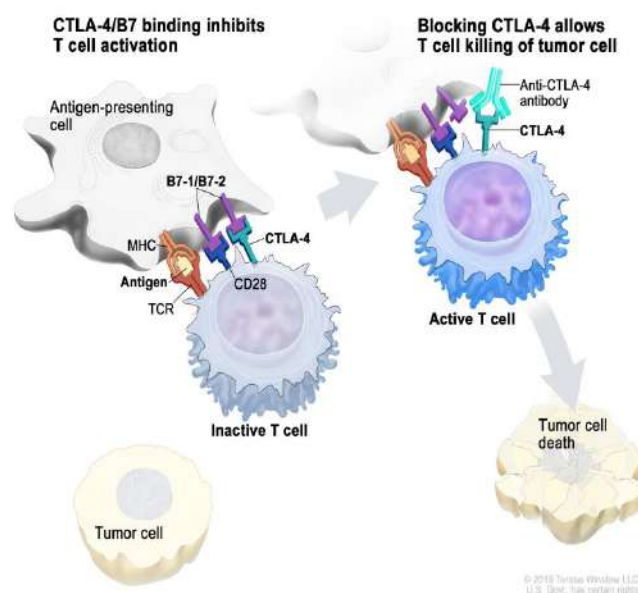


Figure 25. Inhibiteur de point de contrôle immunitaire ([Definition of CTLA-4 - NCI Dictionary of Cancer Terms - NCI](#)).

Les points de contrôle immunitaires, tels que B7-1/B7-2 et CTLA-4, jouent un rôle clé dans la régulation de la réponse immunitaire. Lorsque B7-1/B7-2 se lie à CTLA-4 sur les lymphocytes T, cela limite leur capacité à attaquer les cellules tumorales. En bloquant cette interaction grâce à des inhibiteurs spécifiques (anti-B7-1/B7-2 ou anti-CTLA-4), la réponse immunitaire est réactivée, permettant aux lymphocytes T de cibler et d'éliminer les cellules cancéreuses.

6.5. NFκB1

NFκB1 est un facteur de transcription qui régule l'expression de gènes impliqués dans des processus clés tels que la prolifération cellulaire, la survie, l'invasion et l'angiogénèse. Son activation est souvent liée à des conditions inflammatoires et peut contribuer à la progression tumorale. Dans le contexte du cancer, une activation aberrante de NFκB1 est fréquente et favorise un microenvironnement tumoral propice à la croissance tumorale et à la résistance aux traitements (**Figure 26**).

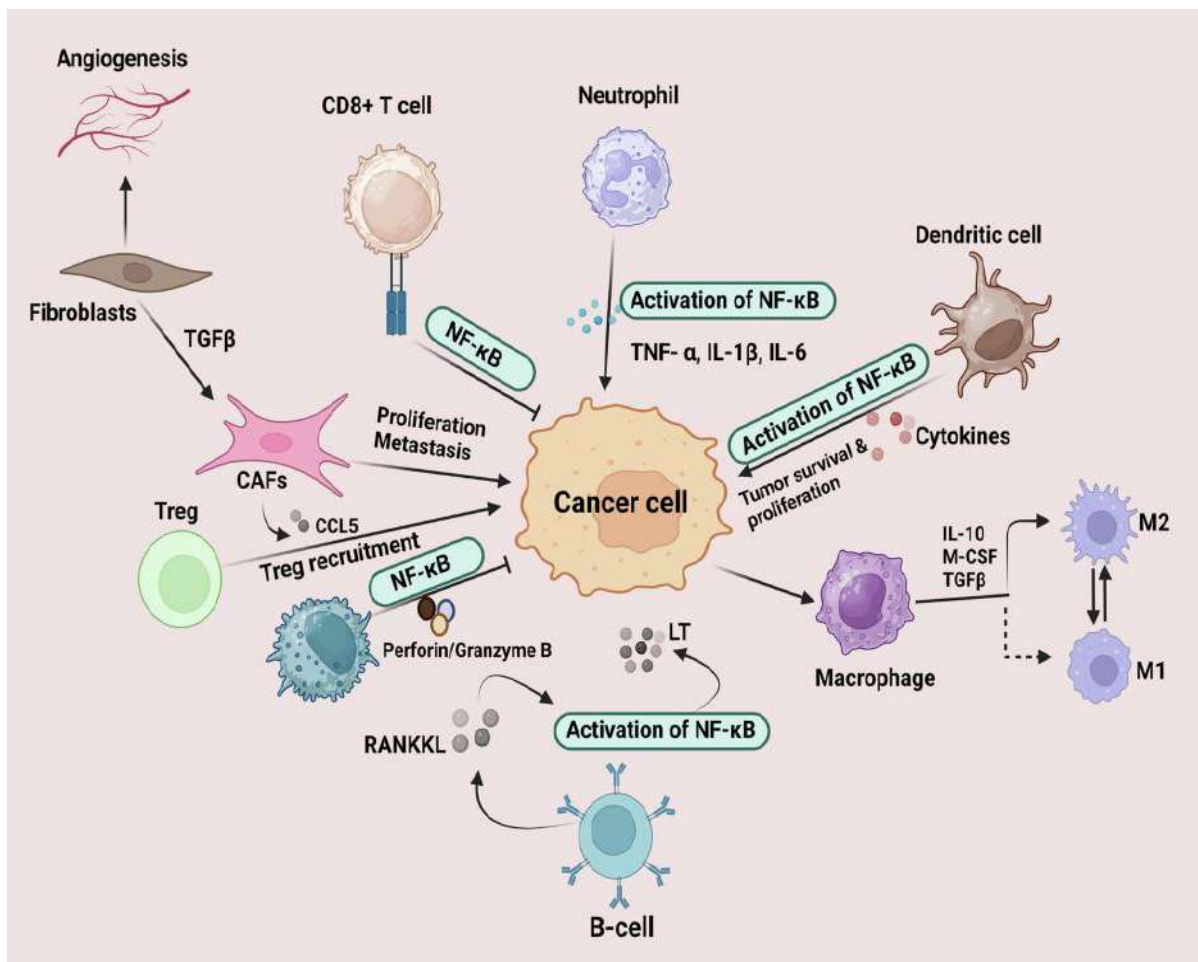


Figure 26. La voie NFκB régule le microenvironnement tumoral (Khan, 2024).

Par exemple, NFκB1 peut induire l'expression de gènes anti-apoptotiques, protégeant ainsi les cellules tumorales contre les thérapies cytotoxiques. De plus, NFκB1 interagit avec d'autres voies de signalisation immunitaire ; il peut moduler l'expression de protéines inhibitrices telles que PD-1 et CTLA-4, influençant directement l'activation des lymphocytes T. En régulant ces points de contrôle immunitaires, NFκB1 joue un rôle essentiel dans l'équilibre entre activation et répression des réponses immunitaires. Une compréhension approfondie de son rôle pourrait conduire à des approches plus personnalisées et efficaces en oncologie (**Brown et al., 2008 ; Concetti et Wilson, 2018 ; Rotte, 2019 ; Ebrahimi et al., 2024**).

6.6. Interactions entre facteurs immunitaires et hormonaux

Dans le sein en cycle normal chez les femmes préménopausées, les signaux médiés par les œstrogènes et la progestérone dans les cellules épithéliales et stromales induisent la production de facteur de stimulation des colonies 1 (CSF-1), de TGFβ et d'IL-10, recrutant ainsi diverses populations de cellules immunitaires. Pendant la phase lutéale du cycle menstruel et au cours de la grossesse, ce processus favorise la morphogenèse mammaire, mais au détriment de la surveillance immunitaire. Par exemple, les macrophages privilégient alors leurs fonctions de remodelage tissulaire et réduisent l'expression des marqueurs de présentation antigénique ainsi que la production de cytokines inflammatoires (**Need et al., 2014**). Ce microenvironnement accru en tolérance immunitaire pourrait contribuer au développement du cancer du sein. En effet, des études suggèrent que les œstrogènes favorisent un environnement immunosuppresseur dans le MET en inhibant l'activité antitumorale des cellules NK et des lymphocytes T cytotoxiques (CTL) *via* la régulation positive de l'inhibiteur de protéase-9 (PI9) dans les cellules tumorales (**Somasundaram et al., 2020**). Par ailleurs, l'exposition *in utero* au xénoestrogène bisphénol A (BPA) perturbe la production de cytokines immunorégulatrices, contribuant potentiellement à ce microenvironnement immunosuppresseur (**Fischer et al., 2016**). De même, l'utilisation de contraceptifs oraux pourrait influencer les populations de cellules immunitaires (**Vincent et Salamonsen, 2000 ; Isfoss et al., 2018**).

Ainsi, le statut des récepteurs hormonaux dans les tumeurs mammaires influence fortement le profil d'infiltration immunitaire (**Figure 27**).

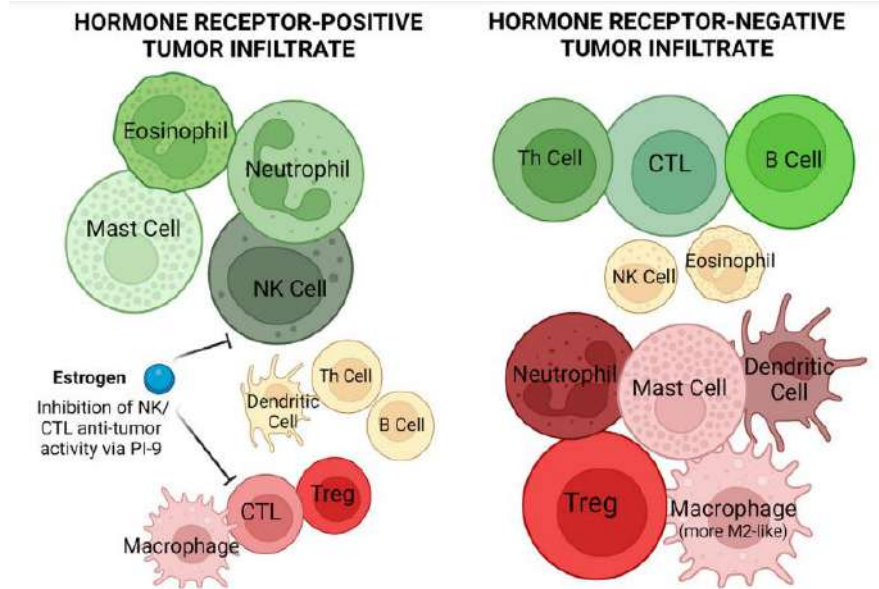


Figure 27. Composition de l'infiltrat immunitaire dans le cancer du sein en fonction du statut des récepteurs hormonaux (Boyle, 2023).

L'infiltrat immunitaire des tumeurs mammaires négatives aux récepteurs hormonaux diffère significativement de celui des tumeurs positives, ces dernières présentant une abondance accrue de mastocytes, de cellules NK, de neutrophiles et d'éosinophiles, tandis que les tumeurs négatives montrent une prédominance de macrophages de type M2, de cellules dendritiques, de mastocytes, de lymphocytes $T\gamma\delta$, de lymphocytes Tregs et de neutrophiles, avec une infiltration marquée CTL, de lymphocytes Th et de lymphocytes B associée à un meilleur pronostic. La taille des cellules représente leur abondance (élevée ou faible) et leur couleur indique leur valeur pronostique : vert pour les populations associées à un pronostic favorable, rouge pour celles liées à un pronostic défavorable, et jaune pour les populations moins abondantes dont le rôle pronostique n'est pas clairement établi.

Les tumeurs positives aux récepteurs hormonaux, particulièrement les tumeurs ER+, présentent une proportion plus élevée de mastocytes, de cellules NK antitumorales et de neutrophiles comparativement aux tumeurs négatives pour ces récepteurs, tandis que les taux d'éosinophiles et de monocytes constituent un indicateur de bonne réponse à la chimiothérapie (Segovia-Mendoza et Morales-Montor, 2019 ; Pyla et al., 2020). Une analyse comparative des différentes pathologies des cancers du sein ER+ révèle que les carcinomes lobulaires invasifs se caractérisent par une moindre abondance de cellules Th folliculaires, de lymphocytes $\gamma\delta$ T et de macrophages, alors que les carcinomes canaux invasifs présentent des fréquences accrues de lymphocytes B, de monocytes et de lymphocytes T cytotoxiques (CTLs) (Desmedt et al., 2018). Dans les tumeurs HER2+, les populations immunitaires dominantes comprennent les cellules dendritiques, mastocytes, lymphocytes $\gamma\delta$ T, lymphocytes T régulateurs (Tregs) et neutrophiles, toutes associées à un pronostic défavorable et une survie réduite (Segovia-Mendoza et

Morales-Montor, 2019). Les tumeurs ER- et leurs métastases présentent quant à elles des proportions réduites de populations immunitaires antitumorales comparativement aux tumeurs ER+, avec une prédominance de populations immunosuppressives comme les Tregs et les macrophages associés aux tumeurs (TAMs), corrélées à un moins bon pronostic (**Segovia-Mendoza et Morales-Montor, 2019 ; Williams et al., 2021**). Par ailleurs, les cellules HER2+ ou triple-négatives (TNBC) montrent une sensibilité significativement moindre à la cytotoxicité des cellules NK que les cellules de cancer du sein positives aux récepteurs hormonaux (**Goldberg et al., 2021**).

La compréhension des interactions entre les facteurs immunitaires et hormonaux influençant le développement et la progression du cancer du sein, peut conduire à de nouvelles stratégies de traitement, car les hormones peuvent moduler le système immunitaire et, inversement, certains facteurs immunitaires peuvent réguler la production et l'activité des hormones. Ces relations complexes ouvrent des perspectives prometteuses pour la recherche et le développement de thérapies ((**Shi et al., 2020 ; Moaz et al., 2023**), identifiant ainsi des cibles potentielles pour des interventions thérapeutiques combinées (**Rolfo et al., 2023**). En associant des approches de techniques de biologie moléculaire, d'immunohistochimie, de séquençage génétique et de spectrométrie de masse, les chercheurs ont pu identifier des signatures moléculaires spécifiques associées à différents types de cancer, ce qui a permis de mieux comprendre les mécanismes sous-jacents et d'ouvrir de nouvelles pistes thérapeutiques (**Gulati et Singh, 2024**).

7. EPIGENETIQUE ET CANCER

7.1. Définition de l'épigénétique

Par définition, l'épigénétique regroupe « *l'ensemble des modifications de l'expression d'un gène, héréditaires lors de la division cellulaire (mitose ou méiose), mais ne résultant pas d'altérations de la séquence d'ADN* » (**Van Seuning, 2009**).

Ainsi, l'épigénétique est un phénomène dans lequel la fonction des gènes produit des variations héréditaires sans modifier la séquence de l'ADN, ce qui entraîne en fin de compte des changements dans le phénotype d'un organisme (**Lee et Kim, 2022**).

Les processus épigénétiques englobent principalement trois catégories : la méthylation de l'ADN, la modification covalente des histones (méthylation, acétylation, phosphorylation...), et la régulation des ARN non codants (ARNnc) (**Fitz-James et Cavalli, 2022**). Le contrôle de la machinerie épigénétique permet ainsi de moduler l'accessibilité de l'ADN aux divers processus

nucléaires telle que la transcription de gènes spécifiques et de maintenir un état d'activité (actif ou silencieux) d'une génération cellulaire à l'autre (Wachsman, 1997 ; Torrisani, 2003).

La méthylation de l'ADN est, de façon générale, associée à l'hétérochromatine et à l'inhibition de l'expression génique en situations normales et pathologiques. Elle joue un rôle important notamment dans l'établissement de l'empreinte parentale, l'inactivation du chromosome X chez les mammifères femelles et le maintien de la stabilité chromosomique (Van Seuning, 2009).

Les altérations des modifications épigénétiques dans le cancer régulent diverses réponses cellulaires, y compris la prolifération cellulaire, l'apoptose, l'invasion et la sénescence (Figure 28) (Cheng et al., 2019).

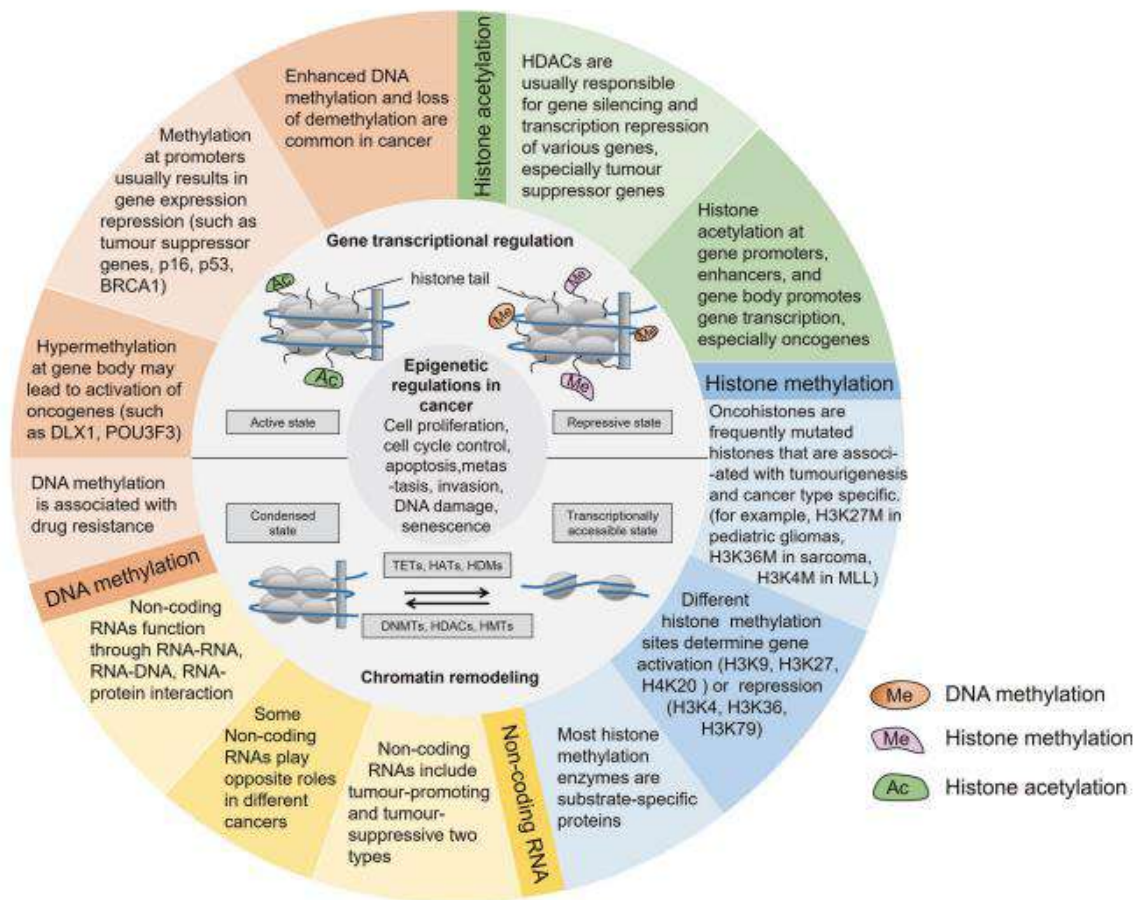


Figure 28. Régulations épigénétiques dans le cancer (Cheng et al., 2019).

7.2. Méthylation de l'ADN

La méthylation de l'ADN est un processus biochimique réversible qui implique l'ajout d'un groupe méthyle (-CH₃) à une base d'ADN à la position 5' du cycle pyrimidique de la cytosine de l'îlot CpG (Figure 29A) et maintenu au cours des divisions cellulaires. Cette modification n'affecte que les cytosines suivies d'une guanine dans la séquence d'ADN (les sites ou

dinucléotides CpG) (Galbraith et Snuderl, 2022). Deux types de réactions de méthylation de l'ADN sont décrites : les réactions de méthylation *de novo* et les réactions de méthylation de maintenance (Figure 29B). Les réactions de méthylation *de novo* de l'ADN équivalent à l'ajout de groupements méthyles sur les cytosines des dinucléotides CpG localisés sur les 2 brins d'ADN. Ce type de réaction est principalement à l'origine de l'hyperméthylation des gènes rencontrés dans le cancer (Cartrona et al., 2015).

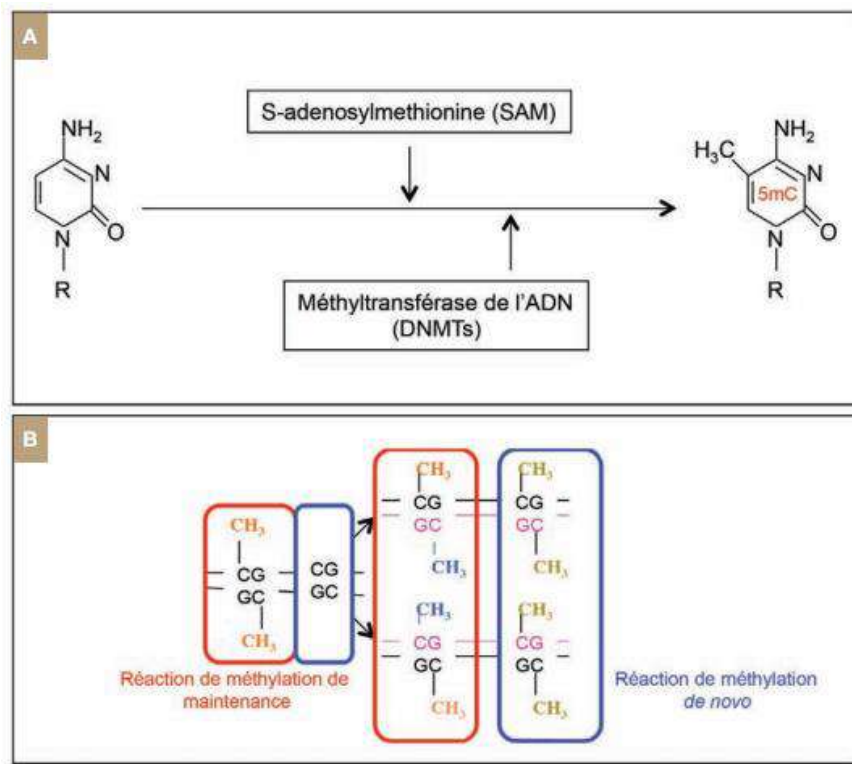


Figure 29. Méthylation et ADN méthyltransférases (DNMTs) (Cartrona et al., 2015).

A. Schéma de la modification chimique d'une base cytosine par ajout d'un groupement méthyle (CH₃). Il y a un groupement donneur de méthyle (SAM) et une enzyme qui ajoute ce groupement méthyle à la Cytosine (DNMT). B. Représentation schématique de la méthylation de maintenance et de la méthylation de novo médiées par les DNMTs sur les dinucléotides CpG.

Ce processus est catalysé par des enzymes appelées ADN méthyltransférases (DNMT), qui utilisent la S-adenosylméthionine comme donneur de méthyle (Deng et al., 2023). Dans les cellules de mammifères, les DNMT peuvent être classées en trois types principaux : DNMT1, DNMT3a et DNMT3b. La DNMT1 est responsable du maintien de l'état de méthylation de l'ADN existant, ainsi que de la catalyse de l'ADN hémiméthylé et de participer à la réparation de la méthylation de l'ADN (Yan et al., 2023).

D'autre part, les DNMT3a et DNMT3b catalysent l'ajout de nouveaux groupes méthyles à l'ADN non méthylé, le convertissant ainsi en ADN méthylé (Lyko, 2018). La méthylation de l'ADN joue un rôle crucial dans la régulation de divers processus biologiques au sein du génome des mammifères. Ces processus comprennent la transcription et le traitement post-transcriptionnel, la modification post-traductionnelle, le remodelage de la chromatine, l'empreinte génomique, l'inactivation du chromosome X et l'inhibition des éléments répétitifs de l'ADN (Saghafinia et al., 2018).

Au niveau des cellules normales, la plupart des sites CpG sont méthylés. Au contraire, les îlots CpG, associés à la région 5' et au premier exon de 50 à 60% des gènes de mammifères ne sont pas méthylés. Le génome d'une cellule tumorale subit une perte considérable de méthylation et présente 20 à 60% moins de cytosines méthylées qu'une cellule normale (Costello et Plass, 2001 ; Esteller, 2007).

7.3. Méthylation de l'ADN et cancers

Les mutations de certains gènes, y compris les gènes suppresseurs de tumeurs et les proto-oncogènes, contribuent au développement du cancer (Riaud et al., 2024 ; Tuval et al., 2024). Cependant, des recherches récentes ont révélé que les anomalies épigénétiques (méthylation de l'ADN) qui constituent une voie alternative aux mutations, peuvent également conduire à l'inactivation de gènes suppresseurs de tumeurs et l'activation de proto-oncogènes, jouant ainsi un rôle crucial dans l'initiation, la progression, l'invasion et la propagation du cancer (Figure 30) (Feinberg et Levchenko, 2023). Plus de 65% des cancers humains sont liés à la méthylation de l'ADN (Jones, 1996 ; Herman et Baylin, 2003).

Les progrès des techniques de séquençage à haut débit ont permis de mieux comprendre les mécanismes épigénétiques dans le développement du cancer et de révéler de nombreux biomarqueurs épigénétiques spécifiques au cancer qui offrent un potentiel d'indicateurs effectifs pour évaluer le risque tumoral élevé, faciliter le diagnostic précoce, orienter les décisions thérapeutiques et prédire le pronostic (Lauschke et al., 2019).

Au cours de la tumorigenèse, on observe généralement une diminution du niveau global de méthylation de l'ADN intracellulaire. Cependant, il y a une élévation anormale de la méthylation de l'ADN dans les régions des îlots CpG de certains gènes spécifiques (Ghosh et al., 2022). Ce schéma de méthylation anormal entraîne à une instabilité à l'échelle du génome et à une altération des profils d'expression génique (Liang et Weisenberger, 2017).

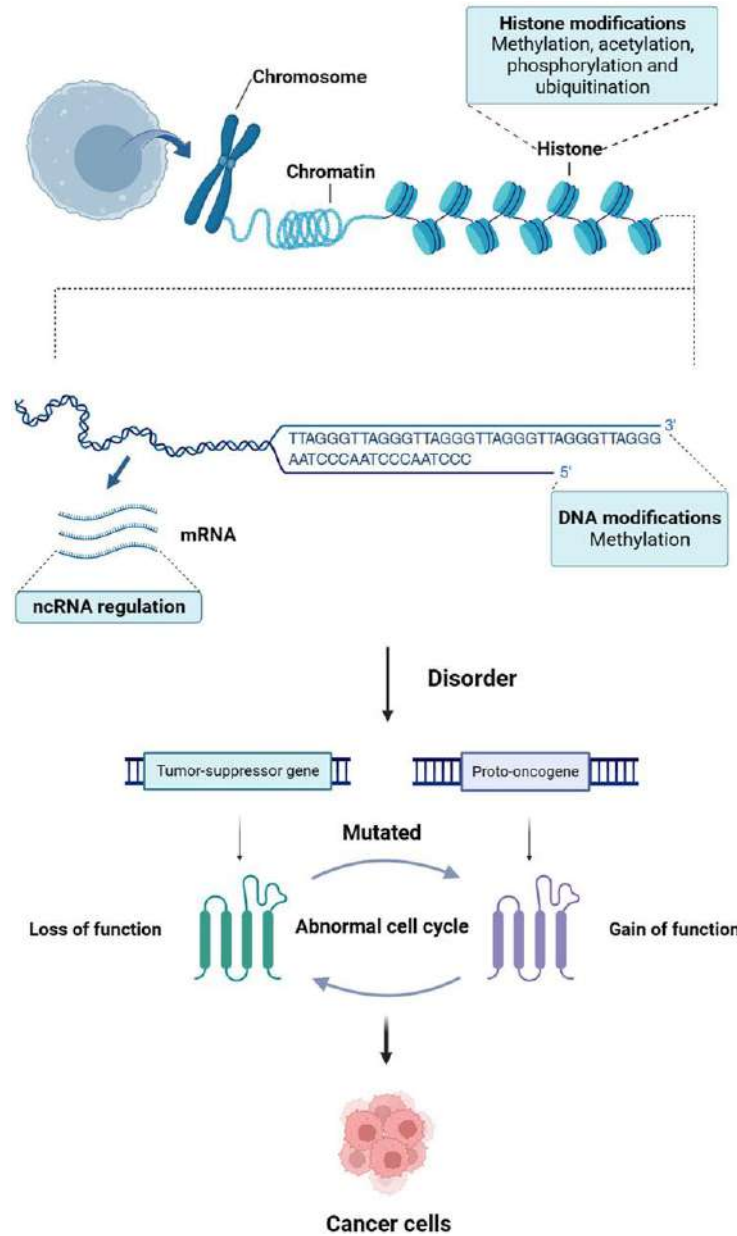


Figure 30. Lien entre l'épigénétique et le cancer (Gao et al., 2024).

Les troubles épigénétiques peuvent entraîner l'inactivation de gènes suppresseurs de tumeurs et l'activation de proto-oncogènes, qui jouent un rôle crucial dans l'initiation, la progression, l'invasion et la propagation du cancer.

Il est intéressant de noter que les proto-oncogènes tendent à avoir des niveaux de méthylation de l'ADN réduits au cours de la tumorigenèse. Cette réduction de la méthylation favorise l'expression élevée de ces proto-oncogènes, contribuant potentiellement au développement de la tumeur (Casalino et Verde, 2020). Inversement, les gènes suppresseurs de tumeurs présentent souvent des niveaux accrus de méthylation de l'ADN, ce qui inhibe leur expression et favorise la

croissance de la tumeur. Des études récentes ont également montré que des schémas de méthylation spécifiques peuvent être détectés dans les cellules tumorales circulantes, le sang, l'urine et d'autres liquides biologiques (**Palanca-Ballester et al., 2021**). Ces profils de méthylation, connus comme des gènes de méthylation spécifiques à la tumeur, sont devenus des outils précieux pour le diagnostic précoce et le pronostic des tumeurs (**Oliver et al., 2022**). Par exemple, de nombreuses études ont examiné la méthylation de l'ADN comme un biomarqueur diagnostique potentiel dans le cancer de l'ovaire à un stade précoce (**Singh et al., 2019**).

Les deux formes d'anomalies de la méthylation de l'ADN les plus connues sont l'hyperméthylation des îlots CpG au niveau de certains promoteurs de gènes suppresseurs de tumeurs et l'hypométhylation globale du génome résultant en une instabilité chromosomique ou en l'expression d'oncogènes. (**Jones, 1996 ; Egger et al., 2004; Digel et Lubbert, 2005**).

7.3.1. Hypométhylation globale

L'hypométhylation est un processus qui affecte principalement les régions intergéniques (**Wilson et al., 2007**). Il s'agit d'une diminution globale de la méthylation (baisse de 20 à 60 % de la teneur en 5-méthylcytosines) dans le génome entier (hypométhylation globale) ou une déméthylation relative localisée dans différentes régions du génome, telles que les régions promotrices des proto-oncogènes ou les séquences répétées normalement hautement méthylées (**Feinberg et Vogelstein, 1983**). De manière paradoxale, cette hypométhylation générale est fréquemment associée à une hyperméthylation localisée et spécifique de certains promoteurs de gènes, essentiellement au niveau des îlots CpG. Elle entraîne la répression d'une grande variété de gènes aux fonctions anti-prolifératives appartenant à une nouvelle génération de gènes suppresseurs de tumeurs (**Esteller, 2006**).

L'hypométhylation globale semble contribuer à la tumorigenèse selon trois mécanismes : la réactivation des transposons, l'instabilité des chromosomes et la perte d'empreinte parentale (**Gaudet et al., 2003**). En effet, certaines études ont démontré que dans les tumeurs, le degré d'hypométhylation des éléments répétitifs est corrélé avec l'instabilité génomique et la progression tumorale (**Narayan et al., 1998 ; Qu, 1999**).

Dans les cancers, l'hypométhylation peut aussi avoir lieu dans une région promotrice CpG d'un proto-oncogène, par exemple, entraînant sa surexpression qui induit l'augmentation de la prolifération cellulaire (**Ghazi et al., 1992**).

7.2.2. Hyperméthylation des gènes suppresseurs de tumeurs

L'anomalie épigénétique la plus souvent décrite dans les régions promotrices est l'hyperméthylation. Elle est retrouvée dans la majorité des néoplasmes et est associée à l'inhibition de la transcription de plusieurs gènes suppresseurs de tumeurs (**Herman et Baylin 2003**). L'hyperméthylation des îlots CpG est à l'origine de la suppression transcriptionnelle de gènes suppresseurs de tumeurs jouant un rôle clé dans diverses fonctions indispensables pour la cellule, telles que la régulation du cycle cellulaire, la réparation de l'ADN, l'adhésion cellulaire, l'apoptose et plusieurs autres fonctions (**Esteller et al., 2000**). Selon le type de cancer, les gènes méthylés peuvent être des marqueurs de détection dans différents types de tumeurs (**Torrisani et al., 2003**). D'autre part, une élévation du taux des DNMTs a été observée dans différents types de tumeurs. Cette augmentation pourrait être responsable de l'hyperméthylation de promoteurs souvent observée dans certaines tumeurs (**Robertson et al., 1999 ; Girault et al., 2003**).

8. INTELLIGENCE ARTIFICIELLE ET IMPORTANCE DANS LES PATHOLOGIES

8.1. Systèmes de Support à la Décision Clinique (SSDC)

Les systèmes de support à la décision clinique (SSDC) sont des outils informatiques conçus pour assister les professionnels de la santé dans la prise de décisions cliniques en fournissant une aide informatique pour gérer les informations des patients. Ces systèmes sont essentiels dans les environnements de santé modernes pour améliorer la qualité des soins, réduire les erreurs médicales et accroître l'efficacité. Les SSDC utilisent des règles de décision et des algorithmes sophistiqués pour générer des recommandations précises, telles que des alertes sur les interactions médicamenteuses, des suggestions de tests diagnostiques et des rappels pour des mesures préventives. Ils intègrent ces recommandations dans le flux de travail clinique, facilitant ainsi des décisions informées et opportunes. Certains SSDC offrent également un soutien diagnostique et thérapeutique en analysant les symptômes et les caractéristiques des patients pour proposer des diagnostics potentiels et des options de traitement adaptées. De plus, les fonctionnalités des SSDC incluent la surveillance des résultats des patients dans le temps, permettant d'affiner les recommandations et d'améliorer les soins. Certains systèmes intègrent des composants visant à engager les patients en offrant des informations personnalisées et des rappels pour la gestion de la santé. Enfin, les SSDC génèrent des données précieuses pour la recherche et l'amélioration de la qualité des soins, permettant une évaluation continue des pratiques cliniques et de l'adhérence aux lignes directrices (**Chen et al., 2023**).

8.2. L'IA dans le domaine sanitaire, la médecine de précision et la gestion des ressources

Au fil des ans, la biotechnologie a considérablement évolué. Cette évolution s'accompagne de progrès majeurs en informatique : les ordinateurs sont devenus plus rapides et plus compacts, tandis que les ensembles de données voient leur hétérogénéité et leur volume augmenter de façon exponentielle. Ces avancées alimentent le développement de l'intelligence artificielle (IA), qui permet de découvrir des techniques innovantes pour résoudre des problèmes complexes dans presque tous les domaines, comme le diagnostic médical assisté par IA ou la modélisation moléculaire en biotechnologie (Ahmed et al., 2020).

Ainsi, l'IA permet de traiter les problèmes clés du système de santé (Figure 31) (les erreurs de diagnostic, les traitements excessifs, les approches standardisées, les tâches répétitives, la baisse de productivité, la sous-utilisation des données, les coûts et dépenses élevés) et identifier des biomarqueurs pertinents pour proposer des traitements économiques et personnalisés, grâce à une analyse intelligente de données hétérogènes (Makary et Daniel, 2016).

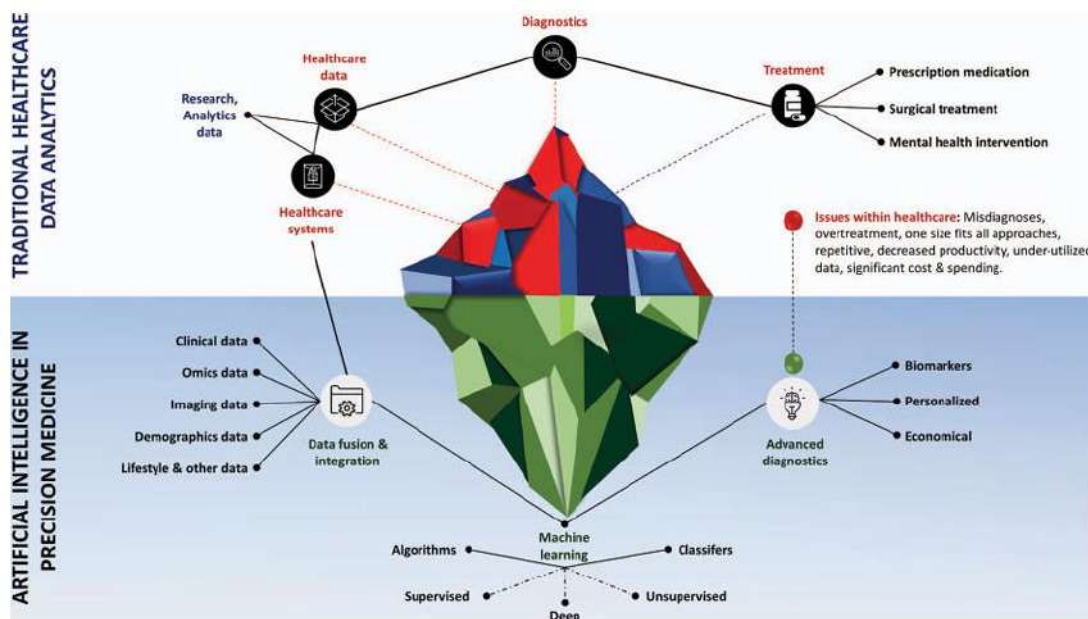


Figure 31. Rôle de l'intelligence artificielle dans l'analyse traditionnelle des données de santé et dans la médecine de précision (Ahmed et al., 2020).

L'IA est une branche de l'informatique visant à doter les machines de capacités d'imitation, voire de dépassement, de l'intelligence humaine (Shalev-Shwartz et Ben-David, 2014). L'IA est utilisée dans de nombreux domaines, dont la médecine, la communication, les transports, les finances, etc... (Jordan et Mitchell, 2015). Dans le domaine médical, l'IA est principalement

utilisée pour le diagnostic, le traitement et la prédiction du pronostic des maladies, en soutenant la prise de décision en temps réel et en proposant des solutions à des problèmes complexes via une analyse computationnelle intensive (Ahmed et al., 2020). L'IA médicale comporte deux branches principales : l'IA virtuelle et l'IA physique. La partie virtuelle de l'IA repose sur l'apprentissage profond (DL) et de l'apprentissage automatique (ML), qui offrent des moyens potentiels pour développer des approches robustes assistées par ordinateur. La partie physique de l'IA englobe, quant à elle, les robots et les appareils médicaux (Hamet et Tremblay, 2017). Des études récentes (Chen et al., 2021 ; Li et al., 2021 ; Yang et al., 2022) soulignent son efficacité, notamment en oncologie, où elle améliore le diagnostic et la prise en charge thérapeutique précoce, ainsi que le pronostic des patients grâce à des technologies avancées.

L'analyse utilisant IA et ML dans la médecine de précision (Figure 30) repose sur : (i) des approches, incluant Support Vector Machine, Deep Learning, Régression logistique, Analyse discriminante, Arbres de décision, Random Forest, Régression linéaire, Naïve Bayes, K-Nearest Neighbors, modèles de Markov cachés et algorithmes génétiques ; (ii) examen multifactoriel, couplé à une base de connaissances et un système d'aide à la décision pour la classification, le clustering et l'analyse de régression ; (iii) une gestion optimisée des ressources pour le stockage des données et l'analyse computationnelle.

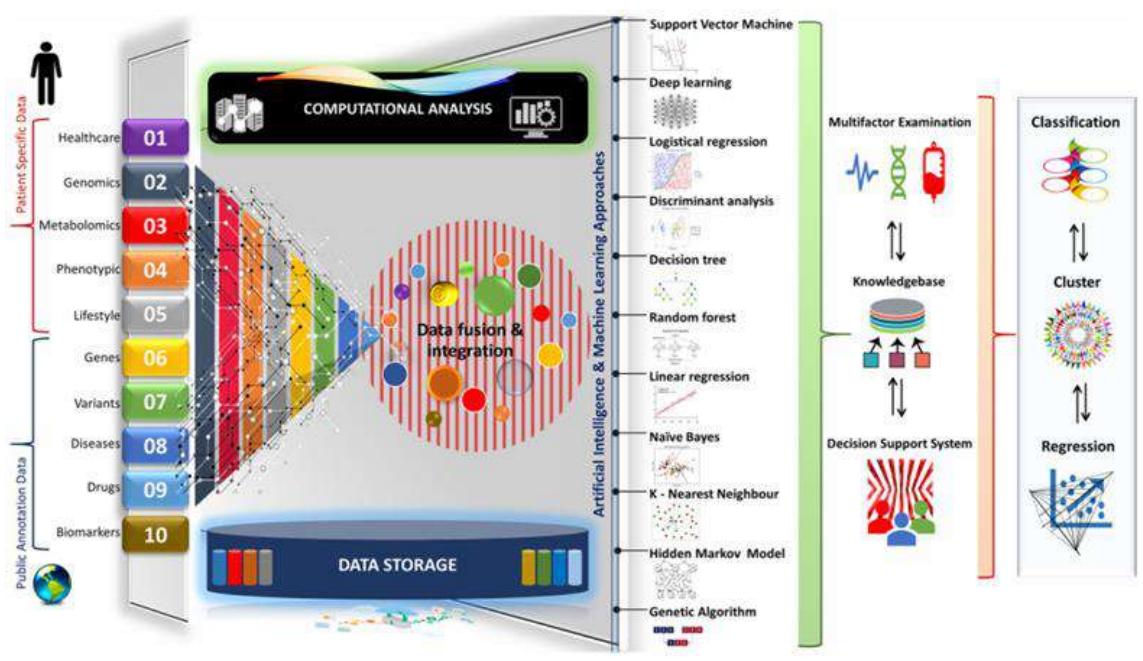


Figure 32. Modélisation et analyse de données hétérogènes intégrant des informations sur les soins de santé, la génomique, la métabolomique, le phénotype et le mode de vie propres au patient, ainsi que des données d'annotation accessibles au public, y compris les gènes, les variants, les maladies, les médicaments et les biomarqueurs.

La médecine de précision, telle que définie par Collins et Varmus (2015), représente une avancée majeure par rapport aux pratiques médicales traditionnelles, permettant des interventions thérapeutiques plus précoces grâce au développement de diagnostics prédictifs et de traitements personnalisés (Perkins et al., 2018). Néanmoins, comme le démontrent Schüssler-Fiorenza et al. (2019), son implémentation clinique nécessite une approche multidisciplinaire complexe ainsi que la mise en place d'un écosystème de santé innovant, capable d'intégrer et d'analyser l'ensemble des données des patients. Ceci inclut notamment les dossiers médicaux électroniques, les profils multi-omiques (génomiques, métabolomiques, protéomiques) et les facteurs liés au mode de vie (Hou et al., 2020).

8.3. Avantages de l'Intelligence Artificielle (IA)

L'intelligence artificielle (IA) enrichit les systèmes de soutien à la décision clinique (SSDC) en apportant des capacités d'analyse avancées, une personnalisation accrue et une automatisation améliorée. En intégrant des algorithmes d'IA dans les SSDC, les professionnels de la santé bénéficient d'outils plus puissants pour la prise de décisions cliniques, l'amélioration des soins aux patients et l'optimisation des flux de travail. Le choix de ces algorithmes est essentiel pour assurer le bon fonctionnement des SSDC. Les techniques d'apprentissage automatique (machine learning) et d'apprentissage profond (deep learning) permettent aux SSDC de traiter et d'analyser de grandes quantités de données, d'identifier des modèles complexes, de fournir des recommandations plus personnalisées et d'améliorer la précision des diagnostics. L'intégration de l'IA dans les SSDC optimise leur efficacité en rendant les systèmes plus adaptatifs et en offrant des insights plus approfondis, contribuant ainsi à une meilleure qualité des soins et à une gestion plus efficace des décisions cliniques (Elhaddad et Hamam, 2024).

9. OBJECTIFS DE LA THESE

9.1. Objectif principal

L'objectif principal de cette thèse est le développement d'une approche intégrative combinant l'analyse de données cliniques, moléculaires et multiomiques (immunologiques, hormonales et épigénétiques) pour améliorer le diagnostic, la stratification et le pronostic des patientes atteintes d'un cancer du sein, en exploitant des méthodes avancées d'intelligence artificielle (apprentissage supervisé et non supervisé) et en les comparant aux approches statistiques classiques. Cette démarche vise à optimiser la médecine personnalisée et les stratégies théranostiques, en associant recherche translationnelle et innovations en IA.

9.2. Objectifs spécifiques

Les objectifs spécifiques s'articulent autour de trois axes distincts :

- ❖ Un premier axe qui concerne une étude prospective menée sur une cohorte de patientes prises en charge pour un cancer du sein au Centre Anti-Cancer (CAC) d'Annaba, en Algérie. L'analyse repose sur l'exploitation systématique de données multidimensionnelles, incluant les paramètres cliniques et les profils immunohistochimiques, afin d'identifier les facteurs diagnostics et pronostics pertinents dans cette population spécifique.

L'application de l'intelligence artificielle (IA) offre des opportunités majeures pour optimiser l'analyse de ces données. Cette étude propose une méthodologie comparative de quatre modèles d'apprentissage supervisé (Régression Logistique, Random Forest, LightGBM et CatBoost) pour prédire le statut HER2 à partir de données préalablement cités.

- ❖ Un deuxième axe porte sur une étude rétrospective approfondie exploitant les données multiomiques et cliniques fondés sur des données publiques de la base de données *The Cancer Genome Atlas* (TCGA). Notre étude est basée sur 10 biomarqueurs clés soigneusement sélectionnés pour leur implication dans les mécanismes immunitaires et hormonaux. Un total de 6 biomarqueurs immunitaires (PD-1, PD-L1, TLR4, TLR9, NF- κ B1 et CTLA4) ont été choisis pour leur rôle central dans l'échappement immunitaire et la réponse aux immunothérapies, ainsi que 4 biomarqueurs hormonaux (HER2, PR, ER et l'Aromatase) ont été inclus pour leur importance dans les cancers hormono-

dépendants et les thérapies ciblées. Notre approche intégrative a combiné : (1) l'analyse épigénétique par profilage de la méthylation de l'ADN (2) l'évaluation de leur expression et (3) l'exploitation systématique des données cliniques associées.

Grâce à des méthodes avancées de clustering non supervisé, notre objectif était de stratifier la population en sous-groupes moléculaires pertinents pour une médecine personnalisée associant profil épigénétique et expression génique, et d'établir des corrélations significatives entre ces profils moléculaires et les paramètres cliniques. Ceci permettra une vision intégrée des interactions entre ces marqueurs et leur implication potentielle dans la progression tumorale et la réponse aux traitements, ouvrant ainsi des perspectives pour le développement de stratégies théranostiques personnalisées.

- ❖ Un troisième axe qui intègre une approche comparative innovante entre les méthodes statistiques traditionnelles et l'intelligence artificielle, afin d'élaborer un modèle prédictif optimal de la survie globale. Le modèle de Cox-Hazard, référence en analyse de survie, permet d'évaluer l'impact multivarié des facteurs pronostiques. Parallèlement, les architectures de deep learning (ANN et CNN) sont exploitées pour leur capacité à détecter, dans les données, des motifs complexes non perceptibles par l'humain potentiellement invisibles aux modèles classiques. Les ANN, inspirés des réseaux neuronaux biologiques, traitent les relations globales entre variables, tandis que les CNN excellent dans l'extraction de caractéristiques spatiales, particulièrement utiles pour l'analyse d'imagerie ou de données structurées.

L'IA peut transformer notre étude en permettant une analyse plus rapide, plus précise et personnalisée, avec un gain de temps significatif dans l'analyse des données. Cette double approche vise non seulement à améliorer la précision pronostique en oncologie, mais aussi à favoriser le développement d'un outil d'aide à la décision pour les oncologues, permettant une personnalisation thérapeutique optimisée. La comparaison des performances respectives du modèle de Cox-Hazard et des réseaux neuronaux ouvre des perspectives prometteuses pour le développement d'outils décisionnels en médecine de précision.



Matériel et Méthodes

Dans cette section nous présentons le matériel et les méthodologies utilisées dans deux études complémentaires : une étude prospective et une étude rétrospective ayant pour objectif d'identifier les déterminants influençant les résultats du cancer du sein en intégrant des données cliniques et biologiques.

La première partie constitue une cohorte prospective menée au Centre Anti-Cancer (CAC) en Algérie, tandis que la seconde est une étude rétrospective qui adopte une approche multiomique en exploitant les données de *The Cancer Genome Atlas* (TCGA). Enfin, la troisième partie sera consacrée à l'élaboration de modèles pronostic en fonction de la survie des malades.

Ensemble, ces études offriront une compréhension approfondie et multidimensionnelle des facteurs cliniques et biologiques impactant le pronostic et le traitement du cancer du sein, contribuant ainsi au développement de stratégies de médecine personnalisée.



PREMIERE PARTIE

ETUDE PROSPECTIVE : MODELISATION PREDICTIVE SUPERVISEE DU STATUT HER2

1. DESCRIPTION DE LA POPULATION D'ETUDE

Notre étude est de type prospective, visant à explorer l'utilisation des techniques d'apprentissage automatique pour prédire le statut HER2 dans le cancer du sein, en analysant divers paramètres cliniques et biologiques. Nous avons collecté des biopsies tumorales des patientes incluses dans l'étude et réalisé une analyse immuno-histochimique sur ces échantillons afin de mesurer l'expression de quatre biomarqueurs : les récepteurs hormonaux des œstrogènes (ER) et de la progestérone (PR), HER2, et l'indice de prolifération cellulaire Ki-67. Ces biomarqueurs sont essentiels pour la caractérisation moléculaire des sous-types du cancer du sein et pour orienter les stratégies thérapeutiques.

Les données cliniques et biologiques ainsi obtenues ont été combinées à des algorithmes d'apprentissage automatique pour établir des modèles prédictifs du statut HER2, afin d'améliorer la précision du diagnostic et de favoriser des approches thérapeutiques personnalisées.

Cette étude a recruté, entre 2018 et 2020, un total de 141 patientes atteintes du cancer du sein à l'EHS Abdellah Nouaouria, Hôpital El-Bouni-Annaba, et au Centre Anti-Cancer (CAC), Hôpital Universitaire d'Annaba, Algérie. Après application des critères d'inclusions et d'exclusions le nombre des patientes a été réduit à 109 patientes.

1.1. Les critères d'inclusion

- **Diagnostic Confirmé** : un diagnostic confirmé histologiquement de cancer du sein.
- **Données Médicales Complètes** : disponibilité de données médicales complètes incluant les résultats des tests diagnostiques, les traitements administrés, et les résultats du suivi.
- **Caractéristiques Cliniques et Pathologiques** : informations suffisantes sur les caractéristiques cliniques et pathologiques telles que le stade de la maladie, le type histologique, les statuts des récepteurs hormonaux (ER/PR) et HER2.

1.2. Les critères d'exclusion

- **Présence d'Autres Cancers** : présence d'autres types de cancer en plus du cancer du sein.
- **Données Incomplètes** : données médicales incomplètes ou dossiers manquants.
- **Antécédents Médicaux** : antécédents de maladies ou traitements pouvant potentiellement affecter les résultats de l'étude.

➤ **Comorbidités Significatives** : comorbidités significatives ou conditions de santé sous-jacentes.

Les informations démographiques et cliniques ont été extraites des dossiers médicaux à l'aide d'un formulaire standardisé (**Annexe 1**), incluant :

- Âge ;
- Taille de la tumeur ;
- Nombre de ganglions lymphatiques impliqués ;
- Présence de métastases ;
- Type histologique ;
- Stade de la maladie ;
- Statut des récepteurs hormonaux (ER/PR), Ki-67 et HER2 ;

* **Approbation éthique et consentement à participer**

La présente étude analyse de manière prospective les données extraites des dossiers médicaux. Elle adhère aux principes énoncés dans la Déclaration d'Helsinki et a reçu l'approbation du comité d'éthique local du CHU Ibn Roched (No. 1726). Un consentement éclairé a été obtenu de tous les participants avant leur inclusion dans l'étude.

2. COLLECTE DES DONNEES

La collecte des données de cette première étude prospective a été réalisée à l'aide d'un formulaire standardisé, extrayant les informations démographiques et cliniques pertinentes à partir des dossiers médicaux.

3. IMMUNO-HISTOCHIMIE DES BIOMARQUEURS (ER, PR, HER2 et Ki-67)

Le statut HER2 des patientes a été déterminé par immunohistochimie (IHC) et, lorsque nécessaire, confirmé par fluorescence *in situ hybridization* (FISH). HER2 est un récepteur tyrosine kinase dont la surexpression est associée à une agressivité accrue du cancer du sein. La classification HER2 a été réalisée conformément aux recommandations cliniques standardisées, ce qui est crucial pour la stratification thérapeutique et le pronostic des patientes.

3.1. Préparation des échantillons

Les échantillons tumoraux ont été obtenus par biopsie de patientes atteintes de cancer du sein. Après leur prélèvement, les échantillons ont été fixés dans du formol tamponné à 10 % pendant 24 heures, puis inclus dans de la paraffine. Des coupes de tissu de 3 à 4 μm d'épaisseur ont été réalisées à l'aide d'un microtome et montées sur des lames de verre.

3.2. Technique Immuno-Histochimique (IHC)

L'analyse immuno-histochimique a été réalisée pour détecter l'expression des biomarqueurs suivants : récepteur des œstrogènes (ER), récepteur de la progestérone (PR), HER2 (récepteur 2 du facteur de croissance épidermique humain), et l'indice de prolifération Ki-67. Les étapes de la procédure IHC incluent :

1. **Déparaffinage et réhydratation** : les lames contenant les coupes de tissu ont été déparaffinées à l'aide de bains successifs de xylène, puis réhydratées par des passages dans des concentrations décroissantes d'alcool (100 %, 95 %, 70 %) et enfin rincées à l'eau distillée.
2. **Récupération de l'antigène** : afin de restaurer les épitopes antigéniques masqués par la fixation, un traitement de récupération de l'antigène a été réalisé. Les lames ont été chauffées dans une solution tamponnée de citrate à pH 6.0 pendant 20 minutes à 95-100°C.
3. **Incubation avec les anticorps primaires** : les coupes ont été incubées avec des anticorps monoclonaux primaires spécifiques pour chaque biomarqueur :
 - Anti-ER (récepteur des œstrogènes)
 - Anti-PR (récepteur de la progestérone)
 - Anti-HER2
 - Anti-Ki-67

Les anticorps ont été dilués selon les recommandations du fabricant et incubés pendant 30 à 60 minutes à température ambiante.

4. **Révélation et contre-coloration** : la détection des complexes antigène-anticorps a été réalisée en utilisant un système de détection à base de streptavidine-biotin-peroxydase. Le substrat 3,3'-diaminobenzidine (DAB) a été utilisé pour produire un dépôt de couleur brune, permettant de visualiser l'expression des biomarqueurs. Les lames ont ensuite été contre-colorées à l'hématoxyline pour mieux identifier les noyaux cellulaires.

5. **Interprétation des résultats** : les résultats de l'immunohistochimie ont été interprétés par un pathologiste expérimenté du CHU. L'expression des biomarqueurs a été évaluée de la manière suivante :

- **ER et PR** : le pourcentage de cellules tumorales présentant une coloration nucléaire positive a été rapporté, et une expression $\geq 1\%$ a été considérée comme positive.
- **HER2** : la surexpression de HER2 a été évaluée selon une échelle de 0 à 3+, où 0 et 1+ sont considérés comme négatifs, 2+ comme équivoque (requérant une analyse par FISH), et 3+ comme surexpression positive.
- **Ki-67** : l'indice Ki-67 a été calculé comme le pourcentage de cellules tumorales présentant une coloration nucléaire positive. Un seuil de 14 % a été utilisé pour différencier les tumeurs à prolifération faible ou élevée.

4. TRAITEMENT DES DONNEES PAR METHODE SUPERVISEE

Toutes les analyses et modélisations ont été réalisées en utilisant le logiciel Python (3. 10. 0), ce qui a permis une automatisation efficace et une intégration transparente des différentes étapes de traitement et d'analyse des données. Python est un langage de programmation polyvalent et puissant, particulièrement apprécié pour sa simplicité et sa lisibilité. Il est largement utilisé dans les domaines de la science des données, de l'analyse statistique et de l'intelligence artificielle grâce à sa vaste bibliothèque de modules, tels que NumPy, Pandas, Scikit-learn et Matplotlib, qui facilitent le traitement et la visualisation des données. Cela permet aux chercheurs de développer rapidement des solutions pour des analyses complexes tout en maintenant une grande flexibilité. En complément de l'analyse immunohistochimique des biomarqueurs ER, PR, HER2, et Ki-67, nous avons intégré les données cliniques (âge des patientes, statut pathologique, stade du cancer, etc.) ; Pour analyser et prédire le statut HER2 ainsi que d'autres résultats cliniques associés.

4.1. Prétraitement des données

Le prétraitement des données est essentiel pour assurer la qualité, la cohérence et la préparation appropriée des données pour les analyses prédictives utilisant des techniques d'apprentissage automatique et de deep learning.

4.1.1. Nettoyage des données

Enregistrements des données médicaux des patientes atteintes de cancer du sein pour identifier les données pertinentes sur le statut HER2 ainsi que d'autres variables cliniques et moléculaires. Ensuite éliminer les données incomplètes ou erronées qui pourraient compromettre la qualité des données.

4.1.2. Sélection des caractéristiques

Identification des variables pertinentes telles que l'âge, le stade de la maladie, le type histologique, les récepteurs hormonaux (ER/PR), Ki67 et PGR.

4.1.3. Transformation des données

Normalisation ou standardisation des variables numériques pour éviter les biais liés aux unités de mesure.

4.1.4. Imputation des données manquantes

Les dossiers médicaux contenant des valeurs manquantes pour les variables pertinentes, y compris le statut HER2 et d'autres caractéristiques cliniques et moléculaires, ont été exclus de l'analyse. Cette décision a été prise pour garantir la fiabilité et la qualité des résultats de notre étude.

4.1.5. Partitionnement des données (Train/Test)

Division en ensembles d'entraînement (80%) et de test (20%) de manière aléatoire pour éviter tout biais de sélection et pour évaluer les performances des modèles ML.

4.1.6. Validation croisées

Mise en place d'une validation croisée pour garantir la robustesse des modèles et éviter le surajustement.

4.1.7. Préparation des Données pour l'analyse ML

Encodage des variables catégorielles et séparation des variables explicatives de la variable cible (statut HER2).

4.1.8. Évaluation des performances

Définition des mesures de performance telles que la précision, le rappel et le F1-score pour évaluer les modèles ML.

4.2. Description des algorithmes utilisés pour la prédiction du statut HER2

Dans cette étude, nous développons des modèles prédictifs utilisant des algorithmes d'apprentissage supervisé pour identifier un biomarqueur pronostique du cancer du sein (CS). L'apprentissage supervisé, méthode la plus répandue en apprentissage automatique, consiste à entraîner un algorithme sur un ensemble de données où chaque observation est associée à une réponse connue. L'objectif est d'établir une relation entre les caractéristiques des données et les résultats, permettant ainsi à l'algorithme de réaliser des prédictions sur de nouveaux cas. Cette approche est couramment utilisée dans de nombreuses applications, allant de la reconnaissance d'images aux diagnostics médicaux, grâce à la disponibilité de données étiquetées qui permettent d'obtenir des prédictions fiables. Nous avons comparé plusieurs algorithmes d'apprentissage supervisé, à savoir la régression logistique, le Random Forest et les algorithmes de boosting, pour identifier l'approche la plus adaptée à la prédiction du biomarqueur HER2 dans le cancer du sein. Les différentes étapes de ce processus expérimental sont représentées graphiquement dans le diagramme de flux (**Figure 33**). Toutes les mesures ont été effectuées avec le même test de division d'entraînement (80 %: Train et 20 % Test) assurant une approche plus objective et reproductible.

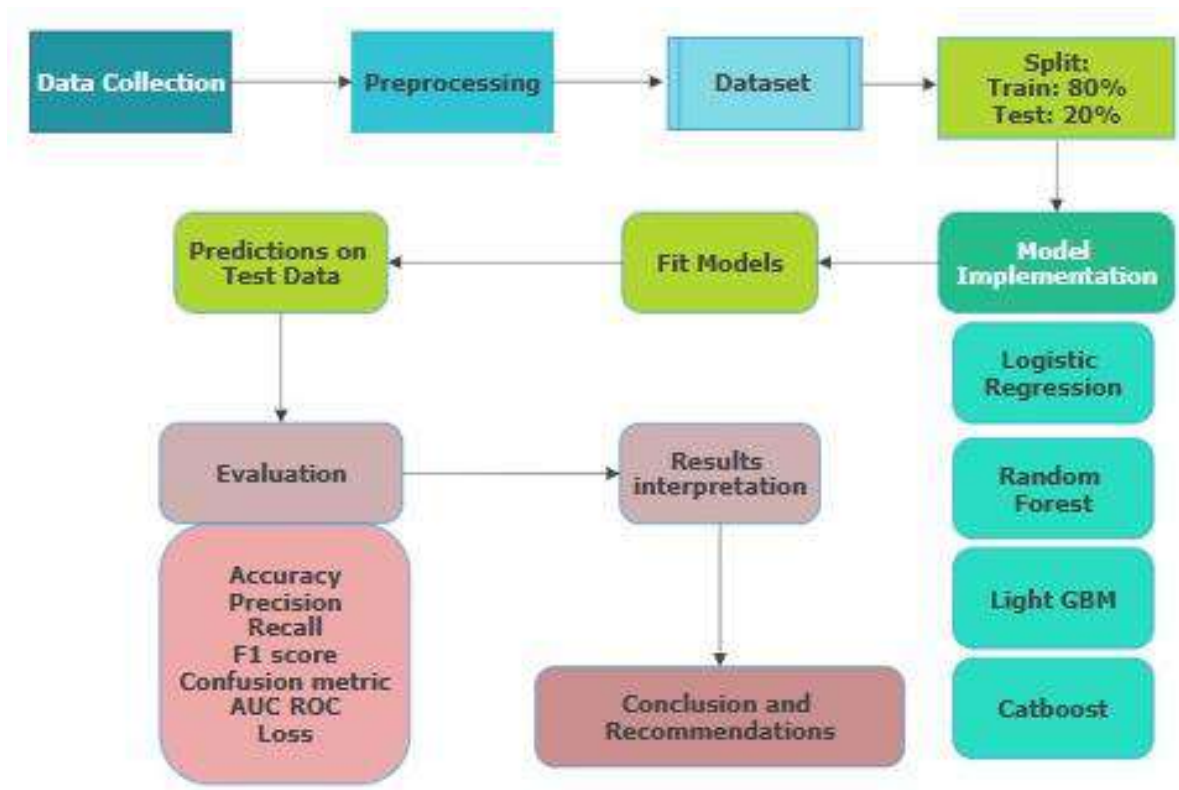


Figure 33. Méthodologie et évaluation des performances des modèles prédictifs de classification du statut HER2 dans le cancer du sein.

4.2.1. Régression Logistique

Nous avons étudié la performance prédictive de la régression logistique (RL), un algorithme de classification binaire qui estime la relation entre les variables indépendantes et la variable dépendante pour convertir le résultat en une valeur de probabilité. Les concepts de base de la régression logistique comprennent des éléments tels que la variable dépendante binaire, les covariables, les scores de probabilité et les seuils de décision. Les fondements mathématiques de la régression logistique reposent sur la fonction sigmoïde, qui permet de transformer la régression linéaire en un modèle de probabilité (**Figure 34**). En apprentissage automatique (ML), la régression logistique est utilisée pour prédire la probabilité qu'un résultat soit l'une des deux classes possibles, étant donné une ou plusieurs variables indépendantes. Dans notre étude, cet algorithme nous permet de prédire la présence de HER2 dans le cancer du sein (**Das, 2014**).

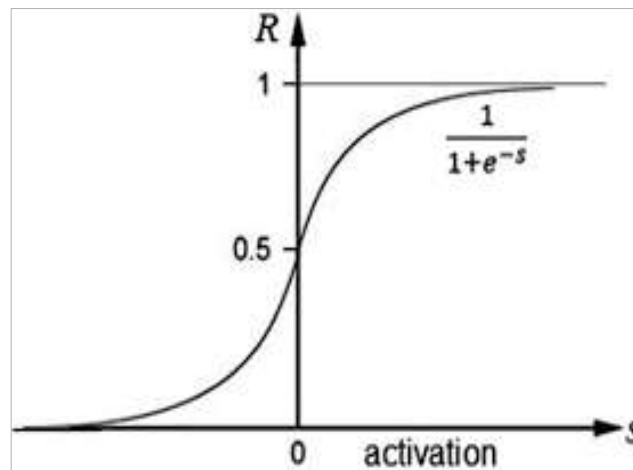


Figure 34. Fonction sigmoïde (Zhang, 2021).

4.2.2. Random Forest

La Forêt d'Arbres Décisionnels Random Forest (RF) est une méthode d'apprentissage ensemble. Basée sur des arbres de décision qui peut être appliquée aux problèmes de régression et de classification. Elle construit une multitude d'arbres de décision et combine leurs prédictions pour obtenir la prédiction finale. Chaque arbre de décision est construit sur un sous-ensemble aléatoire des données d'entraînement, et la prédiction finale est obtenue par l'agrégation des sorties de tous les arbres pour arriver à une prédiction finale (**Figure 35**). Une prédiction plus précise et stable qu'un seul arbre de décision est créée en combinant les sorties de chaque arbre de décision. Le modèle RF a été ajusté pour obtenir la meilleure performance possible dans la prédiction de la variable cible (**Breiman, 2001**).

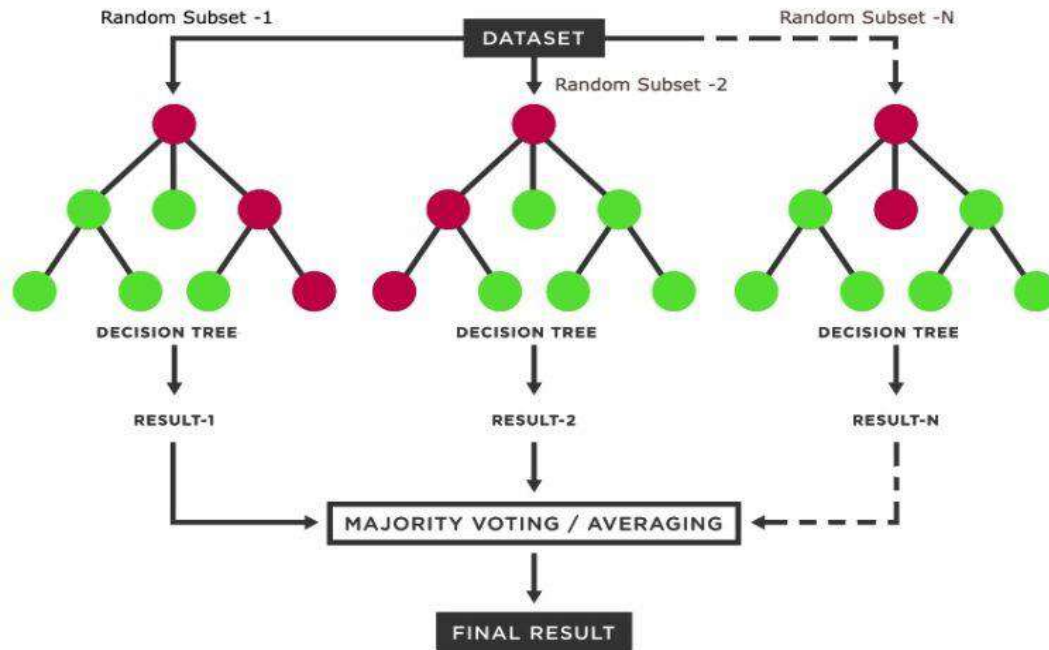


Figure 35. Random forest algorithm dans le machine learning (Yang et al., 2024).

4.2.3. LightGBM

Le modèle Light Gradient Boosting Machine (LGBM) est un cadre de boosting de gradient utilisant l'algorithme de boosting de gradient pour construire un ensemble d'arbres de décision. Elle est conçue pour être efficace et évolutive, ce qui la rend bien adaptée à la gestion de grands ensembles de données (Figure 36). LGBM utilise une stratégie de croissance d'arbre feuille-par-feuille et met en œuvre des fonctionnalités telles que le binning basé sur des histogrammes et le regroupement exclusif de fonctionnalités (Wang et al., 2023). Les paramètres utilisés pour notre modèle LGBM incluent :

- Type de boosting : 'gbdt' (Gradient Boosting Decision Tree);
- Métrique : 'binary_logloss' (perte logarithmique binaire);
- Nombre de feuilles;
- Taux d'apprentissage;
- Fraction de fonctionnalités;
- Verbose;

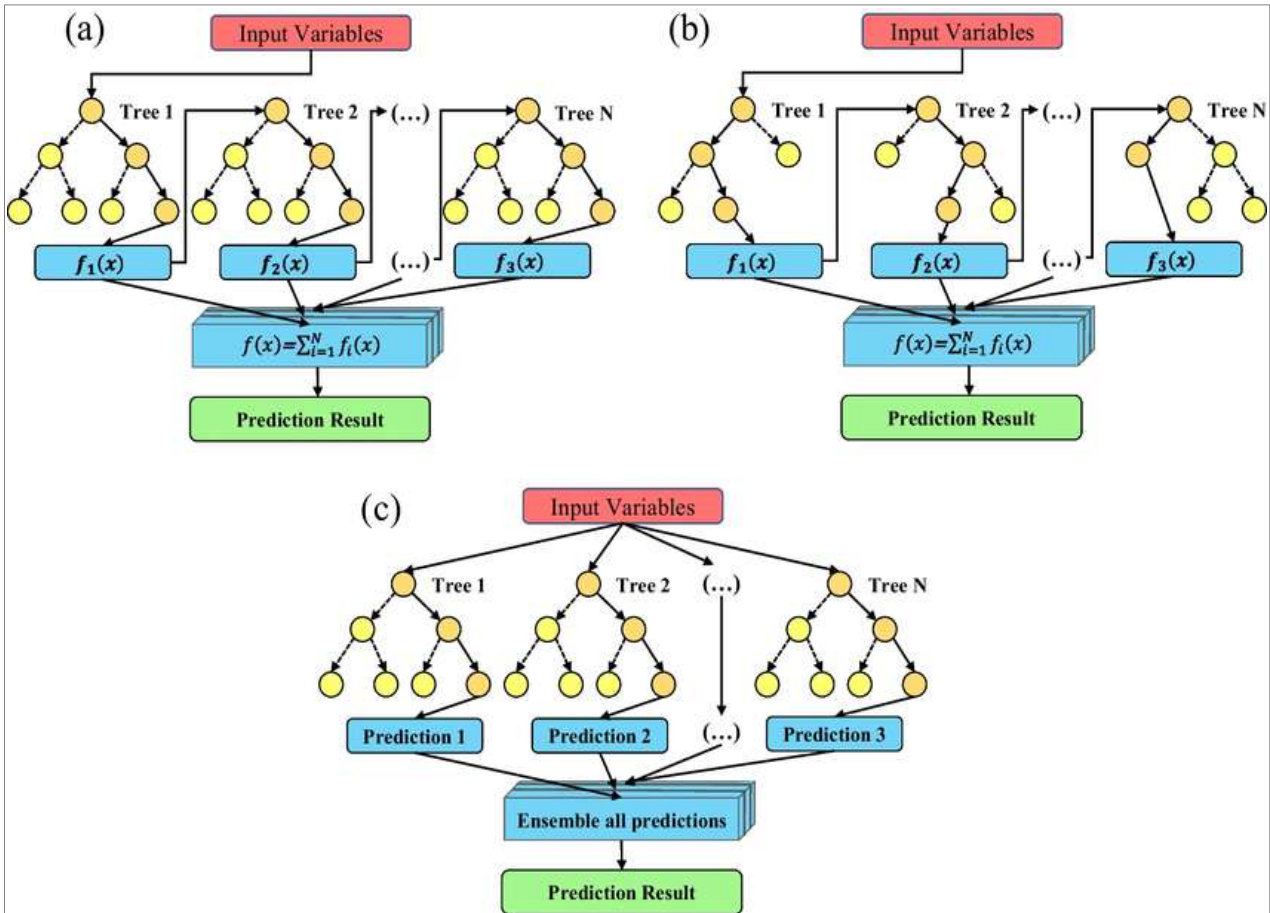


Figure 36. LightGBM Histogramme algorithm (Wang et al., 2023).

4.2.4. CatBoost

CatBoost (CB) est un autre algorithme de boosting de gradient particulièrement efficace pour gérer les variables catégorielles. Il utilise un algorithme de boosting ordonné qui incorpore des techniques nouvelles pour traiter efficacement les fonctionnalités catégorielles (Figure 37). CB convertit automatiquement les variables catégorielles en représentations numériques et gère efficacement les valeurs manquantes (Pandey et al 2023).

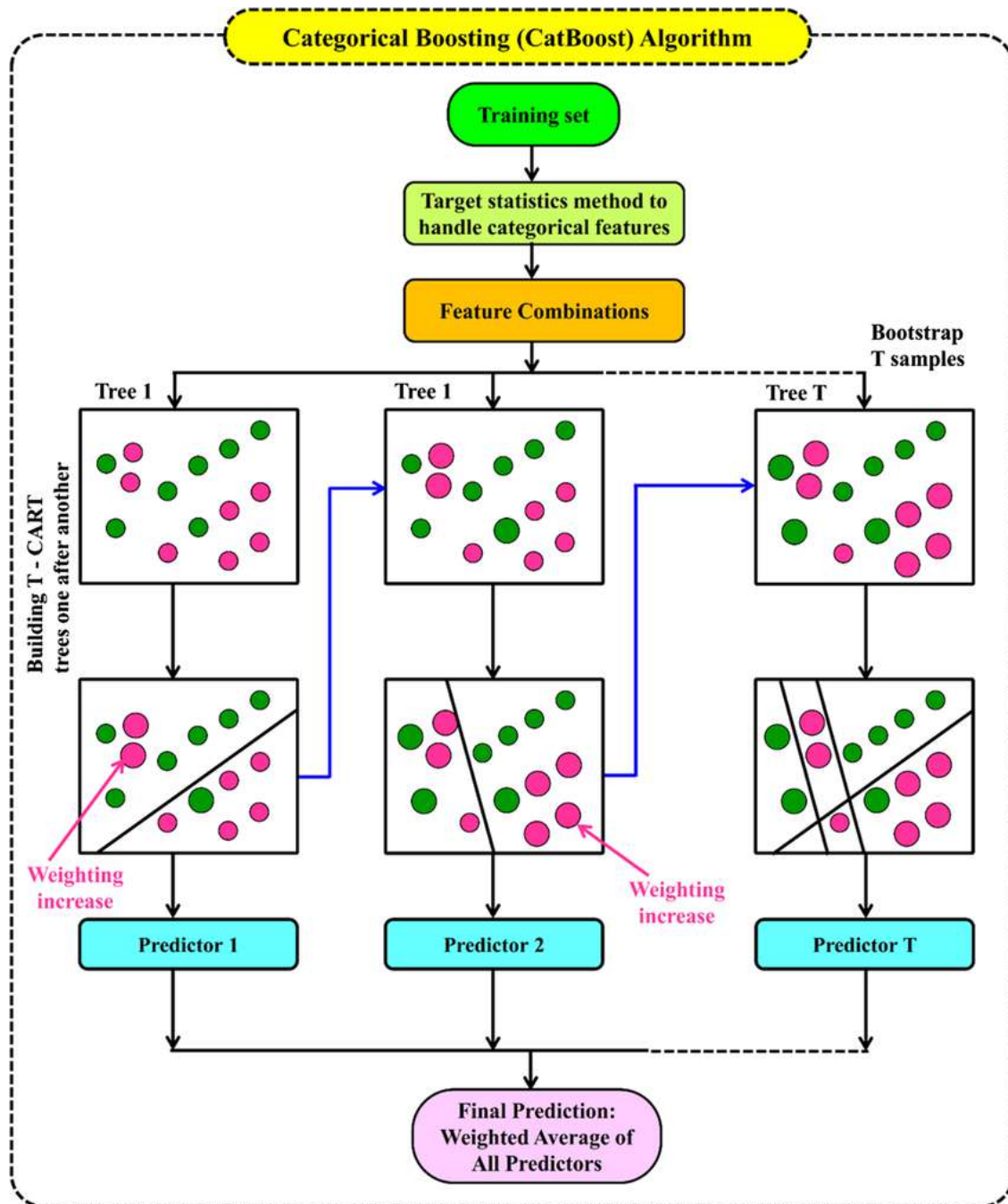


Figure 37. CatBoost Algorithm (Pandey et al., 2023).

4.3. Évaluation des performances des modèles

L'évaluation des performances du modèle s'est principalement concentrée sur la métrique de précision, qui offre une mesure complète de la précision globale des prédictions du modèle par rapport aux étiquettes réelles.

La précision est calculée comme suit :

$$\text{Accuracy} = \frac{TP + TN}{TP + FP + FN + TN} \times 100$$

En entraînant les modèles sur un ensemble de données complet, l'étude a exploité les capacités de ces algorithmes pour obtenir des classifications et des prédictions précises. Des métriques supplémentaires, telles que la précision (precision), le rappel (recall), le F1-score, la sensibilité (sensitivity) et la spécificité (specificity) ont également été calculées pour obtenir une compréhension plus approfondie des performances du modèle, en tenant compte des vrais positifs, des faux positifs et des faux négatifs.

$$\text{Recall} = \frac{TP}{TP + FN} \times 100$$

$$\text{Precision} = \frac{TP}{TP + FP} \times 100$$

$$\text{F-measure} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

La matrice de confusion a été calculée pour évaluer plus en détail les performances du modèle.

$$\text{ER} = \frac{FN + FP}{TP + FP + FN + TN} = 1 - \text{Accuracy}$$

La sensibilité, également appelée taux de vrais positifs (true positive rate), est une métrique de performance cruciale dans les modèles de classification binaire. Elle quantifie la capacité du modèle à identifier correctement les exemples positifs.

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

La spécificité, également appelée taux de vrais négatifs (true negative rate), est une autre mesure importante dans la classification binaire. Elle évalue la capacité du modèle à identifier correctement les exemples négatifs.

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

Nous avons également utilisé deux métriques essentielles pour évaluer les performances de nos modèles de classification : l'AUC-ROC (Surface sous la courbe ROC) et la fonction de perte. L'AUC-ROC est une mesure couramment utilisée pour évaluer la capacité discriminatoire d'un modèle en quantifiant sa capacité à distinguer entre les classes positives et négatives. Elle varie de 0.5 à 1.0, une valeur de 1.0 indiquant une performance parfaite. De plus, la fonction de perte quantifie l'écart entre les prédictions du modèle et les valeurs réelles, une valeur de perte plus faible indiquant une meilleure correspondance entre les prédictions et la vérité terrain.

L'approche courante pour résoudre les tâches d'apprentissage supervisé consiste à minimiser la fonction de perte L :

$$L(f(x), y) = \sum[w_i \times l(f(x_i), y_i)] + J(f)$$

Où :

- $L(f(x), y)$ est la valeur de la fonction de perte au point (x, y) ;
- w_i est le poids du i -ème objet ;
- $J(f)$ est le terme de régularisation ;

La fonction de perte L mesure l'écart entre la sortie prédite $f(x)$ et l'étiquette réelle y pour une entrée donnée x . Elle est calculée en sommant les pertes individuelles pondérées $L(f(x_i), y_i)$ sur tous les exemples d'entraînement, où le poids de chaque exemple w_i reflète son importance relative (**Saber M, 2023**).



DEUXIEME PARTIE

ETUDE RETROSPECTIVE :

**ANALYSE MULTIOMIQUE (TCGA) PAR
STRATIFICATION MOLECULAIRE ET
ENRICHISSEMENT DES VOIES KEGG**

Cette partie de notre étude est de type rétrospectif, et vise à explorer l'intégration des données cliniques et génomiques dans le cancer du sein à l'aide d'une approche multiomique. Cette approche inclut l'analyse des données de méthylation de l'ADN, d'expression génique, et des paramètres cliniques, associée à des techniques avancées de machine learning et de deep learning pour prédire des résultats cliniques et identifier des biomarqueurs clés.

1. POPULATION D'ETUDE ET SOURCE DE DONNEES

Nous avons adopté une approche multiomique pour analyser les données des patientes atteintes de cancer du sein, en intégrant des techniques avancées de machine learning et de deep learning. Les données ont été extraites du dataset de cancer du sein (BRCA) de *The Cancer Genome Atlas* (TCGA), qui comprend des informations cliniques et génomiques complètes. Les variables clés incluent les données cliniques et les données de méthylation de l'ADN provenant de la plateforme Illumina HumanMethylation450 BeadChip (450k). Sur une cohorte initiale de 1 247 patientes, 531 ont été sélectionnées selon les critères suivants :

1.1. Les critères d'inclusion

- Patientes avec un diagnostic confirmé de cancer du sein ;
- Disponibilité d'échantillons tumoraux primaires ;
- Disponibilité des données de méthylation de l'ADN de la plateforme 450k ;
- Disponibilité de données cliniques, y compris le statut de ménopause ;
- Patientes de sexe féminin ;

1.2. Les critères d'exclusion

- Échantillons classés comme tissu normal solide ou métastatique ;
- Données incomplètes ou manquantes pour les variables clés (par exemple, classification du sous-type PAM50, résultats cliniques) ;
- Patients mâles ;

Les données brutes du TCGA ont été extraites à l'aide de l'outil Xena, une plateforme permettant d'accéder et d'explorer des données multiomiques, puis traitées pour :

1. **Imputation des Valeurs Manquantes** : utilisation de SimpleImputer pour traiter les valeurs manquantes.
2. **Normalisation** : application de StandardScaler pour normaliser les niveaux de méthylation.
3. **Élimination des Effets de Lot** : méthode ComBat pour corriger les variations techniques.

1.3. Variables clés

L'étude a intégré des données cliniques et de méthylation de l'ADN ainsi que son expression, incluant les gènes des biomarqueurs tels que *PDCD1*, *CD274*, *TLR4*, *TLR9*, *CTLA4*, *NFKB1*, *ESR1*, *CYP19A1*, *PGR*, et *ERBB2*, ainsi que des paramètres clinico-pathologiques comme l'âge, la taille tumorale, l'envahissement ganglionnaire, les métastases, le statut ménopausique et le stade pathologique.

2. COLLECTE ET PRETRAITEMENT DES DONNEES

2.1. Extraction des données

Les données brutes du TCGA ont été extraites *via* Xena, une plateforme offrant un accès unifié aux données multiomiques.

2.2. Imputation des valeurs manquantes

L'imputation des valeurs manquantes a été réalisée avec l'algorithme SimpleImputer.

2.3. Normalisation et correction des effets de Lot

StandardScaler a été utilisé pour la normalisation des niveaux de méthylation.

La méthode ComBat a été appliquée pour corriger les effets de lot développée pour les données omiques.

3. CLUSTERING ET METHODES D'ANALYSE

3.1. Analyse descriptive

Des statistiques descriptives (moyennes, médianes, écarts-types) ont été calculées pour les caractéristiques cliniques et moléculaires.

3.2. Clustering Non Supervisé

Dans le but d'identifier des sous-groupes homogènes parmi les patientes atteintes de cancer du sein, deux méthodes de clustering non supervisé ont été appliquées :

- **K-Means**, un algorithme partitionnel reposant sur la minimisation de la variance intra-cluster.
- **Le Clustering Agglomératif**, une approche hiérarchique basée sur la fusion progressive de groupes selon une mesure de similarité.

Le nombre optimal de clusters a été fixé à deux, sur la base de deux critères complémentaires :

- La méthode du coude (« *elbow method* »), qui consiste à représenter l'inertie intra-cluster en fonction du nombre de groupes, puis à identifier le point d'inflexion où la diminution de cette inertie ralentit nettement.
- Le coefficient de silhouette, qui évalue la qualité du regroupement en tenant compte à la fois de la cohésion interne des clusters et de leur séparation.

Les données utilisées comprenaient à la fois des paramètres multi-omiques (biomarqueurs méthylés et niveaux d'expression génique) et des variables cliniques. Celles-ci ont été importées depuis un fichier Excel. Une sélection de variables pertinentes a été réalisée afin de centrer l'analyse sur les éléments les plus significatifs sur le plan biologique.

Le prétraitement des données a comporté plusieurs étapes :

- L'imputation des valeurs manquantes par la moyenne.
- L'encodage des variables catégorielles (par exemple, les stades pathologiques) via une méthode one-hot à l'aide d'un ColumnTransformer.
- La normalisation des variables numériques par la méthode StandardScaler.

Afin de réduire la dimensionnalité des données tout en conservant l'essentiel de l'information, une analyse en composantes principales (PCA) a été appliquée, conservant 95 % de la variance totale. En complément, une projection non linéaire en deux dimensions a été réalisée à l'aide de la méthode t-SNE (*t-distributed Stochastic Neighbor Embedding*), facilitant la visualisation des regroupements dans un espace réduit.

Le clustering K-Means a ensuite été exécuté sur les données projetées en 2D, avec un paramètre $k = 2$. L'inertie a été examinée pour évaluer la compacité des clusters, critère essentiel pour juger la qualité du partitionnement.

Les clusters obtenus ont été visualisés dans l'espace t-SNE, et une analyse de leurs caractéristiques a été menée. En particulier, la distribution des sous-types moléculaires PAM50

a été examinée dans chaque groupe, à l'aide de graphiques de type countplot. Cette étape a permis de relier les groupes identifiés à des profils biologiques connus, facilitant ainsi leur interprétation clinique.

3.3. Analyses statistiques et visualisation

La distribution des variables cliniques (âge, stades T/N/M, statut de ménopause, sous-types moléculaires PAM50) selon les clusters a été analysée à l'aide de graphiques de densité, de diagrammes en barres et de boxplots générés avec les bibliothèques Seaborn et Matplotlib. Les différences entre clusters ont été explorées à l'aide de diagrammes de Venn, permettant d'identifier les gènes et facteurs partagés ou spécifiques à chaque groupe. Les comparaisons intergroupes ont été réalisées à l'aide des tests de Mann-Whitney, Chi² ou ANOVA, selon le type de variable. Une analyse de survie de type Kaplan-Meier a été réalisée pour comparer la survie globale (Overall Survival, OS) entre les deux clusters, et la significativité statistique a été évaluée à l'aide du test de log-rank. Les relations entre la méthylation et l'expression génique ont été évaluées à l'aide du coefficient de corrélation de Spearman. L'ensemble de l'analyse a été effectué avec le langage Python (version 3.10), en utilisant les bibliothèques Pandas, NumPy, Scikit-learn, Lifelines, Seaborn et Matplotlib.

4. METHODOLOGIE DE L'ANALYSE KEGG

4.1. Analyse des voies KEGG

L'analyse des voies KEGG (*Kyoto Encyclopedia of Genes and Genomes*) a été réalisée afin d'identifier les processus biologiques et les voies moléculaires enrichis dans les clusters définis par l'analyse de clustering. Cette méthodologie permet de mettre en évidence les mécanismes sous-jacents impliqués dans les différences observées entre les groupes de patients.

4.2. Préparation des données

Les données utilisées incluent les niveaux de méthylation et d'expression génique des biomarqueurs étudiés. Les patients ont été classés en groupes (clusters) sur la base de leur profil multiomique. Les gènes différentiellement exprimés et/ou méthylés entre les clusters ont été sélectionnés avec un seuil de significativité basé sur un p -value < 0.05 et un seuil de log₂FC (log fold change).

4.3. Enrichissement fonctionnel avec KEGG

Les données ont été extraites de la base KEGG (*Kyoto Encyclopedia of Genes and Genomes*), une ressource de référence recensant les voies métaboliques et de signalisation cellulaires. L'analyse d'enrichissement fonctionnel a été réalisée selon deux approches complémentaires : (1) une analyse par sur-représentation génique (*Over-Representation Analysis*, ORA) et (2) une méthode basée sur les scores (*Gene Set Enrichment Analysis*, GSEA). Les p-values ont été calculées à l'aide du test hypergéométrique, avec un ajustement pour les tests multiples via la correction de Benjamini-Hochberg (seuil FDR < 0.05). La significativité statistique des résultats a été visualisée sous forme de $-\log_{10}(\text{p-value})$, permettant une représentation graphique claire des voies biologiques enrichies.



TROISIEME PARTIE

**MODÉLISATION DE LA SURVIE :
BENCHMARK COMPARATIF ENTRE
COX (APPRENTISSAGE HYBRIDE),
ANN ET CNN**

Dans le cadre de cette étude, la modélisation de la survie a été entreprise afin d'évaluer la capacité prédictive de différentes approches statistiques et d'apprentissage automatique à partir de données cliniques et multiomiques issues de patientes atteintes de cancer du sein. Un benchmark comparatif a été établi entre le modèle de Cox à apprentissage hybride, les réseaux de neurones artificiels (ANN) et les réseaux de neurones convolutionnels (CNN). Le modèle de Cox a été appliqué d'abord à l'ensemble de la cohorte, puis spécifiquement à chacun des clusters identifiés par l'analyse non supervisée. En revanche, en raison du nombre restreint de patients par cluster, les modèles ANN et CNN n'ont été exploités que sur l'ensemble de la cohorte, leur complexité nécessitant des volumes de données plus importants pour un apprentissage fiable. Cette démarche vise à comparer la robustesse et la pertinence de chaque méthode dans la prédiction de la survie globale (Overall Survival), en tenant compte des spécificités des données et des contraintes liées à leur structuration.

1. PRÉTRAITEMENT DES DONNÉES POUR LA MODÉLISATION DE LA SURVIE

Les données ont été extraites de la base publique TCGA (*The Cancer Genome Atlas*) via la plateforme UCSC Xena. L'ensemble extrait comprenait des données cliniques, des niveaux de méthylation de l'ADN et des profils d'expression génique.

Les données ont été importées depuis un fichier Excel, puis converties en format numérique lorsque nécessaire. Un nettoyage a été effectué : les valeurs manquantes ont été traitées par suppression ou imputation, selon le type et la fréquence des données manquantes. Les données d'expression génique ont été normalisées pour assurer une comparabilité entre individus. Les variables catégorielles, telles que le stade pathologique et le statut de ménopause, ont été transformées en variables indicatrices (one-hot encoding). Les variables explicatives (features) ont ensuite été séparées de la variable cible, correspondant à la survie globale (Overall Survival, OS).

Une division en deux ensembles, entraînement (80 %) et test (20 %), a été réalisée uniquement pour les modèles supervisés de type réseaux de neurones artificiels (ANN) et réseaux convolutifs (CNN). En revanche, le modèle de Cox, fondé sur une approche semi-paramétrique, a été appliqué directement sur l'ensemble des données, ainsi que sur les sous-groupes définis par clustering, afin de maximiser le nombre d'observations et de préserver la puissance statistique.

2. MODÉLISATION GLOBALE PAR COX HAZARD MODEL

Dans le cadre de cette étude, la variable cible était définie par la colonne OS, représentant la survie globale. Pour l'analyse de survie, nous avons utilisé le modèle de Cox à risques proportionnels, qui permet d'estimer l'effet de plusieurs variables explicatives sur la survie tout en tenant compte de la censure des données. Cette approche est particulièrement adaptée aux études médicales, où la de survie des patients dépend de multiples facteurs cliniques et biologiques.

2.1. Indicateurs statistiques calculés

Dans l'analyse de survie. Le coefficient ($\log(\text{HR})$) reflète la direction et l'intensité de l'effet : une valeur positive indique une augmentation du risque, tandis qu'une valeur négative suggère un effet protecteur. L' $\exp(\text{coef})$ correspond au Hazard Ratio (HR), c'est-à-dire l'effet relatif de la variable sur le risque : une valeur supérieure à 1 indique un risque accru, inférieure à 1 un risque réduit. L'intervalle de confiance à 95% autour du HR permet d'évaluer la précision de l'estimation et d'examiner si l'effet est significatif (si l'intervalle exclut 1). La valeur de z et la p-value sont utilisées pour tester l'hypothèse nulle selon laquelle la variable n'a pas d'effet sur le risque. Une p-value inférieure à un seuil (souvent 0,05) suggère un effet statistiquement significatif. Enfin, la transformation $-\log_2(p)$ permet une meilleure visualisation comparative des p-values sur des graphiques, notamment en accentuant les différences de significativité entre variables.

2.2. Analyse par clusters à partir de la méthylation et de l'expression des gènes

Afin d'explorer des profils tumoraux plus homogènes, une analyse de régression de Cox multivariée a été réalisée séparément au sein de deux clusters de patientes, identifiés préalablement par la méthode de clustering non supervisée (K-means).

Cette classification a été établie à partir des données intégrées de méthylation et d'expression génique des biomarqueurs immunitaire et hormonaux, ainsi que de l'âge au moment du diagnostic pathologique et du statut ménopausique. Cette approche a permis d'analyser l'effet combiné de ces variables sur la survie spécifique à la maladie, tout en tenant compte de l'hétérogénéité moléculaire et clinique au sein de la population étudiée.

3. MODÉLISATION PAR APPRENTISSAGE PROFOND (DEEP LEARNING)

Dans cette section, nous avons exploré le potentiel des approches d'apprentissage profond pour modéliser la survie des patientes atteintes de cancer du sein à partir de données cliniques et multiomiques. Deux architectures ont été testées : un réseau de neurones artificiels (ANN) entièrement connecté, et un réseau de neurones convolutionnel (CNN) conçu pour exploiter les relations locales entre variables, malgré la structure tabulaire des données. Ces modèles ont été entraînés pour prédire la survie globale (OS) sous forme binaire (décès ou non) à partir d'un ensemble de caractéristiques combinant âge, stades pathologiques, biomarqueurs épigénétiques et transcriptomiques, ainsi que le statut ménopausique. L'entraînement a été réalisé sur 80 % des données, avec validation sur les 20 % restantes. Les performances des modèles ont ensuite été comparées à celles du modèle de Cox à travers plusieurs métriques d'évaluation, afin d'examiner la pertinence des approches profondes dans le contexte complexe et multidimensionnel des données biologiques.

3.1. Réseau de Neurone Artificiel (ANN)

Le modèle utilisé repose sur une architecture de réseau de neurones artificiels comprenant trois couches principales (**Figure 38**). La couche d'entrée est constituée de n neurones, chacun correspondant à une variable explicative issue du jeu de données. Elle est suivie d'une couche cachée comportant 20 neurones activés par la fonction ReLU (*Rectified Linear Unit*), introduisant une non-linéarité essentielle pour améliorer la capacité d'apprentissage du modèle. Enfin, la couche de sortie comprend un unique neurone doté d'une fonction d'activation sigmoïde, adaptée à une tâche de classification binaire.

L'entraînement du modèle a été réalisé à l'aide de la descente de gradient comme algorithme d'optimisation, associée à une régularisation L2 afin de limiter le surapprentissage. La fonction de perte utilisée est l'entropie croisée binaire (*Binary Cross-Entropy*), également intégrée avec la régularisation L2. Le modèle a été entraîné sur 2000 époques, avec un taux d'apprentissage fixé à 0,01

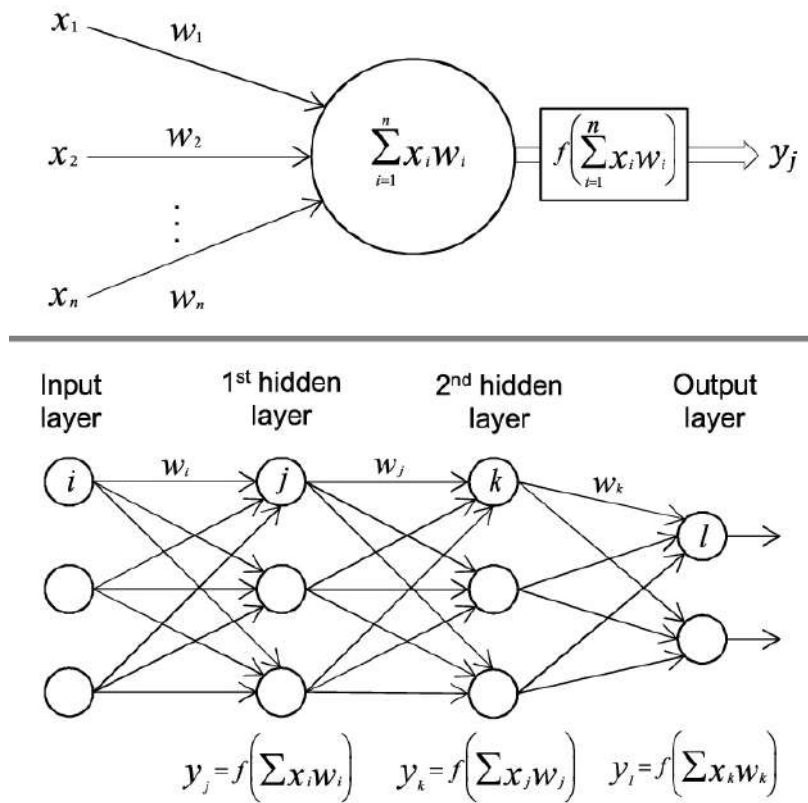


Figure 38. Réseau de Neurone Artificiel (Zhang et al., 2021)

3.2. Réseau de Neurones Convolutif (CNN)

Le modèle mis en place repose sur une architecture de réseau de neurones convolutionnel (CNN), adaptée aux tâches de classification binaire (Figure 39). Il débute par une première couche convolutionnelle comportant 32 filtres de taille 3×3 , suivie d'une fonction d'activation ReLU, introduisant une non-linéarité nécessaire à l'extraction des caractéristiques complexes. Une deuxième couche convolutionnelle est ensuite appliquée, avec 64 filtres de taille 3×3 , également activée par ReLU. Pour réduire la dimensionnalité tout en conservant les informations essentielles, un MaxPooling de taille 2×2 est utilisé.

Les cartes de caractéristiques résultantes sont ensuite aplaties (flattening) en un vecteur unidimensionnel, connecté à une couche dense composée de 64 neurones avec activation ReLU. Enfin, la couche de sortie est constituée d'un unique neurone activé par une fonction sigmoïde, permettant de générer une probabilité pour la classification binaire.

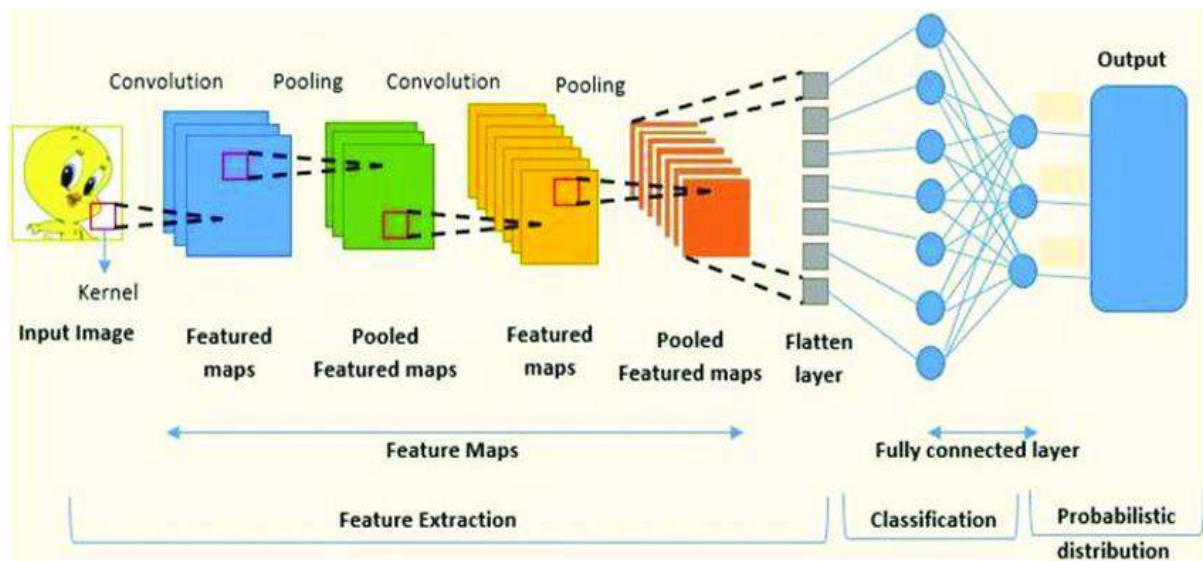


Figure 39. Conventional Neural Network (Eze et al., 2023)

L'entraînement du modèle a été réalisé à l'aide de l'optimiseur Adam, réputé pour sa vitesse de convergence et sa robustesse. La fonction de perte utilisée est l'entropie croisée binaire (Binary Cross-Entropy), adaptée à ce type de tâche. Le modèle a été entraîné sur un total de 50 époques, constituant un compromis entre performance et temps de calcul.

4. ÉVALUATION DES MODELES

L'évaluation des performances des modèles a été réalisée à l'aide de plusieurs métriques complémentaires afin de fournir une appréciation globale et détaillée de leur capacité de classification. La précision globale (Accuracy) a permis de mesurer la proportion d'observations correctement classées. Le rapport de classification a fourni des indicateurs plus fins, incluant la précision (Precision), le rappel (Recall) et le F1-score, ce dernier représentant un compromis entre précision et rappel. L'aire sous la courbe ROC (AUC-ROC) a été utilisée pour évaluer la performance discriminante du modèle, en tenant compte de différents seuils de classification. Enfin, une matrice de confusion a été générée afin d'analyser les erreurs de classification, en distinguant les vrais positifs, vrais négatifs, faux positifs et faux négatifs.



Résultats



PREMIERE PARTIE

ETUDE PROSPECTIVE : MODELISATION PREDICTIVE SUPERVISEE DU STATUT HER2

1. DESCRIPTION CLINICO-PATHOLOGIQUE DE LA POPULATION ETUDIÉE

Le **Tableau 6** présente la répartition de 109 patientes atteintes de cancer du sein selon le statut HER2 et en fonction de plusieurs paramètres cliniques et biologiques.

* L'âge médian au moment du diagnostic était de 47 ans (31-75 ans). La majorité des patientes ont 50 ans ou moins (77.64 %). Ces pourcentages sont comparables dans les deux groupes : 68.23 % chez les HER2+ et 79.16 % chez les HER2-. Les patientes de plus de 50 ans représentent 29.36 % au total, avec une proportion légèrement plus élevée chez les HER2+ (31.76 %) que chez les HER2- (20.83 %).

* La majorité des patientes sont de grade II (71.56 %), tant chez les HER2+ (74.11 %) que chez les HER2- (62.50 %).

* L'analyse biopsique de HER2, a montré que 85 (77,98 %) patientes ont une expression HER2 positive, tandis que 24 (22,01 %) patientes n'ont révélé aucune expression de HER2. Les grades III sont plus fréquents chez les HER2- (33.33 %) que chez les HER2+ (22.35 %), ce qui pourrait indiquer une orientation vers une maladie plus agressive chez les HER2-. Les grades I sont rares dans les deux groupes (3.52 et 4.16 % pour HER2+ et HER2-, respectivement).

Tableau 6. Distribution des patientes selon le statut HER2.

Paramètres	Total Patientes (n = 109) n (%)	HER2+ (n = 85) n (%)	HER2- (n = 24) n (%)
Age moyen (Année)	47 (31-75)	-	-
≤ 50	77 (77.64)	58 (68.23)	19 (79.16)
> 50	32 (29.36)	27 (31.76)	5 (20.83)
NTGM Grade			
I	4 (3.67)	3 (3.52)	1 (4.16)
II	78 (71.56)	63 (74.11)	15 (62.50)
III	27 (24.77)	19 (22.35)	8 (33.33)
Sous type Histologique			
Ductal carcinoma	95 (87.16)	72 (84.70)	23 (95.83)
Lobular carcinoma	14 (12.84)	10 (11.76)	4 (16.66)
Sous type Hormonal			
Luminal A	65 (59.63)	61 (71.76)	4 (16.66)
Luminal B	23 (21.10)	12 (14.11)	11 (45.83)
HER2 +	9 (8.26)	0 (0.00)	9 (37.50)
Basal Like	12 (11.01)	12 (14.11)	0 (0.00)

* Les résultats histologiques ont montré que le carcinome canalaire (ductal) est de loin le plus fréquent (87,16 % au total), aussi bien chez les HER2+ (84,70 %) que chez les HER2- (95,83 %). Le carcinome lobulaire est plus rare, mais un peu plus fréquent chez les HER2- (16,66 %) que chez les HER2+ (11,76 %).

* En ce qui concerne le sous-type hormonal, le sous-type Luminal A domine chez l'ensemble des patientes (59,63 %) et particulièrement chez les HER2+ (71,76 %), alors qu'il est rare chez les HER2- (16,66 %). Le sous-type Luminal B est plus fréquent chez les HER2- (45,83 %) que chez les HER2+ (12,94 %). Le sous-type HER2+ (hormonal) n'est présent que chez les HER2- (37,50 %), ce qui semble paradoxal et peut être dû à une classification spécifique ou à une erreur de transcription. Le sous-type Basal Like est rare, uniquement observé chez les HER2+ (14,11 %).

Les éléments principaux à retenir sont :

- ❖ Les patientes sont majoritairement jeunes (≤ 50 ans) dans les deux groupes.
- ❖ Le grade II prédomine, mais le grade III est plus fréquent chez les HER2-.
- ❖ Le carcinome canalaire est le type histologique principal.
- ❖ Le sous-type Luminal A est surtout retrouvé chez les HER2+, alors que Luminal B et HER2+ (hormonal) sont plus fréquents chez les HER2-.

Ainsi, le statut HER2 est associé à des différences dans la répartition des grades, des sous-types histologiques et hormonaux. Les HER2+ sont majoritairement de type Luminal A et de grade II, alors que les HER2- présentent plus souvent des grades élevés (III) et des sous-types Luminal B ou HER2+ (hormonal). Ces différences peuvent avoir un impact sur le pronostic et les choix thérapeutiques.

Les Boxplots de la **Figure 40** montrent la variation de l'âge selon le statut HER2 et le sous-type hormonal

Les patientes HER2- présentent une distribution d'âge plus large, avec un âge médian plus élevé que les HER2+. Les patientes HER2+ semblent être diagnostiquées à un âge légèrement plus jeune, avec une médiane plus basse et une dispersion plus réduite. Les extrêmes (âges minimum et maximum) sont plus étendus chez les HER2-, tandis que les HER2+ se concentrent dans une tranche d'âge plus restreinte.

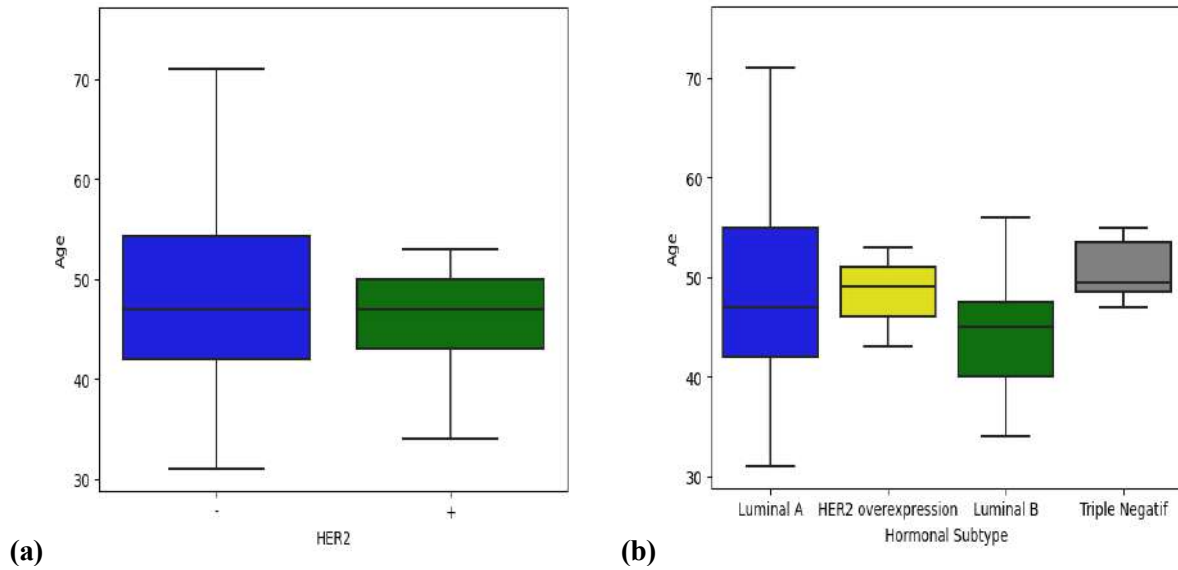


Figure 40. Variation de la distribution par âge dépendante (a) du statut HER2 et (b) du sous-type hormonal

D'autre part, cette figure détaille l'âge selon quatre sous-types hormonaux : Luminal A, HER2 surexprimé, Luminal B, et Triple Négatif. Les patientes du sous-type Luminal A ont la plus grande dispersion d'âge, avec une médiane relativement élevée. Le sous-type HER2 surexprimé présente une distribution plus restreinte avec une médiane légèrement inférieure à celle du Luminal A. Les patientes Luminal B sont diagnostiquées à un âge plus jeune en moyenne, avec une médiane plus basse et une plage d'âge plus réduite. Les patientes Triple Négatif ont une médiane d'âge située entre celles des Luminal A et Luminal B, mais avec une dispersion modérée.

Les éléments principaux à retenir sont :

- ❖ Les cancers du sein HER2+ et certains sous-types hormonaux (notamment Luminal B) touchent des patientes plus jeunes en moyenne.
- ❖ Les sous-types Luminal A et Triple Négatif concernent une population plus âgée ou plus hétérogène en termes d'âge.
- ❖ Ces variations suggèrent que le statut HER2 et le sous-type hormonal influencent l'âge au diagnostic, ce qui peut avoir des implications pronostiques et thérapeutiques.

2. ASSOCIATION DES CARACTÉRISTIQUES CLINIQUES ET BIOLOGIQUES SELON LE STATUT HER2

Le **Tableau 7** présente les résultats d’une analyse de régression logistique (LR) évaluant l’association entre diverses caractéristiques cliniques et biologiques des patientes en fonction du statut HER2.

Tableau 7. Analyse de régression logistique (LR) évaluant l’association entre les caractéristiques cliniques et biologiques des patientes en fonction du statut HER2.

Variables	Coefficient	p-value	OR	(IC95%)
Age	-0.0343	0.2617	0.9661	[0.9334 - 1.0000]
IMC	-1.1764	0.0098	0.3083	[0.0956 - 0.9998]
Grade NGHT	0.2294	0.8371	1.2578	[0.9999 - 1.5835]
Hormonal Subtype	0.7388	0.2330	2.0935	[1.0004 - 4.3847]
T	-0.2308	0.5368	0.7938	[0.6312 - 0.9999]
N	-0.1923	0.6263	0.8250	[0.6811 - 0.9999]
M	-0.0680	0.8884	0.9341	[0.8738 - 1.0000]
Histological Type	-0.1359	0.6833	0.8728	[0.7629 - 1.0000]
Ki67	0.1082	0.8236	1.1142	[0.9329 - 0.9999]
PR	0.3119	0.2986	1.3660	[1.0004 - 1.8653]
ER	-0.4549	0.5622	0.6344	[0.4023 - 0.9992]

Seule la variable IMC (Indice de Masse Corporelle) est significativement associée :

- ✓ **Coefficient** : -1.1764
- ✓ **p-value** : 0.0098
- ✓ **Odds Ratio (OR)** : 0.3083 (IC95% : [0.0956 - 0.9998])

Dans ce modèle, seul l’IMC est un facteur prédictif indépendant et significatif de l’issue IA analysée : une augmentation de l’IMC est associée à une moindre probabilité de survenue. Les autres paramètres cliniques et biologiques testés n’ont pas montré d’association significative.

L'analyse révèle que chaque augmentation d'une unité de l'IMC, diminue de 69,17 % (1 - 0.3083) les chances de la variable dépendante.

Ainsi, un BMI plus élevé est associé à une diminution significative de la probabilité d'obtenir un résultat IA positif (ou l'issue analysée), puisque l'OR est inférieur à 1 et l'intervalle de confiance n'inclut pas 1. Le coefficient négatif confirme cette relation inverse : plus le BMI augmente, plus la probabilité de l'issue IA diminue. Les autres variables (âge, grade, sous-type hormonal, T, N, M, type histologique, Ki67, PR, ER) n'ont pas de p-value significative (toutes > 0.05), ce qui signifie qu'aucune n'est associée de façon statistiquement significative à l'issue IA dans ce modèle.

En résumé, dans ce modèle, seul l'IMC est un facteur prédictif indépendant et significatif de l'issue IA analysée : une augmentation de l'IMC est associée à une moindre probabilité de survenue de cette issue. Les autres paramètres cliniques et biologiques testés n'ont pas montré d'association significative.

La **Figure 41** représente l'importance relative de différentes variables cliniques et biologiques dans la prédiction du statut HER2 chez les patientes atteintes de cancer du sein, selon un modèle d'intelligence artificielle Random Forest (RF).

Les scores d'importance (Importance Score) des variables obtenus dans le modèle RF permettent d'identifier les facteurs les plus influents pour la prédiction du modèle. Plus le score est élevé, plus la variable contribue à la prédiction du statut HER2 par le modèle.

Le Sous-type hormonal a donné un score d'importance de 0,30, le sous-type hormonal se distingue comme la variable la plus influente dans les prédictions du modèle RF. Cela signifie que le modèle accorde un poids considérable à cette variable lors de la classification ou de la prédiction des résultats.

a/ Sous-type hormonal :

C'est la variable la plus importante pour prédire le statut HER2, avec un score d'importance nettement supérieur aux autres (>0,30). Cela signifie que le modèle IA s'appuie principalement sur le sous-type hormonal pour différencier les patientes HER2+ des autres.

b/ Âge :

Deuxième variable la plus importante, avec un score autour de 0,20. L'âge des patientes a donc un poids significatif dans la prédiction du statut HER2.

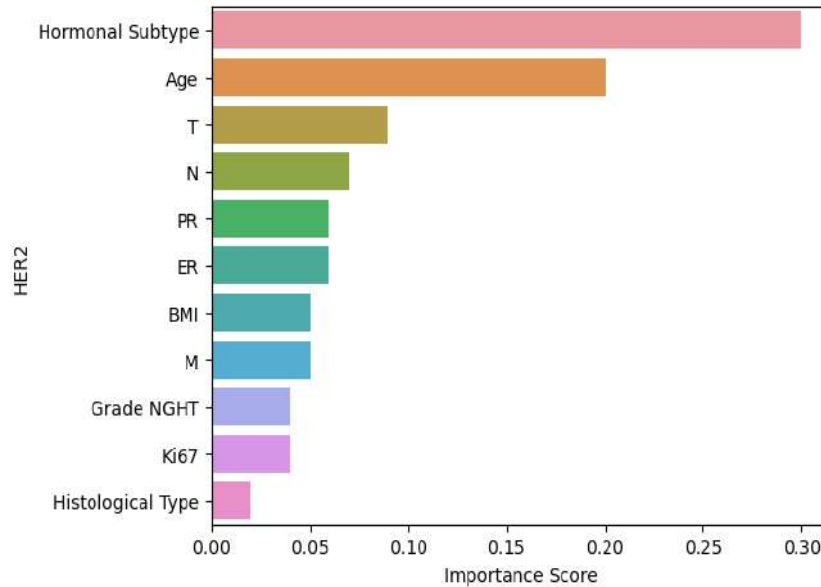


Figure 41. Importance des variables clinico-pathologiques dans la prédiction du statut HER2 par RF.

c/ Variables TNM (T, N, M) :

Les paramètres T (taille tumorale), N (atteinte ganglionnaire) et M (métastases) ont une importance modérée. Cela indique que les caractéristiques anatomiques de la tumeur jouent aussi un rôle, mais moindre que les deux premiers facteurs. Toutefois, le nombre de ganglions lymphatiques atteints (N) est un indicateur clé de la dissémination de la maladie.

d/ Récepteurs hormonaux (PR, ER) :

Les statuts des récepteurs de la progestérone (PR) et des œstrogènes (ER) sont également pris en compte, mais avec une importance moindre.

e/ IMC, Grade NGHT, Ki67, Type histologique :

Ces variables ont une contribution relativement faible à la prédiction du statut HER2 dans ce modèle (inférieurs à 0,04). Le type histologique est la variable la moins contributive. L'IMC, en tant que facteur de risque potentiel, peut avoir un impact indirect sur la progression du cancer du sein *via* l'inflammation et les déséquilibres hormonaux.

Ainsi, le sous-type hormonal et l'âge sont les deux facteurs majeurs pour la prédiction du statut HER2 selon ce modèle d'IA. Les autres variables (T, N, PR, ER, BMI, M, grade, Ki67, type histologique) jouent un rôle secondaire.

En pratique clinique, cela confirme l'importance de bien caractériser le profil hormonal et l'âge de la patiente pour évaluer le statut HER2, tandis que les autres facteurs, bien que pertinents, sont moins déterminants pour cette prédiction spécifique.

Ce graphique met en évidence le poids prépondérant du sous-type hormonal et de l'âge dans la prédiction du statut HER2 par l'intelligence artificielle, ce qui peut guider les cliniciens dans l'interprétation des résultats et l'orientation des stratégies diagnostiques.

L'âge et la taille de la tumeur (T) obtiennent des scores respectifs de 0,20 et 0,09, les plaçant parmi les variables les plus importantes après le sous-type hormonal. Cela reflète l'impact bien documenté de l'âge et de la taille tumorale sur le pronostic et les décisions thérapeutiques. Par exemple, l'âge peut influencer les facteurs hormonaux et le profil génomique, tandis que la taille de la tumeur (T) est directement liée à l'avancement de la maladie.

Des variables comme le nombre de ganglions lymphatiques (N), le récepteur de la progestérone (PR), le récepteur de l'œstrogène (ER), et l'IMC montrent des scores d'importance situés entre 0,05 et 0,07. Bien qu'elles soient légèrement moins influentes que les trois premières variables, leur contribution reste significative: Les récepteurs hormonaux (PR et ER) sont bien connus pour leur rôle dans la réponse aux traitements hormonaux.

Les résultats obtenus suggèrent que bien que ces variables puissent jouer un rôle clinique pertinent, elles ont une influence moindre dans les prédictions de ce modèle spécifique. Cela peut indiquer que ces facteurs, bien qu'utilisés couramment pour la stratification des risques, ne sont pas aussi discriminants pour les prédictions du modèle RF que les autres variables.

3. PERFORMANCES COMPAREES DES DIFFERENTS MODELES DE MACHINE LEARNING POUR PREDIRE LE STATUT HER2

La **Figure 42** compare les performances de quatre modèles de machine learning (LR, RF, LightGBM et CB) pour la prédiction du statut HER2 à l'aide de plusieurs indicateurs statistiques.

3.1. Modèle de régression logistique (LR)

Le modèle régression logistique a montré une précision globale de 90,90 %, ce qui est un bon indicateur de performance. Les métriques détaillées révèlent :

- ❖ **Précision** : 77,77 % signifie que parmi les prédictions positives faites par le modèle, environ 78 % étaient correctes.
- ❖ **Rappel** : 100 % reflète la capacité du modèle à identifier correctement tous les cas positifs, sans en manquer.

- ❖ **F1-score** : 87,5 % indique un équilibre entre précision et rappel, suggérant que le modèle est capable de bien capturer les vrais cas HER2 positifs tout en maintenant un bon taux de précision.

Le rappel parfait (100 %) du modèle LR montre qu'il n'a manqué aucun cas positif de HER2, ce qui est crucial dans les cas où la détection précoce et précise est nécessaire pour adapter les traitements ciblés. Cependant, la précision de 77,77 % montre que le modèle a classé certains cas incorrectement comme positifs, nécessitant des améliorations pour réduire les faux positifs.

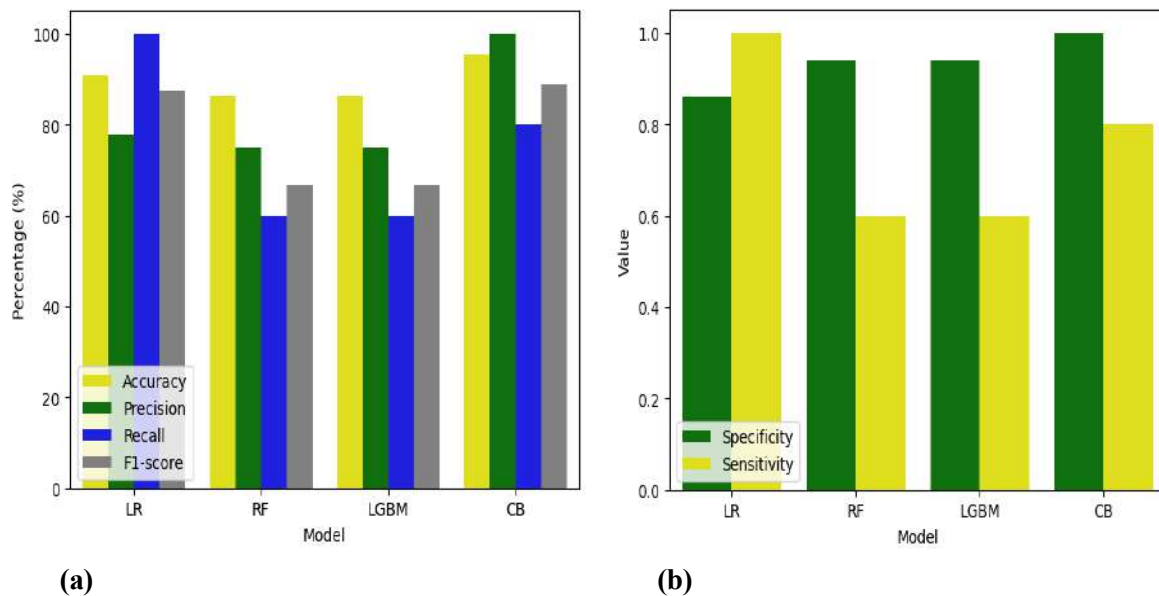


Figure 42. Indicateurs de performance (a) Acc., Recall, Prec. et F1-Score; (b) Spécificité et Sensibilité des modèles proposés.

3.2. Modèles Random Forest (RF) et LightGBM

Ces modèles ont tous deux obtenu une précision de 86,36 %, légèrement inférieure à celle du modèle LR. Les métriques détaillées révèlent :

- ❖ **Précision** : cela signifie que 75 % des prédictions positives étaient correctes.
- ❖ **Rappel** : 60 % montre que les modèles n'ont identifié que 60 % des vrais cas positifs.
- ❖ **F1-score** : 66,67 % reflète un compromis moins favorable entre la précision et le rappel par rapport au modèle LR.

Bien que ces deux modèles montrent une bonne capacité de prédiction générale (spécificité de 94 %), leur rappel relativement faible (60 %) indique qu'ils manquent plusieurs cas positifs. Ces

résultats suggèrent que RF et LightGBM sont plus conservateurs dans leurs prédictions positives, augmentant le risque de ne pas identifier certains patients HER2 positifs.

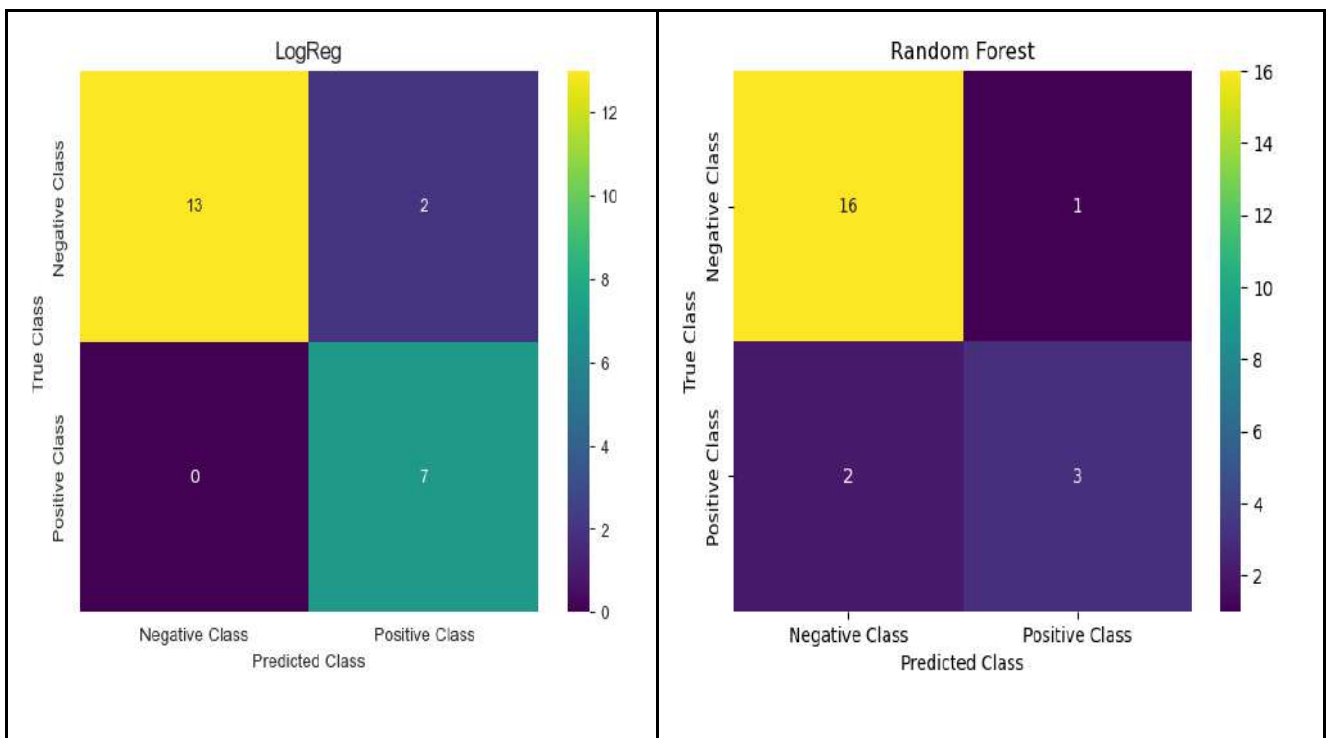
3.3. Modèle CatBoost (CB)

Ce modèle a surpassé tous les autres en obtenant une précision de 95,45 %, ce qui en fait le modèle le plus performant pour la prédiction du statut HER2 dans cette étude. Ses performances détaillées sont les suivantes :

- ❖ **Précision** : 100 % signifie que toutes les prédictions positives étaient correctes, sans aucun faux positif.
- ❖ **Rappel** : 80 % montre que le modèle a manqué 20 % des cas HER2 positifs.
- ❖ **F1-score** : 88,89 % souligne que ce modèle a bien équilibré la capacité à identifier les vrais positifs tout en évitant les faux positifs.

3.4. Matrices de confusion des modèles LR, RF, LGBM et CB

La spécificité et sensibilité élevées de CatBoost (100 % et 80 % respectivement) montrent une excellente performance dans la détection à la fois des cas positifs et négatifs. Ce modèle offre donc une capacité de prédiction robuste qui pourrait être particulièrement utile dans des scénarios cliniques où l'identification précise des patients HER2 positifs est essentielle pour guider le traitement.



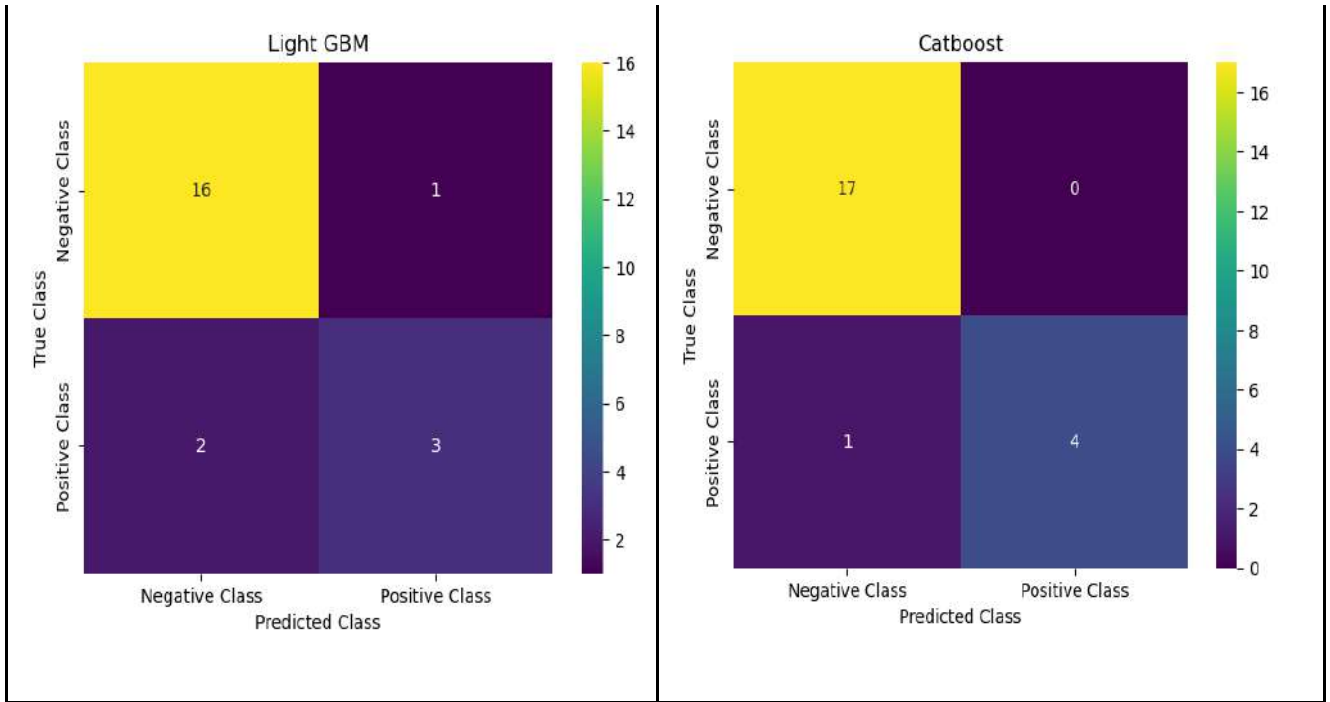


Figure 43. Matrices de confusion des modèles LR, RF, LGBM et CB.

La Figure 43 présente les matrices de confusion des modèles, et permet de visualiser la répartition des prédictions correctes et incorrectes pour chaque modèle.

a/ Modèle de régression logistique (LR) : Le modèle LR, avec une précision de 90,90 %, a bien performé dans la prédiction du statut HER2. La matrice de confusion pour ce modèle montre :

- **13 vrais négatifs** (les cas négatifs correctement prédits),
- **7 vrais positifs** (les cas positifs correctement prédits),
- **2 faux positifs** (les cas négatifs incorrectement classés comme positifs),
- **2 faux négatifs** (les cas positifs incorrectement classés comme négatifs).

Ces résultats suggèrent que le modèle LR a une bonne capacité à identifier à la fois les cas positifs et négatifs, mais avec un petit nombre d'erreurs dans chaque catégorie.

b/ Modèles Random Forest (RF) et Light Gradient Boosting Machine (LGBM) : Les modèles RF et LGBM, ayant tous deux atteint une précision de 86,36 %, montrent une performance légèrement inférieure à celle du modèle LR. Leurs matrices de confusion présentent :

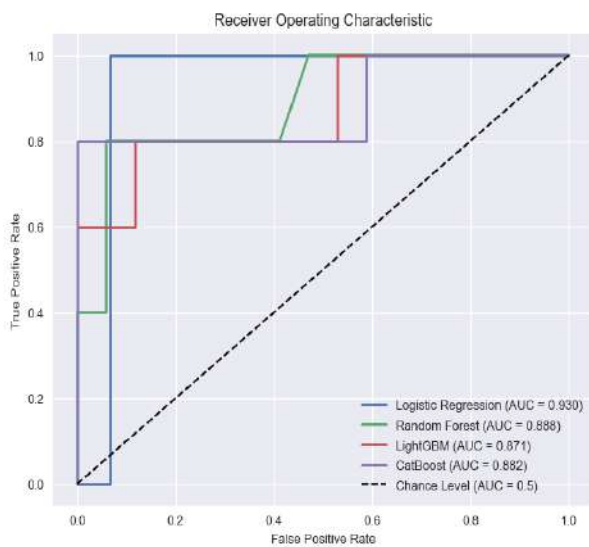
- **16 vrais négatifs**,
- **3 vrais positifs**,
- **0 faux positif**,
- **1 faux négatif**.

Ces modèles ont été capables d'identifier correctement la majorité des cas négatifs, sans produire de faux positifs. Cependant, leur capacité à identifier les cas positifs est moins bonne, avec seulement 3 vrais positifs et 1 faux négatif. Cela suggère que ces modèles pourraient sous-estimer les cas HER2 positifs.

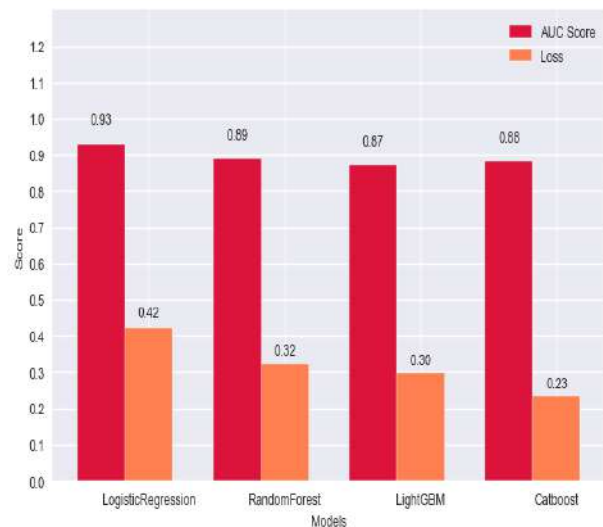
*c/ **Modèle CatBoost (CB)*** : Le modèle CatBoost (CB), avec une précision impressionnante de 95,45 %, est celui qui a obtenu les meilleurs résultats dans la prédiction du statut HER2. Sa matrice de confusion montre :

- 17 vrais négatifs,
- 4 vrais positifs,
- 1 faux positif,
- 0 faux négatif.

Le modèle CB a correctement identifié presque tous les cas, y compris tous les cas positifs (aucun faux négatif), ce qui est un résultat très favorable pour un diagnostic clinique. La présence d'un seul faux positif n'altère pas significativement la performance globale du modèle.



(a)



(b)

Figure 44. Aire Sous la Courbe (AUC) et Perte pour chaque Modèle
(a) Courbe ROC avec scores AUC ; **(b)** Perte pour les modèles implémentés

Les résultats obtenus mettent en évidence la capacité des modèles de classification à bien prédire le statut HER2 dans le cancer du sein, en se basant sur des performances solides en termes d'AUC-ROC et de fonction de perte (**Figure 44**).

L'AUC-ROC est une métrique clé pour évaluer la capacité des modèles à discriminer les patientes en fonction de leur statut HER2. Plus l'AUC-ROC est proche de 1, meilleure est la capacité du modèle à séparer les cas positifs et négatifs :

- Le **modèle de régression logistique (LR)** a atteint une AUC-ROC de 0,930, indiquant une excellente capacité discriminative.
- Les **modèles Random Forest (RF)** et **Light Gradient Boosting Machine (LGBM)** ont respectivement obtenu des AUC-ROC de 0,89 et 0,88, ce qui montre également une très bonne capacité à discriminer les cas HER2 positifs et négatifs, bien que légèrement inférieure à celle du modèle LR.
- Le **modèle CatBoost (CB)** a atteint une AUC-ROC de 0,87, indiquant une bonne performance mais légèrement inférieure à celle des autres modèles.

Ces résultats suggèrent que le modèle LR est le plus performant en termes de discrimination, tandis que les autres modèles restent compétitifs avec des AUC-ROC supérieures à 0,85, ce qui est un seuil généralement considéré comme bon dans des contextes cliniques.

* **Analyse de la fonction de perte** : en complément des AUC-ROC, la fonction de perte permet d'évaluer l'écart entre les valeurs prédites par les modèles et les valeurs réelles du statut HER2. Une faible perte indique une meilleure précision des prédictions :

1. Le **modèle LR** a présenté une **perte de 0,422**, ce qui montre qu'il y a un écart modéré entre les prédictions et les valeurs réelles.
2. Le **modèle RF** a affiché une **perte de 0,321**, indiquant une légère augmentation de la dispersion des prédictions par rapport aux valeurs réelles, mais dans un intervalle acceptable.
3. Les **modèles LGBM et CB** ont démontré des **pertes plus faibles de 0,297 et 0,235**, respectivement. Ces résultats signalent un degré élevé d'accord entre les prédictions et les valeurs réelles, en particulier pour le modèle **CB**, qui a la plus faible fonction de perte.

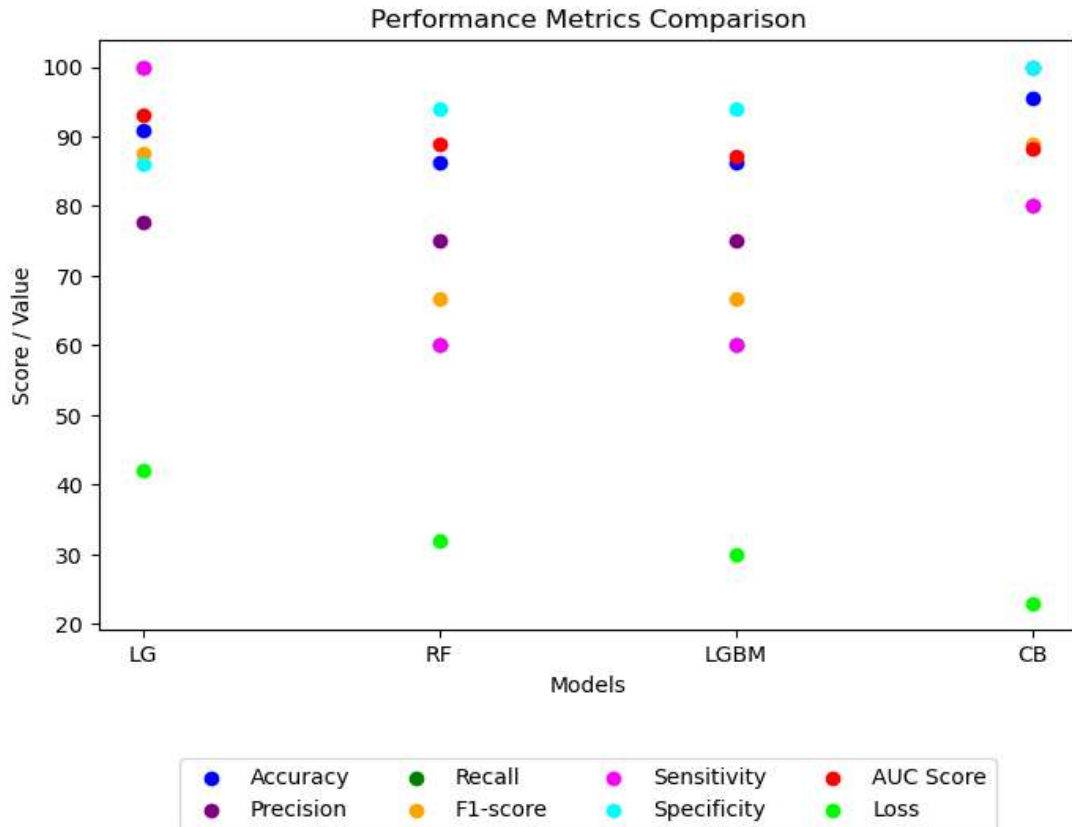


Figure 45. Score/valeur de validation croisée en fonction du nombre de caractéristiques à chaque itération.

Globalement, les résultats obtenus (**Figure 45**) permettent de dégager un classement clair des modèles de classification utilisés pour prédire le statut HER2 dans le CS.

Parmi ces modèles, CatBoost (CB) se distingue comme le meilleur, offrant des performances exceptionnelles sur plusieurs métriques. Il a montré une combinaison équilibrée de précision, de rappel, de spécificité, et de F1-score, ce qui suggère sa robustesse dans le contexte de classification binaire :

- **Précision : 95,45 %**, ce qui signifie que CB a correctement prédit le statut HER2 dans la majorité des cas.
- **Rappel : 80 %**, indiquant que le modèle a bien identifié les patients HER2 positifs.
- **AUC-ROC : 0,87**, montrant une bonne capacité discriminative entre les classes positives et négatives.
- **Perte : 0,235**, la plus faible parmi les modèles, ce qui traduit une excellente concordance entre les prédictions et les valeurs réelles du statut HER2.

Le modèle de régression logistique (LR) se place également parmi les meilleurs, notamment en termes de précision et d'AUC-ROC :

- **Précision : 90,90 %**, ce qui montre sa capacité à fournir des prédictions fiables.
- **Rappel : 100 %**, indiquant que LR n'a pas manqué un seul cas de HER2 positif dans ce jeu de données, ce qui est crucial dans le cadre du cancer où la détection des cas positifs est primordiale.
- **AUC-ROC : 0,930**, le plus élevé parmi tous les modèles, ce qui reflète sa capacité exceptionnelle à distinguer les patientes positives et négatives pour le statut HER2.
- **Perte : 0,422**, légèrement plus élevée que celle des autres modèles, mais restant dans des limites acceptables.

Le modèle LR se distingue également par sa forte capacité discriminative et son rappel parfait, bien que sa fonction de perte soit un peu plus élevée que celle de CB.

Le modèle LightGBM (LGBM) se situe juste derrière, avec de bonnes performances en termes d'AUC-ROC et une faible perte, bien qu'il ne soit pas aussi performant que CB et LR dans l'ensemble :

- **AUC-ROC : 0,88**, montrant une capacité compétitive à discriminer les cas HER2 positifs et négatifs.
- **Perte : 0,297**, l'une des plus faibles, ce qui indique une bonne précision dans ses prédictions.

Cependant, le modèle LGBM a obtenu des résultats légèrement moins bons en termes de précision et de rappel, ce qui le place en troisième position.

Quant au Random Forest (RF), bien qu'il ait montré une bonne capacité à identifier les cas négatifs avec une précision élevée, il a obtenu :

- **AUC-ROC : 0,89**, indiquant une bonne capacité discriminative.
- **Perte : 0,321**, légèrement plus élevée que les autres modèles.

Ses performances générales sont un peu moins élevées que celles de CB et LR, ce qui le place en quatrième position.

Récapitulation :

- ❖ **CatBoost (CB)** est le modèle le plus performant sur l'ensemble des indicateurs (Accuracy, Precision, Recall, F1-score, Spécificité, Sensibilité).
- ❖ **Random Forest (RF)** offre également de très bonnes performances, mais légèrement inférieures à CB.
- ❖ **Logistic Regression (LR)** est très sensible mais moins spécifique, ce qui peut conduire à plus de faux positifs.
- ❖ **LightGBM (LGBM)** est moins performant que les autres modèles.

Le modèle CatBoost est le plus fiable pour cette la prédiction, offrant à la fois une excellente capacité à détecter les cas positifs (sensibilité) et à éviter les faux positifs (spécificité), ainsi que des scores globaux très élevés sur tous les autres critères de performance. Ce modèle serait donc recommandé pour une utilisation en pratique clinique ou en recherche.



DEUXIEME PARTIE

ETUDE RETROSPECTIVE :

**ANALYSE MULTIOMIQUE (TCGA) PAR
STRATIFICATION MOLECULAIRE ET
ENRICHISSEMENT DES VOIES KEGG**

1. CARACTERISTIQUES DEMOGRAPHIQUES ET CLINICO-PATHOLOGIQUES DES PATIENTES

Dans cette partie de la thèse, nous avons utilisé des moyens modernes de bio-informatique, à savoir l'intelligence artificielle, et des ressources de bases de données, en exploitant la base de données TCGA pour sélectionner des gènes immunitaires et hormonaux surexprimés dans le cancer du sein. Les résultats sont présentés dans le **Tableau 8**.

Tableau 8. Caractéristiques clinico-pathologiques des patientes.

Caractéristiques	Patientes (n = 531) n (%)
Age (ans)	
Moyenne ± DS	57.56 ± 13.1177
min - max	26 - 90
Décès	
Moyenne ± DS	1733.81 ± 1176.33
min - max	160.0 - 6593.0
Sous type PAM50	
LuminalA	275 (51.79)
LuminalB	123 (23.16)
Basal-Like	85 (16.01)
HER2	31 (5.84)
Normal-Like	17 (3.20)
TNM	
T Pathologique:	
T1	144 (5.27)
T2	308 (58.00)
T3	60 (11.30)
T4	19 (0.38)
N Pathologique:	
N0	235 (44.25)
N1	187 (35.23)
N2	74 (13.93)
N3	35 (6.59)
M Pathologique:	
M0	461 (86.82)
M1	70 (13.18)
Statut Ménopausique	
Post	342 (64.41)
Pré	131 (24.67)
Indéterminé	15 (2.82)
Péri	15 (2.82)
Stade Pathologique	
I	88 (16.57)
II	301 (56.69)
III	130 (24.48)
IV	10 (1.88)

La cohorte de cette étude prospective comprenait 531 femmes atteintes du cancer du sein, sélectionnées à partir de la database TCGA selon les critères d'inclusion et d'exclusion cités dans la partie matériel et méthodes.

1.1. Répartition selon l'âge

La **Figure 46** représente une distribution normale des patientes correspondant à un âge moyen observé de $57,6 \pm 13,1$ ans (intervalle : 26 - 90 ans). Ces résultats reflètent une incidence accrue du cancer du sein chez les femmes post-ménopausées, souvent associée aux sous-types hormono-dépendants.

La faible fréquence des diagnostics chez les patientes plus jeunes (<40 ans) peut être liée à des cancers plus agressifs ou des mutations génétiques, tandis que les cas plus rares chez les femmes âgées (>80 ans) pourraient résulter d'un diagnostic tardif ou de comorbidités. Ces résultats soulignent l'importance du dépistage ciblé entre 50 et 70 ans pour maximiser l'efficacité des prises en charge précoces.

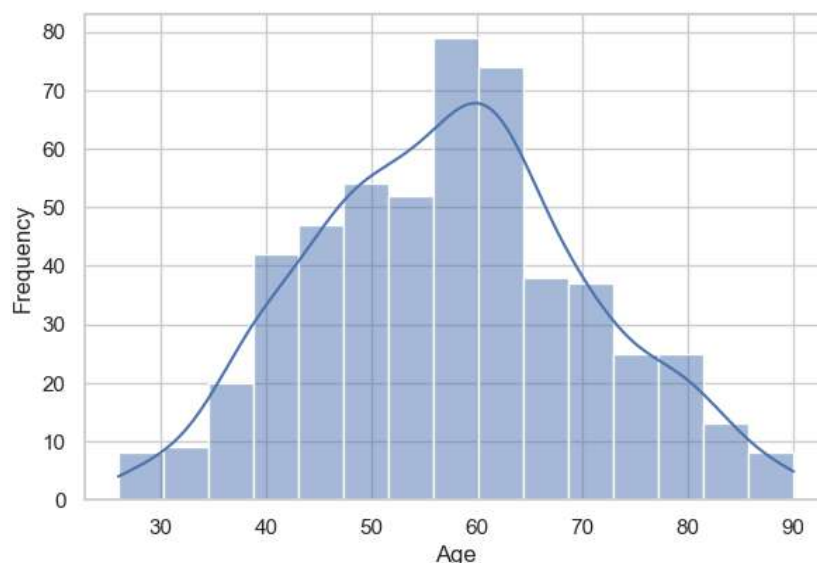


Figure 46. Distribution de l'âge au diagnostic pathologique initial des patientes.

1.2. Répartition selon le stade pathologique et le stade TNM

La répartition des sous-types moléculaires selon la classification PAM50 a mis en évidence une prédominance des sous-types hormono-dépendants suivants : Luminal A (n = 275 ; 51,8 %) et Luminal B (n = 123 ; 23,2%). Les sous-types plus agressifs, à savoir Basal-Like (n = 85 ;

16.0 %) et HER2-enrichi (n = 31 ; 8.0 %), sont moins fréquents, tandis que le sous-type Normal-Like (n = 17 ; 3.2 %) est minoritaire (**Tableau 8 et Figure 47**).

La distribution des stades TNM a montré une prédominance du stade T2 (n = 308 ; 58.00 %), N0 (n = 235 ; 44.25 %) et M0 (n = 461 ; 86.82 %).

Cette distribution reflète l'hétérogénéité biologique du cancer du sein et ses implications sur le pronostic et les options thérapeutiques.

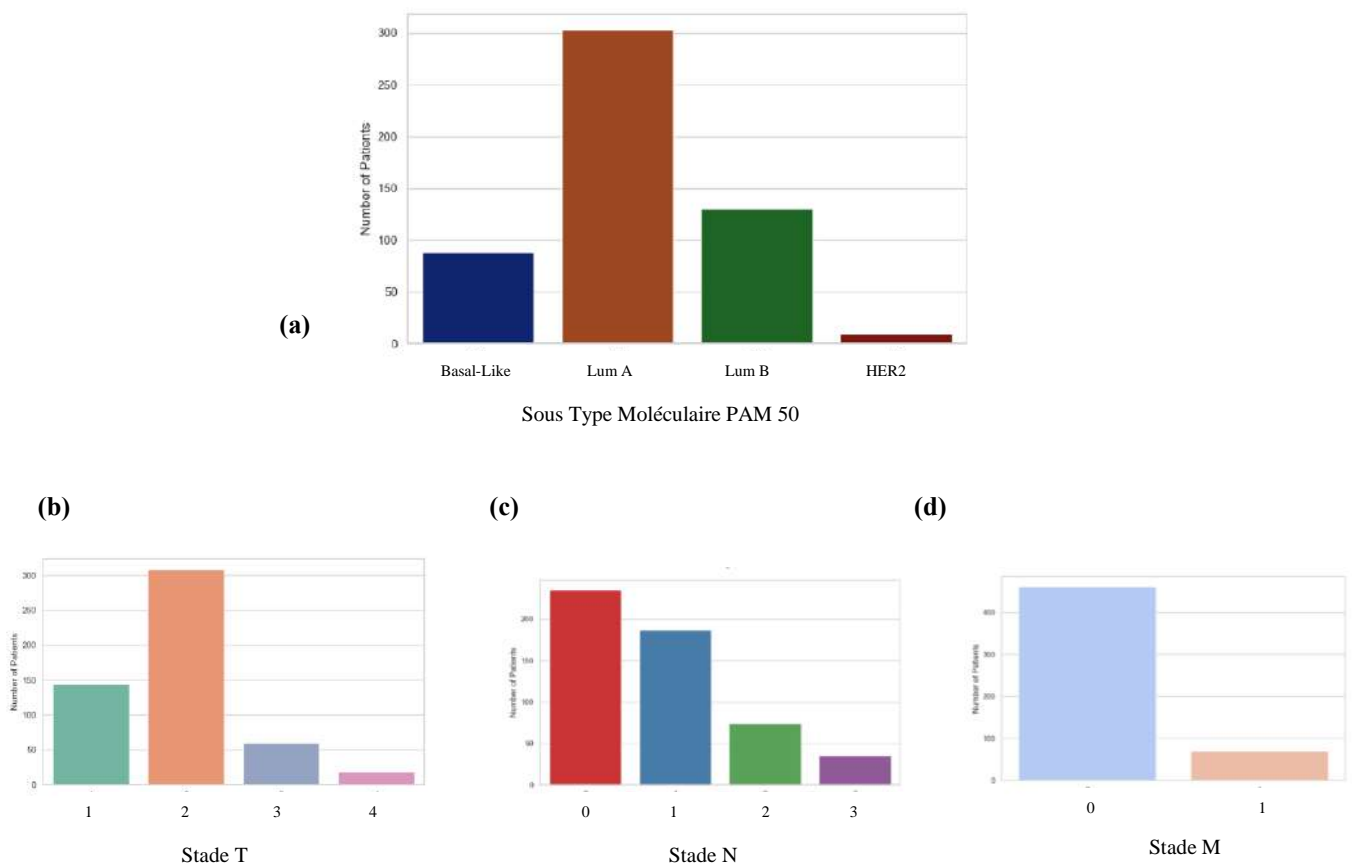


Figure 47. Distribution en fonction des sous type moléculaire PAM 50 (a) et des stades T (b), N (c) et M (d) dans la population étudiée.

1.3. Répartition selon le stade ménopausique et la survie

En termes de statut ménopausique, la majorité des patientes étaient post-ménopausées (n = 342 ; 64,4 %), suivies de patientes pré-ménopausées (n = 131 ; 24,7 %). Les catégories péri-ménopausées (n = 15 ; 2,8 %) et indéterminées (n = 15 ; 2,8 %) étaient marginales (**Figure 48 a**). Cette répartition met en évidence le rôle potentiel des facteurs hormonaux dans le développement et la progression du cancer du sein. Tous les échantillons provenaient de tumeurs primaires, garantissant une cohérence dans l'analyse histopathologique et moléculaire.

Concernant la survie, une variabilité importante a été observée dans le délai avant le décès, avec une durée moyenne de $1733,8 \pm 1176,3$ jours (intervalle : 160 à 6593 jours), suggérant des disparités considérables dans les trajectoires cliniques des patientes (**Figure 48 b**).

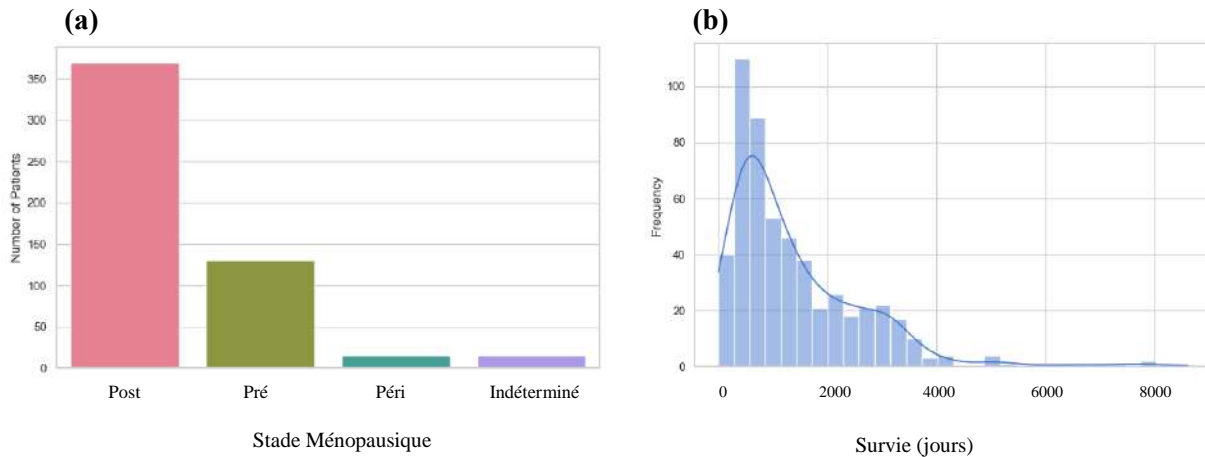


Figure 48. Distribution en fonction du stade ménopausique (a) et de la survie (b) dans la population étudiée.

2. DISTRIBUTION DES GENES METHYLES ET EXPRIMES (IMMUNITAIRES ET HORMONAUX) DANS LA POPULATION TOTALE

2.1. Gènes méthylés

Les résultats de la distribution des gènes méthylés immunitaires (*CD274*, *NFKB1*, *PDCD1*, *CTLA4*, *TLR4* et *TLR9*) et hormonaux (*PGR*, *ESR1*, *CYP19A1* et *ERBB2*) sont représentés dans la **Figure 34 et 35**, et montrent des variations significatives des niveaux de méthylation d'ADN des différents gènes.

2.1.1. Gènes immunitaires

Les résultats de la distribution des gènes immunitaires méthylés (*CD274*, *NFKB1*, *PDCD1*, *CTLA4*, *TLR4* et *TLR9*) dans la population totale sont représentés dans la **Figure 49 (a,c) et 50**.

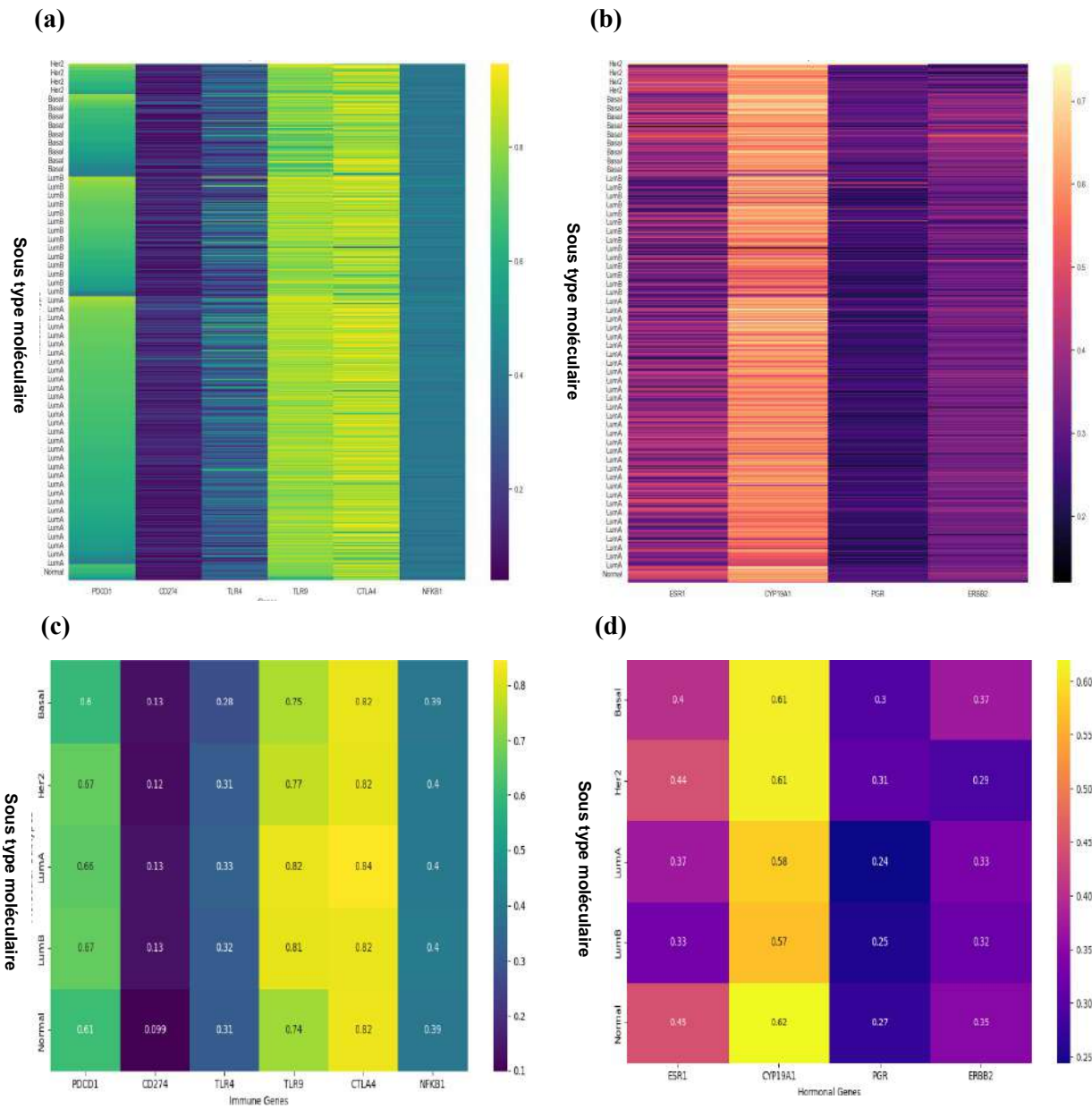


Figure 49. Heatmaps des profils de méthylation d'ADN des gènes immunitaires (a, c) et hormonaux (b,d) selon : (a, b) toutes les patientes et (c, d) les groupes de sous-types moléculaires.

a/ Hypométhylation de CD274, NFKB1 et PDCD1 : les niveaux faibles de méthylation observés pour ces gènes, particulièrement l’hypométhylation importante de *CD274*, suggèrent une activation accrue de ces derniers. L'hypométhylation est généralement associée à une augmentation de l'expression génique, ce qui pourrait refléter une réponse immunitaire exacerbée ou une inflammation accrue. Le gène *CD274* (codant pour PD-L1), régulateur clé de l'inhibition immunitaire, pourrait indiquer une perturbation du contrôle de la réponse immunitaire, influençant ainsi la progression tumorale et la réponse aux traitements.

b/ Hyperméthylation de CTLA4 : une tendance à l'hyperméthylation de CTLA4 suggère une répression accrue de ce gène. L'hyperméthylation est souvent liée à une réduction de l'activité immunitaire, ce qui pourrait limiter la capacité du système immunitaire à cibler et détruire les cellules tumorales.

c/ Variabilité des niveaux de méthylation de TLR4 : la variabilité observée dans les niveaux de méthylation de TLR4, comprenant à la fois des régions hyperméthylées et hypométhylées, reflète une régulation complexe de ce gène. Ces fluctuations peuvent correspondre à des changements dynamiques dans l'expression de TLR4, essentiel pour la détection des pathogènes et l'activation des réponses immunitaires innées.

d/ Hypométhylation de TLR9 : la tendance à l'hypométhylation de TLR9 pourrait indiquer une activation accrue de ce récepteur, impliqué dans la reconnaissance des motifs moléculaires associés aux pathogènes.

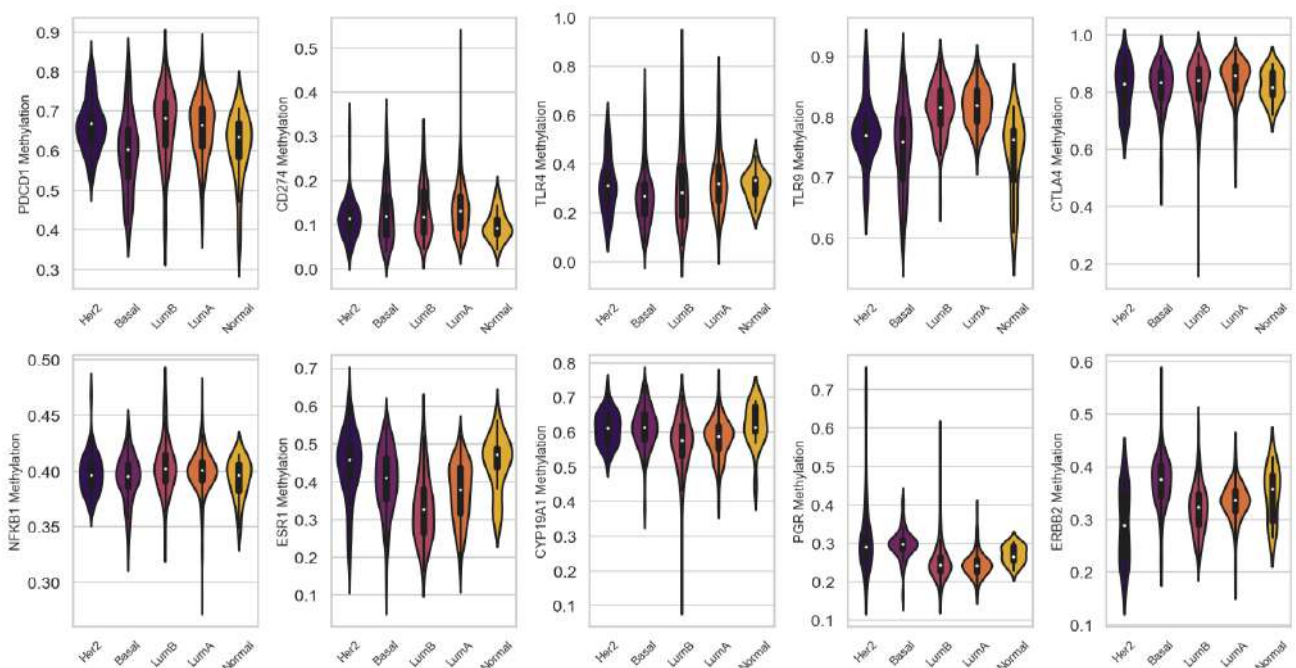


Figure 50. Distribution des gènes méthylés, immunitaires et hormonaux, selon les sous-types moléculaires PAM 50 du cancer du sein.

2.1.2. Gènes hormonaux

Les résultats de la distribution des gènes hormonaux méthylés (*PGR*, *ESR1*, *CYP19A1* et *ERBB2*) dans la population totale sont représentés dans la **Figure 49 (b,d) et 50**.

a/ **Hypométhylation de PGR, ERBB2 et ESR1** : l'hypométhylation observée de *PGR* (récepteur de la progestérone), de *ESR1* (récepteurs des œstrogènes) et *ERBB2* (gène codant HER2), suggère une activation accrue de ces gènes. Cette activation pourrait refléter une dépendance hormonale marquée dans certains sous-types tumoraux.

b/ **Hypométhylation de CYP19A1** : l'hypométhylation cohérente de *CYP19A1* à travers différents sous-types moléculaires souligne son rôle potentiel dans l'altération de la biosynthèse des œstrogènes. *CYP19A1* (aromatase) est une enzyme clé dans la conversion des androgènes en œstrogènes, ce qui peut influencer la croissance tumorale dans les cancers du sein dépendants des hormones.

- ❖ L'hypométhylation des gènes immunitaires tels que *CD274* et *TLR9* pourrait favoriser une réponse inflammatoire ou immunitaire accrue, tandis que l'hyperméthylation de *CTLA4* suggère une suppression immunitaire. Du côté hormonal, l'hypométhylation des gènes clés comme *ESR1* et *CYP19A1* souligne leur rôle central dans la croissance tumorale dépendante des œstrogènes.

2.2. Expression des gènes méthylés

Les résultats de la distribution de l'expression des gènes méthylés immunitaires (*CD274*, *NFKB1*, *PDCD1*, *CTLA4*, *TLR4* et *TLR9*) et hormonaux (*PGR*, *ESR1*, *CYP19A1* et *ERBB2*) sont représentés dans la **Figure 51 et 52**. Ces résultats montrent des variations significatives des niveaux d'expression des gènes immunitaires et hormonaux selon les sous-types moléculaires du cancer du sein. Ils mettent en évidence des profils moléculaires distincts des gènes immunitaires et hormonaux selon les sous-types (Basal-Like, Her2, LumA, LumB, et Normal-Like).

2.2.1. Expression des gènes immunitaires

Les résultats de la distribution de l'expression des gènes immunitaires méthylés (*CD274*, *NFKB1*, *PDCD1*, *CTLA4*, *TLR4* et *TLR9*) dans la population totale sont représentés dans la **Figure 51 a,c et 52**.

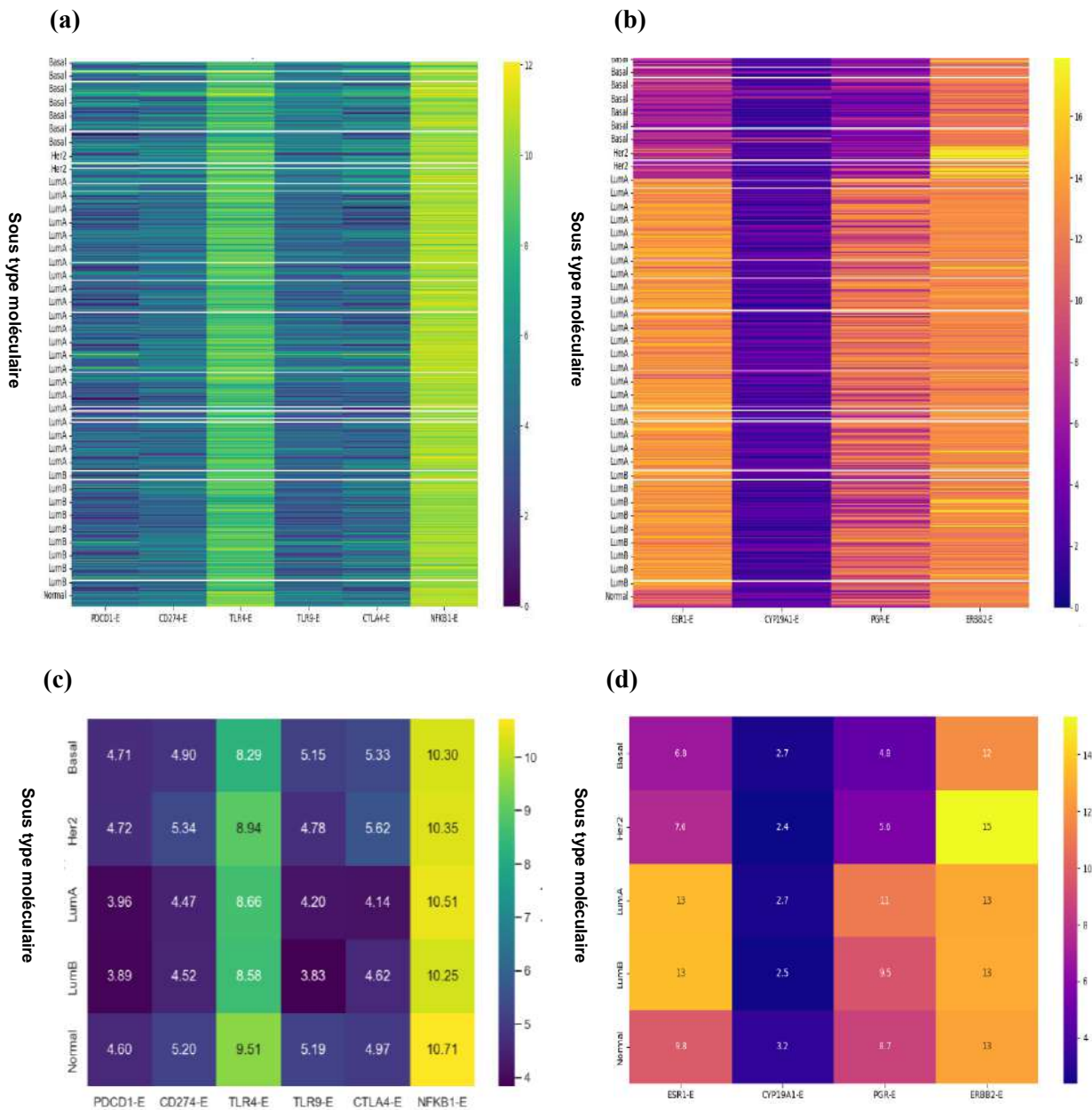


Figure 51. Heatmaps des profils d’expression des gènes immunitaires (a, c) et hormonaux (b,d) selon : (a, b) toutes les patientes et (c, d) les groupes de sous-types moléculaires.

a/ PDCD1-E (PD-1) et CD274-E (PD-L1) : l’expression élevée de CD274 (gène exprimant PD-L1) dans les sous-types HER2 et Basal-Like reflète une activation immunitaire accrue, tout en soulignant un rôle potentiel de l’évasion immunitaire, caractéristiques de sous-types plus agressifs. Cette expression de PD-L1 est souvent observée dans les cancers triples négatifs, qui utilisent cette voie pour échapper au système immunitaire. Les patientes pourraient répondre favorablement aux inhibiteurs de points de contrôle immunitaire (anti-PD-1/PD-L1).

b/ TLR4 et TLR9 : ces récepteurs sont impliqués dans l'inflammation et la progression tumorale, avec une expression variable selon les sous-types. L'expression renforcée de *TLR4* dans ces mêmes sous-types indique une réponse immunitaire innée active, tandis que *TLR9* montre une expression plus faible dans le sous type LumB, suggérant une moindre activation immunitaire.

c/ NFKB1 : l'expression relativement stable de *NFKB1* dans tous les sous-types met en lumière son rôle central dans l'inflammation. Son activation peut induire une résistance aux traitements et favoriser une inflammation chronique dans le microenvironnement tumoral.

d/ CTLA4 : son expression élevée dans les sous-types HER2 et Basal-Like reflète une inhibition de l'immunité anti-tumorale et indique une régulation immunosuppressive, rendant les cancers plus agressifs pouvant justifier l'usage d'inhibiteurs de CTLA-4 en immunothérapie. En revanche, le sous-type normal présente des niveaux d'expression modérés pour les gènes immunitaires, traduisant un environnement stable.

2.2.2. Expression des gènes hormonaux

Les résultats de la distribution de l'expression des gènes hormonaux méthylés (*PGR*, *ESR1*, *CYP19A1* et *ERBB2*) dans la population totale sont représentés dans la **Figure 51 (b,d) et 52**.

a/ ESR1 et PGR : leur expression est élevée dans les sous-types Luminal A et B, confirmant leur dépendance hormonale et leur sensibilité potentielle aux thérapies endocriniennes. Dans les cancers triples négatifs (TNBC) et certains HER2+, l'expression de ces gènes est faible expliquant leur résistance aux thérapies hormonales. Une discordance ESR1+/PGR- peut suggérer une résistance aux traitements hormonaux.

b/ ERBB2 : le sous-type HER2, en revanche, montre une expression accrue de *ERBB2*, valide comme cible thérapeutique clé. Le sous-type Basal-Like, marqué par une faible expression des gènes hormonaux, confirme son statut de cancer triple négatif et l'absence de dépendance hormonale.

c/ CYP19A1 (Aromatase) : ce gène est responsable de la conversion des androgènes en œstrogènes. Sa surexpression serait un marqueur de dépendance hormonale et pourrait indiquer une sensibilité aux inhibiteurs de l'aromatase. L'expression relativement homogène de *CYP19A1* dans tous les sous-types, avec un léger pic dans le sous-type Normal-Like, reflète une activité hormonale de base.

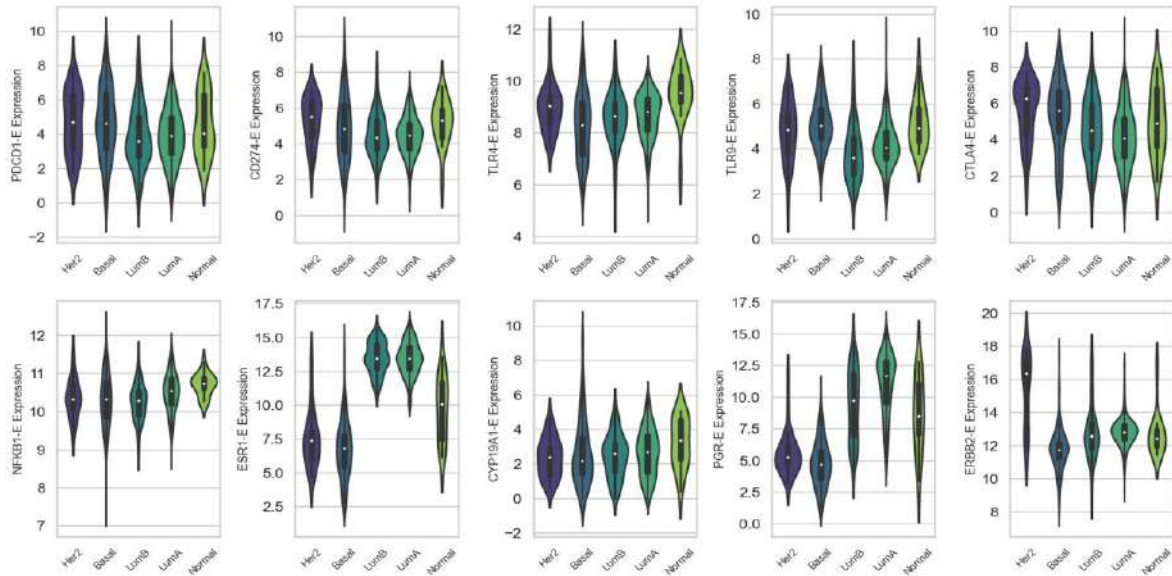


Figure 52. Distribution de l’expression des gènes méthylés, immunitaires et hormonaux, selon les sous-types moléculaires PAM 50 du cancer du sein.

- ❖ Ces différences d'expression et de méthylation révèlent des stratégies biologiques spécifiques d'évasion tumorale, de dépendance hormonale et d'activation immunitaire, qui ont des implications cliniques et thérapeutiques majeures.

3. METHODE DE CLUSTERING PAR INTELLIGENCE ARTIFICIELLE : K-MEANS ET METHODE AGGLOMERATIVE

Deux méthodes de clustering, K-Means et Clustering Hiérarchique Agglomératif, ont été appliquées pour segmenter les patientes atteintes de cancer du sein en deux clusters. Ces analyses se sont basées sur des données cliniques et moléculaires, incluant la méthylation de l'ADN, l'expression génique et des paramètres cliniques à savoir, l'âge au diagnostic, la survie, le statut pathologique et le statut ménopausique. Les résultats sont illustrés dans la **Figure 53**.

Les résultats obtenus ont montré les silhouette score suivants :

* Meilleur K-Means avec k=2: Silhouette Score = **0.397**

* Meilleur Agglomerative Clustering avec k=2: Silhouette Score = **0.402**

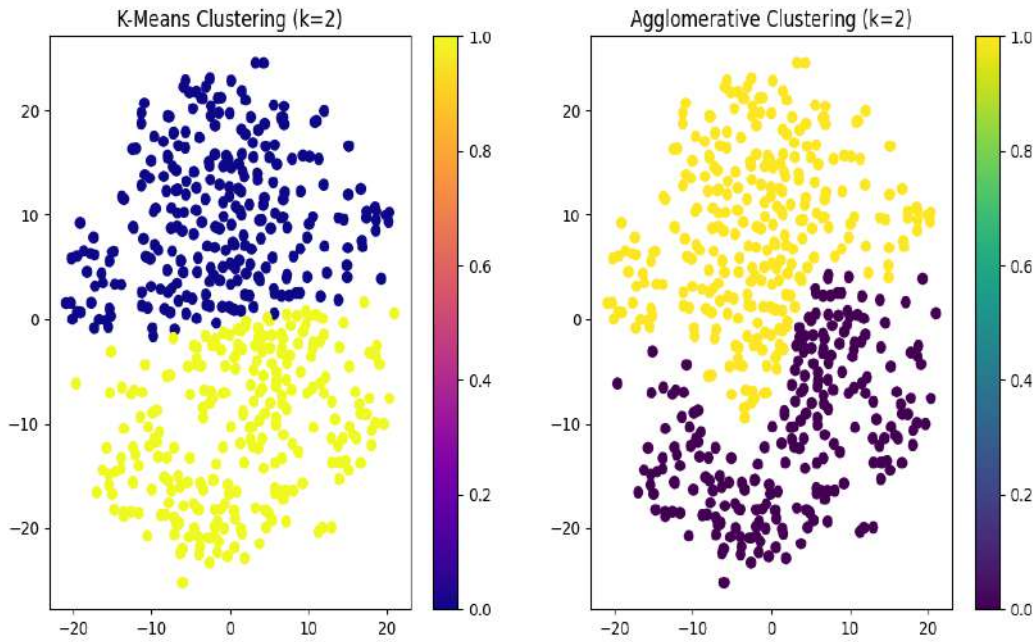


Figure 53. Segmentation des patientes en deux clusters basées sur la méthylation de l'ADN, l'expression génique des biomarqueurs immunitaires et hormonaux et les caractéristiques clinico-pathologiques.

Les deux méthodes ont identifié un nombre optimal de deux clusters, reflétant des profils cliniques et biologiques distincts. Ces clusters mettent en évidence la complexité des données cliniques et moléculaires des patientes, tout en offrant des pistes pour une stratification permettant de prendre les décisions thérapeutiques appropriées.

3.1. Clustering par K-Means

L'analyse par K-Means a distingué deux clusters principaux (**Tableau 9**) :

1/ Cluster 0 :

- **Caractéristiques cliniques** : Âge moyen plus élevé (61,58 ans), survie légèrement inférieure (1711 jours), tumeurs plus volumineuses ($T = 2,38$) et envahissement ganglionnaire plus marqué ($N = 1,65$).
- **Profil moléculaire** : Méthylation élevée de *PDCD1*, *CD274*, *TLR4*, *CTLA4*. Forte expression des gènes hormonaux *ESR1*, *CYP19A1*, *PGR*, *ERBB2*, indiquant une dépendance hormonale et une sensibilité aux thérapies endocriniennes.

2/ Cluster 1 :

- **Caractéristiques cliniques** : Âge moyen plus jeune (53,39 ans), survie légèrement meilleure (1756 jours), tumeurs moins invasives (T = 2,15, N = 1,53).
- **Profil moléculaire** : Méthylation et expression réduites des gènes *PDCDI*, *CD274*, *ESR1*, *CYP19A1*, indiquant un profil tumoral plus agressif et une moindre dépendance hormonale.

Tableau 9. Clustering par méthode K-Means.

	Cluster 0	Cluster 1
Âge au Diagnostic Initial	61,58 ans	53,39 ans
Survie	1711 jours.	1756 jours
Paramètres Pathologiques :		
Taille de la tumeur (T)	2,38	2,15
Envahissement ganglionnaire (N)	1,65	1,53
Stade pathologique global	2,56	2,30
Méthylation / Expression	- Méthylation élevée de <i>PDCDI</i> , <i>CD274</i> , <i>TLR4</i> , <i>CTLA4</i> . - Expression élevée de <i>ESR1</i> , <i>CYP19A1</i> , <i>PGR</i> , <i>ERBB2</i> .	- Méthylation réduites de <i>PDCDI</i> , <i>CD274</i> , <i>ESR1</i> , <i>CYP19A1</i> . - Expression modérée de <i>CYP19A1</i> -E et <i>ESR1</i> -E.
	Niveau élevé de méthylation, suggérant un contrôle épigénétique accru.	Méthylation plus faible, permettant des comportements tumoraux plus agressifs.

3.2. Clustering Hiérarchique Agglomératif

L'analyse hiérarchique a identifié également deux clusters distincts (**Tableau 10**) :

1/ Cluster 0 : regroupe des patientes plus jeunes (âge moyen : 52,05 ans), présentant une survie légèrement meilleure (1767 jours) et des tumeurs de taille plus petite (T = 2,15) avec un envahissement ganglionnaire réduit (N = 1,53). Sur le plan moléculaire, ce cluster se distingue par une méthylation élevée de *PDCDI*, *CD274*, *TLR4*, *CTLA4* et une forte expression des gènes hormonaux (*ESR1*, *CYP19A1*, *PGR*, *ERBB2*), indiquant une dépendance hormonale marquée et une sensibilité potentielle aux thérapies endocriniennes.

2/ **Cluster 1** : regroupe des patientes plus âgées (âge moyen : 62,03 ans), avec une survie légèrement inférieure (1706 jours), des tumeurs plus volumineuses (T = 2,38) et un envahissement ganglionnaire plus important (N = 1,65). Ce cluster est caractérisé par une méthylation et une expression réduite des gènes *PDCD1*, *CD274*, *TLR4*, traduisant une régulation génétique accrue, une évasion immunitaire prononcée, et une dépendance hormonale limitée, indiquant un profil tumoral plus agressif.

Tableau 10. Clustering Hiérarchique Agglomératif.

	Cluster 0	Cluster 1
Âge au Diagnostic Initial	52,05 ans.	62,03 ans
Survie (days_to_death)	1776 jours.	1706 jours
Paramètres Pathologiques :		
Taille de la tumeur (T)	2,15	2,38
Envahissement ganglionnaire (N)	1,53	1,65
Stade pathologique global	2,30	2,56
Méthylation/Expression	- Méthylation élevée de <i>PDCD1</i> , <i>CD274</i> , <i>TLR4</i> , <i>CTLA4</i> . - Expression élevée de <i>ESR1</i> , <i>CYP19A1</i> , <i>PGR</i> , <i>ERBB2</i> .	- Méthylation réduites de <i>PDCD1</i> , <i>CD274</i> , <i>TLR4</i> . - Expression modérée de <i>ESR1</i> -E et <i>CYP19A1</i> -E.
	Niveau élevé de méthylation, indiquant une évasion immunitaire prononcée.	Méthylation plus élevée, associée à une régulation génétique accrue et un profil tumoral agressif

Cette méthode de clustering permet une segmentation prometteuse des patientes atteintes de cancer du sein, en distinguant des groupes avec des caractéristiques cliniques et moléculaires différenciées. Le Cluster 0 semble associé à des patientes présentant un profil moins agressif mais dépendant des hormones, tandis que le Cluster 1 regroupe des patientes avec des profils plus agressifs et une moindre dépendance hormonale.

3.3. Similitudes et différences entre le Cluster 0 et le Cluster 1

Les cercles montrent les gènes et facteurs cliniques communs et spécifiques entre les deux clusters du cancer du sein (**Figure 54**). Les gènes *PDCDI* et *CD274* sont partagés entre les deux clusters, mais ils présentent des différences significatives en termes de méthylation et d'expression. *PDCDI* et *CD274* sont méthylés de manière accrue dans le Cluster 0, tandis que dans le Cluster 1, ils montrent une méthylation réduite. En termes d'expression, ces gènes sont exprimés de manière plus marquée dans le Cluster 0 par rapport au Cluster 1.

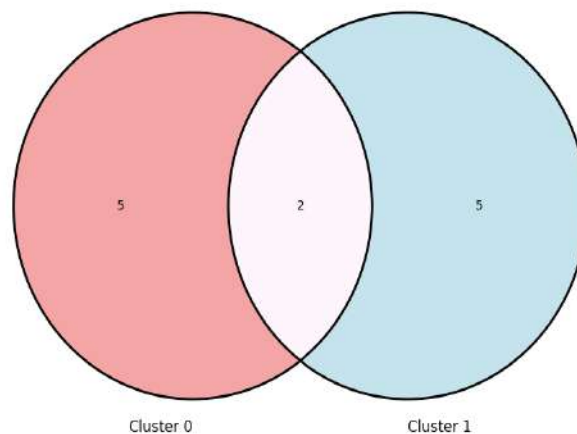


Figure 54. Diagramme de Venn représentant les similitudes et les différences entre le Cluster 0 et le Cluster 1.

Les facteurs cliniques communs aux deux clusters incluent le stade pathologique, l'âge au diagnostic, et le temps avant décès. Le Cluster 0 se distingue par une méthylation accrue de *PDCDI*, *CD274*, et *TLR4*, ainsi qu'une expression accrue de *ESR1* et *CTLA4*. Le Cluster 1, quant à lui, présente une expression accrue de *NFKB1* et *CYP19A1*, et une méthylation réduite de *PDCDI*, *CD274*, et *TLR9*.

3.4. Justification du choix de l'algorithme de clustering K-means

Dans le cadre de cette étude, l'algorithme K-means a été retenu pour réaliser la classification non supervisée des patientes atteintes de cancer du sein, en alternative aux méthodes hiérarchiques telles que le clustering agglomératif. Ce choix repose sur plusieurs considérations méthodologiques :

* le K-means est particulièrement adapté à des jeux de données de grande dimension, comme ceux issus de plateformes multi-omiques (expression génique, méthylation, paramètres cliniques), car il offre une meilleure efficacité computationnelle que les méthodes hiérarchiques, dont la complexité algorithmique est plus élevée.

* l'algorithme K-means génère des clusters distincts, centrés sur des moyennes, ce qui facilite l'analyse comparative post-clustering. Cette propriété a été exploitée dans ce travail pour dégager des tendances différentielles entre les deux groupes de patientes, notamment en termes d'âge au diagnostic, de stade tumoral, de niveaux de méthylation et d'expression des gènes ciblés.

* K-means s'intègre aisément dans une approche exploratoire multi-omique, où les variables sont préalablement standardisées afin de permettre une contribution équitable de chaque type de donnée à la formation des clusters.

- ❖ Ainsi, le recours à l'algorithme K-means s'est imposé comme un choix judicieux, tant pour sa simplicité d'implémentation que pour sa pertinence dans l'analyse des données hétérogènes issues du cancer du sein.

4. DISTRIBUTION DES PARAMETRES DEMOGRAPHIQUES ET CLINIQUES PAR CLUSTERS (K-MEANS)

4.1. Distribution de l'âge en fonction des clusters

La répartition en fonction de l'âge au diagnostic initial pour les deux clusters 0 et 1 est représentée par la **Figure 55**. Le Cluster 0 présente un âge moyen légèrement plus élevé (58,59 ans), avec un pic autour de 55-65 ans. La distribution est légèrement étalée vers les âges plus avancés, reflétant une proportion plus importante de patientes âgées dans ce groupe.

En revanche, le Cluster 1, avec un âge moyen de 56,35 ans, est principalement composé de patientes légèrement plus jeunes, mais sans un pic marqué autour de 40-45 ans comme mentionné précédemment. Les distributions des deux clusters restent globalement similaires, bien que le Cluster 0 ait une représentation légèrement plus importante des tranches d'âge avancé.

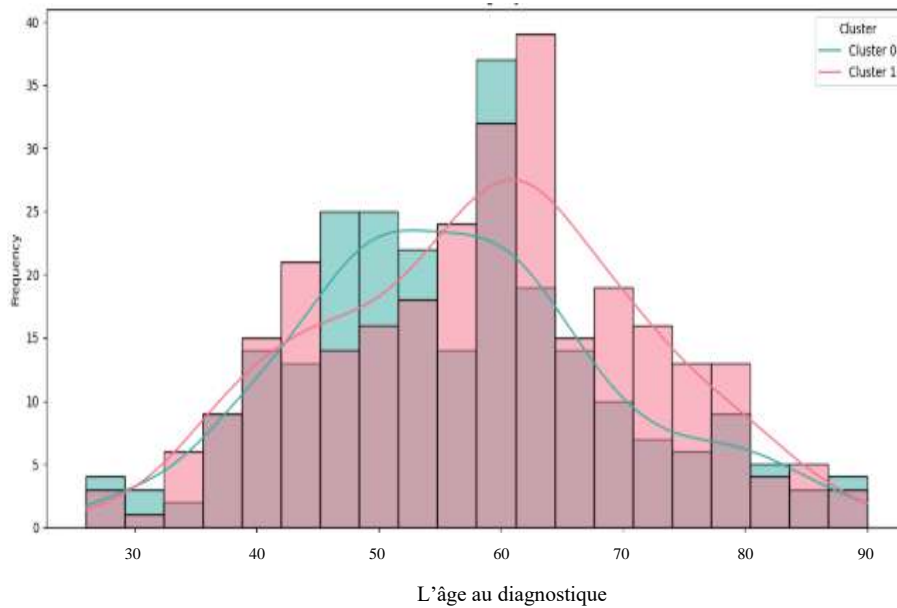


Figure 55. Distribution de l'âge par cluster (K-Means).

4.2. Distribution des stades TNM en fonction des clusters

La Figure 56 illustre la répartition des stades T, N et M en fonction des clusters 0 et 1.

1/ Répartition du stade pathologique T :

- **Cluster 0** : la majorité des patientes se trouvent dans la catégorie T2, représentant des tumeurs de taille intermédiaire. Il existe plus de patientes T1/T2 dans ce Cluster, suggérant un groupe avec des tumeurs plus petites et localisées. La distribution des patientes montre une légère décroissance vers les tailles de tumeurs plus grandes (T3) et plus petites (T1).
- **Cluster 1** : comprend moins de patientes T3/T4 par rapport au cluster 0, indiquant un groupe associé à des tumeurs plus volumineuses ou invasives. Toutefois, la distribution des patientes suit une tendance similaire au cluster 0 avec un pic également au stade T2. La fréquence des tumeurs de plus petite taille (T1) est légèrement plus marquée comparée au Cluster 0.

Différences entre clusters :

Le Cluster 1 semble lié à des stades T avancés, ce qui pourrait refléter un risque accru de progression ou de métastases. Le Cluster 0 pourrait correspondre à des tumeurs de meilleur pronostic, nécessitant potentiellement des traitements moins agressifs.

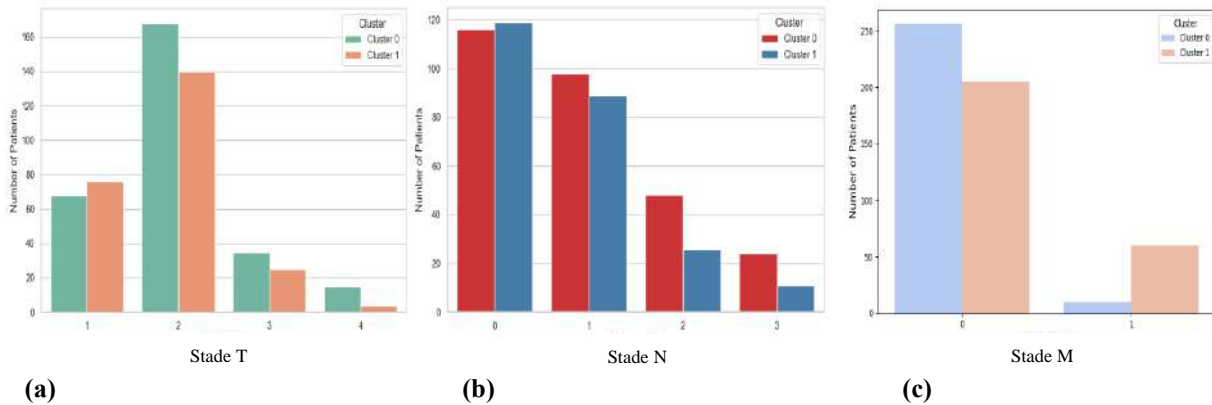


Figure 56. Distribution des patientes selon les clusters (K-Means) des stades T (a), N (b) et M (c).

2/ Répartition du stade pathologique N :

- **Cluster 0** : la majorité des patientes du cluster 0 présente le stade N0 (absence d'envahissement ganglionnaire). Ainsi, on remarque une nette dominance des stades N0/N1, suggérant un groupe à faible charge tumorale ganglionnaire. Les catégories des stades N2 et N3 sont moins représentées.
- **Cluster 1** : le cluster 1 présente une répartition similaire, avec une concentration importante du stade N0. Toutefois, la proportion du stade N1 est légèrement plus élevée que dans le Cluster 0, tandis que les stades N2 et N3 demeurent rares, mais associés à une maladie ganglionnaire étendue.

Différences entre clusters :

Le Cluster 1 pourrait identifier des patientes à haut risque de rechute, nécessitant un traitement adjuvant intensifié (chimiothérapie ou radiothérapie). Le Cluster 0 pourrait correspondre à des cas avec un pronostic plus favorable, où des thérapies ciblées ou une surveillance accrue seraient suffisantes.

3/ Répartition du stade pathologique M :

- **Cluster 0** : la grande majorité des patientes dans les deux clusters présente le stade M0, ce qui signifie l'absence de métastases à distance. Cela est visible par la haute fréquence de cette valeur.
- **Cluster 1** : on observe quelques patientes présentant le stade M1 (présence de métastases), mais en faible proportion.

Différences entre clusters :

La distribution montre que le stade M est un facteur discriminant entre les deux clusters. Le cluster 0 est principalement composé de patientes sans métastases, tandis que Cluster 1 regroupe une proportion significative de patientes avec métastases, ce qui pourrait indiquer un profil clinique ou pronostique différent entre les deux groupes.

4.3. Distribution du stade pathologique par cluster

La **Figure 57** illustre la répartition des stades pathologiques en fonction des clusters 0 et 1.

- **Cluster 0** : le nombre de patientes est plus élevé dans le stade II ; il diminue progressivement pour le stade III et de manière sévère pour le stade IV. Ce cluster pourrait représenter un groupe de patientes avec une progression moins agressive de la maladie (stade I et II), où la majorité des cas sont diagnostiqués à des stades précoces. Cela suggère que le Cluster 0 pourrait être associé à des tumeurs de meilleur pronostic avec une réponse favorable aux traitements.
- **Cluster 1** : le nombre de patientes est faible dans le stade I et augmente significativement pour le stade II. Le nombre de patientes diminue progressivement au stade avancé III, mais de façon prononcée au stade IV. Ce cluster domine aux stades II et III, cela pourrait indiquer qu'il capture des profils tumoraux plus avancés ou à haut risque.

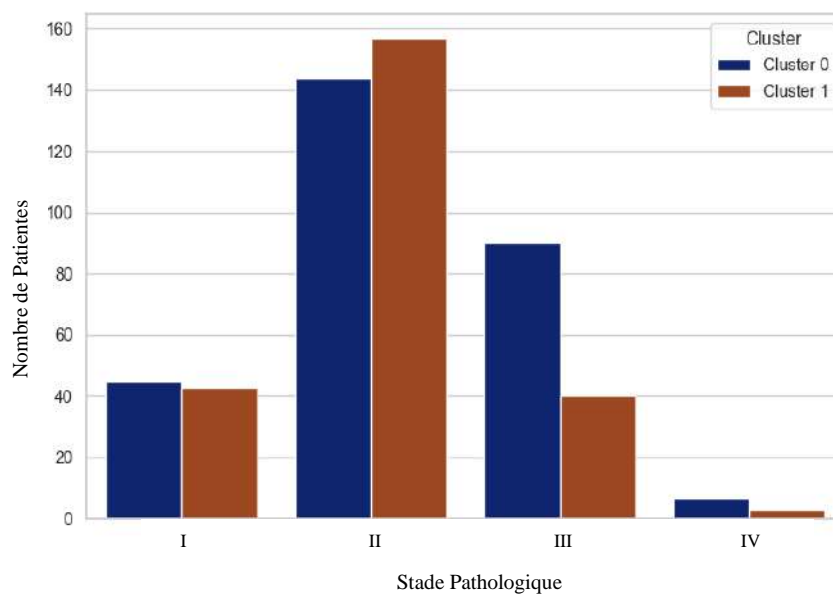


Figure 57. Distribution du stade pathologique par cluster (K-Means) chez les patientes.

- ❖ Il y a peu de patientes avec un stade pathologique avancé (stade IV), ce qui est cohérent avec la répartition générale des cancers du sein où les stades avancés sont moins fréquents au moment du diagnostic.

Il semble que le cluster 0 ait une proportion assez élevée de stades III, suggérant probablement une association avec une forme plus avancée du cancer.

4.4. Distribution des sous-types moléculaires PAM 50 par cluster

La **Figure 58** illustre la répartition des patientes selon les sous types moléculaires PAM 50 et en fonction des clusters 0 et 1. Cette démarche permettra de comprendre si les clusters K-Means reflètent des différences biologiques significatives entre les sous-types PAM 50, ce qui pourrait avoir des implications pour la classification des tumeurs ou les stratégies thérapeutiques.

- **Cluster 0** : présente un profil hétérogène, dominé par LumA (36 %), avec une proportion importante de Basal-Like (31 %) et HER2 (11 %). Ce cluster est caractérisé par des tumeurs hormono-dépendantes (LumA/LumB) et agressives (Basal-Like/HER2).
- **Cluster 1** : ce cluster est dominé par les sous types LumA (61.6 %) et LumB (28.3 %), soit plus de 89 % du sous-type Luminal. Les sous types Basal-like (6.7 %) et HER2 (2.4 %) sont présents à très faible proportion.

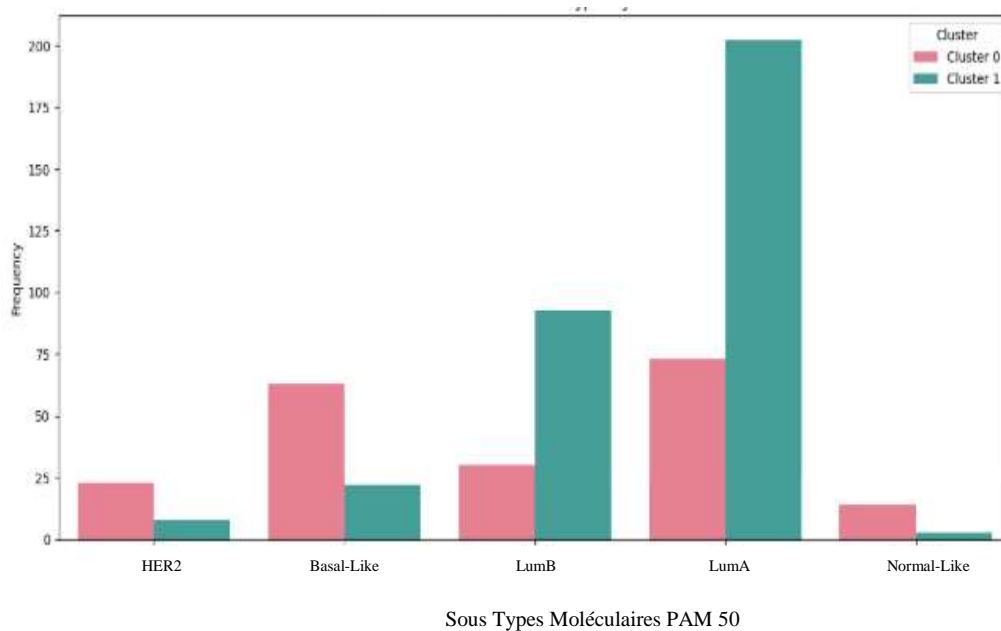


Figure 58. Distribution des sous-types moléculaires PAM 50 par cluster (K-Means).

4.5. Distribution du statut ménopausique par cluster

La **Figure 59** montre la répartition des patientes selon statut ménopausique et en fonction des clusters 0 et 1.

- **Cluster 0** : présente une prédominance de patientes en post-ménopause (n = 195), suivi par les patientes en pré-ménopause (n = 73). Ce Cluster semble associé à des caractéristiques cliniques plus favorables, suggérant des tumeurs hormono-dépendantes et de meilleur pronostic.
- **Cluster 1** : la majorité des patientes sont également en post-ménopause (n = 147), mais le nombre de patientes en pré-ménopause est plus faible (n = 58).

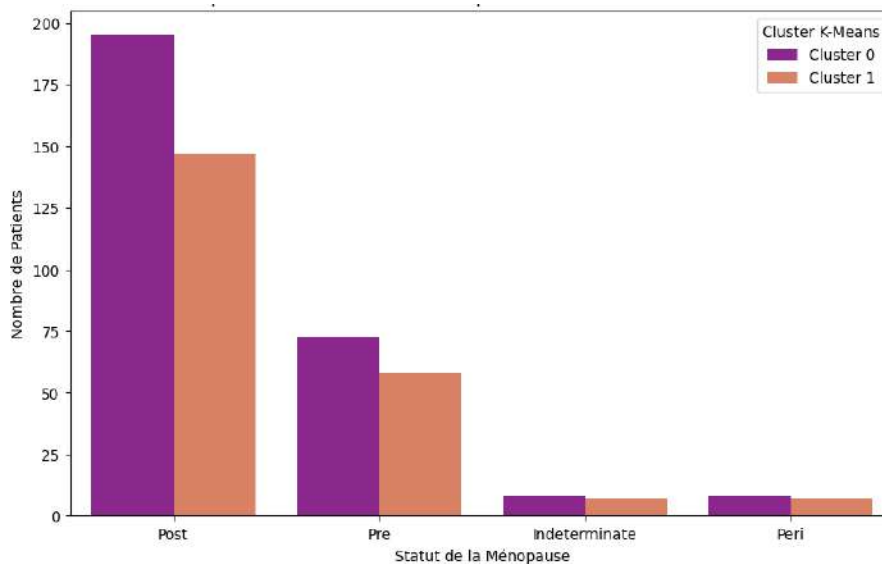


Figure 59. Distribution du statut ménopausique par Clusters (K-Means).

4.6. Distribution de la survie par cluster

La courbe de Kaplan-Meier ci-dessous (**Figure 60**) illustre la probabilité de survie globale (Overall Survival : OS) des patientes atteintes de cancer du sein, réparties en deux groupes distincts issus d'une analyse de clustering. Chaque courbe représente l'évolution de la survie au fil du temps (en jours), accompagnée d'un intervalle de confiance à 95 % (zones ombrées).

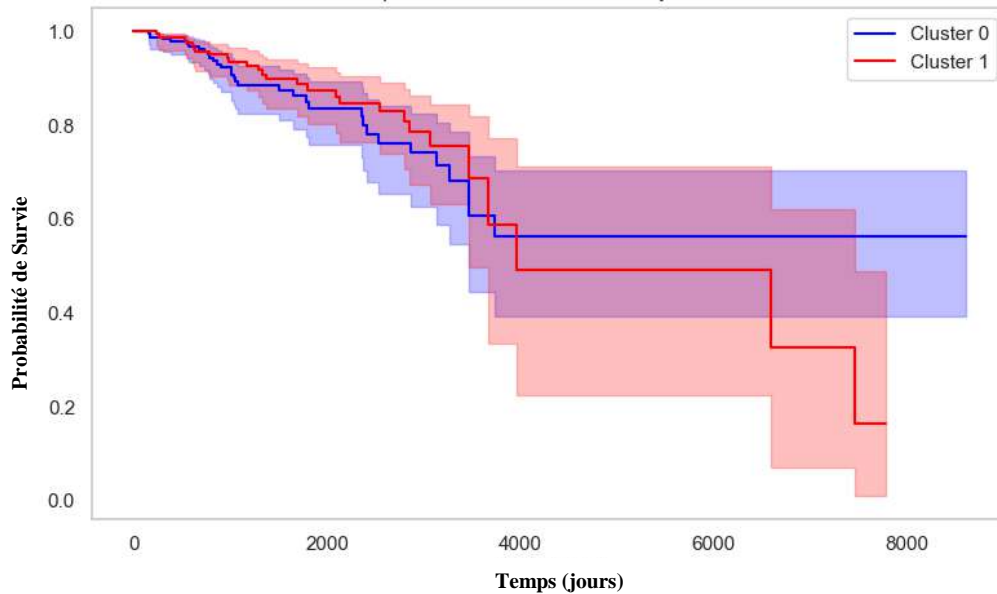


Figure 60. Représentation de la survie globale par Kaplan Meyer en fonction des clusters.

L'observation montre que le Cluster 0 présente une meilleure probabilité de survie à long terme comparativement au Cluster 1. En effet, la courbe du Cluster 0 demeure plus élevée que celle du Cluster 1 au-delà de 4000 jours, traduisant un pronostic plus favorable pour ces patientes. À l'inverse, les patientes du Cluster 1 présentent une chute plus marquée de la survie avec le temps, atteignant une probabilité de survie inférieure à 20 % après 8000 jours, ce qui suggère une forme plus agressive de la maladie ou une moins bonne réponse au traitement.

La divergence entre les courbes témoigne ainsi de différences cliniques ou biologiques significatives entre les deux groupes. Cette hétérogénéité pourrait être attribuée à des facteurs comme l'âge au diagnostic, le stade tumoral, l'état hormonal ou encore l'expression et la méthylation différentielle de certains biomarqueurs immunitaires et hormonaux, comme ceux analysés dans notre étude.

5. DISTRIBUTION DES BIOMARQUEURS IMMUNITAIRES ET HORMONAUX PAR CLUSTER

Les **Figure 61 - 64** ci-dessous montrent les distributions des différentes signatures épigénétiques ainsi que leurs expressions entre les deux clusters. On y observe des profils distincts selon les gènes, avec des variations de la moyenne, de la dispersion et des asymétries dans les distributions, illustrant une hétérogénéité fonctionnelle entre les groupes tumoraux.

5.1. Analyse des gènes méthylés

5.1.1. Gènes Immunitaires

L'analyse conjointe des niveaux de méthylation dans les deux clusters (**Figure 61**) révèle des signatures épigénétiques contrastées, reflétant des états immunologiques tumoraux différenciés.

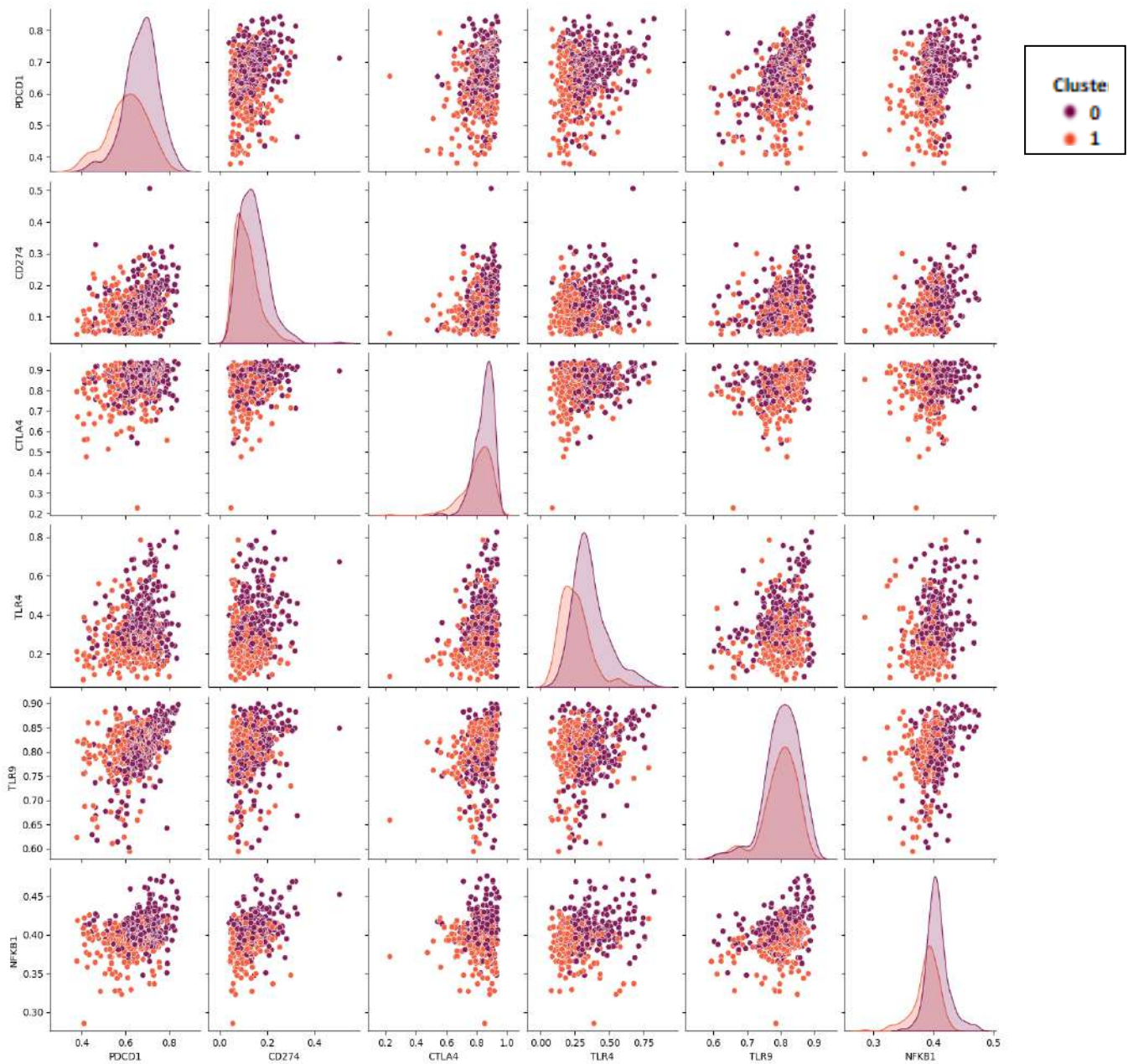


Figure 61. Distribution des gènes immunitaire méthylés selon les deux clusters.

a/ PDCDI

Le Cluster 0 présente un niveau moyen de méthylation de *PDCDI* plus élevé ($0,673 \pm 0,076$) que le Cluster 1 ($0,608 \pm 0,091$). Cette hyperméthylation, relativement homogène dans le Cluster 0, suggère une répression transcriptionnelle stable de ce gène, impliqué dans l'épuisement des lymphocytes T. En revanche, le Cluster 1 montre une méthylation plus faible mais plus variable, traduisant une hétérogénéité inter-tumorale. Cela pourrait refléter la coexistence de sous-populations tumorales aux potentiels immunologiques distincts, ouvrant des perspectives pour des traitements personnalisés ciblant la voie PD-1.

b/ CD274

De même, *CD274* est plus méthylé dans le Cluster 0 ($0,142 \pm \text{NC}$) que dans le Cluster 1 ($0,113 \pm \text{NC}$), ce qui indique une co-régulation épigénétique potentielle avec *PDCDI*. Cette configuration dans le Cluster 0 est compatible avec une suppression coordonnée de l'axe PD-1/PD-L1, suggérant un microenvironnement plus immunosuppresseur et possiblement moins réactif aux immunothérapies. À l'inverse, la moindre méthylation de *CD274* dans le Cluster 1 pourrait permettre une meilleure accessibilité transcriptionnelle et rendre ces tumeurs plus sensibles aux traitements immunomodulateurs.

c/ TLR9

Les deux clusters présentent des valeurs proches aux percentiles élevés, mais globalement, le Cluster 0 montre une méthylation légèrement plus marquée de *TLR9*. Cette régulation pourrait inhiber la détection de l'ADN tumoral par les cellules immunitaires innées, limitant l'activation des réponses immunitaires dans ce groupe. Le Cluster 1, avec une méthylation plus faible, conserverait un meilleur potentiel de signalisation innée via *TLR9*, contribuant à une vigilance immunitaire accrue.

d/ NFKB1

La méthylation de *NFKB1* est légèrement plus élevée dans le Cluster 0 ($0,406 \pm 0,019$) comparé au Cluster 1 ($0,389 \pm 0,021$). Cette différence, bien que modeste, peut refléter une réduction de l'activité transcriptionnelle de cette voie clé de l'inflammation et de la réponse immunitaire. Ainsi, le Cluster 0 pourrait être associé à un état inflammatoire affaibli, tandis que le Cluster 1 conserverait une plus grande capacité d'activation de la voie NF- κ B, en soutien à la réponse antitumorale.

➤ Ces résultats mettent en lumière des profils épigénétiques divergents entre les deux clusters. Le Cluster 0 présente une hyperméthylation coordonnée de *PDCDI* et *CD274*, une modulation inhibitrice de *TLR9* et *NFKB1*, traduisant un état d'échappement immunitaire stable et un microenvironnement tumoral plus silencieux sur le plan immunologique. Ce profil suggère une résistance potentielle aux immunothérapies classiques et un besoin de stratégies combinées ou alternatives.

En revanche, le Cluster 1 se distingue par des niveaux de méthylation plus faibles et plus hétérogènes, surtout pour *PDCDI*, ce qui pourrait refléter une plus grande plasticité tumorale et une réactivité immunitaire mieux conservée. Cette configuration ouvre la voie à des approches thérapeutiques plus ciblées, notamment par les inhibiteurs des points de contrôle (*checkpoints*) immunitaires.

5.1.2. Gènes hormonaux

L'évaluation des profils de méthylation entre les deux clusters met en évidence des divergences épigénétiques nettes concernant les gènes liés à la signalisation hormonale, soulignant des contextes biologiques distincts (**Figure 62**).

a/ ESR1

Le niveau moyen de méthylation d'*ESR1* est significativement plus élevé dans le Cluster 0 ($0,410 \pm 0,076$) que dans le Cluster 1 ($0,317 \pm 0,073$), les deux groupes présentant une dispersion comparable. Cette hyperméthylation pourrait indiquer une répression transcriptionnelle d'*ESR1* paradoxale, compte tenu du contexte Luminal souvent associé à une expression élevée. Elle suggère une régulation compensatoire ou instable, typique de certains profils Luminal B. En revanche, dans le Cluster 1, l'hypométhylation relative suggère une moindre contrainte épigénétique sur *ESR1*, bien que cela n'implique pas nécessairement une activation fonctionnelle, en particulier si d'autres voies régulatrices sont altérées.

b/ PGR

Le gène *PGR* suit une tendance similaire : le Cluster 0 présente une méthylation plus marquée, 0,276 contre 0,235 dans le Cluster 1. Cette concordance avec *ESR1* confirme une régulation épigénétique concertée des voies hormonales. Le profil du Cluster 0 semble refléter une hormono-dépendance modulée par l'épigénétique, potentiellement exploitable par des thérapies endocriniennes. À l'inverse, la méthylation plus faible dans le Cluster 1 pourrait signaler soit

une expression maintenue mais dysfonctionnelle, soit une perte de régulation hormonale au sein d'un contexte tumoral plus agressif.

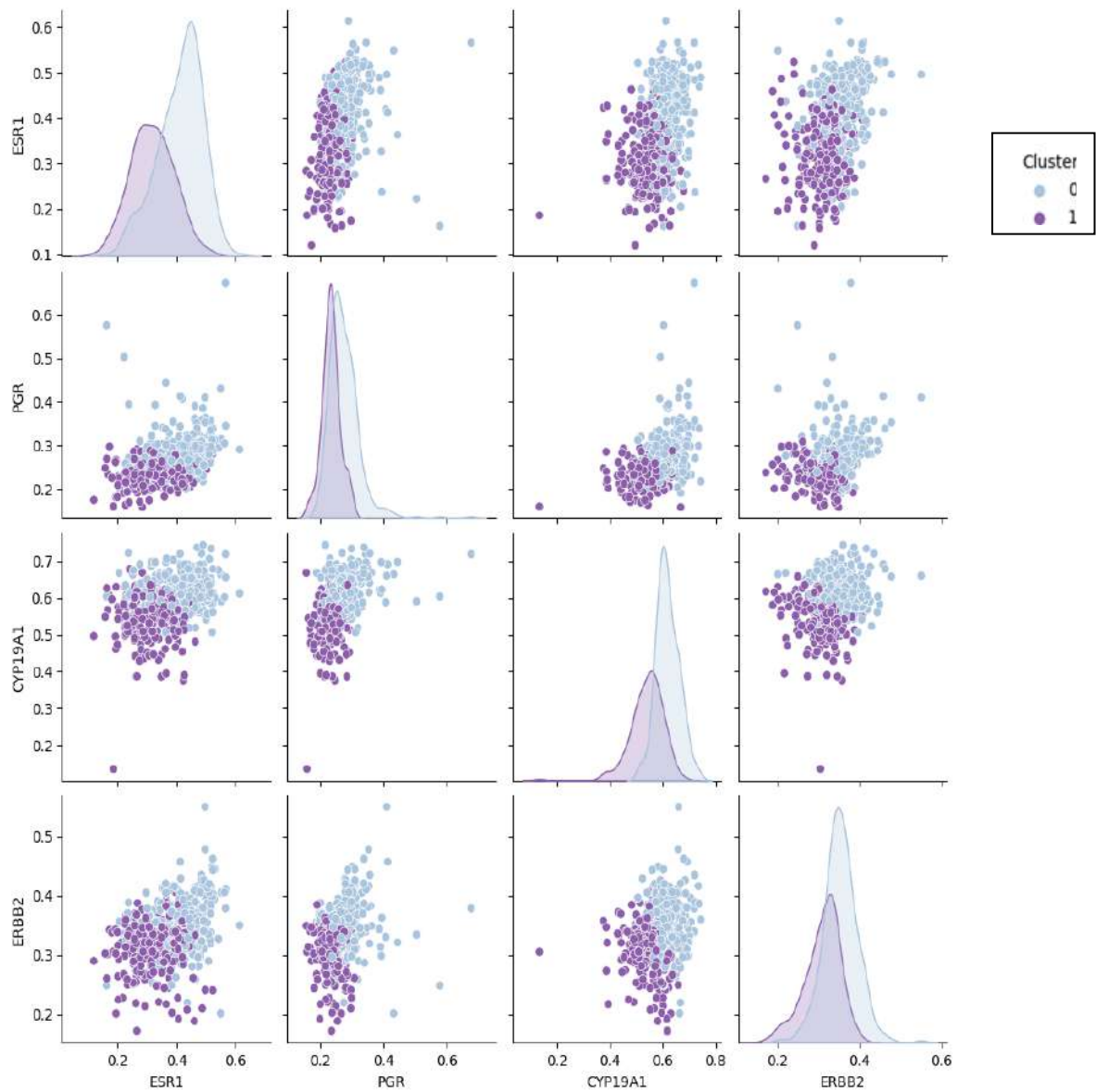


Figure 62. Distribution des gènes hormonaux méthylés selon les deux clusters.

c/ CYP19A1

Le percentile supérieur de méthylation de *CYP19A1* est plus élevé dans le Cluster 0 (0,647 contre 0,580 dans le Cluster 1), témoignant d'une modulation accrue de ce gène impliqué dans la biosynthèse des œstrogènes. Cette régulation pourrait favoriser un microenvironnement riche en œstrogènes, compatible avec une forte dépendance hormonale. À l'opposé, dans le Cluster 1,

la méthylation moindre pourrait refléter une dérégulation du métabolisme des œstrogènes, propice à l'émergence de phénotypes tumoraux moins sensibles aux inhibiteurs de l'aromatase.

d/ ERBB2

Enfin, le gène *ERBB2* présente également un niveau de méthylation plus élevé dans le Cluster 0 ($0,352 \pm 0,041$) comparé au Cluster 1 ($0,309 \pm 0,043$). Bien que ces valeurs restent inférieures au seuil habituellement associé à une répression transcriptionnelle franche, cette différence suggère une régulation modérée du locus *HER2*. Cela pourrait correspondre à un statut HER2 intermédiaire, sans surexpression, souvent observé dans les sous-types luminal B. Dans le Cluster 1, la méthylation est plus faible, mais l'amplitude des variations reste modeste, ne permettant pas de conclure quant à une surexpression effective du récepteur.

- Le **Cluster 0** présente une signature Luminale affirmée, avec une régulation épigénétique active des gènes *ESR1*, *PGR* et *CYP19A1*, traduisant une forte hormono-dépendance. Ce profil évoque un sous-type Luminal A si l'activité proliférative est faible, ou Luminal B en cas de prolifération accrue. La méthylation modérée d'*ERBB2*, sans franchissement du seuil de répression, appuie l'absence de surexpression *HER2*, cohérente avec ces sous-types.

Le **Cluster 1**, quant à lui, montre une atténuation des marqueurs hormonaux, évoquant soit une perte de différenciation Luminale, soit une transition vers un phénotype plus agressif et moins dépendant des voies endocriniennes. Ce profil pourrait correspondre à des tumeurs Luminal-Like atypiques ou en transition vers un phénotype basal-Like, avec une hétérogénéité tumorale plus marquée et un potentiel moindre de réponse aux traitements hormonaux classiques.

5.2. Analyse de l'expression des gènes méthylés

5.2.1. Gènes Immunitaires

L'analyse conjointe des niveaux de l'expression des gènes immunitaires méthylés dans les deux clusters est représentée dans la **Figure 63**.

a. PDCDI-E

L'expression de *PDCDI-E* est significativement plus élevée dans le Cluster 0 ($5,54 \pm 1,37$) que dans le Cluster 1 ($3,13 \pm 1,17$). Cette surexpression suggère un microenvironnement immunosuppresseur dans le Cluster 0, en cohérence avec le rôle de *PD-1* dans l'épuisement des lymphocytes T, contribuant potentiellement à une résistance à la surveillance immunitaire.

b/ CD274-E

Une tendance similaire est observée pour *CD274-E*, dont l'expression dans le Cluster 0 ($5,48 \pm$ NC) est supérieure à celle du Cluster 1 ($4,02 \pm$ NC). L'expression concomitante de *PDCD1-E* et *CD274-E* soutient l'hypothèse d'un axe *PD-1/PD-L1* actif, typique d'un environnement tumoral permissif à l'évasion immunitaire, et possiblement moins réactif aux immunothérapies ciblant ce mécanisme.

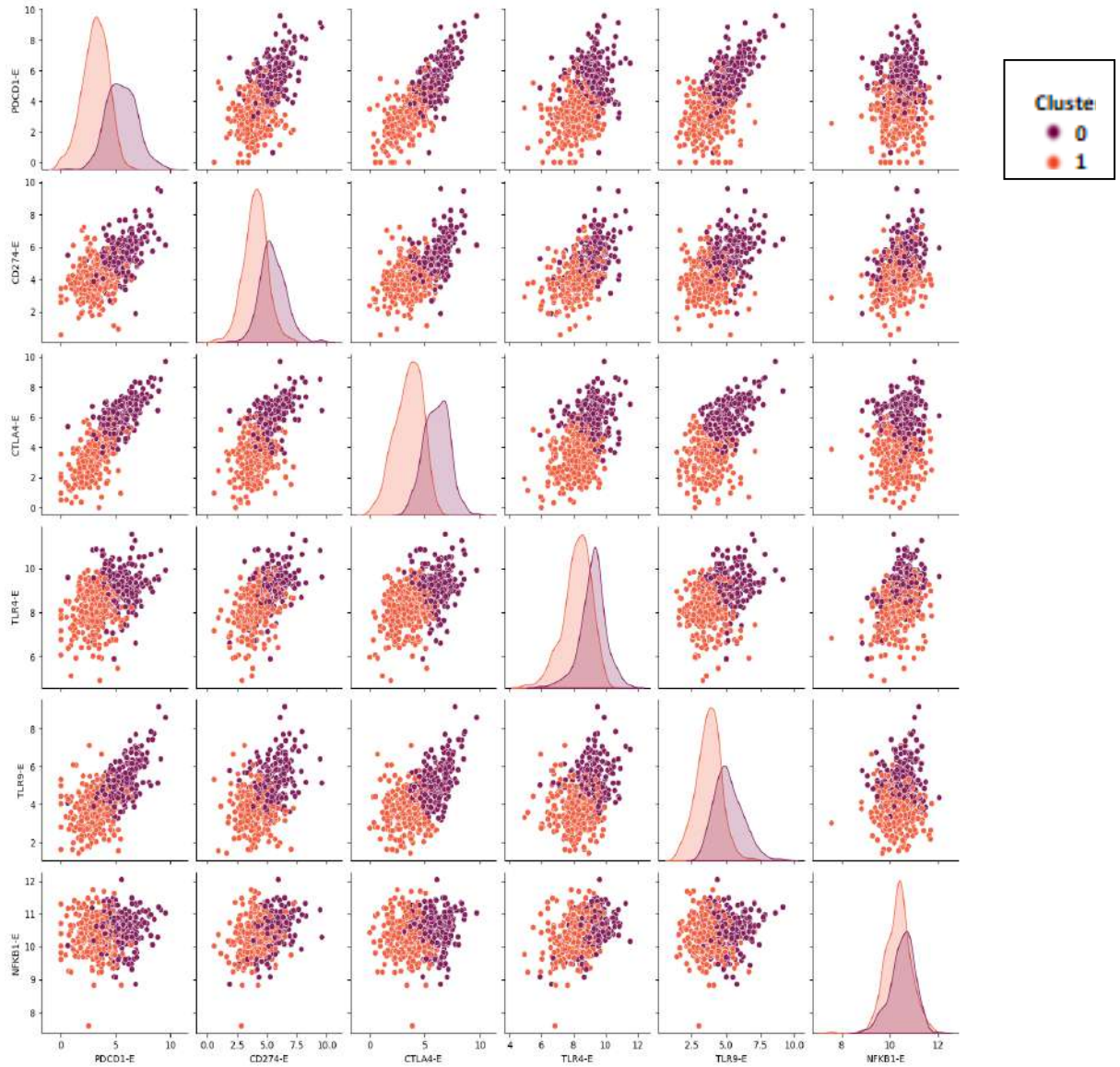


Figure 63. Distribution de l'expression des gènes immunitaires selon les deux clusters.

c. TLR9-E

Les valeurs d'expression de *TLR9-E* dans le Cluster 0 dépassent les 75^e percentiles (5,82 vs 4,33) observés dans le Cluster 1. Cela suggère une activation accrue des voies de l'immunité innée *via* la reconnaissance de l'ADN tumoral, reflétant potentiellement un microenvironnement immunologiquement plus dynamique ou réactif dans ce cluster.

d. NF-κB1-E

Bien que les niveaux d'expression de *NF-κB1-E* soient relativement proches entre les deux groupes (Cluster 0 = $10,52 \pm 0,51$; Cluster 1 = $10,34 \pm 0,53$), la légère élévation dans le Cluster 0 pourrait indiquer une activation modérée de la voie NF-κB. Cette voie, associée à des fonctions pro-inflammatoires et de survie cellulaire, pourrait contribuer à un phénotype tumoral plus inflammatoire et potentiellement pro-tumoral.

- Le Cluster 0 présente un profil immunitaire caractérisé par une surexpression des points de contrôle immunitaires (*PDCD1-E* et *CD274-E*) et une activation possible de l'immunité innée (*TLR9-E*), dessinant un microenvironnement à la fois immunosuppresseur et inflammatoire. Ce contexte pourrait favoriser l'évasion immunitaire, rendant les tumeurs plus résistantes aux thérapies immuno-modulatrices conventionnelles.

À l'inverse, le Cluster 1, par l'expression plus faible de ces marqueurs, semble moins immunosuppresseur, ce qui pourrait le rendre plus réactif à certaines immunothérapies ou à des stratégies combinées.

5.2.2. Gènes hormonaux

L'analyse conjointe des niveaux de l'expression des gènes hormonaux méthylés dans les deux clusters est représentée dans la **Figure 64**.

a. ESRI-E

Le Cluster 1 présente une expression significativement plus élevée de *ESRI-E* ($13,36 \pm 1,35$) par rapport au Cluster 0 ($8,37 \pm 3,08$), avec une distribution plus homogène (écart-type plus faible). Cette stabilité d'expression est typique des tumeurs luminales ER+, bien différenciées et fortement hormonodépendantes. À l'inverse, le Cluster 0 semble regrouper des tumeurs partiellement ER+ ou hétérogènes, suggérant une dépendance hormonale plus faible ou instable.

b. PGR-E

Une tendance parallèle est observée pour *PGR-E*, avec des niveaux plus élevés dans le Cluster 1 (10,60 vs 6,17). Cette co-expression élevée de *PGR-E* et *ESR1-E* dans le Cluster 1 renforce l’hypothèse d’un profil luminal classique. En revanche, la plus grande variabilité dans le Cluster 0 reflète une hétérogénéité phénotypique et possiblement une perte partielle de réponse hormonale.

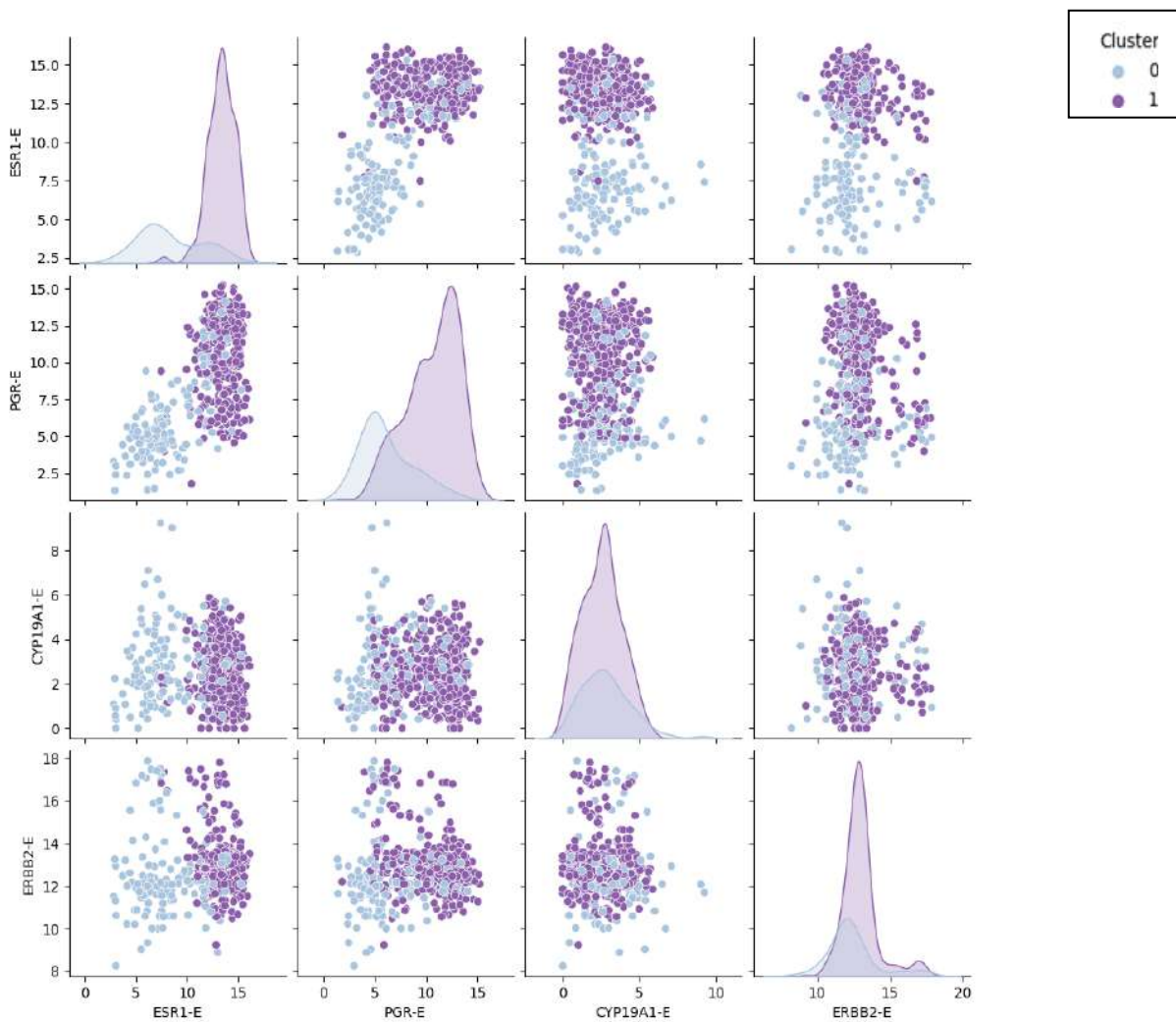


Figure 64. Distribution de l’expression des gènes hormonaux selon les deux clusters.

c. CYP19A1-E

Le Cluster 0 montre des valeurs de *CYP19A1-E* plus élevées au 75e percentile (3,67 vs 3,38), ce qui peut indiquer une activité accrue de l’aromatase. Cela suggère que certaines tumeurs de ce cluster pourraient compenser une faible expression des récepteurs hormonaux par une synthèse

locale accrue d'œstrogènes, favorisant ainsi la croissance tumorale et une résistance partielle à l'hormonothérapie conventionnelle.

d. ERBB2-E

Les niveaux d'expression d'*ERBB2-E* sont similaires entre les deux clusters (Cluster 0 : $12,98 \pm 1,25$; Cluster 1 : $12,42 \pm 1,76$). Ces résultats indiquent que les deux groupes sont principalement composés de tumeurs HER2-négatives, et donc peu susceptibles de répondre aux thérapies ciblées anti-HER2 comme le trastuzumab.

- Le **Cluster 1** se caractérise par un profil transcriptionnel compatible avec des tumeurs lumorales hormonodépendantes (ER+/PR+), bien différenciées, avec une expression stable et élevée des récepteurs hormonaux. Ce groupe pourrait être plus réceptif aux traitements anti-œstrogéniques classiques.

En revanche, le **Cluster 0** présente une plus grande hétérogénéité hormonale, marquée par une expression plus variable, une possible hyperméthylation et une activation alternative de la voie œstrogénique via l'aromatase. Ce profil suggère une sensibilité réduite aux traitements hormonaux standards, nécessitant des stratégies personnalisées ou combinées.

5.3. Moyenne des biomarqueurs immunitaires et hormonaux en fonction des deux clusters

Les graphiques illustrant la moyenne des biomarqueurs par cluster (**Figure 65**) mettent en évidence des différences notables dans les profils d'expression et de méthylation entre le Cluster 0 et le Cluster 1. Ces moyennes soulignent les spécificités moléculaires de chaque groupe, traduisant des voies biologiques potentiellement divergentes et des implications pronostiques ou thérapeutiques distinctes.

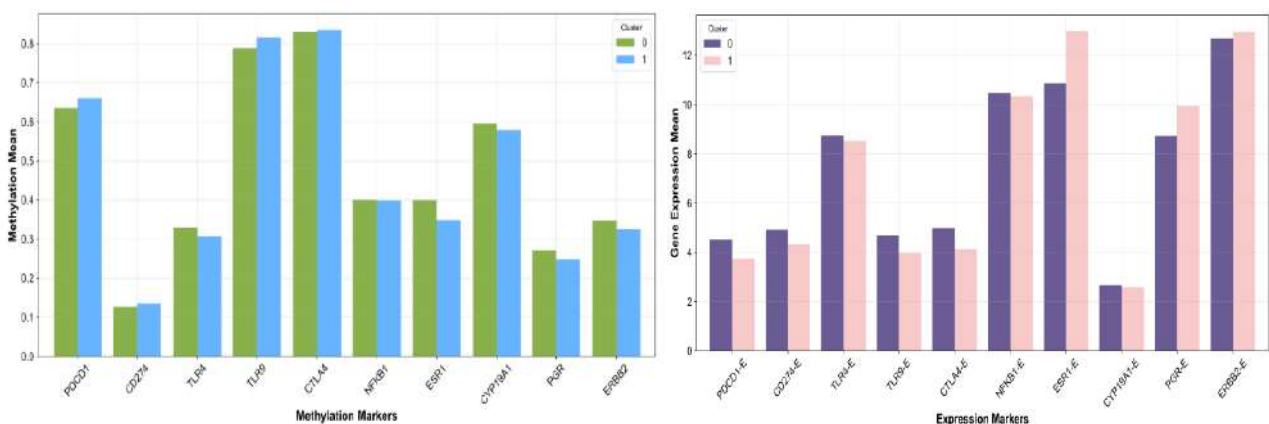


Figure 65. Comparaison des moyennes de la méthylation des biomarqueurs et des niveaux d'expression génique entre les deux clusters (K-means).

6. ETUDE DES CORRELATIONS

6.1. Corrélation entre les gènes méthylés et leurs expressions

Afin de réaliser cette étude, nous avons utilisé le Test de Spearman (**Tableau 11 ; Figure 66**) qui est recommandé dans ce cas, car les données de méthylation sont asymétriques et l'expression génique suit une échelle logarithmique.

Tableau 11. Corrélation entre les gènes méthylés et leurs expressions dans chaque cluster.

Clusters	Genes	Expression	Coef Spearman r	p-value
0	PDCD1	PDCD1-E	-0.053415	0.375
	CD274	CD274-E	-0.251602	2.19e-05***
	TLR4	TLR4-E	-0.048067	0.425
	TLR9	TLR9-E	-0.264180	8.03e-06***
	CTLA4	CTLA4-E	-0.189970	0.00146**
	NFKB1	NFKB1-E	-0.260456	1.09e-05***
	ESR1	ESR1-E	-0.491039	2.81e-18***
	CYP19A1	CYP19A1-E	0.073845	0.220
	PGR	PGR-E	-0.609693	1.09e-29***
	ERBB2	ERBB2-E	-0.277210	2.69e-06***
1	PDCD1	PDCD1-E	0.002449	0.969
	CD274	CD274-E	-0.303384	8.74e-07***
	TLR4	TLR4-E	0.156772	0.0125*
	TLR9	TLR9-E	-0.469709	2.74e-15***
	CTLA4	CTLA4-E	-0.381403	3.49e-10***
	NFKB1	NFKB1-E	0.025840	0.683
	ESR1	ESR1-E	-0.176163	0.00495**
	CYP19A1	CYP19A1-E	0.066629	0.291
	PGR	PGR-E	-0.650123	8.83e-32***
	ERBB2	ERBB2-E	-0.574037	1.42e-23***

- **Cluster 0** : la plupart des corrélations entre la méthylation et l'expression sont négatives, ce qui correspond à l'hypothèse biologique selon laquelle une méthylation accrue peut réduire l'expression génique.
 - * ***PGR*** ($r = -0.609$; $p < 1e-29$) : forte corrélation négative et très significative, indiquant un rôle clé de la méthylation dans la régulation de l'expression de ce gène.
 - * ***ESR1*** ($r = -0.491$; $p < 1e-18$) : corrélation négative significative, cohérente avec son rôle hormonal dans le cancer du sein.
 - * ***ERBB2*** ($r = -0.277$; $p = 2.7e-06$) : corrélation modérée, significative, suggérant une interaction complexe entre la méthylation et l'expression.
 - * ***CYP19A1*** ($r = 0.073$; $p = 0.219$) : absence de corrélation significative, ce qui peut refléter une régulation indépendante de la méthylation.
- **Cluster 1** : les résultats montrent des corrélations plus variées, avec des coefficients positifs et négatifs. Certains gènes montrent des corrélations fortes et significatives similaires au cluster 0, tandis que d'autres affichent des tendances opposées.
 - * ***PGR*** ($r = -0.650$; $p < 1e-32$) : corrélation très forte et significative, renforçant son importance dans les deux clusters.
 - * ***ERBB2*** ($r = -0.574$; $p < 1e-23$) : corrélation négative très forte, suggérant un rôle plus marqué dans ce cluster par rapport au cluster 0.
 - * ***TLR9*** ($r = -0.469$; $p < 1e-15$) : corrélation négative significative, mettant en évidence une potentielle implication immunitaire dans ce cluster.
 - * ***CYP19A1*** ($r = 0.066$; $p = 0.291$) : comme dans le cluster 0, absence de corrélation significative.
 - * ***NFκB1*** ($r = 0.025$; $p = 0.682$) : absence de corrélation, malgré son rôle attendu dans les réponses inflammatoires.

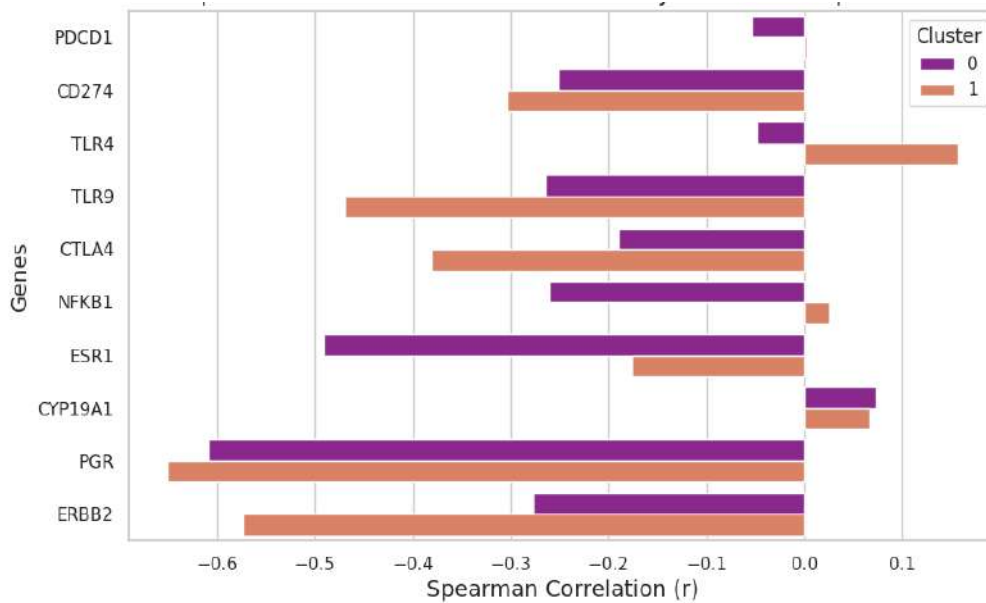


Figure 66. Corrélation entre les niveaux de méthylation de l'ADN méthylé et son expression selon les deux Clusters (K-means).

Les gènes tels que *PGR*, *ESR1*, et *ERBB2* montrent des corrélations négatives significatives dans les deux clusters, ce qui pourrait indiquer des mécanismes similaires de régulation *via* la méthylation.

Les corrélations pour certains gènes immunitaires (*TLR9*, *CTLA4*) sont plus marquées dans le cluster 1, reflétant potentiellement des différences biologiques entre les clusters, telles qu'une activité immunitaire accrue ou une réponse inflammatoire.

6.2. Corrélation entre les expressions géniques des marqueurs immunitaires et hormonaux

L'analyse intégrée des corrélations génétiques (**Figure 67**) et des valeurs de p ajustées (**Tableau 12**) révèle des interactions complexes et intéressantes entre les voies d'expression des gènes immunitaires et hormonaux, mettant en lumière des associations potentielles qui pourraient influencer la réponse immunitaire et le profil hormonal des tumeurs.

Les résultats montrent des associations significatives entre plusieurs gènes clés : les marqueurs immunosuppresseurs *CTLA4-E* et *PDCD1-E* présentent des corrélations négatives élevées avec *ESR1-E* (récepteur d'œstrogènes) et *PGR-E* (récepteur de progestérone), avec des valeurs de p hautement significatives ($p < 0,0001$). Ces données suggèrent que les tumeurs hormono-dépendantes (Luminal) maintiennent un microenvironnement immunitaire moins inhibé, ce qui pourrait expliquer leur faible réponse à l'immunothérapie.

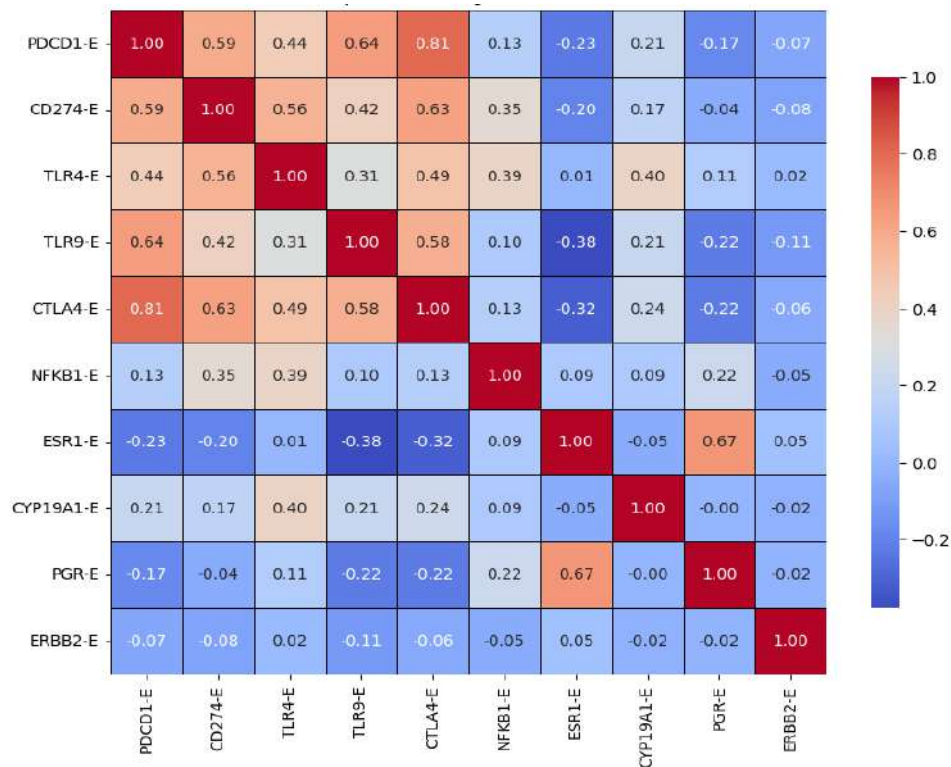


Figure 67. Matrice de corrélation entre les expressions géniques des marqueurs immunitaires et hormonaux.

Parallèlement, on observe des corrélations positives significatives entre PDCD1-E et CYP19A1-E (aromatase), indiquant un lien potentiel entre l'épuisement immunitaire et la production d'œstrogènes. Fait intéressant, ERBB2-E (HER2) ne montre aucune association significative avec les marqueurs immunitaires, un résultat inattendu qui nécessite des investigations supplémentaires.

Ces résultats mettent en lumière deux profils tumoraux distincts : un groupe à dominante hormonale avec faible expression des points de contrôle immunitaires, et un groupe immunosuppresseur marqué par l'expression simultanée de *PD-1*, *PD-L1* et *CTLA-4*.

Ces aspects pourraient guider des stratégies thérapeutiques personnalisées, combinant par exemple des inhibiteurs d'aromatase avec des immunothérapies pour les tumeurs exprimant fortement *CYP19A1*, tout en soulignant la nécessité de valider ces corrélations dans des cohortes cliniques prospectives pour établir leur pertinence pronostique et prédictive.

Tableau 12. Valeurs de p ajustées pour les corrélations entre l'expression des gènes immunitaires et hormonaux.

	ESR1-E	CYP19A1-E	PGR-E	ERBB2-E
PDCD1-E	0.0	0.000003	0.000104	0.144073
CD274-E	0.00001	0.000136	0.337633	0.085203
TLR4-E	0.90184	0.0	0.013155	0.752488
TLR9-E	0.0	0.000003	0.000001	0.013155
CTLA4-E	0.0	0.0	0.000001	0.234068
NFKB1-E	0.044726	0.044726	0.000001	0.333174

Récapitulation :

1/ Corrélations Fortes (≥ 0.5 ou ≤ -0.5)

- ✓ PDCD1-E montre des corrélations positives fortes avec CD274-E (0.59), TLR9-E (0.64), et CTLA4-E (0.81), suggérant une co-régulation possible de ces marqueurs immunitaires.
- ✓ CD274-E (PD-L1) est également fortement corrélé avec CTLA4-E (0.63), ce qui pourrait indiquer une synergie dans les voies de signalisation immunosuppressives.
- ✓ ESR1-E (récepteur des œstrogènes) présente une corrélation négative modérée avec TLR9-E (-0.38) et CTLA4-E (-0.32), suggérant que les tumeurs hormonodépendantes pourraient avoir une expression réduite de certains gènes immunitaires.
- ✓ PGR-E (récepteur de la progestérone) est fortement corrélé avec ESR1-E (0.67), reflétant leur lien connu dans les tumeurs hormono-sensibles.

2/ Corrélations Faibles ou Nulles

- ✓ NFKB1-E (impliqué dans l'inflammation) présente des corrélations faibles avec la plupart des autres gènes, sauf une légère association avec TLR4-E (0.39) et CD274-E (0.35).
- ✓ ERBB2-E (HER2) ne montre aucune corrélation significative avec les autres marqueurs, ce qui est surprenant étant donné son rôle dans certains sous-types agressifs.

Ces résultats présentent des implications cliniques intéressantes :

- * En immunothérapie, les fortes corrélations entre PDCD1-E, CD274-E (PD-L1), et CTLA4-E pourraient indiquer une réponse potentielle aux inhibiteurs de points de contrôle immunitaires

(anti-PD-1/PD-L1, anti-CTLA4). Des molécules anti-PD-1/PD-L1 sont actuellement utilisées en immunothérapie.

* Dans le cas des tumeurs hormono-dépendantes, la corrélation négative entre ESR1-E/PGR-E et certains gènes immunitaires (TLR9-E, CTLA4-E) pourrait expliquer la moindre efficacité de l'immunothérapie dans les cancers Lumineux.

* Lors des stratifications thérapeutiques, les patientes avec une expression élevée de CD274-E et CTLA4-E pourraient être candidates à une combinaison d'immunothérapie et de traitements ciblés.

Les valeurs $p < 0.001$ pour les corrélations entre PDCD1-E, CD274-E, TLR9-E, et CTLA4-E avec ESR1-E, CYP19A1-E, et PGR-E, indiquent une association très significative entre ces gènes.

La corrélation entre PDCD1-E et ESR1-E est très hautement significative.

L'expression des gènes hormonaux *ESR1*, *CYP19A1*, *PGR* montrent des associations significatives avec plusieurs gènes immunitaires, suggérant que l'expression de ces gènes pourrait influencer l'activité immunitaire ou vice versa.

7. ANALYSE DE LA MÉTHYLATION ET L'EXPRESSION DES GÈNES IMMUNITAIRES ET HORMONAUX SELON LE STATUT MÉNOPAUSIQUE

7.1. Analyse différentielle de la méthylation des gènes et leur expression selon le statut ménopausique

Afin d'évaluer l'effet du statut ménopausique sur la méthylation des gènes immunitaires méthylés, impliqués dans le cancer du sein, une analyse de normalité (test de Shapiro-Wilk) a été effectuée pour chaque groupe (Pré-ménopause, Péri-ménopause, Indéterminé et Post-ménopause), suivie d'un test non paramétrique de Kruskal-Wallis.

7.1.1. Gènes Méthylés

a/ Gènes immunitaires

La majorité des distributions de méthylation ne suivent pas la loi normale, justifiant le recours au test de Kruskal-Wallis pour la comparaison entre groupes. Parmi les gènes étudiés, seul *TLR4*

présente une différence statistiquement significative de méthylation selon le statut de ménopause ($p = 0.0388$). Cela suggère un lien potentiel entre l'activité épigénétique de *TLR4* et les modifications hormonales liées à la ménopause, pouvant jouer un rôle dans la régulation immunitaire au cours de la tumorigenèse mammaire (**Tableau 13**).

Tableau 13. Résultats des tests de normalité (Shapiro-Wilk) et du test de Kruskal-Wallis pour la méthylation des gènes selon le statut ménopausique.

Gènes	Shapiro p (Pré)	Shapiro p (Péri)	Shapiro p (Indét.)	Shapiro p (Post)	Kruskal-Wallis p
<i>PDCD1</i>	0.000	0.109	0.057	0.192	0.1200
<i>CD274</i>	0.000	0.000	0.135	0.587	0.2119
<i>TLR4</i>	0.000	0.000	0.107	0.008	0.0388
<i>TLR9</i>	0.000	0.000	0.084	0.197	0.7338
<i>CTLA4</i>	0.000	0.000	0.012	0.173	0.7159
<i>NF-κB1</i>	0.000	0.000	0.863	0.116	0.4176

b/ Gènes hormonaux

Parmi les gènes hormonaux étudiés (**Tableau 14**), seul *ESR1* présente une variation significative de méthylation selon le statut ménopausique ($p < 0.0001$).

Tableau 14. Résultats des tests de normalité (Shapiro-Wilk) et du test de Kruskal-Wallis pour la méthylation des gènes hormonaux selon le statut de ménopausique.

Gène	Shapiro p (Pré)	Shapiro p (Péri)	Shapiro p (Indét.)	Shapiro p (Post)	Kruskal-Wallis p
<i>ESR1</i>	0.026	0.000	0.880	0.667	0.0000
<i>CYP19A1</i>	0.000	0.156	0.834	0.633	0.7257
<i>PGR</i>	0.000	0.000	0.302	0.320	0.4565
<i>ERBB2</i>	0.000	0.038	0.177	0.030	0.7636

Cette observation souligne une possible régulation épigénétique du récepteur aux œstrogènes liée aux modifications hormonales autour de la ménopause, avec un impact potentiel sur la réponse aux traitements endocriniens. Les trois autres gènes (*CYP19A1*, *PGR*, *ERBB2*) ne montrent pas de différences significatives de méthylation entre les groupes.

7.1.2. Gènes Exprimés

a/ Gènes immunitaires

Parmi les gènes immunitaires analysés, seul TLR4 présente une expression significativement différente selon le statut de ménopause ($p = 0.0023$) selon test de Kruskal-Wallis, suggérant une implication potentielle de la voie TLR4 dans les changements immunitaires associés à la ménopause (**Tableau 15**). Les gènes *PDCD1*, *CD274*, *TLR9*, *CTLA4* et *NFKB1* ne montrent pas de variations significatives de leur expression entre les groupes, ce qui peut refléter une stabilité relative de leur régulation transcriptionnelle vis-à-vis des changements hormonaux.

Tableau 15. Tests de normalité et test de Kruskal-Wallis pour l'expression des gènes immunitaires selon le statut ménopausique.

Gène	Shapiro p (Pré)	Shapiro p (Péri)	Shapiro p (Indét.)	Shapiro p (Post)	Kruskal-Wallis p
PDCD1-E	0.123	0.208	0.005	0.317	0.3163
CD274-E	0.015	0.542	0.651	0.309	0.8698
TLR4-E	0.000	0.168	0.838	0.514	0.0023
TLR9-E	0.001	0.475	0.535	0.543	0.8342
CTLA4-E	0.406	0.007	0.153	0.178	0.2833
NFKB1-E	0.000	0.133	0.378	0.186	0.3634

b/ Gènes hormonaux

Seule l'expression du gène *ESR1*, codant pour le récepteur des œstrogènes, varie significativement selon le statut ménopausique ($p < 0.0001$) selon test de Kruskal-Wallis, ce qui peut refléter une régulation hormonale marquée de ce gène avec l'âge et le statut hormonal. En revanche, *CYP19A1-E* (aromatase) et *ERBB2-E* ne montrent aucune variation significative entre les groupes (**Tableau 16**).

Tableau 16. Tests de normalité et test de Kruskal-Wallis pour l'expression des gènes hormonaux selon le statut ménopausique.

Gènes	Shapiro p (Pré)	Shapiro p (Péri)	Shapiro p (Indét.)	Shapiro p (Post)	Kruskal-Wallis p
ESR1-E	0.000	0.000	0.028	0.045	0.0000
CYP19A1-E	0.000	0.000	0.520	0.148	0.9244
PGR-E	0.000	0.000	0.084	0.218	0.0694
ERBB2-E	0.000	0.000	0.026	0.152	0.3847

7.2. Analyse des corrélations croisées des gènes méthylés et de leur expression selon le statut ménopausique

Afin de mieux comprendre les mécanismes régulateurs sous-jacents à l'expression des gènes impliqués dans les voies hormonales et immunitaires du cancer du sein, une analyse conjointe de la méthylation de l'ADN et de l'expression génique a été réalisée. Cette analyse a été stratifiée selon le statut ménopausique, en raison de son influence potentielle sur l'environnement hormonal et, par conséquent, sur la régulation transcriptionnelle et épigénétique des gènes. Les résultats obtenus sont présentés dans le **Tableau 17** et permettent de dégager des éléments clés concernant la régulation différenciée de certains biomarqueurs en fonction du contexte hormonal.

L'étude des corrélations entre la méthylation et l'expression génique révèle plusieurs associations significatives, majoritairement négatives, compatibles avec un mécanisme d'inhibition transcriptionnelle induite par l'hyperméthylation. Ces corrélations négatives sont observées pour les gènes *PDCDI* ($r = - 0.14$; $p = 0.0022$), *CD274* ($r = - 0.35$; $p < 0.0001$), *TLR9* ($r = - 0.49$; $p < 0.0001$), *CTLA4* ($r = - 0.32$; $p < 0.0001$), *NF-κB1* ($r = - 0.14$; $p = 0.0023$), *ESR1* ($r = - 0.53$; $p < 0.0001$), *PGR* ($r = - 0.66$; $p < 0.0001$) et *ERBB2* ($r = - 0.48$; $p < 0.0001$). En revanche, une corrélation positive significative a été observée pour *CYP19A1* ($r = 0.10$; $p = 0.0259$), suggérant un mode de régulation distinct pour ce gène.

Tableau 17. Analyse conjointe de la méthylation de l'ADN et de l'expression génique en fonction du statut ménopausique.

Biomarqueurs	Coef. Spearman (r)	p-value
PDCD1	- 0.14	0.0022
CD274	- 0.35	< 0.0001
TLR4	+ 0.08	0.0700
TLR9	- 0.49	< 0.0001
CTLA4	- 0.32	< 0.0001
NFKB1	- 0.14	0.0023
ESR1	- 0.53	< 0.0001
CYP19A1	+ 0.10	0.0259
PGR	- 0.66	< 0.0001
ERBB2	- 0.48	< 0.0001

L'ensemble de ces données met en lumière le rôle central de la méthylation dans la modulation de l'expression génique des facteurs immunitaires et hormonaux impliqués dans le cancer du sein. Plus particulièrement, les résultats convergents observés pour *ESR1* et *TLR4*, gènes à la fois méthylés, exprimés de manière différentielle selon le statut de ménopause et corrélés, renforcent l'hypothèse d'une régulation épigénétique hormono-dépendante. Cette interaction complexe entre environnement hormonal et modifications épigénétiques pourrait avoir un impact majeur sur la progression tumorale ainsi que sur la sensibilité aux traitements ciblés, notamment les thérapies hormonales et immunomodulatrices.

7.3. Distribution des gènes méthylés et leur expression en fonction du statut ménopausique et selon les clusters

7.3.1. Distribution des gènes Immunitaires Méthylés dans le cluster 0

L'analyse statistique par test de Kruskal-Wallis appliquée aux niveaux de méthylation des gènes *PDCD1*, *CD274*, *TLR4*, *TLR9*, *CTLA4* et *NF-κB1* n'a révélé aucune différence significative en fonction du statut ménopausique ($p > 0.05$ pour tous les gènes) (**Figure 68**).

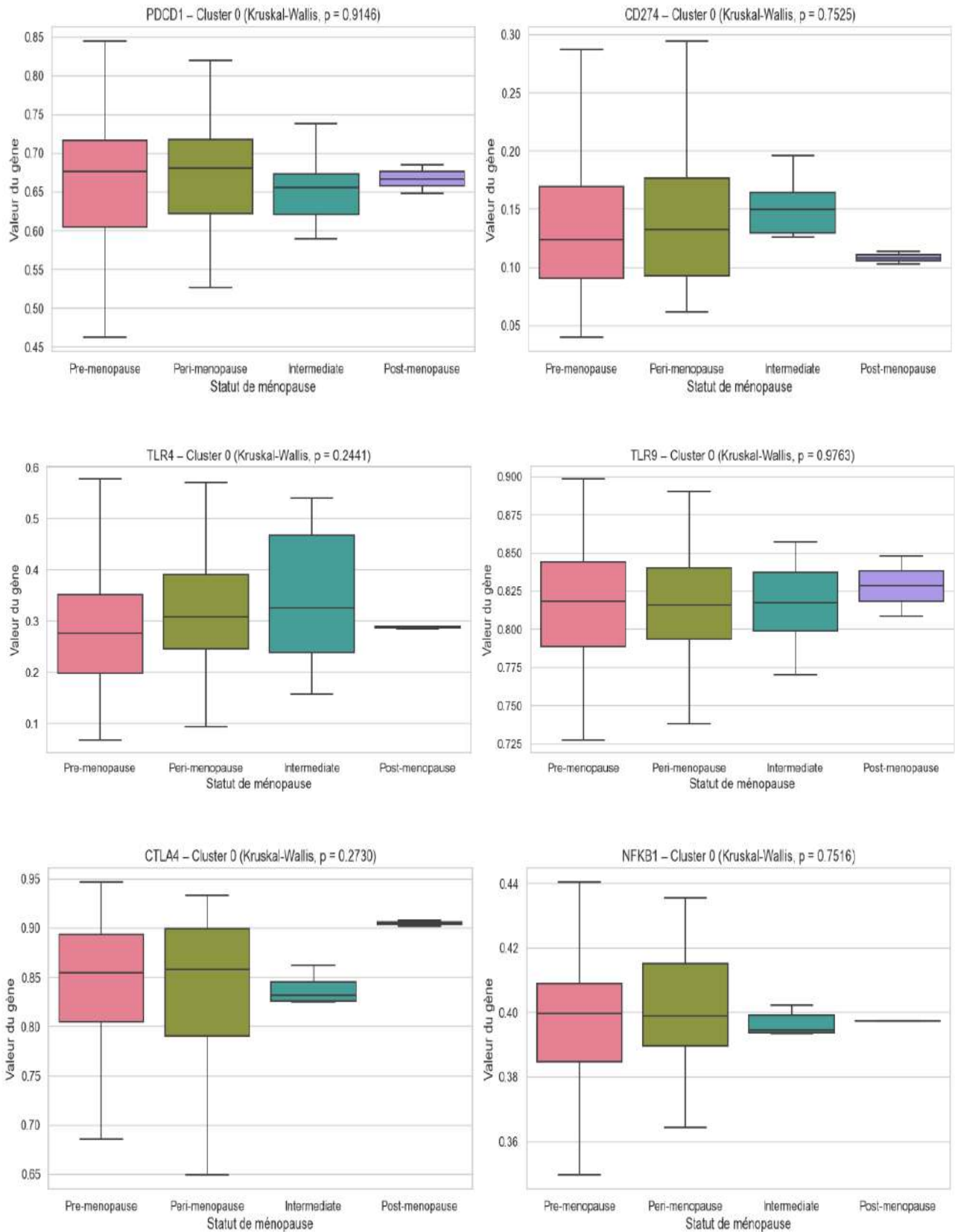
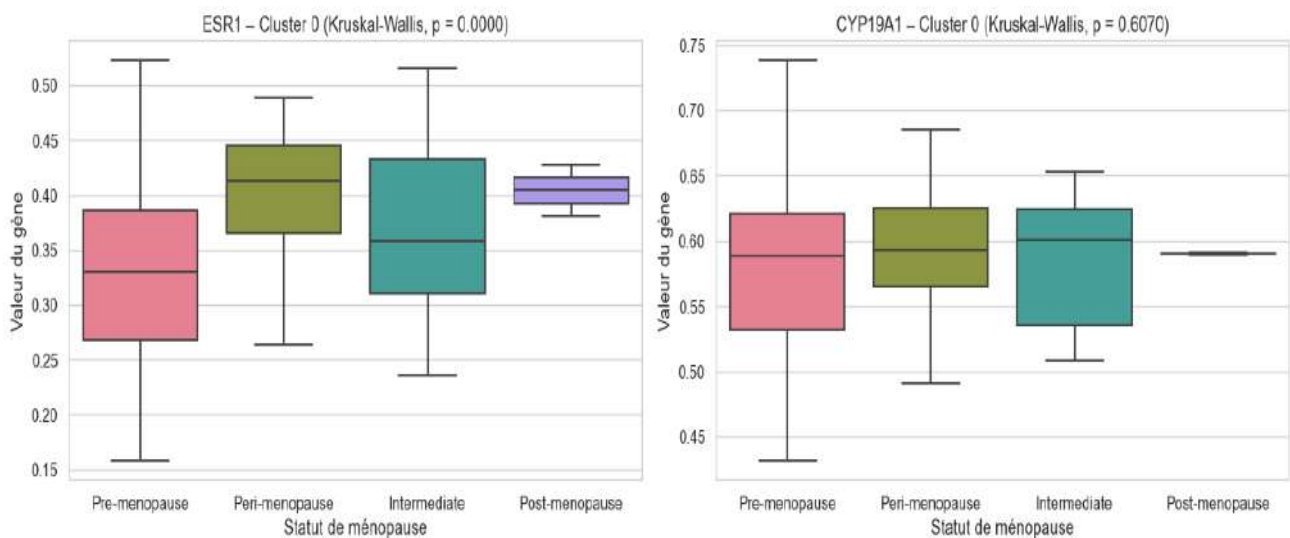


Figure 68. Distribution des gènes immunitaires méthylés selon le statut ménopausique dans le cluster 0.

- Dans le **cluster 0**, les profils de méthylation des gènes immunitaires semblent indépendants du statut hormonal des patientes. Cela suggère que les modifications épigénétiques affectant ces gènes sont probablement liées à d'autres facteurs biologiques (âge, inflammation chronique, environnement tumoral) plutôt qu'aux variations hormonales liées à la ménopause.
- L'absence de variation significative peut également indiquer une méthylation stable ou un profil épigénétique homogène au sein de ce cluster, qui pourrait refléter un état immunologique commun, peu influencé par l'état endocrinien.

7.3.2. Distribution des gènes hormonaux méthylés dans le cluster 0

Dans ce cluster, seul le gène *ESR1* semble sensible aux modifications épigénétiques liées au statut hormonal, ce qui pourrait en faire un marqueur clé de la dépendance hormonale dans cette population. Les autres gènes hormono-dépendants ne montrent pas de variation significative de méthylation, suggérant une hétérogénéité dans la régulation hormonale (**Figure 69**).



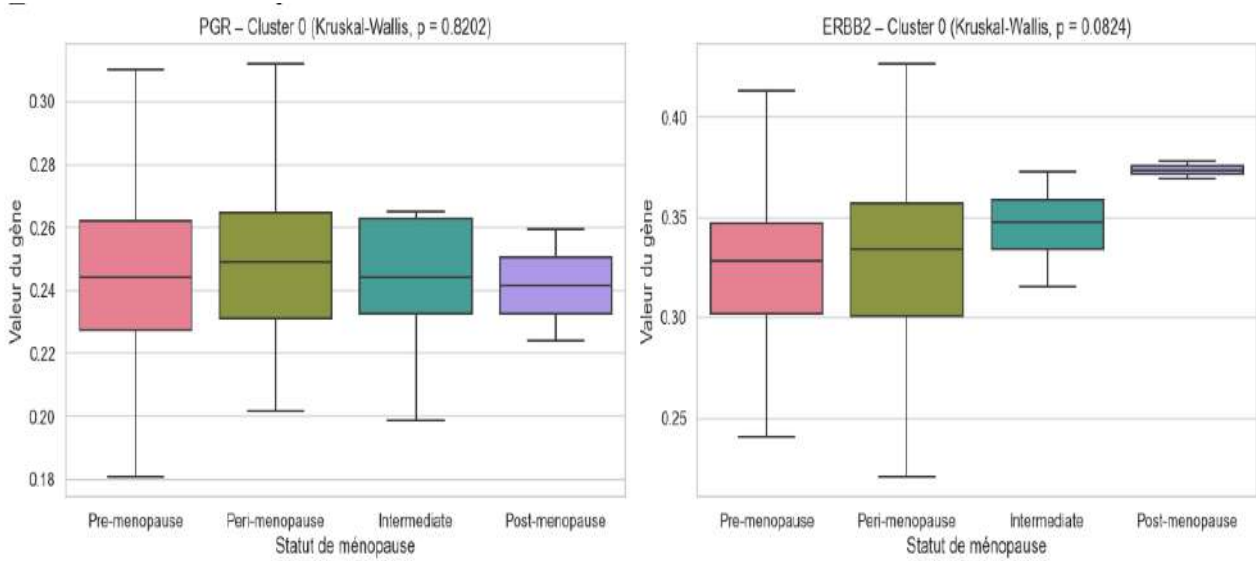
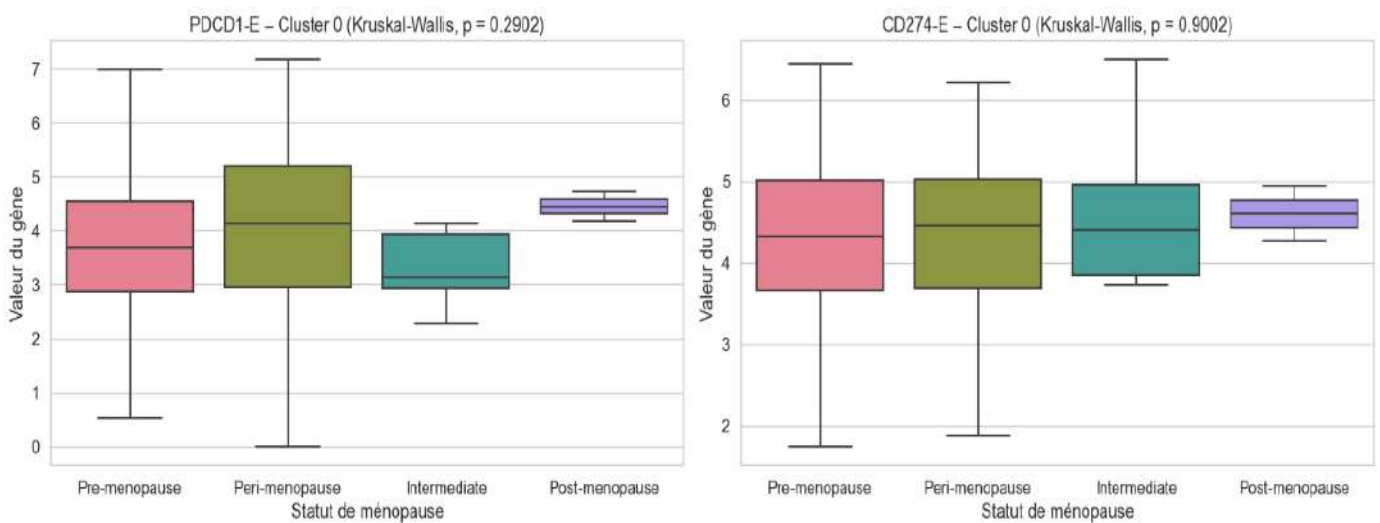


Figure 69. Distribution des gènes hormonaux méthylés selon le statut ménopausique dans le cluster 0.

7.3.3. Distribution des gènes Immunitaires méthylés exprimés dans le cluster 0

Les résultats obtenus (Figure 70) indiquent que dans ce cluster, le système immunitaire anti-tumoral (représenté par ces gènes) ne semble pas influencé par les variations hormonales liées à la ménopause. Cela peut traduire une activité immunitaire homogène quel que soit l'état hormonal des patientes de ce groupe.



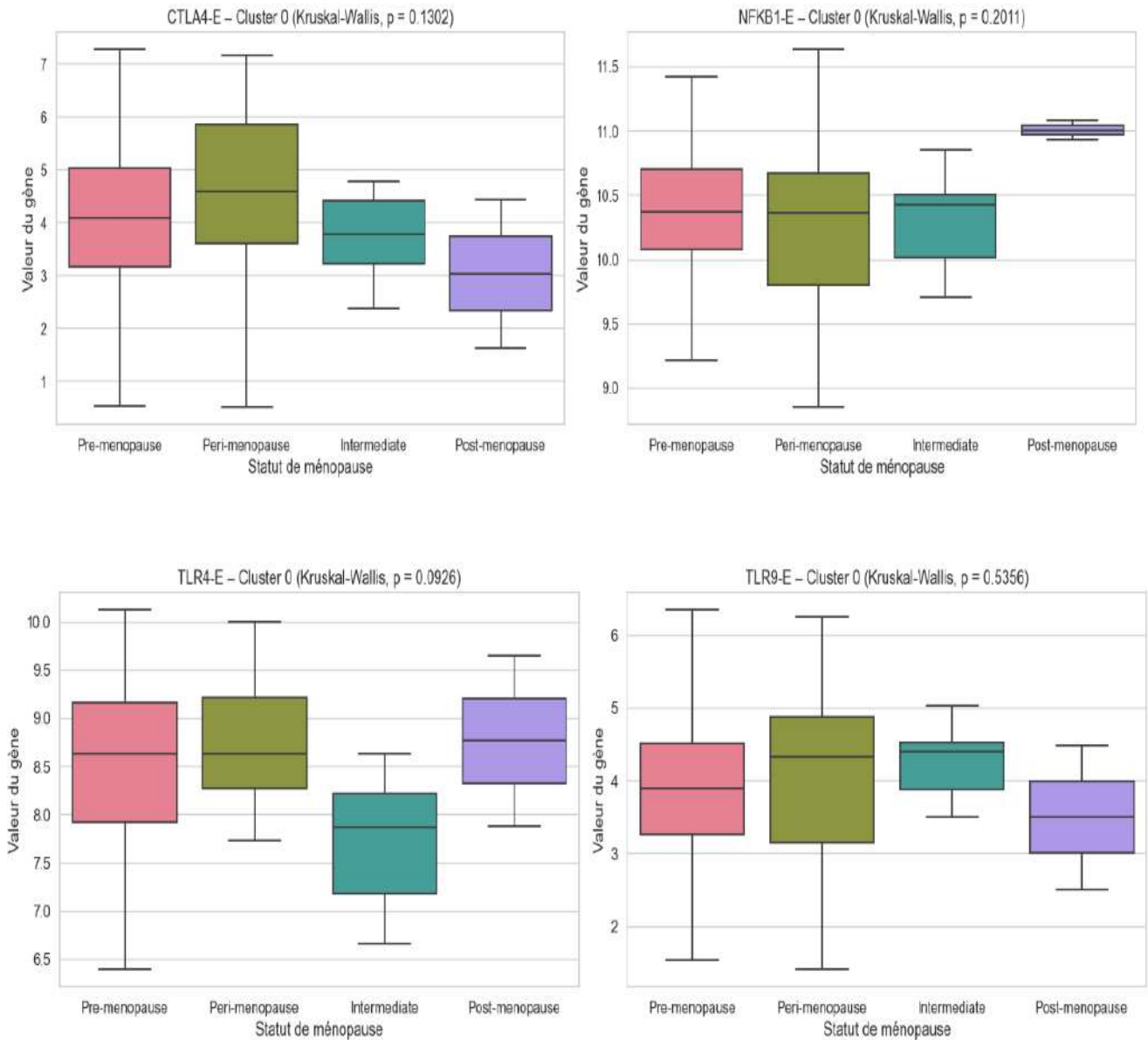


Figure 70. Distribution des gènes immunitaires exprimés selon le statut ménopausique dans le cluster 0.

7.3.4. Distribution des gènes hormonaux méthylés exprimés dans le cluster 0

Parmi les gènes analysés, seule l'expression du gène *ESR1* présente une différence significative ($p = 0.000$) entre les groupes (**Figure 71**). Les trois autres gènes *CYP19A1*, *PGR* et *ERBB2* ne montrent pas de différences significatives quant à leur expression.

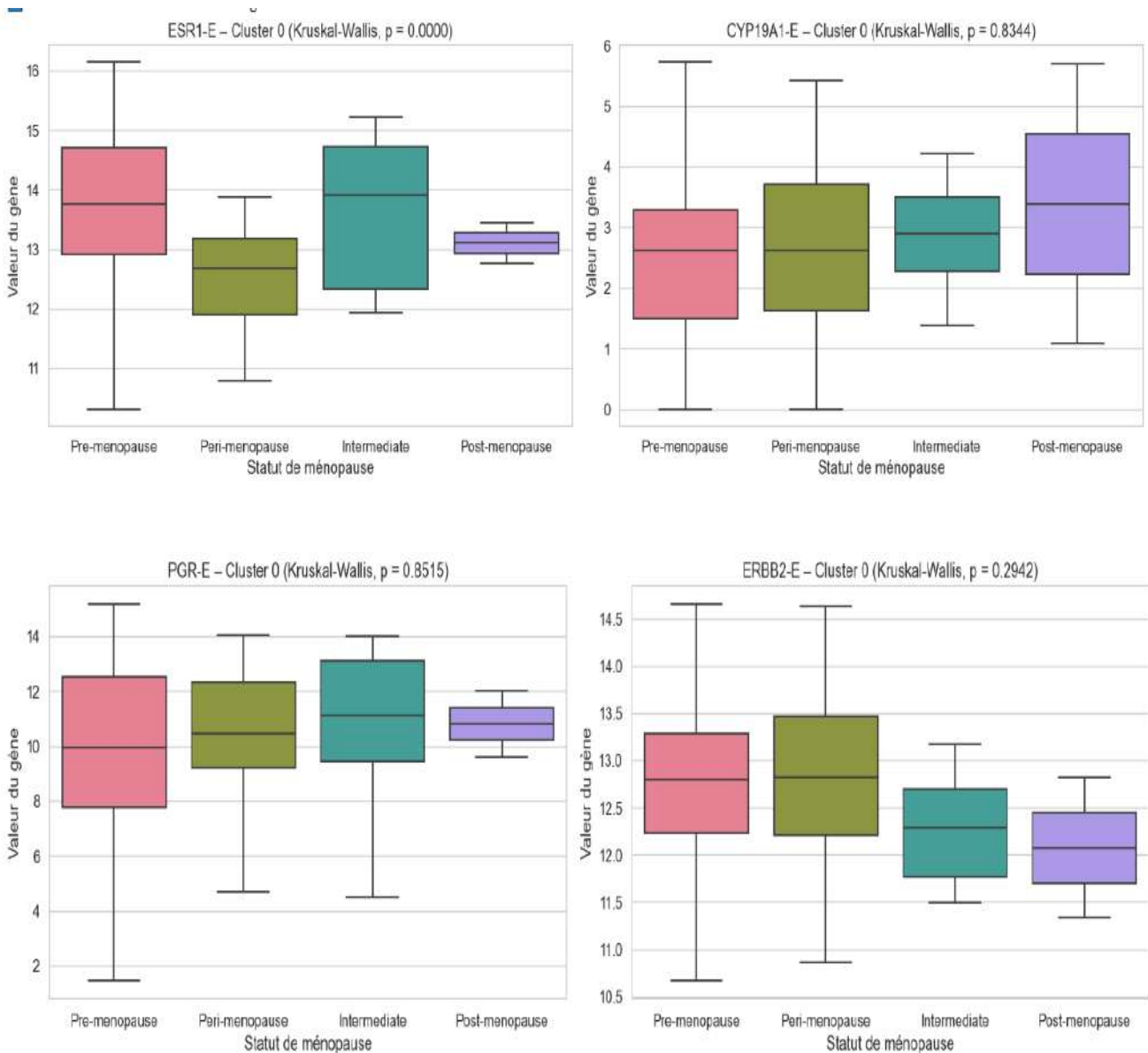


Figure 71. Distribution des gènes hormonaux exprimés selon le statut ménopausique dans le cluster 0.

7.3.5. Distribution des gènes Immunitaires méthylés dans le cluster 1

Aucun des gènes étudiés *PDCD1*, *CD274*, *TLR4*, *TLR9*, *CTLA4* et *NF-κB1* ne montre de différence significative quant à l'expression des gènes entre les différents groupes, suggérant qu'ils n'ont pas de rôle distinctif dans la stratification des groupes dans cette analyse (**Figure 72**).

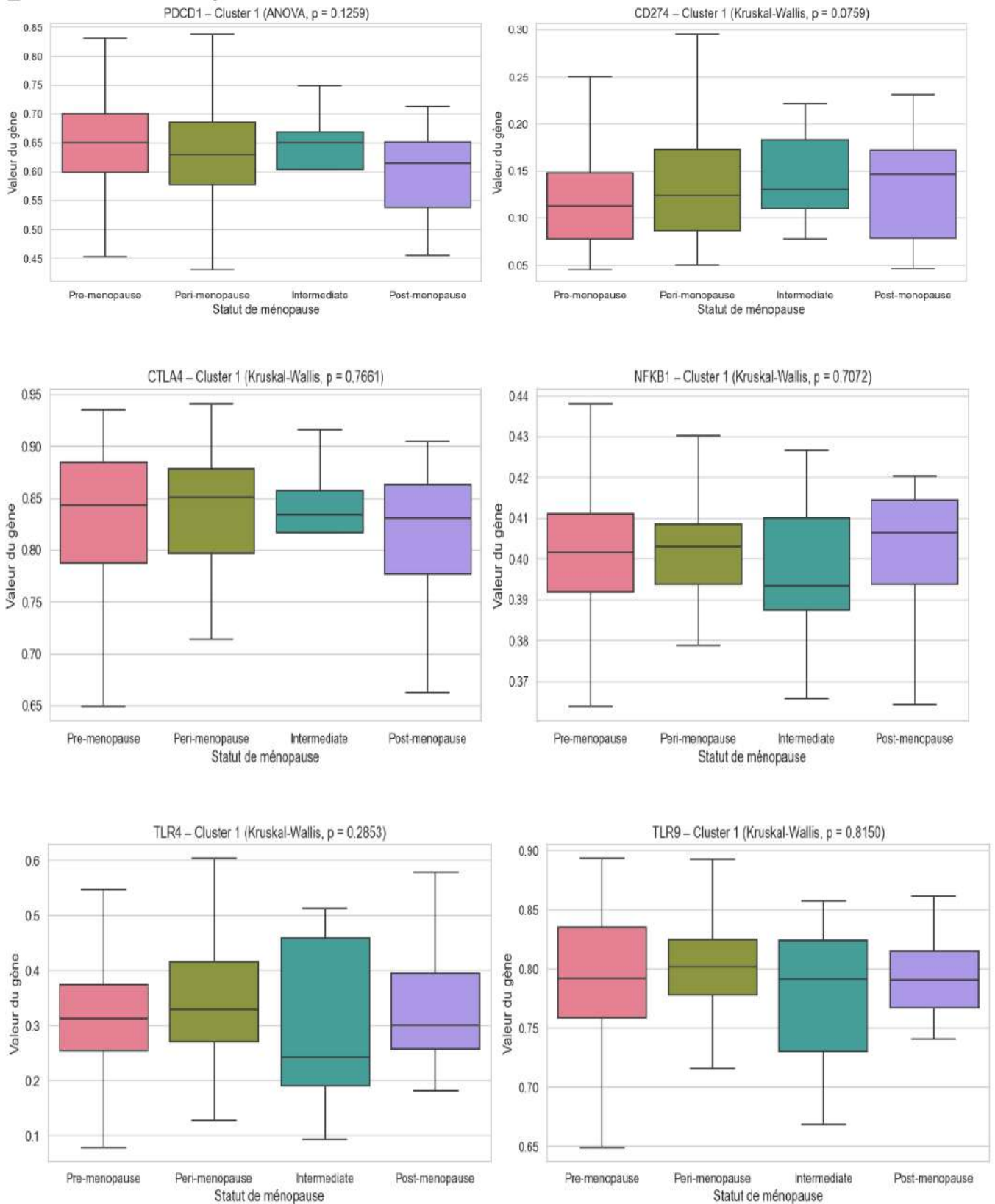


Figure 72. Distribution des gènes immunitaires méthylés selon le statut ménopausique dans le cluster 1.

7.3.6. Distribution des gènes hormonaux méthylés dans le cluster 1

L'expression des gènes *ESR1*, *CYP19A1*, et *ERBB2* ne montre pas de différence significative entre les groupes, ce qui indique que ces gènes ne sont pas des biomarqueurs discriminants dans le contexte étudié. En revanche, l'expression du gène *PGR* présente une différence significative entre les groupes ($p = 0.0006$), ce qui pourrait indiquer un rôle potentiel de ce gène dans la classification ou la réponse au traitement dans le cancer du sein (**Figure 73**). Ces résultats soulignent l'importance d'explorer davantage le rôle de *PGR* dans les sous-types cliniques du cancer du sein.

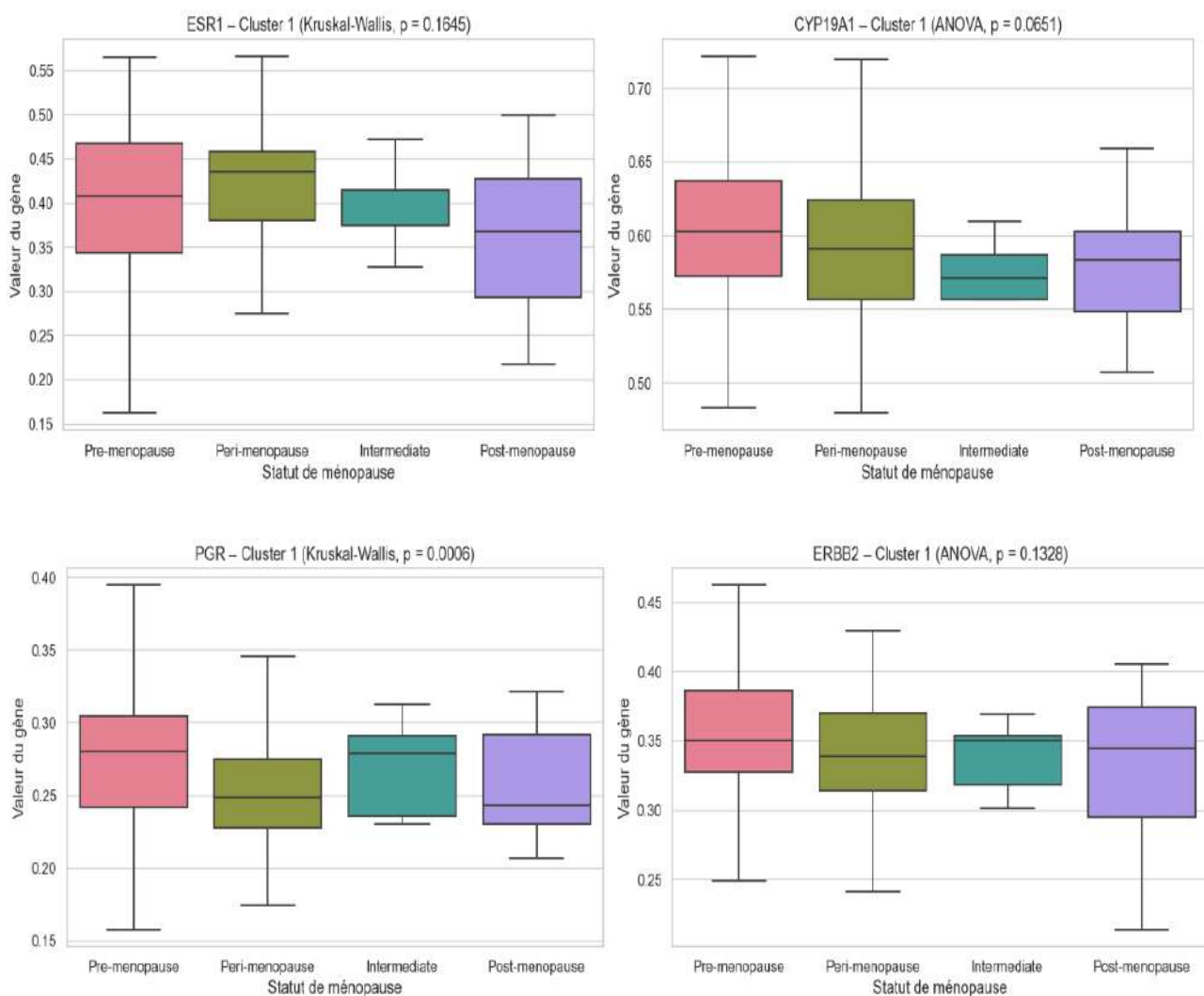
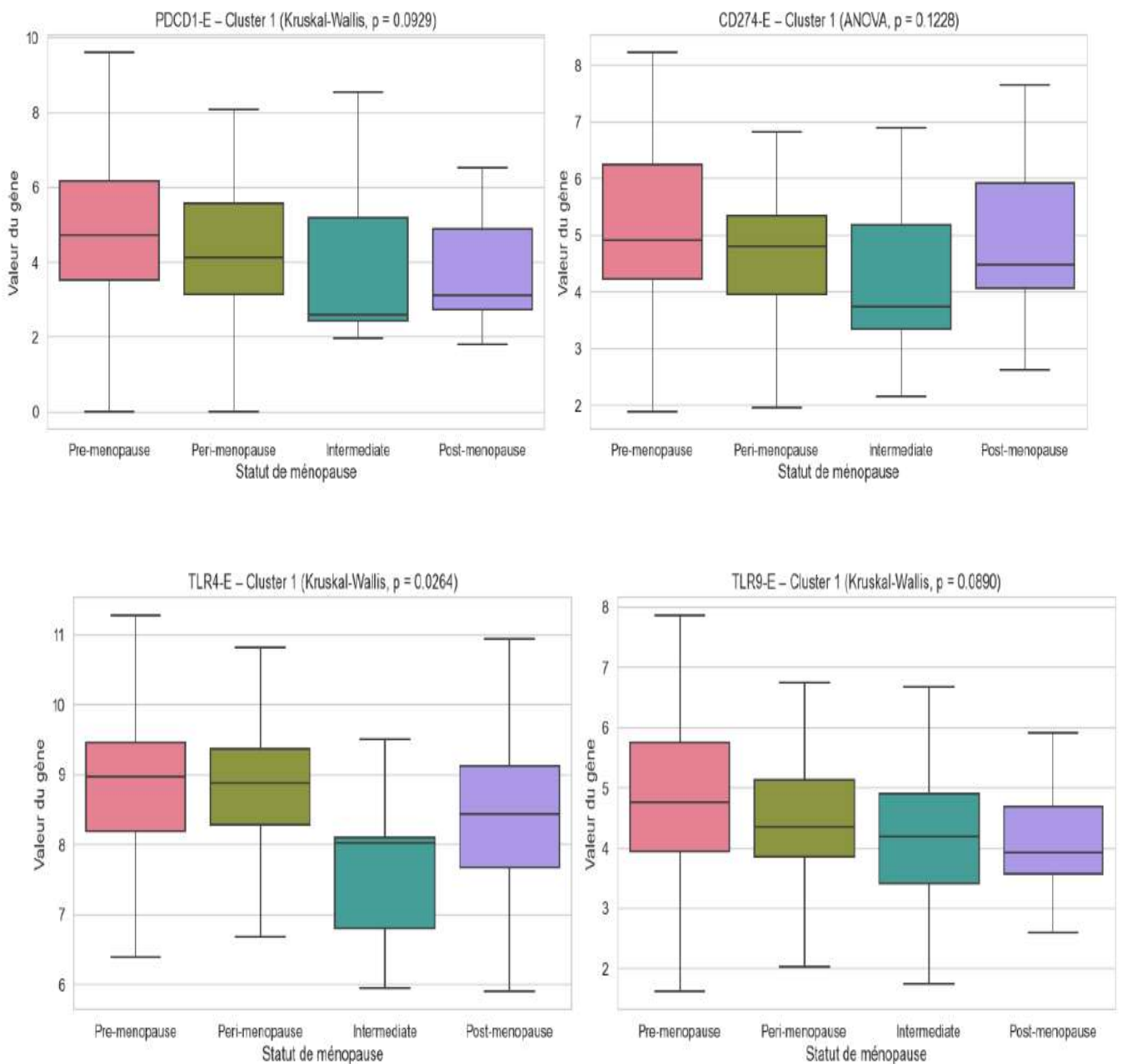


Figure 73. Distribution des gènes hormonaux méthylés selon le statut ménopausique dans le cluster 1.

7.3.7. Distribution des gènes immunitaires méthylés exprimés dans le cluster 1

L'expression des gènes *PDCD1*, *CD274*, *TLR9*, *CTLA4*, et *NF-κB1* ne montre pas de différences significatives entre les groupes. Toutefois, *TLR4* présente une différence significative ($p = 0.0264$), ce qui suggère que ce récepteur pourrait avoir un rôle distinctif dans la régulation de la réponse immunitaire ou d'autres processus biologiques, selon les groupes étudiés (**Figure 74**). Ce résultat met en lumière l'importance de *TLR4* comme biomarqueur potentiel ou cible thérapeutique dans certaines configurations cliniques du cancer du sein.



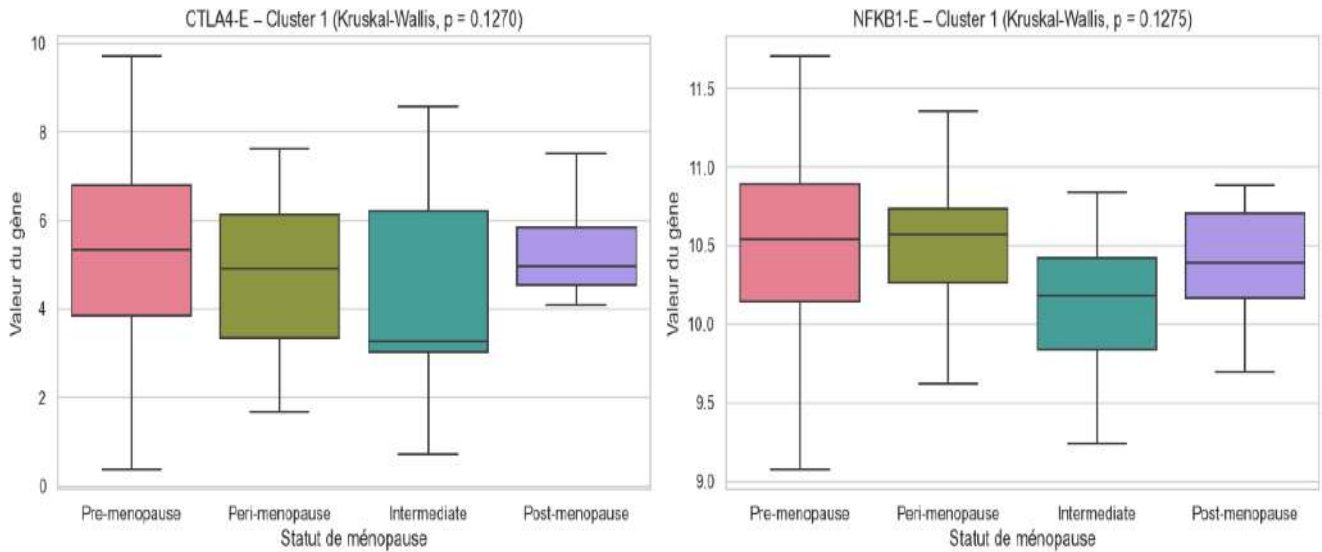
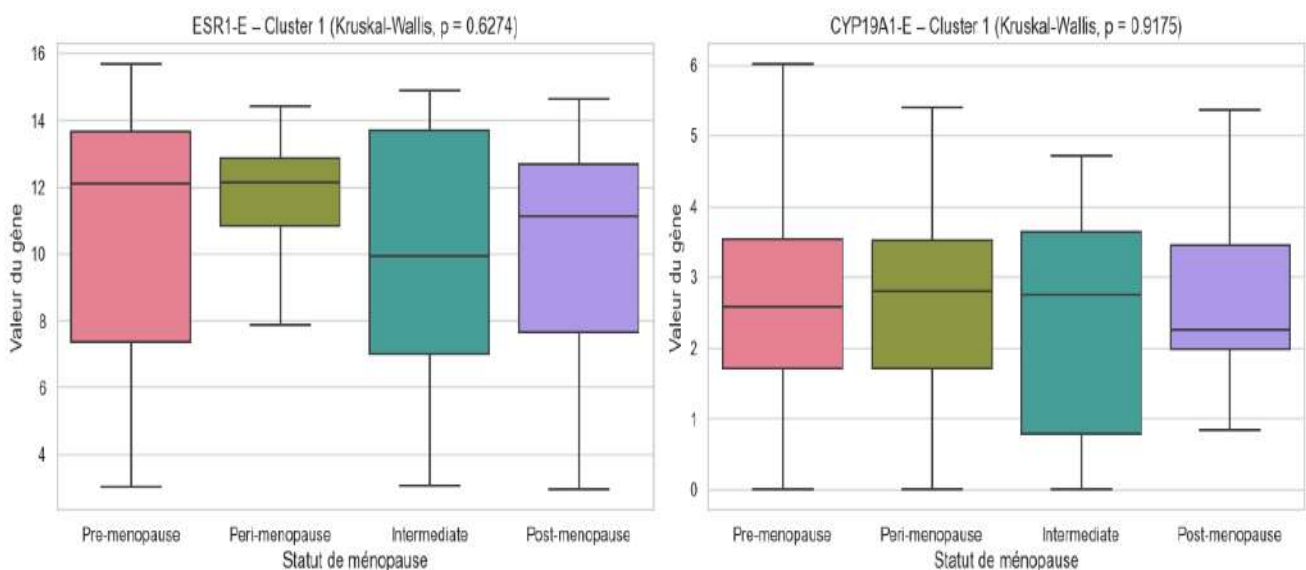


Figure 74. Distribution des gènes immunitaires exprimés selon le statut ménopausique dans le cluster 1.

7.3.8. Distribution des gènes hormonaux méthylés exprimés dans le cluster 1

L'expression des gènes *ESR1*, *CYP19A1*, et *ERBB2* ne montre pas de différence significative entre les groupes, ce qui suggère que ces gènes ne sont pas des marqueurs discriminants dans cette analyse. En revanche, l'expression du gène *PGR* présente une différence significative entre les groupes ($p = 0.0028$), ce qui peut indiquer que le récepteur de la progestérone joue un rôle clé dans la caractérisation des groupes étudiés (Figure 75).



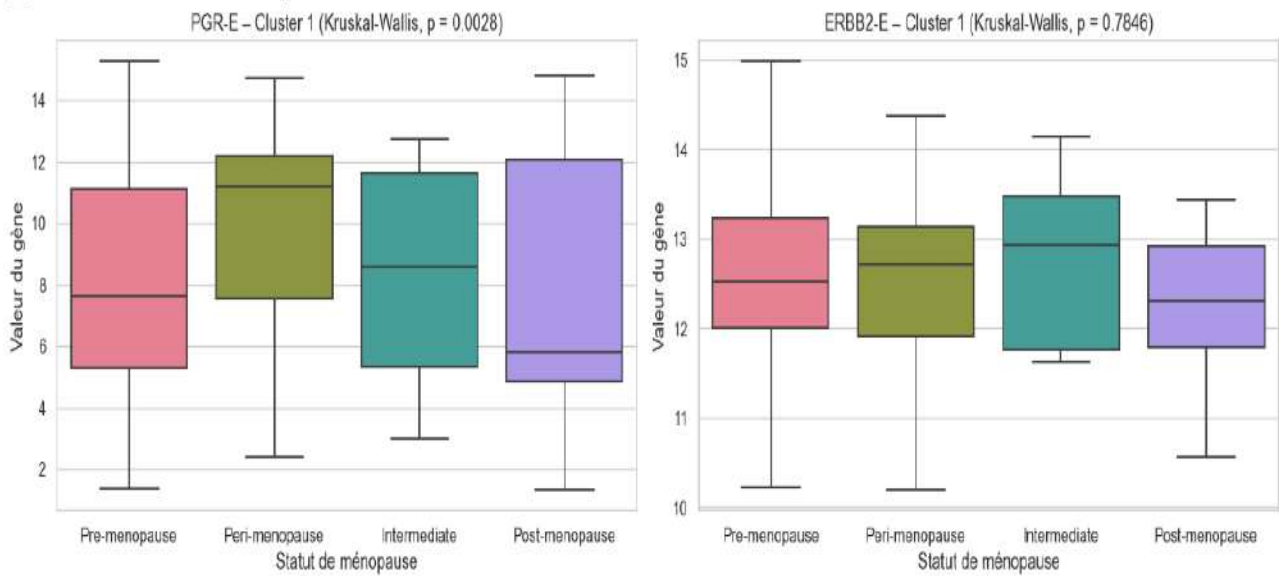


Figure 75. Distribution des gènes hormonaux exprimés selon le statut ménopausique dans le cluster 1.

8. ANALYSE DE LA MÉTHYLATION ET L'EXPRESSION DES GÈNES IMMUNITAIRES ET HORMONAUX SELON LA SURVIE

8.1. Distribution des gènes méthylés et leur expression dans la population totale

8.1.1. Gènes méthylés

La **Figure 76** présente une série de graphiques en violon qui comparent la distribution selon la survie des gènes méthylés (épigénétique) dans la population totale entre deux groupes de patientes (vivantes, décédées). Chaque ligne du graphique montre une paire de distributions pour chaque gène.

L'analyse comparative de la méthylation de l'ADN entre patientes vivantes et décédées n'a révélé aucune différence statistiquement significative parmi les dix gènes étudiés (immunitaires et hormonaux) ($p > 0,05$)

Les graphiques en violon permettent de visualiser à la fois la densité de probabilité des données (la largeur du violon à chaque niveau indique la fréquence des valeurs) et des statistiques de synthèse comme la médiane et l'étendue interquartile. Plus le violon est large à un endroit, plus il y a de valeurs à ce niveau.

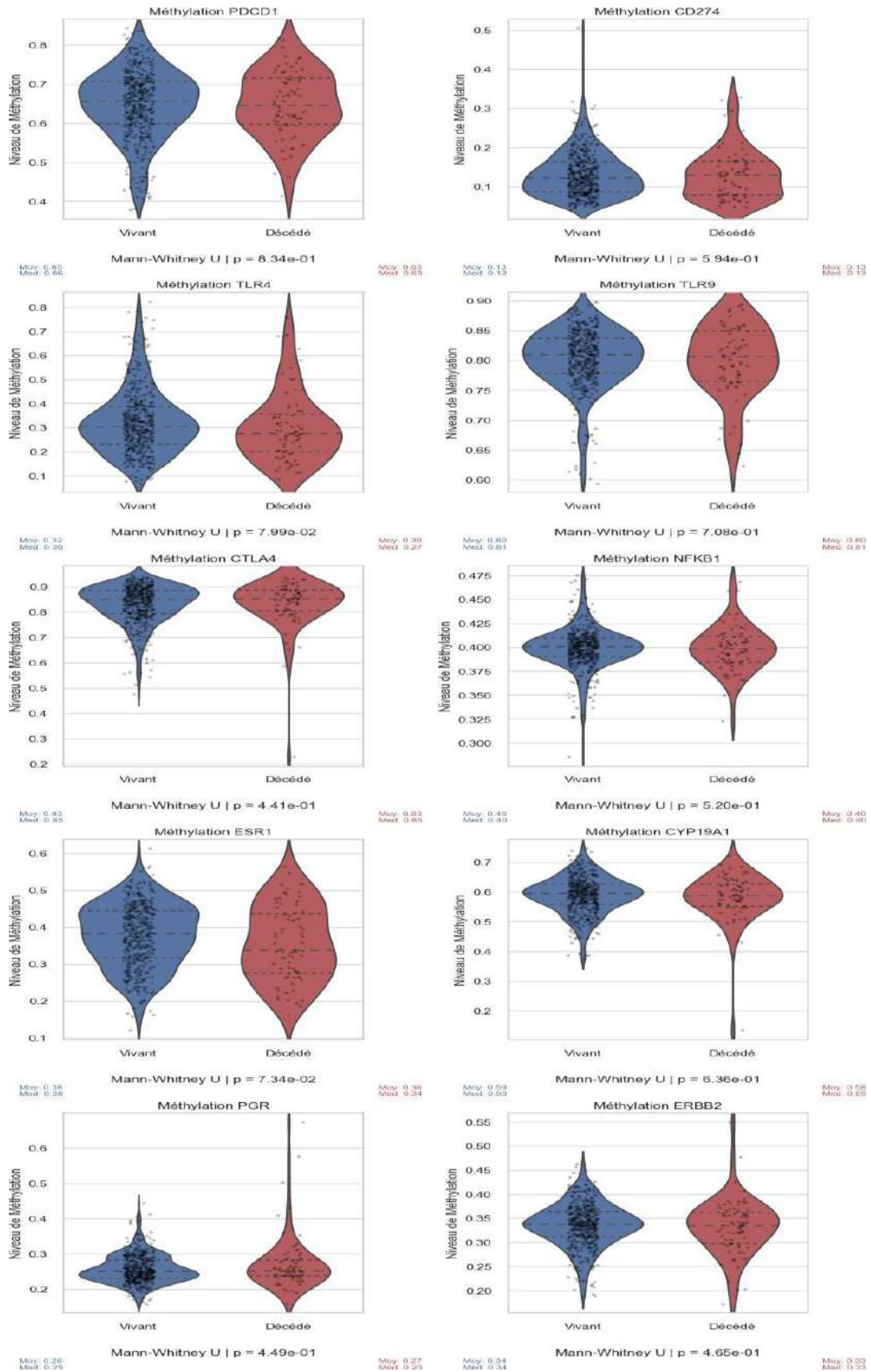


Figure 76. Distribution des gènes méthylés, immunitaires et hormonaux, selon la survie dans la population totale.

8.1.2. Gènes exprimés

L'analyse comparative des niveaux d'expression génique de l'ADN méthylé (**Figure 77**), entre les patientes a mis en évidence trois gènes présentant des différences significatives selon le statut de survie.

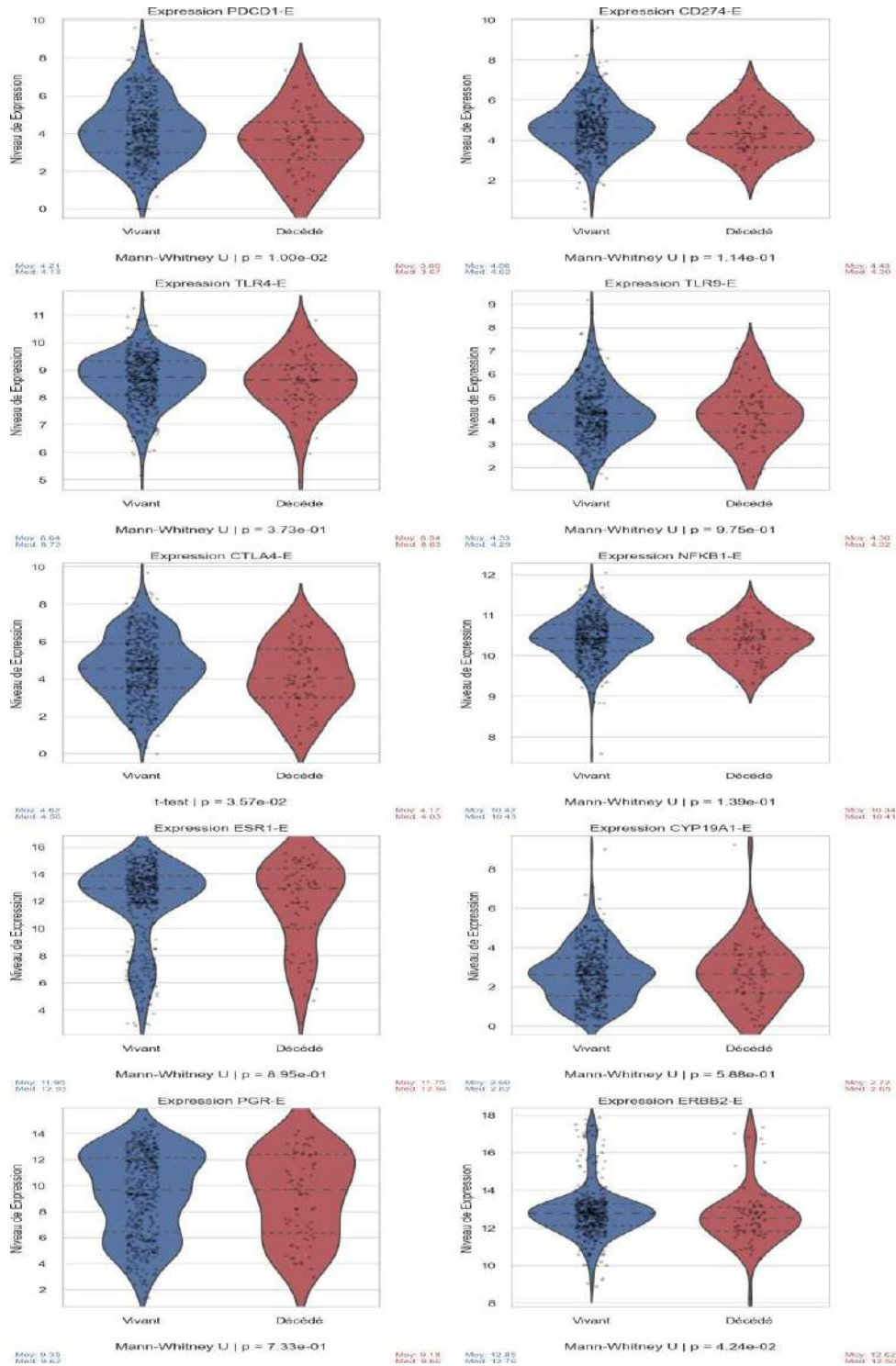


Figure 77. Distribution de l'expression des gènes méthylés, immunitaires et hormonaux, selon la survie dans la population totale.

Les gènes exprimés PDCD1-E ($p = 0,01$) et CTLA4-E ($p = 0,0366$), sont impliqués dans la régulation immunitaire, et sont surexprimés chez les patientes vivantes, suggérant une réponse immunitaire anti-tumorale plus active et potentiellement protectrice.

De manière plus inattendue, ERBB2-E ($p = 0,0424$), codant pour le récepteur HER2, est également plus exprimé chez les patientes vivantes. Cette observation pourrait s'expliquer par l'efficacité des thérapies ciblées anti-HER2, comme le trastuzumab, chez ces patientes.

- ❖ Dans l'ensemble, l'expression génique semble un meilleur discriminant que la méthylation pour différencier les patientes selon leur issue clinique, et constitue une piste pertinente pour le développement d'approches pronostiques et thérapeutiques personnalisées.

8.2. Survie globale en fonction des profils épigénétiques et transcriptomiques

Pour évaluer les chances de survie et le risque de décès, nous avons utilisé la méthode d'analyse de survie de Kaplan-Meier.

8.2.1. Gènes immunitaires : méthylation et expression génique

La **Figure 78** montre des courbes de survie de Kaplan-Meier pour quatre groupes de patientes, classées selon deux critères biologiques :

- ✓ **Meth** : le niveau de méthylation (modification épigénétique de l'ADN)
- ✓ **Exp** : le niveau d'expression (quantité de protéine produite à partir du gène)

Les groupes sont :

- ✓ **High_Meth_Low_Exp** : forte méthylation, faible expression
- ✓ **High_Meth_High_Exp** : forte méthylation, forte expression
- ✓ **Low_Meth_High_Exp** : faible méthylation, forte expression
- ✓ **Low_Meth_Low_Exp** : faible méthylation, faible expression

Chaque courbe représente la probabilité de survie au fil du temps pour l'un des groupes. Les zones colorées autour des courbes indiquent l'intervalle de confiance (la marge d'incertitude) pour chaque estimation de survie. On observe que les courbes descendent progressivement, ce qui signifie que la probabilité de survie diminue avec le temps.

Les différences entre les courbes suggèrent que la combinaison de la méthylation et de l'expression du gène PDCD1 influence la survie des patients.

a/ PDCDI :

Les patients ayant une faible méthylation et une forte expression du gène *PDCDI* (*Low_Meth_High_Exp*) ont la meilleure survie, tandis que ceux avec une forte méthylation et une faible expression (*High_Meth_Low_Exp*) ont la moins bonne survie.

Ces résultats s'interprètent à la lumière du rôle immunomodulateur de *PDCDI* :

* une expression réduite pourrait refléter une incapacité du système immunitaire à détecter ou à contrôler efficacement les cellules tumorales, facilitant ainsi l'échappement immunitaire. Ceci semble être un facteur de mauvais pronostic, associé à une survie réduite.

* à l'inverse, les groupes caractérisés par une expression élevée de *PDCDI* présentent une meilleure survie, indépendamment du niveau de méthylation. Ceci suggère un rôle protecteur ou bénéfique de *PDCDI* dans ce contexte.

Ces résultats suggèrent qu'un contrôle immunitaire renforcé, possiblement induit par une expression accrue de *PDCDI*, pourrait améliorer la réponse immunitaire et ralentir la progression tumorale. En application clinique, ceci présente des avantages :

- ❖ **Stratification du risque** : Ces résultats peuvent aider à identifier les patients à haut risque (*High_Meth_Low_Exp*) qui pourraient bénéficier d'un suivi plus rapproché ou de traitements spécifiques.
- ❖ **Ciblage thérapeutique** : Le statut de méthylation et d'expression de *PDCDI* pourrait guider l'utilisation de thérapies immunomodulatrices ou épigénétiques.
- ❖ **Biomarqueurs pronostiques** : *PDCDI*, Méthylé et Exprimé, pourrait servir de biomarqueur pour prédire la survie et adapter la prise en charge.

b/ CD274 (PD-L1) :

L'analyse combinée des niveaux de méthylation (profil épigénétique) et d'expression de *CD274* (*PD-L1*) influencent fortement la survie des patients. Les patients avec une faible méthylation et une forte expression de *CD274* ont la meilleure survie, tandis que ceux avec une forte méthylation et une faible expression ont la survie la plus courte. Le statut épigénétique (méthylation) et l'expression du gène *CD274* peuvent être utilisés comme biomarqueurs pronostiques pour stratifier les patients et adapter leur prise en charge. Ces informations peuvent aider à personnaliser le suivi et les traitements en fonction du risque.

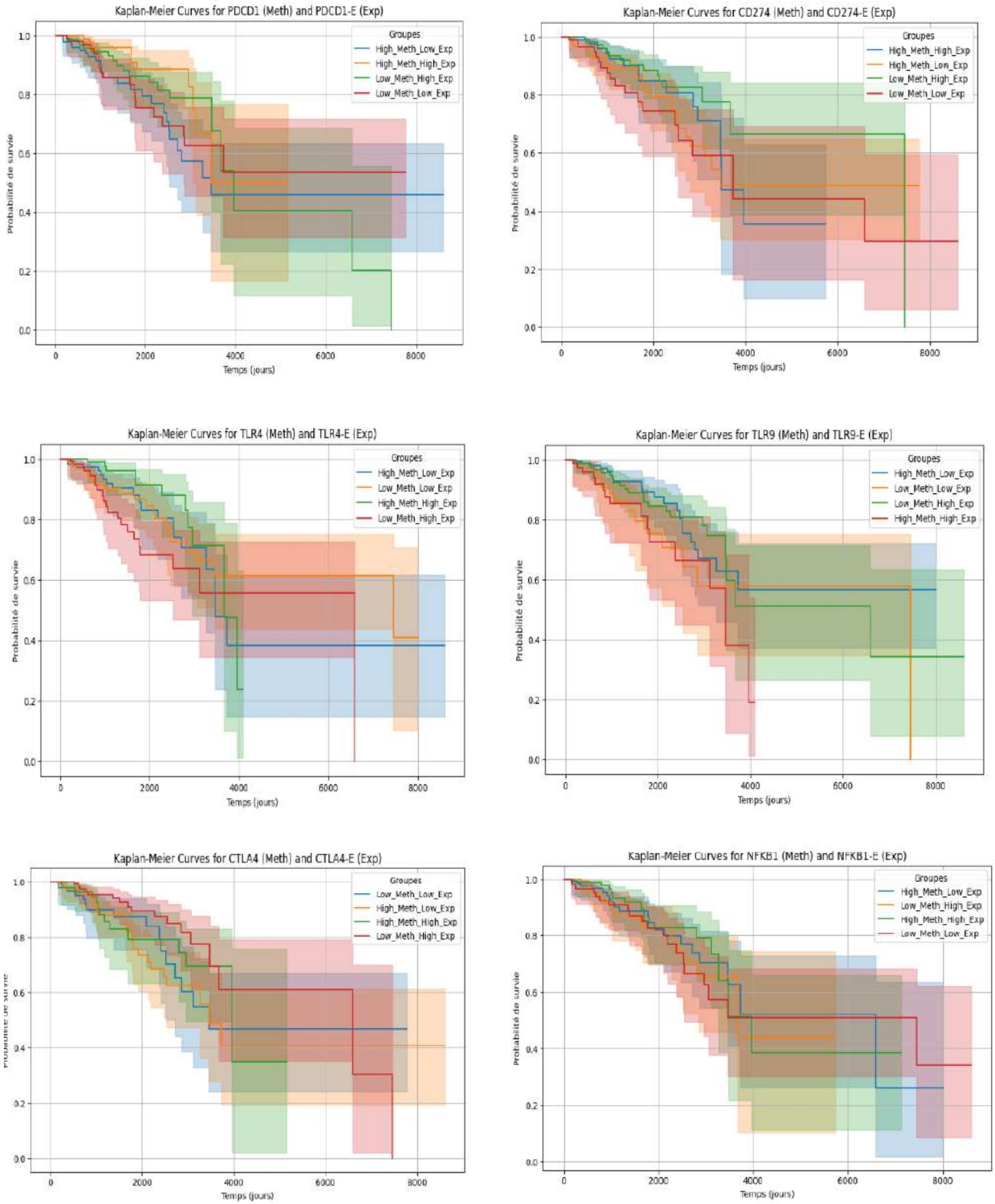


Figure 78. Courbes de survie de Kaplan-Meier en fonction des gènes immunitaires : méthylation et expression génique.

c/ TLR4 :

Les groupes avec une méthylation faible et une expression élevée de *TLR4* montrent une survie globale nettement supérieure. Cette combinaison pourrait indiquer un état immunitaire activé, dans lequel *TLR4* contribue à l'activation de réponses inflammatoires antitumorales. À l'opposé, le groupe *High_Meth_Low_Exp* est associé à la survie la plus faible, suggérant qu'une suppression de *TLR4*, potentiellement par méthylation, pourrait limiter la surveillance immunitaire et aggraver le pronostic.

d/ TLR9 :

Concernant *TLR9*, les patientes du groupe *Low_Meth_High_Exp* bénéficient également d'une meilleure survie, ce qui indique un effet protecteur probable d'une expression conservée du gène. Bien qu'une méthylation élevée soit généralement associée à une réduction d'expression, les patientes du groupe *High_Meth_High_Exp* présentent une survie légèrement améliorée par rapport à celles de *High_Meth_Low_Exp*, ce qui pourrait refléter une compensation partielle par l'expression, malgré une méthylation élevée. Ces résultats renforcent le rôle immunitaire positif de *TLR9* dans le microenvironnement tumoral.

e/ CTLA4 :

La méthylation élevée de *CTLA4* semble associée à une diminution marquée de la survie, en particulier dans le groupe *High_Meth_Low_Exp*. Cette configuration suggère que la répression épigénétique de *CTLA4* pourrait nuire à la régulation fine des réponses immunitaires, contribuant à un microenvironnement tumoral permissif. Le groupe *Low_Meth_High_Exp*, caractérisé par une faible méthylation et une expression élevée, montre la meilleure probabilité de survie à long terme, illustrant une configuration favorable. Les autres groupes présentent des profils intermédiaires, confirmant l'importance de la régulation combinée (épigénétique et transcriptionnelle) dans la modulation du pronostic.

f/ NF-κB1 :

Enfin, les analyses de survie selon la méthylation et l'expression de *NF-κB1* révèlent des différences significatives apparaissant surtout après 4000 jours. Le groupe *High_Meth_Low_Exp* est associé à une survie globale réduite, probablement liée à une réduction de l'activité transcriptionnelle de *NF-κB1* en raison de la méthylation. En revanche, les patientes appartenant au groupe *Low_Meth_High_Exp* présentent une meilleure survie, confirmant l'intérêt potentiel de ce gène en tant que biomarqueur pronostique et cible immuno-thérapeutique.

Le statut combiné de méthylation et d'expression du gène PDCD1 a une valeur pronostique importante.

8.2.2. Gènes hormonaux : méthylation et expression génique

L'analyse des courbes de survie de Kaplan-Meier en fonction de la méthylation et de l'expression génique des marqueurs hormonaux est représentée dans la **Figure 79**.

a/ ERBB2 :

La méthylation élevée du gène *ERBB2* est associée à une expression diminuée et à une survie défavorable, comme observé dans les groupes *High_Meth_Low_Exp* et *High_Meth_High_Exp*. En revanche, le groupe *Low_Meth_High_Exp*, combinant faible méthylation et forte expression, montre une meilleure survie, suggérant que la dé-répression épigénétique d'*ERBB2* pourrait favoriser une réponse thérapeutique ou indiquer une biologie tumorale moins agressive.

b/ ESRI :

Le groupe *Low_Meth_High_Exp* pour *ESRI* présente une meilleure survie, ce qui soutient l'hypothèse que la faible méthylation favorise une expression élevée de ce récepteur hormonal. Ce profil est souvent associé à une sensibilité accrue aux traitements hormonaux, tels que le tamoxifène. À l'inverse, la forte méthylation et la faible expression (*High_Meth_Low_Exp*) sont liées à un pronostic plus sombre, probablement en lien avec une résistance aux thérapies endocriniennes et une perte de la dépendance hormonale.

c/ CYP19A1 :

Les analyses révèlent une meilleure survie dans le groupe *Low_Meth_High_Exp*, suggérant que la faible méthylation et l'expression accrue de *CYP19A1* (gène codant l'aromatase) pourraient favoriser un meilleur contrôle tumoral, notamment dans les cancers hormono-dépendants. Le groupe *High_Meth_Low_Exp* est associé à une survie réduite, renforçant l'idée qu'une répression épigénétique du métabolisme des œstrogènes influence négativement le pronostic. Les variations observées dans les groupes à forte expression, malgré la méthylation élevée, indiquent toutefois une complexité biologique modulée, très probablement, par d'autres facteurs.

d/ PGR :

La survie prolongée observée dans le groupe *Low_Meth_High_Exp* pour *PGR* confirme que l'activation de ce récepteur, en l'absence de méthylation inhibitrice, est bénéfique. À l'opposé, les patients du groupe *High_Meth_Low_Exp* présentent une survie plus courte, traduisant une

possible inactivation fonctionnelle du récepteur à la progestérone. Ces résultats suggèrent que la méthylation du gène *PGR* influence la réponse hormonale, et par conséquent le pronostic.

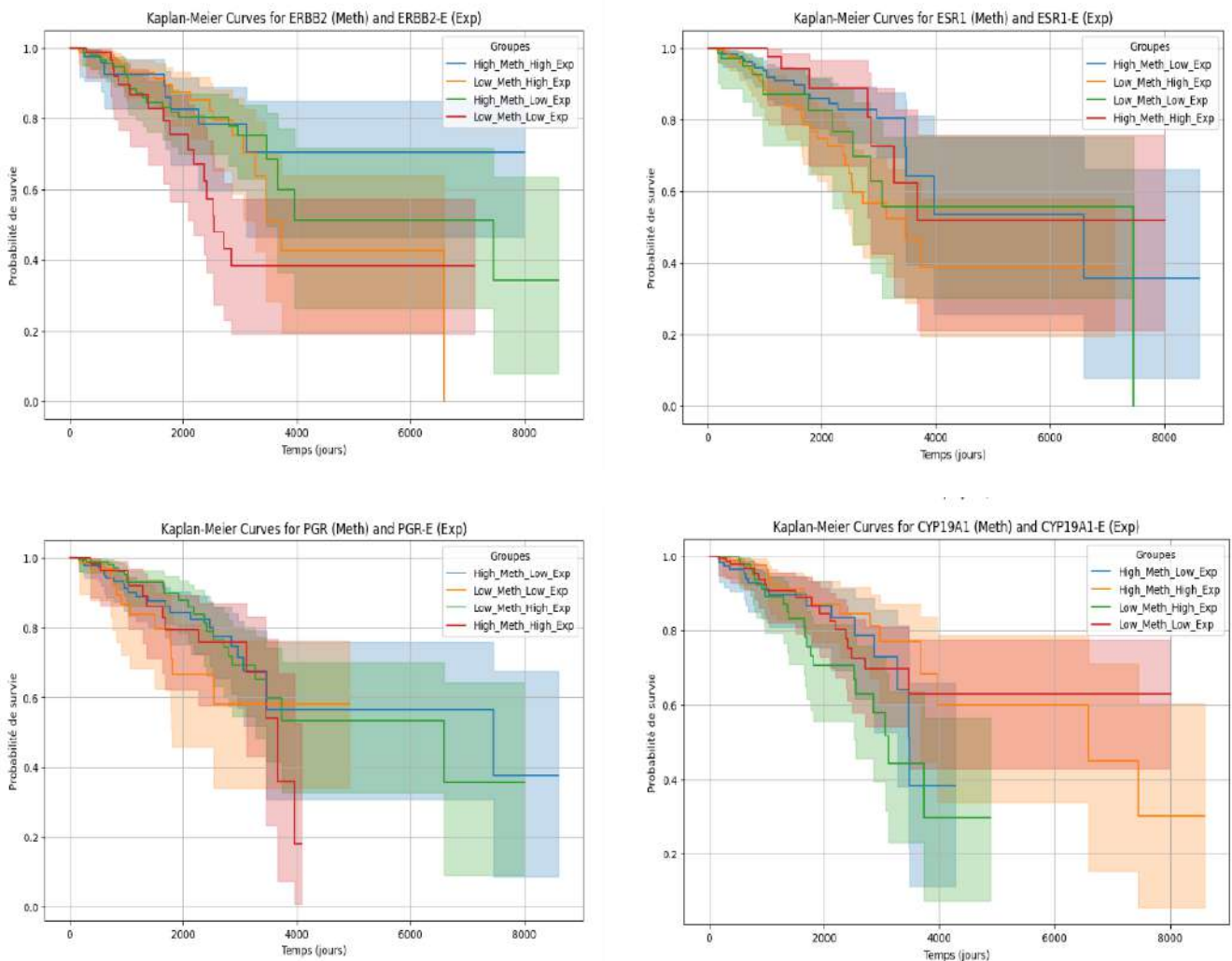


Figure 79. Courbes de survie de Kaplan-Meier en fonction des gènes hormonaux : méthylation et expression génique.

8.2.3. Impact de la Méthylation et de l'Expression Génique sur la Survie des Patients

Cette section présente les résultats combinés de l'analyse ANOVA et des tests de Log-Rank réalisés pour évaluer l'impact de la méthylation de l'ADN et de l'expression génique sur la survie des patientes atteintes de cancer du sein.

a. Analyse de Variance (ANOVA) :

L'ANOVA a été utilisée pour explorer les variations de l'expression génique en fonction des groupes de survie. Les résultats sont synthétisés ci-dessous (**Tableau 18**) :

Tableau 18. Variations de l'expression génique en fonction des groupes de survie.

Gène	F-value	p-value
<i>PDCD1</i>	0.0947	0.9630
<i>TLR4</i>	2.6508	0.0482
<i>TLR9</i>	1.3056	0.2718
<i>CTLA4</i>	1.2747	0.2823
<i>NFKB1</i>	1.6485	0.1773
<i>ESR1</i>	3.9496	0.0084
<i>CYP19A1</i>	2.1234	0.1356
<i>CD274</i>	0.2670	0.8492
<i>PGR</i>	0.8465	0.4689
<i>ERBB2</i>	1.5227	0.2077

Seuls *TLR4* et *ESR1* montrent une variation significative d'expression selon les groupes, suggérant leur implication potentielle dans les mécanismes liés à la survie.

b. Tests de Log-Rank : Association entre Groupes Méthylation/Expression et Survie

Des groupes ont été formés selon les combinaisons de méthylation (haute/basse) et d'expression (haute/basse), et comparés *via* des tests de Log-Rank.

* **PDCD1** : chez les patientes fortement méthylées, une expression élevée de *PDCD1* est associée à une meilleure survie (*High_Meth_Low_Exp* vs *High_Meth_High_Exp* (p = 0.0314)).

* **CD274 (PD-L1)** : dans un contexte de faible méthylation, une expression plus élevée de *CD274* est associée à une meilleure survie (*Low_Meth_High_Exp* vs *Low_Meth_Low_Exp* (p = 0.0321)).

* **TLR4** : la combinaison méthylation/expression suggère un effet protecteur de *TLR4* lorsqu'il est à la fois méthylé et exprimé. (*High_Meth_High_Exp* vs *Low_Meth_High_Exp* (p = 0.0144)).

* **TLR9** : l'expression de *TLR9* dans un contexte de forte méthylation pourrait jouer un rôle dans la modulation de la survie. (*High_Meth_Low_Exp* vs *High_Meth_High_Exp* ($p = 0.0466$) et *Low_Meth_High_Exp* vs *High_Meth_High_Exp* ($p = 0.0529$)).

* **CTLA4** : l'expression accrue de *CTLA4*, même en présence de faible méthylation, pourrait être bénéfique en termes de survie (*High_Meth_Low_Exp* vs *Low_Meth_High_Exp* ($p = 0.0309$)).

* **NFKB1** : aucun test n'est significatif. Le gène *NFKB1* ne semble pas influencer la survie dans ce contexte.

* **ESR1** : l'expression élevée d'*ESR1* dans un contexte de faible méthylation est associée à une meilleure survie (*High_Meth_Low_Exp* vs *Low_Meth_High_Exp* ($p = 0.0180$)).

* **CYP19A1** : (Aromatase) : ce gène hormonal semble important dans la survie, surtout lorsqu'il est faiblement méthylé et fortement exprimé (*High_Meth_Low_Exp* vs *Low_Meth_High_Exp* ($p = 0.0425$)).

* **ERBB2** : l'association positive entre méthylation et expression d'*ERBB2* pourrait indiquer un rôle complexe dans la régulation tumorale (*High_Meth_High_Exp* vs *Low_Meth_High_Exp* ($p = 0.0274$)).

* **PGR** : ce gène semble associé à un meilleur pronostic lorsque son expression est élevée dans un contexte de faible méthylation (*High_Meth_Low_Exp* vs *Low_Meth_High_Exp* ($p = 0.0298$)).

➤ Les résultats mettent en évidence plusieurs gènes dont la méthylation et/ou l'expression influencent la survie des patientes, notamment :

- *TLR4*, *ESR1*, *PDCD1*, *CTLA4*, *CYP19A1*, *ERBB2*, et *PGR*.
- *TLR4* et *ESR1* ressortent avec des résultats significatifs à la fois en ANOVA et en tests de Log-Rank, suggérant un rôle clé dans la biologie tumorale et la survie.

8.3. Analyse de la survie par cluster

Seul le gène *ESR1* montre une différence significative de méthylation dans le cluster 0 entre les patientes vivants et décédés ($p = 0.039$). Cela suggère un lien potentiel entre la méthylation d'*ESR1* et la survie globale dans ce sous-groupe (**Tableau 19**).

Tableau 19. Comparaison de la méthylation des gènes entre les clusters et les statuts de survie.

Gènes	Clusters	OS	Moyenne	Écart-type	Médiane	p-value
<i>PDCD1</i>	0	V	0.660	0.088	0.675	0.560
		D	0.657	0.077	0.669	
	1	V	0.635	0.086	0.643	0.791
		D	0.633	0.095	0.621	
<i>CD274</i>	0	V	0.135	0.059	0.123	0.527
		D	0.131	0.060	0.129	
	1	V	0.126	0.054	0.119	0.790
		D	0.128	0.068	0.128	
<i>TLR4</i>	0	V	0.311	0.140	0.288	0.101
		D	0.287	0.152	0.249	
	1	V	0.331	0.127	0.311	0.680
		D	0.322	0.130	0.306	
<i>TLR9</i>	0	V	0.815	0.038	0.816	0.662
		D	0.811	0.049	0.814	
	1	V	0.789	0.058	0.794	0.402
		D	0.777	0.071	0.787	
<i>CTLA4</i>	0	V	0.835	0.083	0.857	0.727
		D	0.829	0.119	0.862	
	1	V	0.827	0.069	0.841	0.620
		D	0.840	0.051	0.848	
<i>NFKB1</i>	0	V	0.399	0.023	0.399	0.099
		D	0.394	0.024	0.394	
	1	V	0.400	0.019	0.400	0.219
		D	0.405	0.021	0.405	
<i>ESR1</i>	0	V	0.352	0.081	0.352	0.039
		D	0.326	0.087	0.318	
	1	V	0.399	0.082	0.412	0.649
		D	0.401	0.104	0.425	
<i>CYP19A1</i>	0	V	0.581	0.067	0.593	0.379
		D	0.566	0.089	0.578	
	1	V	0.594	0.055	0.593	0.476
		D	0.604	0.049	0.603	
<i>PGR</i>	0	V	0.247	0.032	0.245	0.984
		D	0.248	0.034	0.240	
	1	V	0.266	0.045	0.259	0.086

Gènes	Clusters	OS	Moyenne	Écart-type	Médiane	p-value
		D	0.309	0.109	0.275	
ERBB2	0	V	0.325	0.039	0.333	0.887
		D	0.322	0.049	0.328	
	1	V	0.345	0.047	0.344	0.737
		D	0.345	0.065	0.347	

V : Vivants ; D : Décédés

Une expression significativement plus élevée de *PDCDI* ($p = 0.015$) et *CD274* ($p = 0.004$) est observée chez les patientes vivantes du cluster 0. Cela pourrait indiquer un rôle protecteur de ces gènes immuno-modulateurs dans ce sous-groupe (**Tableau 20**).

Tableau 20. Comparaison de l'expression génique entre les clusters et les statuts de survie.

Gènes	Clusters	OS	Moyenne	DS	Médiane	p-value
PDCDI-E	0	V	3.848	1.441	3.860	0.015
		D	3.199	1.609	3.348	
	1	V	4.552	1.891	4.419	0.401
		D	4.210	1.652	4.156	
CD274-E	0	V	4.406	1.048	4.462	0.004
		D	3.985	0.828	3.964	
	1	V	4.890	1.423	4.768	0.234
		D	5.107	1.148	5.225	
TLR4-E	0	V	8.549	0.881	8.626	0.114
		D	8.330	0.898	8.445	
	1	V	8.726	1.026	8.868	0.455
		D	8.851	1.253	8.912	
TLR9-E	0	V	3.953	0.996	3.905	0.529
		D	4.057	1.179	4.075	
	1	V	4.683	1.243	4.505	0.782
		D	4.681	1.317	4.784	
CTLA4-E	0	V	4.192	1.447	4.250	0.077
		D	3.790	1.627	3.639	
	1	V	5.015	1.834	5.000	0.550

		D	4.755	1.856	4.869	
NFKB1-E	0	V	10.363	0.558	10.410	0.254
		D	10.280	0.441	10.315	
	1	V	10.480	0.512	10.525	0.699
		D	10.42	0.524	10.440	
ESR1-E	0	V	13.077	2.177	13.480	0.582
		D	12.611	2.729	13.355	
	1	V	10.905	3.264	12.145	0.542
		D	10.437	3.446	11.265	
CYP19A1-E	0	V	2.602	1.445	2.617	0.600
		D	2.456	1.431	2.580	
	1	V	2.600	1.304	2.596	0.101
		D	3.117	1.722	2.891	
PGR-E	0	V	9.994	2.944	10.120	0.547
		D	9.634	3.065	10.200	
	1	V	8.742	3.466	9.322	0.666
		D	8.494	3.740	8.223	
ERBB2-E	0	V	12.992	1.349	12.819	0.064
		D	12.733	1.293	12.570	
	1	V	12.722	1.464	12.620	0.139
		D	12.453	1.884	12.280	

V : Vivants ; D : Décédés

9. PRONOSTIC VITAL : COURBE ROC

Afin de mieux comprendre le rôle pronostic des gènes méthylés (Biomarqueurs Epigénétiques) et leurs expressions (Biomarqueurs Transcriptionnels) dans les cas de cancer du sein, une analyse ROC a été réalisée.

La courbe ROC (Receiver Operating Characteristic) est un graphique utilisé pour évaluer la performance d'un test de diagnostic, de pronostic ou d'un système de classification binaire. Il compare le taux de vrais positifs (sensibilité) au taux de faux positifs (1-spécificité) pour différents seuils de décision. Chaque point sur la courbe ROC représente une combinaison de

sensibilité et de spécificité pour un seuil donné. Plus la courbe s'approche du coin supérieur gauche du graphique, meilleure est la performance du test (haute sensibilité et haute spécificité).

L'interprétation de l'AUC (Aire sous la courbe) est comme suit :

- * **AUC = 1** : Test parfait (100% de sensibilité et de spécificité)
- * **AUC > 0,9** : Excellent test
- * **AUC entre 0,7 et 0,9** : Bon test
- * **AUC = 0,5** : Test sans valeur (équivalent à un choix aléatoire)

9.1. Analyse ROC dans la population totale

9.1.1. Courbe ROC des gènes méthylés (Biomarqueurs Epigénétiques)

L'analyse des performances (AUC) est représentée dans la **Figure 80** :

- ✓ ***TLR4*** (AUC = **0.984**) est le meilleur marqueur du panel de gènes analysés. Ce résultat traduit une excellente performance indiquant que la méthylation de ce gène permet une discriminer presque parfaitement dans ce groupe.
- ✓ ***PDCDI*** (AUC = **0.676**), ***CYP19A1*** (AUC = **0.690**) et ***CD274*** (AUC = **0.666**) montrent une bonne capacité de discrimination génique, donc une bonne performance, mais restent en dessous du précédent biomarqueur.
- ✓ ***CTL4A*** (AUC = **0.657**), ***NF-κB1*** (AUC = **0.652**), ***TLR9*** (AUC = **0.605**), ***ESR1*** (AUC = **0.624**), ***PGR*** (AUC = **0.604**) et ***ERBB2*** (AUC = **0.569**). Ces gènes ont un pouvoir discriminant faible à modéré, leur utilité comme biomarqueurs est donc limitée. Leurs performances demeurent faibles à modérées.

Ainsi, *TLR4* et *CD274* sont d'excellents biomarqueurs pour différencier les gènes étudiés (AUC très proches de 1). Ils sont les meilleurs candidats pour le développement de tests diagnostiques basés sur la méthylation dans le cancer du sein. Les gènes *CYP19A1* et *PDCDI* sont de bons candidats, mais leur performance est moindre. Les autres gènes sont moins discriminants et donc moins intéressants pour un diagnostic ou pronostic fiable basé sur la méthylation. Ils pourraient être utilisés en combinaison (panel de biomarqueurs) pour améliorer la sensibilité et la spécificité.

Implications cliniques : Un test basé sur la méthylation de *TLR4* et/ou *CD274* pourrait permettre un diagnostic précoce, un suivi de la maladie ou une stratification des patientes pour des traitements personnalisés.

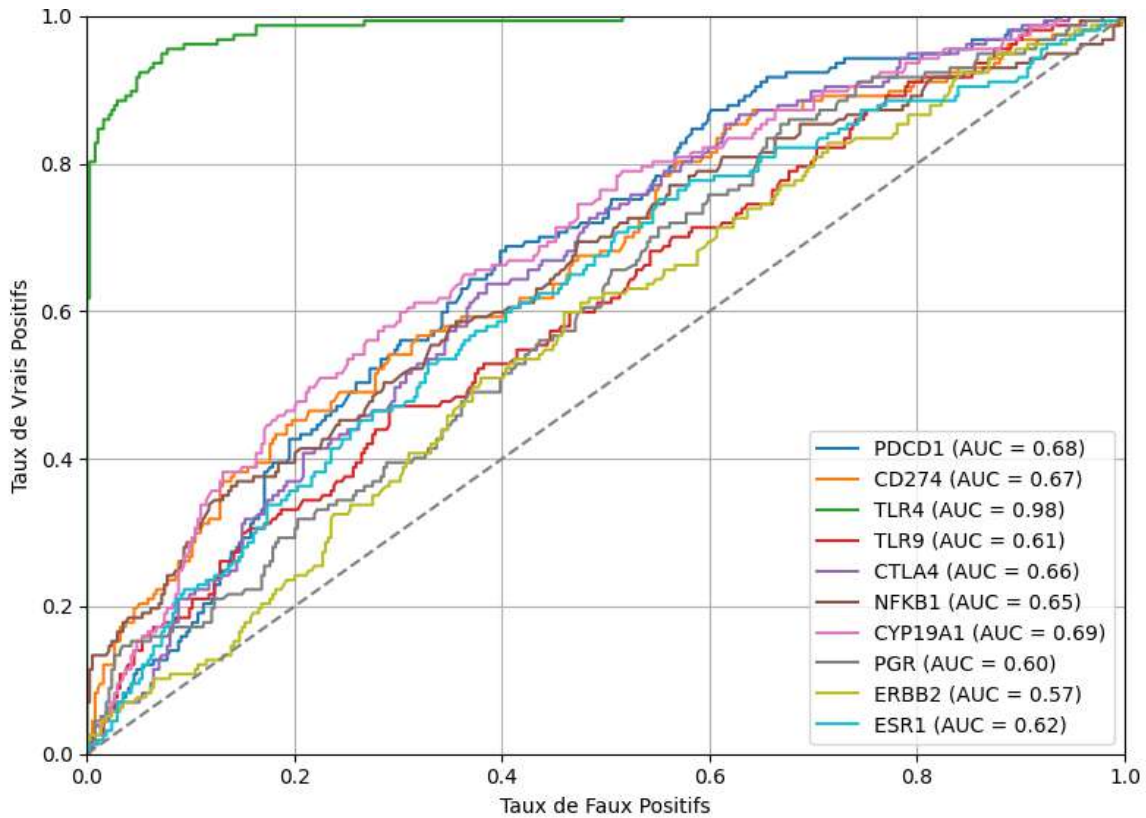


Figure 80. Evaluation de la valeur prédictive des biomarqueurs épigénétiques par la courbe ROC.

Chaque courbe représente un gène méthylé, et l'aire sous la courbe (AUC) indique la capacité discriminante du marqueur. La droite en pointillés représente un test aléatoire (AUC = 0,5).

ROC : Receiver Operating Characteristic ; **AUC** : Area Under Curve.

9.1.2. Courbe ROC des gènes exprimés (Biomarqueurs Transcriptionnels)

L'analyse des performances (AUC) de l'expression des gènes méthylés est représentée dans la **Figure 81** :

- ✓ **ESR1-E (AUC = 0,853)** est le meilleur biomarqueur. Ce résultat traduit une excellente performance indiquant que la forme exprimé de ce gène méthylé permet une discriminer presque parfaitement dans ce groupe de gènes.
- ✓ **PGR-E (AUC = 0,624)** montrent une bonne capacité de discrimination génique, mais demeure modérément performants
- ✓ **ERBB2-E (AUC = 0,587)** est peu discriminant.
- ✓ **CYP19A1-E (AUC = 0,447), NF-κB1-E (AUC = 0,423), TLR4-E (AUC = 0,383), CTLA4-E (AUC = 0,309), CD274-E (AUC = 0,326), PDCD1-E (AUC = 0,325) et TLR9-E (AUC = 0,260)** ont un pouvoir discriminant très faible, voire inexistant. Une AUC

inférieure à 0,5 indique même que le test donne plus souvent de mauvais résultats que le hasard.

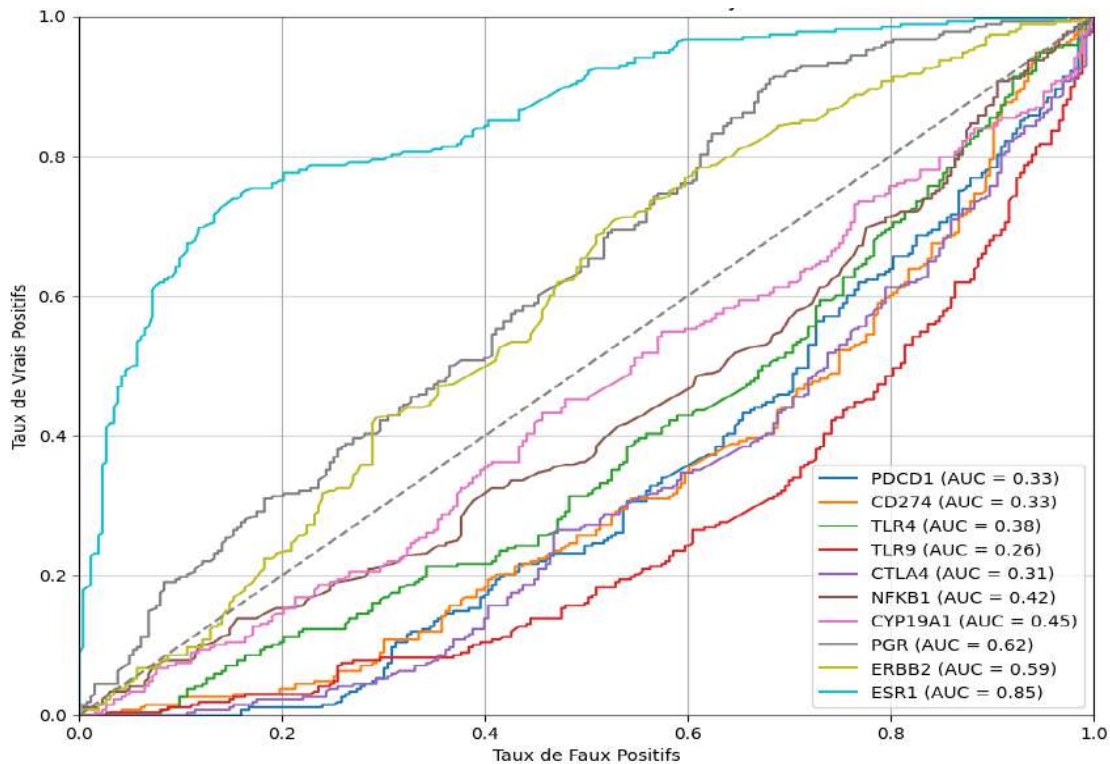


Figure 81. Evaluation de la valeur prédictive des biomarqueurs transcriptionnels par la courbe ROC.

En conclusion, ESR1-E est le seul biomarqueur transcriptionnel fiable pour distinguer les gènes testés. PGR-E et ERBB2-E pourraient être utilisés, mais ils restent moyens.

9.2. Analyse ROC de la combinaison génique dans le groupe des gènes méthylés et celui des gènes exprimés dans la population totale

9.2.1. Cas des gènes méthylés (Biomarqueurs Epigénétiques)

L'analyse des courbes ROC des combinaisons des biomarqueurs méthylés est représentée dans la **Figure 82**.

- ✓ **CD274_PDCD1 (AUC = 0,632)** est la combinaison la plus performante et présente la plus grande AUC, ce qui signifie qu'elle offre la meilleure capacité discriminante parmi toutes celles testées. Elle pourrait servir de base à un test de dépistage ou d'aide au diagnostic du cancer du sein, mais la performance reste modérée ($AUC < 0,7$).
- ✓ **TLR4_CTLA4 (AUC = 0,511)**, cette combinaison a un AUC très proche de 0,5, ce qui indique un pouvoir discriminant très faible.

✓ *CYP19A1_TLR4* (AUC = 0,438), *TLR4_PGR* (AUC = 0,401), *NF-κB1_ERBB2* (AUC = 0,364), *CYP19A1_ERBB2* (AUC = 0,338), *CYP19A1_ESR1* (AUC = 0,234) et *ESR1_PGR* (AUC = 0,196) sont des combinaisons peu ou pas discriminantes.

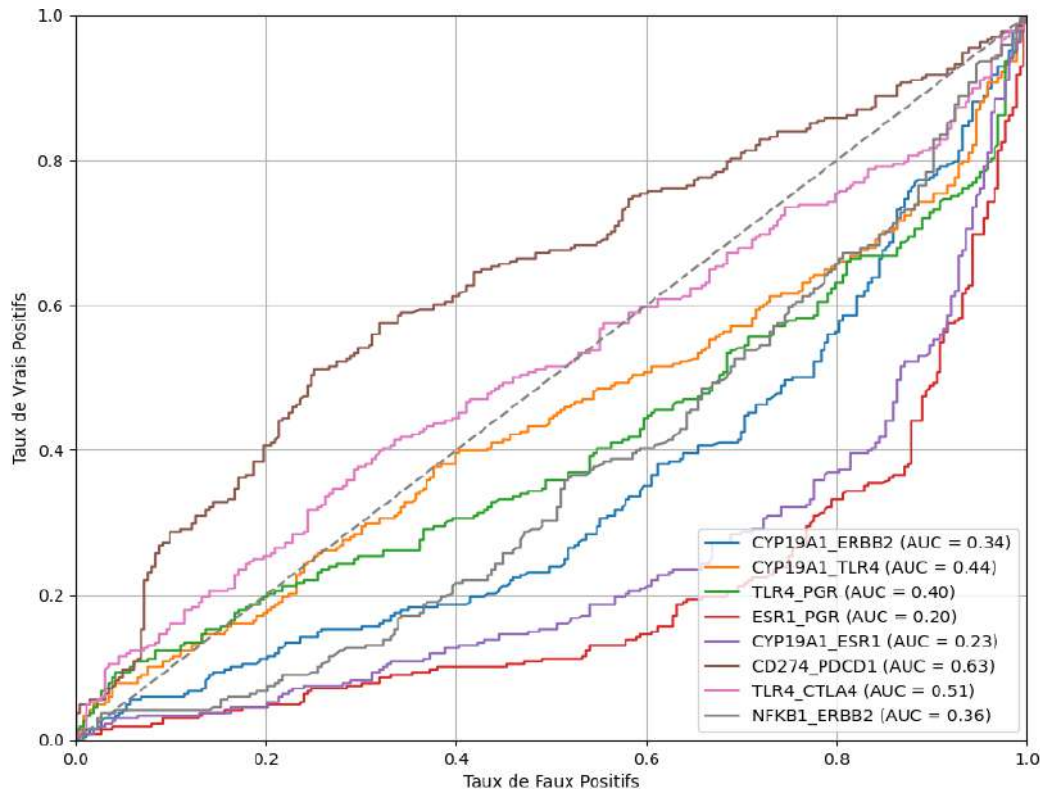


Figure 82. Evaluation de la valeur prédictive de la combinaison des biomarqueurs épigénétiques par la courbe ROC.

Un AUC < 0,5 indique que ces combinaisons ont un pouvoir discriminant inférieur au hasard, ce qui signifie qu'elles ne permettent pas de différencier efficacement les patientes atteintes de cancer du sein des femmes saines. Son utilité clinique est très limitée, voire nulle.

CD274_PDCD1 est la seule combinaison qui montre un intérêt potentiel, mais sa performance reste modérée (AUC = 0,632), puisqu'elle n'atteint pas le seuil généralement recherché pour un bon biomarqueur (AUC > 0,7). Elle pourrait être explorée davantage, éventuellement en l'associant à d'autres biomarqueurs ou méthodes diagnostiques pour améliorer la performance globale.

9.2.2. Cas des gènes exprimés (Biomarqueurs Transcriptionnels)

L'analyse des courbes ROC des combinaisons comparatives des biomarqueurs exprimés est représentée dans la **Figure 83**.

- ✓ **CYP19A1-E_ESR1-E (AUC = 0,762)**, cette combinaison a la plus grande valeur AUC, indiquant une très bonne capacité à distinguer les patientes atteintes de cancer du sein des femmes saines et prédire l'issue clinique. Ces combinaisons présentent une excellente performance et est la plus prometteuse pour un diagnostic ou un dépistage fiable.
- ✓ **ESR1-E_PGR-E (AUC = 0,714)**, cette combinaison est également très discriminante et pourrait être utilisée en clinique, seule ou en panel. Elle présente une très bonne performance. Cela montre que la combinaison des récepteurs des œstrogènes et de la progestérone peut être pertinente pour évaluer le pronostic vital. Cela suggère que l'association entre ESR1-E et PGR-E est efficace pour discriminer différents états cliniques, liés aux mécanismes hormonaux dans le cancer du sein.
- ✓ **TLR4-E_PGR-E (AUC = 0,574)**, **NF-κB1-E_ERBB2-E (AUC = 0,544)**, **CYP19A1-E_ERBB2-E (AUC = 0,507)**, ces combinaisons ont un pouvoir discriminant modéré (performance modérée), mais restent peu pertinentes pour un usage clinique isolé.
- ✓ **CYP19A1-E_TLR4-E (AUC = 0,409)**, **CD274-E_PDCD1-E (AUC = 0,305)** et **TLR4-E_CTLA4-E (AUC = 0,312)**, ces combinaisons à faible performance ne permettent pas de différencier efficacement les patientes atteintes de cancer du sein des femmes saines, leur utilité clinique est donc très limitée, voire nulle.

Ainsi, CYP19A1-E_ESR1-E et ESR1-E_PGR-E sont les meilleures combinaisons de biomarqueurs pour le diagnostic ou le dépistage du cancer du sein dans cette population, avec des AUC nettement supérieures à 0,7. Ces combinaisons montrent une capacité significative à discriminer les états cliniques liés à l'expression hormonale et au métabolisme des œstrogènes. Pour la pratique clinique, il est pertinent de privilégier les panels intégrant ces combinaisons pour améliorer la fiabilité du diagnostic et du pronostic du cancer du sein.

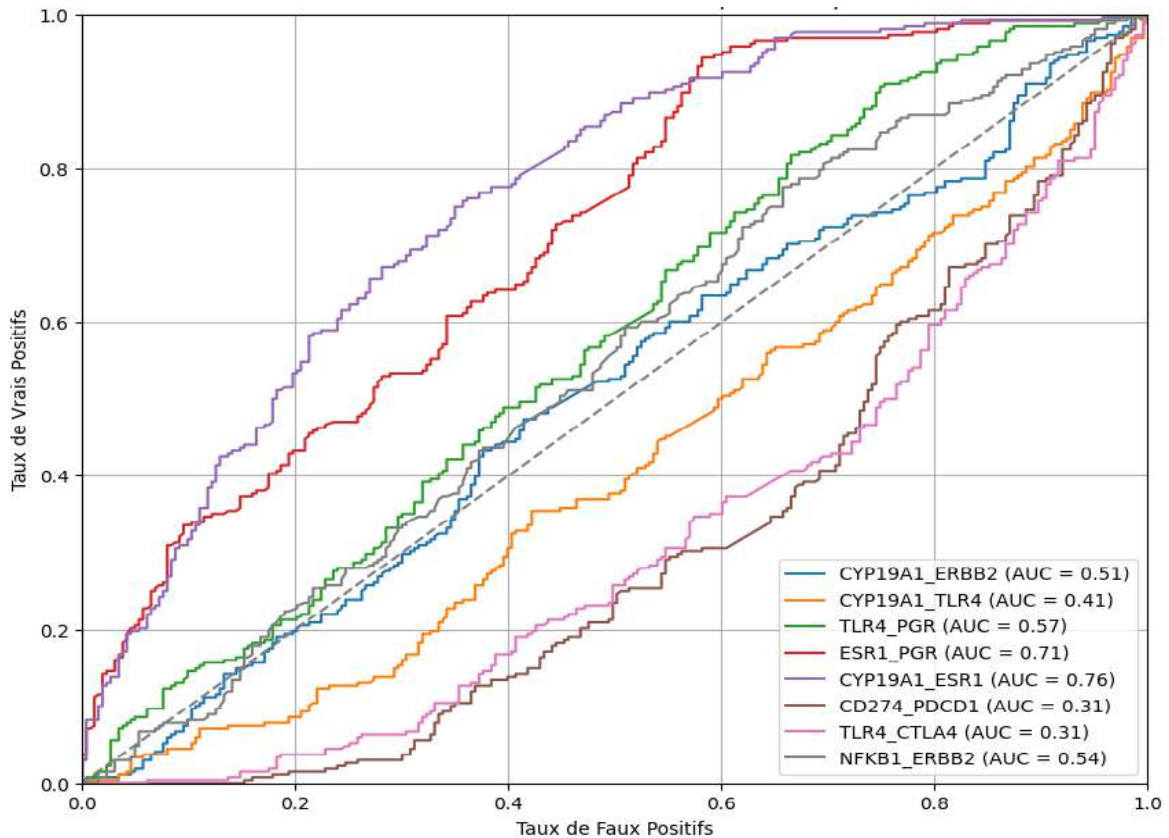


Figure 83. Evaluation de la valeur prédictive de la combinaison des biomarqueurs transcriptionnels par la courbe ROC.

En résumé, les combinaisons impliquant ESR1-E et CYP19A1-E montrent les meilleures performances, probablement en raison de leur rôle central dans la régulation hormonale du cancer du sein. Les autres combinaisons, en particulier celles impliquant CD274 et PDCD1, sont moins performantes, suggérant qu'elles ne sont pas des biomarqueurs discriminants efficaces dans ce contexte.

9.3. Analyse ROC par cluster

9.3.1. Cas du cluster 0

a/ Gènes méthylés (Biomarqueurs Epigénétiques)

Les résultats obtenus pour les biomarqueurs épigénétiques montrent des performances distinctes en termes de capacité prédictive, mesurée par l'AUC (**Figure 84**).

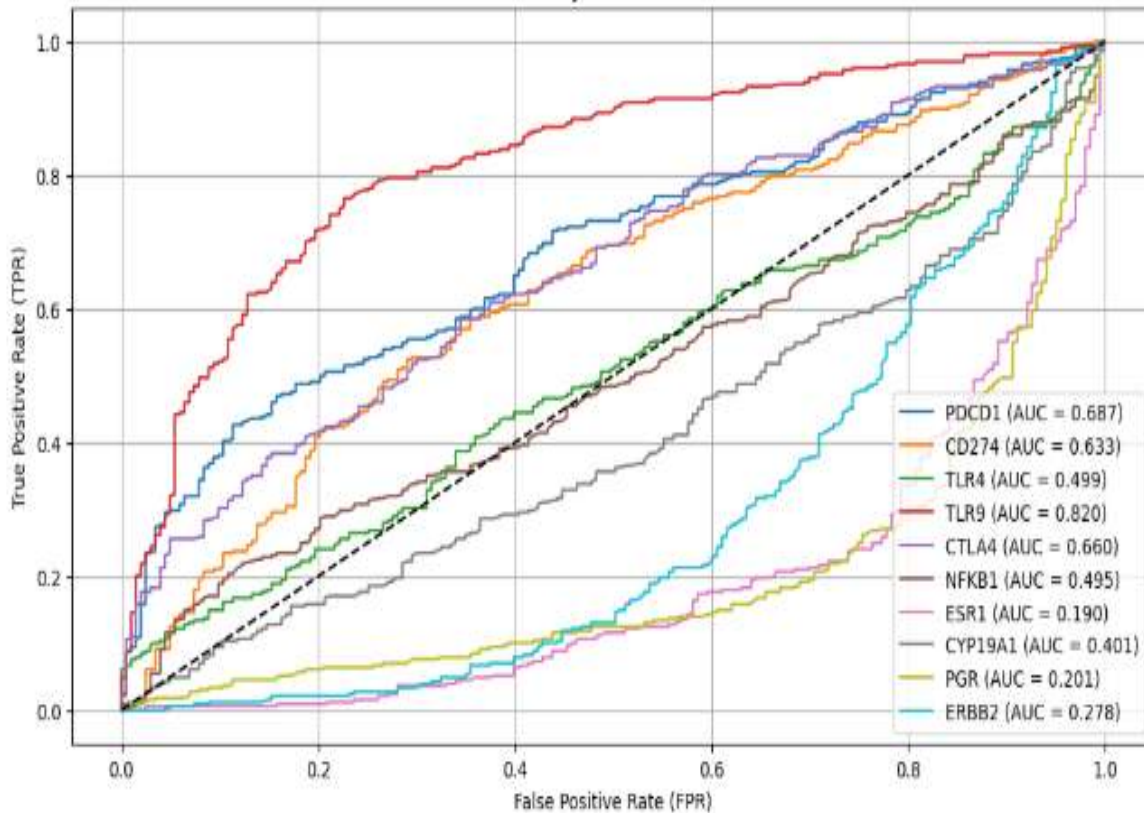


Figure 84. Evaluation de la valeur prédictive des biomarqueurs épigénétiques du cluster 0 par la courbe ROC.

- ✓ ***TLR9*** (AUC = **0.820**), présente un très bon pouvoir discriminant. Ce gène méthylé permet de distinguer efficacement les patientes atteintes de cancer du sein des témoins. *TLR9* est un excellent candidat pour le diagnostic ou le dépistage.
- ✓ ***PDCD1*** (AUC = **0.687**), présente un bon pouvoir discriminant. Ce gène permet une discrimination utile, mais moins marquée que *TLR9*.
- ✓ ***CD274*** (AUC = **0.633**), présente un pouvoir discriminant modéré. Peut contribuer à un panel de biomarqueurs, mais n'est pas suffisant seul.
- ✓ ***CTLA4*** (AUC = **0.660**), présente un pouvoir discriminant modéré. Utile en association avec d'autres marqueurs.
- ✓ ***TLR4*** (AUC = **0.499**) et ***NF-κB1*** (AUC = **0.495**) sont des biomarqueurs peu ou pas discriminants. Leur performance est faible.
- ✓ ***ESR1*** (AUC = **0.190**), ***CYP19A1*** (AUC = **0.401**), ***PGR*** (AUC = **0.201**), ***ERBB2*** (AUC = **0.278**), présente un pouvoir discriminant très faible voir nul.

TLR9 est le biomarqueur méthylé le plus prometteur pour le dépistage, le diagnostic ou le pronostic du cancer du sein dans ce cluster. *TLR9* peut être envisagé comme biomarqueur principal pour la détection du cancer du sein basé sur la méthylation.

PDCD1, *CTLA4* et *CD274* présentent un intérêt modéré et pourraient être intégrés dans un panel pour augmenter la sensibilité et la spécificité du test.

Les autres gènes n'ont pas d'intérêt clinique en tant que biomarqueurs méthylés seuls pour le cancer du sein, selon ces résultats.

b/ Gènes exprimés (Biomarqueurs Transcriptionnels)

Les résultats obtenus pour les biomarqueurs transcriptionnels montrent des performances distinctes en termes de capacité prédictive, mesurée par l'AUC (**Figure 85**).

- ✓ **ESR1-E (AUC = 0.800)** présente la meilleure performance discriminante parmi tous les marqueurs testés. Avec un AUC de 0.800, il montre une excellente capacité à distinguer les deux groupes étudiés. Cela est cohérent avec le rôle central du récepteur des œstrogènes (*ESR1*) dans le cancer du sein, notamment pour les tumeurs hormono-dépendantes.
- ✓ **PGR-E (AUC = 0.685)** et **ERBB2-E (AUC = 0.624)** montrent une capacité de discrimination modérée. PGR (récepteur de la progestérone) et ERBB2 (*HER2*) sont également des marqueurs majeurs dans le sous-typage et la prise en charge du cancer du sein.
- ✓ **CYP19A1-E (AUC = 0.341)**, **NFKB1-E (AUC = 0.380)**, **CD274-E (AUC = 0.166)**, **PDCD1-E (AUC = 0.124)**, **TLR4-E (AUC = 0.236)**, **TLR9-E (AUC = 0.124)** et **CTLA4-E (AUC = 0.097)** présentent des AUC faibles, inférieurs à 0.5, indiquant une faible voire aucune capacité discriminante dans ce contexte. Certains AUC très bas (<0.2) suggèrent même une tendance inverse, mais cela peut aussi refléter un manque de puissance statistique ou une absence d'association réelle.

ESR1-E ressort comme le meilleur biomarqueur discriminant pour le Cluster 0 dans le cancer du sein, suivi par *PGR-E* et *ERBB2-E*. Les autres biomarqueurs immunitaires testés n'ont pas de valeur discriminante notable dans cette analyse.

Ces résultats confirment l'importance des marqueurs hormonaux (ER, PR) et HER2 dans la stratification et la prise en charge du cancer du sein, tandis que les marqueurs immunitaires testés ici n'apportent pas d'information discriminante supplémentaire dans ce contexte.

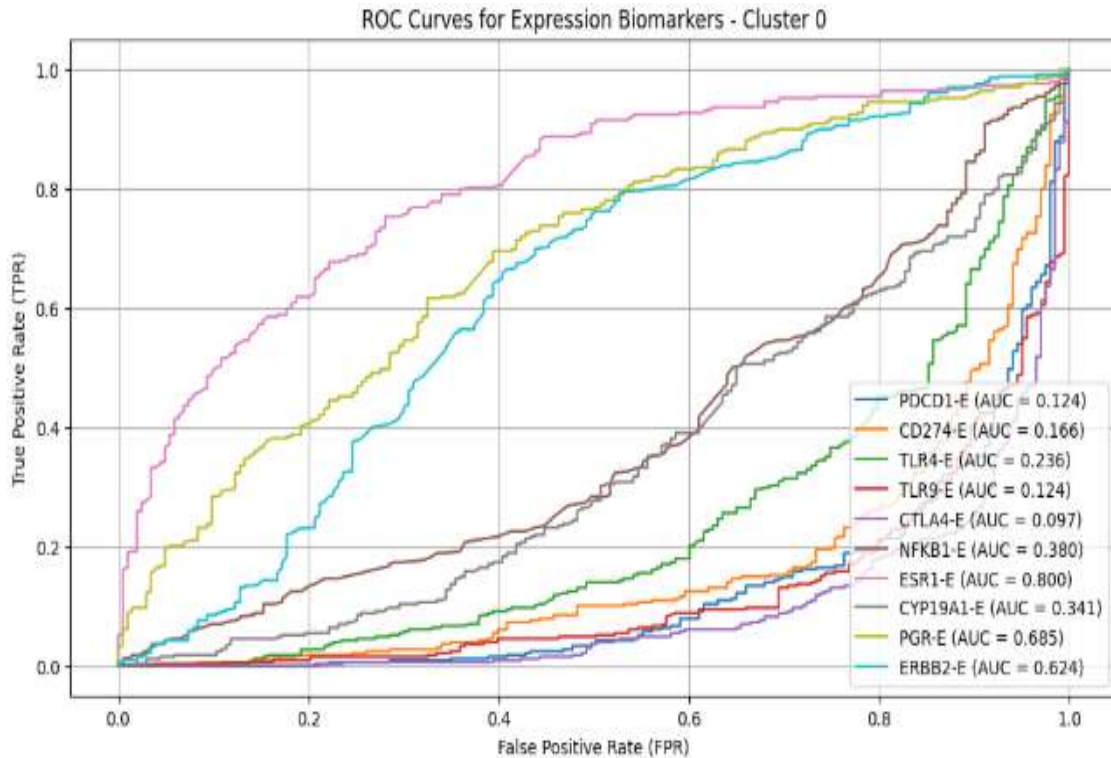


Figure 85. Evaluation de la valeur prédictive des biomarqueurs transcriptionnels du cluster 0 par la courbe ROC.

En pratique clinique, il serait intéressant de privilégier ESR1-E, éventuellement en panel avec PGR-E et ERBB2-E, pour améliorer la détection ou le pronostic du cancer du sein dans ce cluster.

9.3.2. Cas du cluster 1

a/ Gènes méthylés (Biomarqueurs Epigénétiques)

La courbe ROC pour les différents biomarqueurs épigénétiques, permettant d'évaluer leur capacité de prédiction vitale mesurée par l'AUC (**Figure 86**).

- ✓ **ESR1 (AUC = 0.810)** et **PGR (AUC = 0.799)** présentent la meilleure performance discriminante. Un AUC supérieur à 0.8 indique une excellente capacité à distinguer les groupes comparés, ce qui confirme leur rôle central dans la biologie et le diagnostic du cancer du sein.

- ✓ **ERBB2 (AUC = 0.722)** montre également une bonne performance, cohérente avec son importance dans le sous-typage moléculaire du cancer du sein (HER2+).
- ✓ **CYP19A1 (AUC = 0.599)** et **NFKB1 (AUC = 0.505)** présentent une capacité de discrimination faible à modérée, avec des AUC proches de 0.5, ce qui limite leur utilité diagnostique dans ce contexte.
- ✓ **TLR4 (AUC = 0.501)** n'a aucune valeur discriminante (AUC proche de la ligne aléatoire).

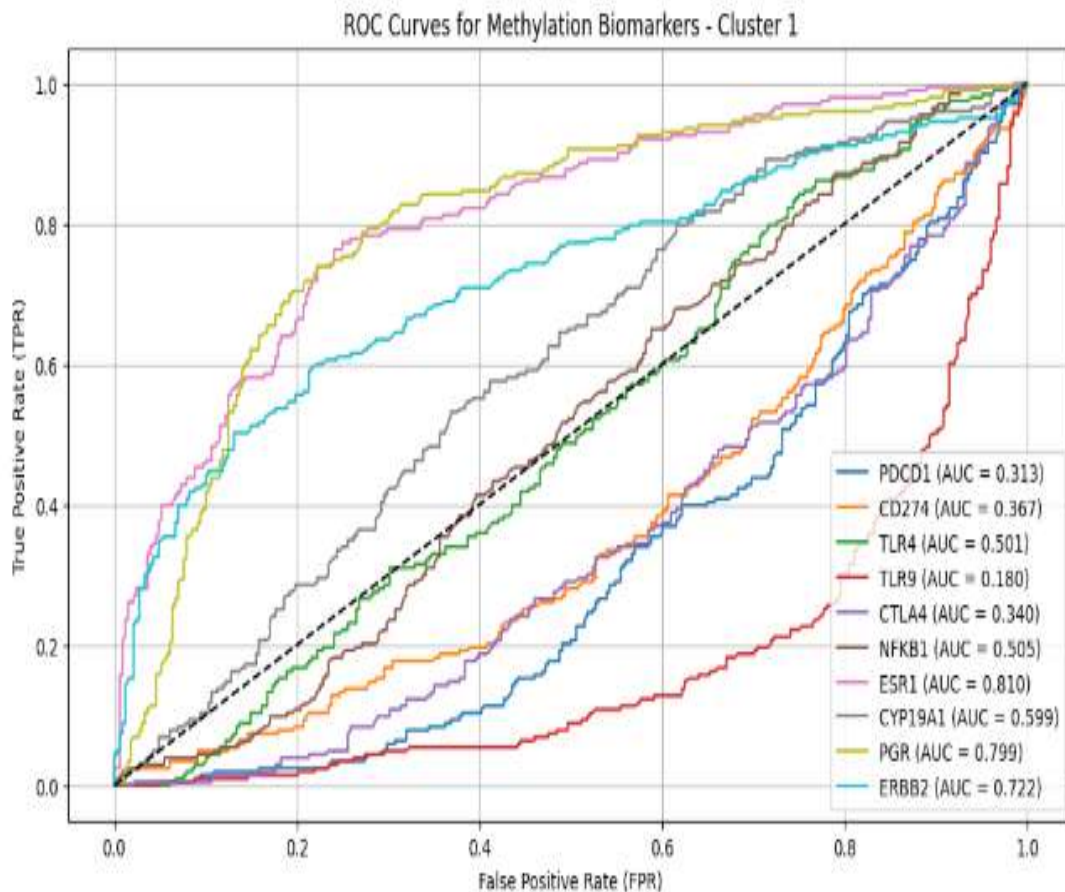


Figure 86. Evaluation de la valeur prédictive des biomarqueurs épigénétiques du Cluster 1 par la courbe ROC.

- ✓ **PDCD1 (AUC = 0.313), CD274 (AUC = 0.367), TLR9 (AUC = 0.180), CTLA4 (AUC = 0.340)** ont des AUC nettement inférieurs à 0.5, ce qui indique non seulement une absence de pouvoir discriminant, mais suggère aussi que leur méthylation pourrait être inversement associée à la pathologie étudiée ou que ces marqueurs ne sont pas pertinents dans ce contexte.

ESR1, PGR et ERBB2 sont les meilleurs biomarqueurs de méthylation pour discriminer les états cliniques dans le **Cluster 1** du cancer du sein, ce qui confirme leur utilité clinique et biologique. Les autres biomarqueurs, notamment immunitaires (*PDCD1*, *CD274*, *TLR9*, *CTLA4*), ne montrent pas de valeur discriminante dans cette analyse.

Ces résultats soulignent l'importance des marqueurs hormonaux et HER2 dans la stratification des cancers du sein, tandis que les marqueurs de l'immunité innée/adaptative testés ici apparaissent moins informatifs pour la méthylation dans ce contexte.

Ces biomarqueurs peuvent être utilisés pour différencier ce sous-groupe de patients et pourraient avoir des applications cliniques pour le diagnostic et la stratification des traitements.

b/ Gènes exprimés (Biomarqueurs Transcriptionnels)

La courbe ROC pour les différents biomarqueurs épigénétiques, permettant d'évaluer leur capacité de prédiction vitale mesurée par l'AUC (**Figure 87**).

- **CTLA4-E (AUC = 0.903), PDCD1-E (AUC = 0.876), TLR9-E (AUC = 0.876), CD274-E (AUC = 0.834)** (biomarqueurs immunitaires) montrent une excellente performance discriminante ($AUC > 0.8$), suggérant qu'ils sont particulièrement pertinents pour différencier les groupes cliniques dans ce cluster.
- **TLR4-E (AUC = 0.764)** et **CYP19A1-E (AUC = 0.659)** présentent une capacité de discrimination modérée.
- **NFKB1-E (AUC = 0.620)** présente une performance modérée.
- **ESR1-E (AUC = 0.200), PGR-E (AUC = 0.315), ERBB2-E (AUC = 0.376)**, ces marqueurs classiques du cancer du sein (hormonaux et HER2) ont des AUC très faibles, indiquant une absence de pouvoir discriminant, voire une association inverse dans ce contexte.

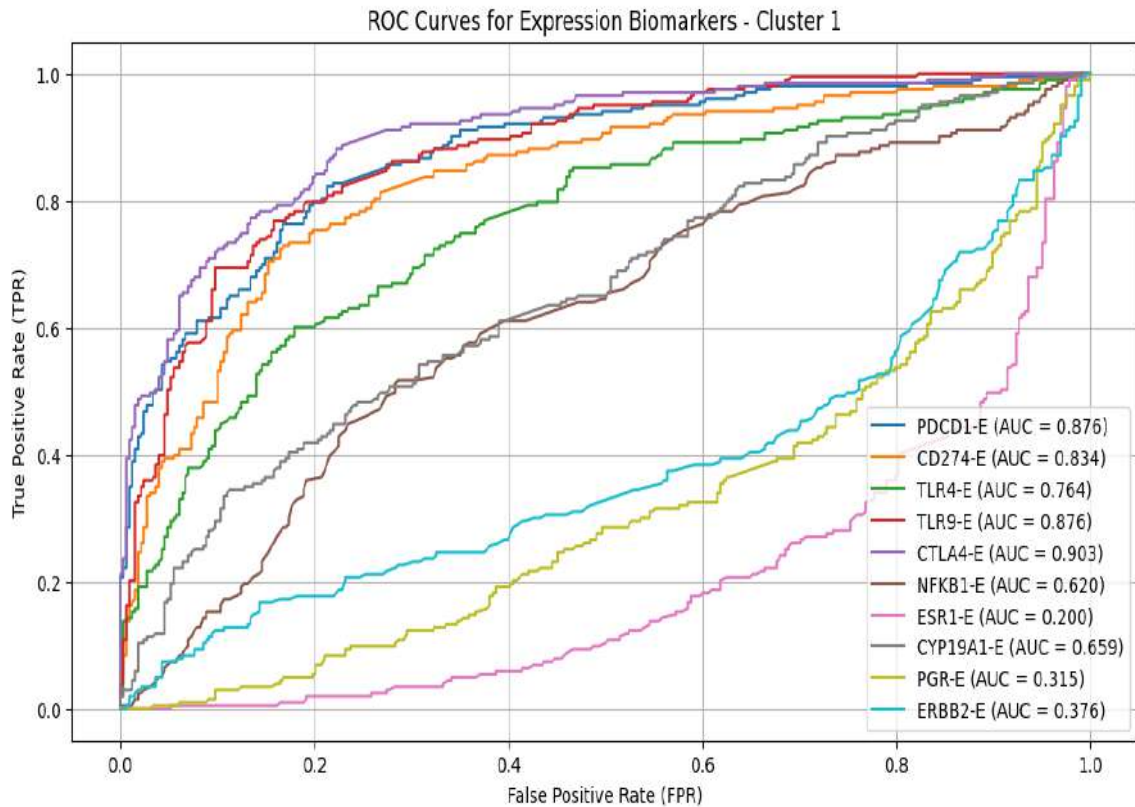


Figure 87. Evaluation de la valeur prédictive des biomarqueurs transcriptionnels du Cluster 1 par la courbe ROC.

Les biomarqueurs immunitaires (CTLA4-E, PDCD1-E, TLR9-E, CD274-E) sont les plus performants pour discriminer les états cliniques du Cluster 1, surpassant largement les marqueurs hormonaux et HER2.

Ces résultats suggèrent que, pour le Cluster 1, l'expression des gènes immunitaires joue un rôle clé dans la stratification tumorale, ouvrant des perspectives pour le développement de signatures pronostiques ou prédictives basées sur l'immunité dans le cancer du sein.

10. ANALYSE PAR CLUSTER DES VOIES KEGG RÉVÉLANT DES SIGNATURES MOLÉCULAIRES TUMORALES

10.1. Cas du cluster 0

L'analyse fonctionnelle des gènes regroupés dans le **Cluster 0** (Figure 88) révèle une forte association avec des processus biologiques clés du développement mammaire et de la physiologie hormonale féminine, notamment la différenciation des glandes mammaires, la maturation de l'endomètre et la signalisation par les récepteurs hormonaux tels que ESR1

(récepteur aux œstrogènes). Ces résultats sont particulièrement pertinents dans le contexte du cancer du sein, où l'expression et l'amplification du gène *ESR1* jouent un rôle central dans la progression tumorale et la réponse aux traitements hormonaux, notamment les anti-œstrogènes comme le tamoxifène.

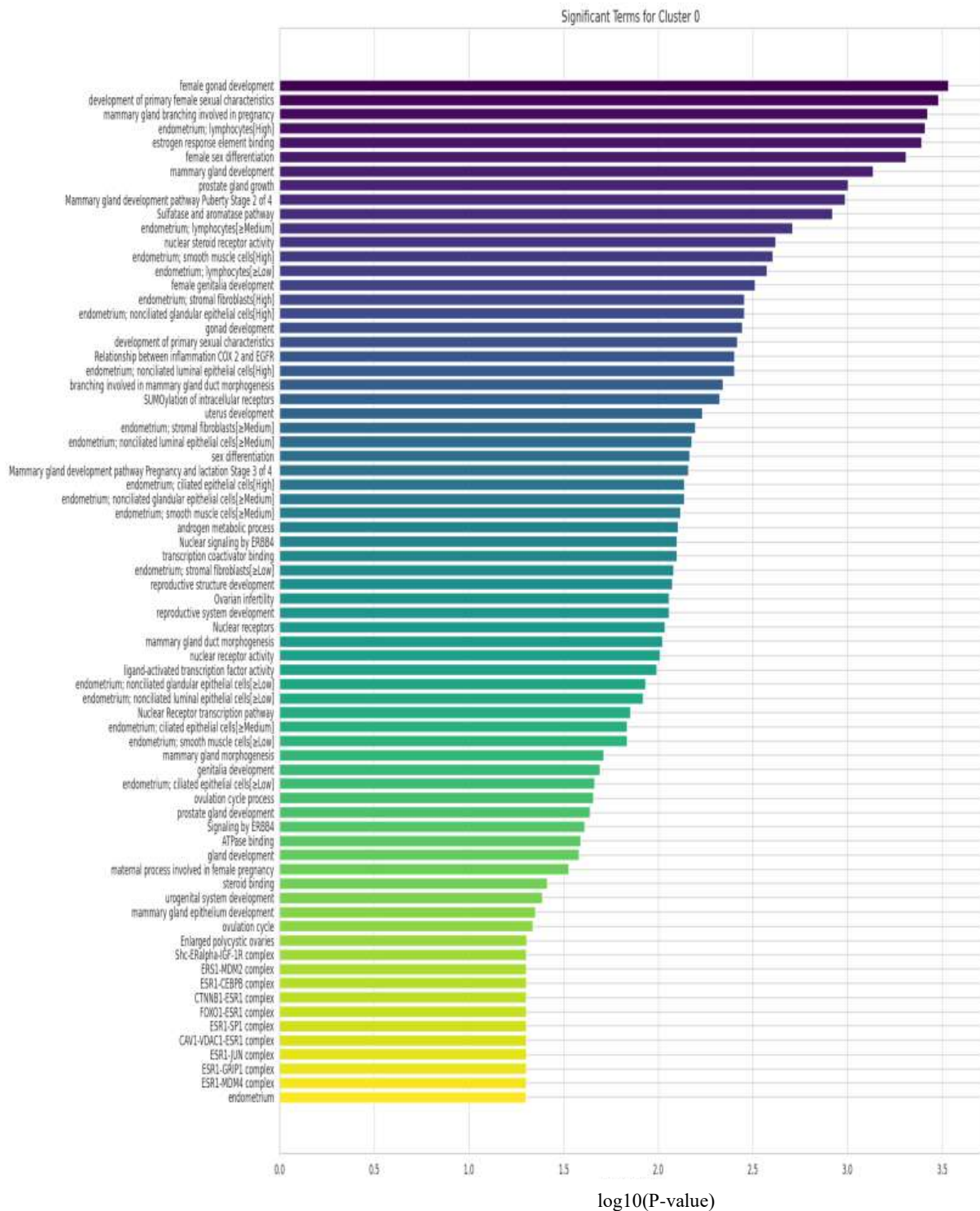


Figure 88. Analyse des voies KEGG des gènes différentiellement exprimés dans le cluster 0.

La **Figure 88** présente les termes biologiques ou fonctionnels les plus significativement enrichis pour les gènes du **Cluster 0**. L'axe des abscisses indique la significativité statistique de chaque terme ($-\log_{10}(p\text{-value})$), ce qui signifie que plus la barre est longue, plus l'enrichissement est significatif.

Les termes enrichis liés à la « *female gonad development* », au « *mammary gland branching involved in pregnancy* » et à la « *response to estrogen stimulus* » soulignent l'importance des voies hormonales dans la régulation de la croissance et de la différenciation des cellules mammaires, processus souvent altérés dans les tumeurs mammaires. Par ailleurs, l'enrichissement en termes liés à la réponse immunitaire locale suggère un rôle potentiel du microenvironnement immunitaire dans la modulation de la progression tumorale.

1. **Gènes et processus associés** : les termes les plus significatifs sont liés à la gonadogenèse féminine, au développement des caractères sexuels primaires féminins, à la ramification de la glande mammaire pendant la grossesse, et à la réponse aux estrogènes (exp. "*estrogen response element binding*").
2. **P-values et $-\log_{10}(P\text{-value})$** : les p-values faibles, inférieures à 0.05, indiquent une forte probabilité que ces processus soient effectivement associés au cluster. Par exemple, des p-values comme **0.000294** ("*female gonad development*") indiquent une association significative. Le **$-\log_{10}(P\text{-value})$** montre la force de l'association : plus il est élevé, plus l'association est significative. Par exemple, une valeur de **3.53** correspond à une p-value de **0.000294**, ce qui montre une forte significativité.
3. **Intersection Size et Gene IDs** : cela montre combien de gènes sont impliqués dans chaque processus et les identifiants des gènes associés. Cela peut vous donner des informations supplémentaires sur les gènes ou complexes spécifiques qui sont associés à chaque processus biologique. Par exemple, l'intersection de **3** pour "*female gonad development*" implique 3 gènes spécifiques.

Cette signature fonctionnelle corrobore les modèles actuels de cancérogenèse mammaire, où les altérations génétiques et épigénétiques affectent les cellules souches mammaires ou leurs progéniteurs, modifiant leur capacité d'auto-renouvellement et leur différenciation, ce qui peut favoriser la formation de cellules cancéreuses capables de proliférer et de métastaser.

En résumé, le **Cluster 0** semble représenter un ensemble de gènes impliqués dans les mécanismes fondamentaux du développement mammaire et de la régulation hormonale, qui sont

au cœur de la physiopathologie du cancer du sein. Ces résultats renforcent l'intérêt d'étudier ces voies pour mieux comprendre les mécanismes de résistance aux traitements hormonaux et pour identifier de nouvelles cibles thérapeutiques.

10.2. Cas du cluster 1

La **Figure 89** présente les termes biologiques ou fonctionnels les plus significativement enrichis pour les gènes du **Cluster 1**. L'axe des abscisses indique la significativité statistique de chaque terme ($-\log_{10}(p\text{-value})$), ce qui signifie que plus la barre est longue, plus l'enrichissement est significatif.

L'analyse d'enrichissement fonctionnel des gènes de ce Cluster 1 met en évidence une prédominance marquée des processus immunitaires, avec une significativité particulièrement forte pour les interactions entre cellules immunitaires et microARNs dans le microenvironnement tumoral, ainsi que pour la voie PD-1/PD-L1, cible majeure des immunothérapies actuelles. D'autres processus tels que la signalisation Toll-like, la prolifération lymphocytaire et la régulation de la production de cytokines confirment l'implication centrale de l'immunité adaptative et innée dans ce Cluster. Ces résultats suggèrent que les tumeurs du sein présentant ce profil moléculaire pourraient bénéficier de stratégies thérapeutiques ciblant l'environnement immunitaire tumoral, notamment les inhibiteurs de points de contrôle immunitaire. Cette signature fonctionnelle pourrait ainsi servir de base à une stratification des patientes en vue de traitements personnalisés

1. **Processus Immunitaires** : ce Cluster est fortement associé à des processus immunitaires comme **l'interaction entre les cellules immunitaires et les microARN**, **l'expression de PD-L1** et **le chemin de la voie de checkpoint PD-1**, et **la régulation de la réponse immune**. Cela suggère que ce cluster pourrait être plus activé sur le plan immunologique.
2. **P-values et $-\log_{10}(P\text{-value})$** : comme pour le cluster 0, les p-values faibles et les **$-\log_{10}(P\text{-value})$** élevés indiquent une association significative. Par exemple, une p-value de **0.000052** pour "*Interactions between immune cells and microRNAs*" et une **valeur de $-\log_{10}(P\text{-value})$ de 4.29** montrent une très forte association.
3. **Gene IDs et Intersection Size** : ces valeurs nous montrent le nombre de gènes impliqués dans chaque processus. Par exemple, "*positive regulation of immune response*" a une intersection de **4**, signifiant que 4 gènes sont associés à ce processus spécifique.

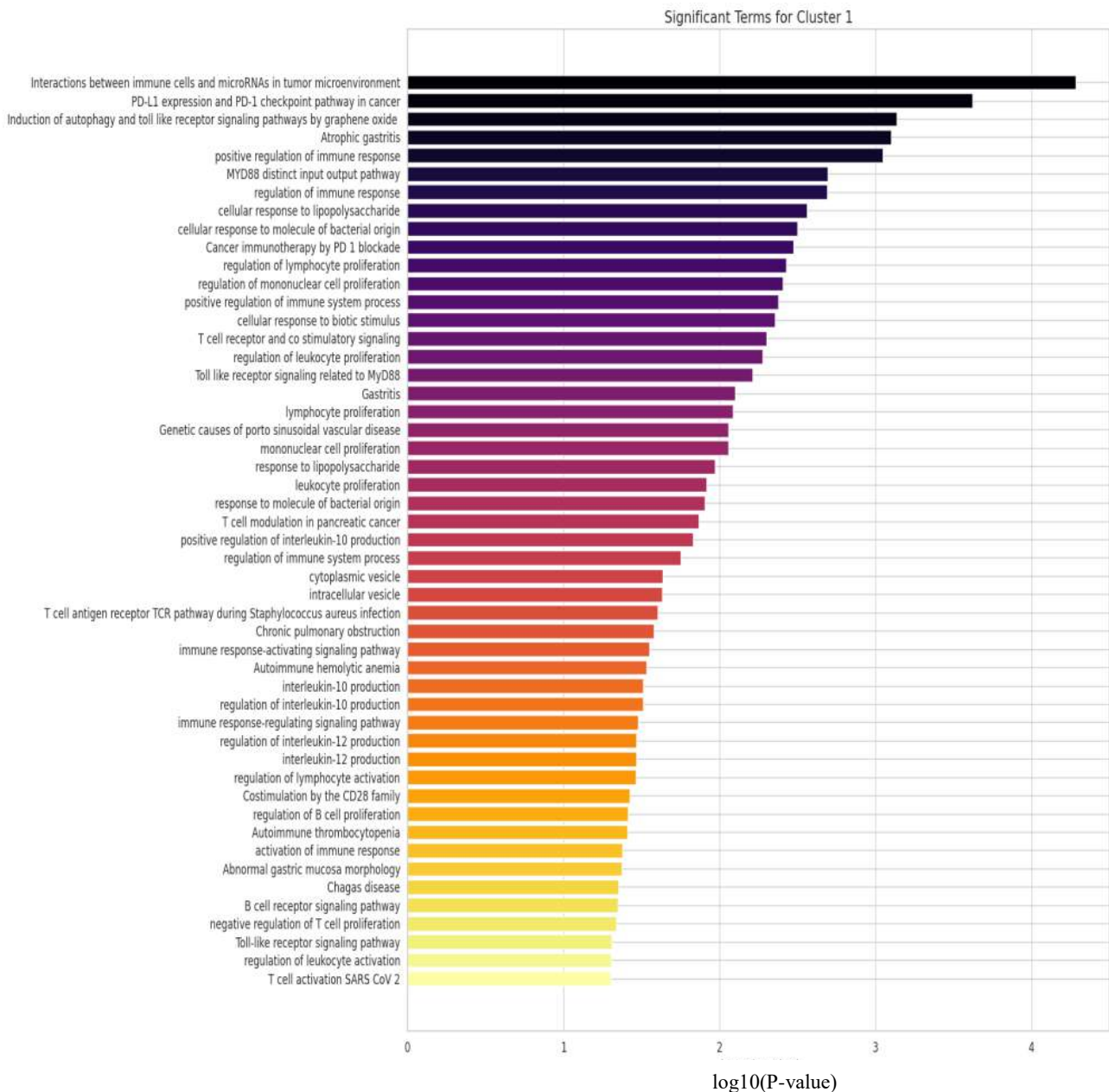


Figure 89. Analyse des voies KEGG des gènes différentiellement exprimés dans le cluster 1.

1. **Le Cluster 1 est fortement associé à la régulation immunitaire :** l'enrichissement en voies PD-1/PD-L1, signalisation Toll-like, activation des lymphocytes T et B, et production de cytokines (interleukines) suggère que les gènes de ce cluster participent à la modulation de la réponse immunitaire dans le microenvironnement tumoral.
2. **Implication thérapeutique :** les voies PD-1/PD-L1 sont des cibles majeures des immunothérapies modernes (anticorps anti-PD-1/PD-L1). Leur enrichissement ici suggère

que les tumeurs caractérisées par ce cluster pourraient être particulièrement sensibles à ces traitements.

3. **Perspective translationnelle** : la prédominance des processus immunitaires dans ce cluster met en avant l'importance de la composante immunitaire dans certains sous-types de cancer du sein, et pourrait orienter vers une stratification des patientes pour des thérapies ciblées sur l'immunité.

Cluster 0 semble être plus lié à des processus hormonaux et de développement (notamment dans la gonadogenèse féminine et les réponses aux estrogènes), ce qui pourrait suggérer une sensibilité hormonale plus marquée, potentiellement avec des traitements hormonaux.

Cluster 1, quant à lui, est davantage lié à des processus immunitaires, comme la régulation de la réponse immunitaire et les interactions immunitaires, ce qui pourrait indiquer un potentiel pour des thérapies immunomodulatrices ou des thérapies basées sur l'immunothérapie, telles que les inhibiteurs de point de contrôle (par exemple, anti-PD1).



TROISIEME PARTIE

**MODÉLISATION DE LA SURVIE :
BENCHMARK COMPARATIF ENTRE
COX (APPRENTISSAGE HYBRIDE),
ANN ET CNN**

1- MODELISATION DE LA SURVIE PAR LE MODELE COX – HAZARD

Dans le cadre de cette étude, une analyse de régression de Cox (**Tableau 21**) a été réalisée afin d'évaluer l'impact de plusieurs variables cliniques et moléculaires sur le risque associé au cancer du sein.

Tableau 21. Analyse de survie par le modèle de Cox-Hazard dans la population totale.

Variable	Coef	Exp (Coef)	Se (Coef)	Coef lower 95%	Coef upper 95%	Exp (Coef) lower 95%	Exp (Coef) upper 95%	z	p	-log2(p)
Âge au diagnostic pathologique	0.03	1.03	0.01	0.01	0.05	1.01	1.05	2.86	<0.005	7.90
Stade pathologique T	0.20	1.23	0.13	-0.06	0.47	0.94	1.60	1.52	0.13	2.96
Stade pathologique N	0.04	1.04	0.13	-0.21	0.29	0.81	1.33	0.30	0.76	0.39
Stade pathologique M	0.59	1.80	0.14	0.31	0.86	1.37	2.37	4.22	<0.005	15.29
Stade pathologique	0.02	1.02	0.07	-0.13	0.16	0.88	1.17	0.23	0.82	0.29
DSS	-10.69	0.00	8.34	-27.03	5.65	0.00	283.35	-1.28	0.20	2.32
PDCD1-E	-0.16	0.85	0.07	-0.30	-0.03	0.74	0.97	-2.35	0.02	5.74
ESR1-E	-0.02	0.98	0.05	-0.12	0.08	0.89	1.08	-0.46	0.65	0.63
CYP19A1-E	-0.02	0.98	0.09	-0.20	0.15	0.82	1.17	-0.26	0.79	0.34
PGR-E	-0.06	0.94	0.05	-0.16	0.04	0.85	1.04	-1.16	0.24	2.03

Coefficient (Coef) et Hazard Ratio (Exp(Coef))

- ✓ Le coefficient représente l'effet log-linéaire de la variable sur le risque.
- ✓ Exp(Coef) est le hazard ratio (HR), c'est-à-dire le facteur multiplicatif du risque associé à une unité d'augmentation de la variable.
- ✓ $HR > 1$ signifie augmentation du risque, $HR < 1$ signifie diminution du risque.

L'évaluation du modèle Cox-Hazard s'appuie sur les paramètres suivants :

- **Concordance** : 0.24
- **AIC partiel** : 1707.70
- **Test de log-vraisemblance** : - 923.56 sur 10 degrés de liberté

La Concordance du modèle indique un pouvoir prédictif très faible. Un bon modèle devrait atteindre une concordance d'au moins 0.5.

L'AIC partiel sert à comparer plusieurs modèles. Plus il est bas, meilleur est l'ajustement.

Le Test de log-vraisemblance confirme un ajustement relativement faible du modèle.

Les résultats montrent que l'âge au diagnostic pathologique est un facteur pronostique significatif, avec un Hazard Ratio (HR) de 1,03 (IC 95% : 1,01–1,05, $p < 0,005$), indiquant une augmentation de 3 % du risque par année supplémentaire.

Par ailleurs, le stade pathologique métastatique (stade M) s'est révélé être un facteur majeur, augmentant significativement le risque de 80 % (HR = 1,80 ; IC 95 % : 1,37–2,37 ; $p < 0,005$). En revanche, les stades T et N n'ont pas montré d'association statistiquement significative dans ce modèle ($p > 0,1$).

Sur le plan moléculaire, l'expression de PDCD1-E a été identifiée comme un facteur protecteur, avec un HR de 0,85 (IC 95 % : 0,74–0,97 ; $p = 0,02$), suggérant que son augmentation est associée à une diminution significative du risque de 15 %. Les autres marqueurs étudiés (ESR1-E, CYP19A1-E, PGR-E) n'ont pas présenté d'effet significatif.

Ces résultats soulignent l'importance combinée des facteurs cliniques et moléculaires dans la stratification du risque et la prise en charge personnalisée des patients. Ils ouvrent également des perspectives pour le développement de biomarqueurs pronostiques et de cibles thérapeutiques.

Bien que certaines variables cliniques (âge, stade M) soient significatives, le pouvoir prédictif du modèle reste faible (concordance 0.24). Il est nécessaire de :

- Explorer d'autres variables cliniques ou biologiques,
- Évaluer différents biomarqueurs,
- Envisager des modèles plus complexes

1.1. Application du modèle Cox-Hazard sur les clusters 0 et 1

1- Biomarqueurs Epigénétiques

Le choix de focaliser l'analyse sur la méthylation découle de considérations biologiques majeures : il est désormais bien établi que les altérations épigénétiques, en particulier l'hyperméthylation des promoteurs de gènes critiques, constituent des événements précoces de la tumorigenèse et sont susceptibles de moduler l'expression génique de manière durable.

L'analyse de la survie globale (OS) a été approfondie par des modèles de régression de Cox multivariée, appliqués séparément aux deux sous-groupes identifiés par la méthode de clustering (Tableau 22 et 23). Ces modèles ont intégré la méthylation de biomarqueurs immuno-hormonaux, ainsi que l'âge au diagnostic et le statut ménopausique, dans le but d'évaluer le rôle pronostique indépendant de l'épigénétique tumorale.

Tableau 22. Analyse de survie par le modèle de Cox-Hazard pour le Cluster 0.

Gènes	Coef	Exp (Coef)	SE (Coef)	Coef 95% CI [Bas]	Coef 95% CI [Haut]	Exp (Coef) 95% CI [Bas]	Exp (Coef) 95% CI [Haut]	z	p	-log2(p)
<i>PDCD1</i>	1.45	4.24	2.42	-3.30	6.19	0.04	489.64	0.60	0.55	0.86
<i>CD274</i>	2.83	16.89	3.21	-3.47	9.12	0.03	9150.16	0.88	0.38	1.40
<i>TLR4</i>	-0.54	0.59	1.16	-2.81	1.74	0.06	5.69	-0.46	0.64	0.63
<i>TLR9</i>	-7.69	0.00	5.59	-18.64	3.26	0.00	26.03	-1.38	0.17	2.57
<i>CTLA4</i>	3.21	24.83	2.69	-2.06	8.49	0.13	4853.45	1.19	0.23	2.10
<i>NFKB1</i>	-9.24	0.00	8.03	-24.98	6.50	0.00	662.34	-1.15	0.25	2.00
<i>ESR1</i>	-4.18	0.02	2.48	-9.03	0.67	0.00	1.96	-1.69	0.09	3.45
<i>CYP19A1</i>	-4.02	0.02	3.32	-10.52	2.49	0.00	12.09	-1.21	0.23	2.14
<i>PGR</i>	7.12	1235.69	6.02	-4.67	18.91	0.01	1.63e+08	1.18	0.24	2.08
<i>ERBB2</i>	-0.63	0.53	4.40	-9.26	8.00	0.00	2971.08	-0.14	0.89	0.17
Age au diagnostic	0.03	1.03	0.02	-0.01	0.06	0.99	1.06	1.63	0.10	3.26
statut Ménopausique	0.12	1.13	0.31	-0.49	0.73	0.61	2.07	0.39	0.70	0.52

Dans ce sous-groupe, aucune variable n'a atteint le seuil de significativité statistique. Toutefois, certaines tendances méritent d'être soulignées :

- La méthylation d'*ESRI* (codant pour le récepteur des œstrogènes) et de *CYP19A1* (aromatase) est associée à un coefficient négatif, ce qui pourrait refléter un effet protecteur de la méthylation, par suppression de voies favorisant la prolifération tumorale hormono-dépendante.
- À l'inverse, l'hyperméthylation du gène *PGR* (récepteur de la progestérone) semble associée à une augmentation marquée du risque de décès (coef = 7,12), bien que l'intervalle de confiance large impose une interprétation prudente.
- Concernant les gènes immunitaires, la méthylation de *PDCDI* et de *CD274* semble, de façon non significative, corrélée à un pronostic défavorable.

Tableau 23. Analyse de survie par le modèle de Cox-Hazard pour le Cluster 1.

Variable	Coef	exp(Coef)	SE(Coef)	Coef 95% CI [Bas]	Coef 95% CI [Haut]	exp(Coef) 95% CI [Bas]	exp(Coef) 95% CI [Haut]	z	p	-log2(p)
<i>PDCDI</i>	-3.23	0.04	2.66	-8.45	1.99	0.00	7.33	-1.21	0.23	2.15
<i>CD274</i>	7.28	1453.30	4.09	-0.74	15.30	0.48	4.43e+06	1.78	0.08	3.73
<i>TLR4</i>	-3.17	0.04	1.69	-6.48	0.13	0.00	1.14	-1.88	0.06	4.07
<i>TLR9</i>	-3.35	0.04	3.51	-10.22	3.52	0.00	33.93	-0.95	0.34	1.56
<i>CTLA4</i>	3.95	52.12	3.62	-3.14	11.05	0.04	62742.96	1.09	0.27	1.86
<i>NFKB1</i>	18.46	1.04e+08	11.49	-4.06	40.99	0.02	6.33e+17	1.61	0.11	3.21
<i>ESRI</i>	-1.44	0.24	2.37	-6.09	3.21	0.00	24.70	-0.61	0.54	0.88
<i>CYP19A1</i>	0.30	1.35	4.37	-8.26	8.87	0.00	7101.54	0.07	0.94	0.08
<i>PGR</i>	8.47	4752.77	3.01	2.56	14.37	12.95	1.74e+06	2.81	<0.005	7.66
<i>ERBB2</i>	-6.61	0.00	3.72	-13.90	0.68	0.00	1.98	-1.78	0.08	3.72
Age au diagnostic	-0.01	0.99	0.02	-0.05	0.03	0.96	1.03	-0.39	0.70	0.51
statut Ménopausique	-0.61	0.54	0.35	-1.30	0.08	0.27	1.09	-1.72	0.09	3.55

Le modèle de Cox pour le Cluster 0 a été ajusté sur 265 patients, parmi lesquels 46 événements (décès) ont été observés. L'analyse fait émerger plusieurs facteurs pronostiques :

- L'hyperméthylation du *PGR* est fortement associée à une augmentation significative du risque de décès ($p < 0,005$). Cette observation suggère que dans ce sous-groupe, la perte d'expression du récepteur de la progestérone, via des mécanismes épigénétiques, pourrait être un marqueur d'agressivité tumorale accrue.
- Les méthylation de *CD274 (PD-L1)* et de *TLR4* tendent vers la significativité. La méthylation accrue de *CD274* est associée à un mauvais pronostic, probablement en raison d'une désactivation du checkpoint immunitaire et d'une évasion immune facilitée.
- Enfin, la méthylation d'*ERBB2* semble associée à un meilleur pronostic (coef négatif), en cohérence avec la littérature suggérant que la perte d'expression d'*ERBB2 (HER2)* est moins fréquemment associée aux formes les plus agressives de cancer du sein

* Performances globales des modèles :

Dans les deux clusters, les modèles de Cox présentent une Concordance de 0,72, attestant d'une bonne capacité prédictive. Le partial AIC est inférieur dans le Cluster 1 (249,37) comparé au Cluster 0 (411,76), ce qui suggère un ajustement plus précis du modèle aux données du Cluster 1. De plus, le test du log- vraisemblance révèle une signification statistique pour le Cluster 1 ($p < 0,02$), contrairement au Cluster 0 ($p = 0,07$), renforçant l'idée que dans le Cluster 1, la méthylation exerce un impact plus directement mesurable sur la survie.

* Considérations biologiques et cliniques :

Ces résultats confirment l'hypothèse selon laquelle l'impact des modifications épigénétiques sur la survie varie selon les caractéristiques moléculaires des tumeurs. Dans les tumeurs du **Cluster 1**, au profil probablement plus agressif, l'hyperméthylation de gènes clés (hormonaux et immunitaires) apparaît comme un déterminant majeur de la survie, tandis que dans le **Cluster 0**, la méthylation semble exercer une influence plus subtile, noyée parmi d'autres mécanismes biologiques.

D'un point de vue translationnel, ces observations suggèrent que la méthylation du *PGR* et du *CD274* pourrait constituer des cibles potentielles pour le développement de nouveaux outils pronostiques ou pour affiner les stratégies thérapeutiques personnalisées, en particulier dans les cancers du sein de biologie agressive.

2- Biomarqueurs Transcriptionnels

Afin d'explorer les facteurs associés à la survie globale (Overall Survival) dans les deux clusters identifiés par l'analyse de regroupement, un modèle de Cox multivarié a été ajusté séparément pour chaque cluster (**Tableau 24 et 25**). Les covariables incluses comprenaient l'expression génique de *PDCD1*, *CD274*, *TLR4*, *TLR9*, *CTLA4*, *NFKB1*, *ESR1*, *CYP19A1*, *PGR* et *ERBB2*, ainsi que l'âge au diagnostic et le statut ménopausique.

Tableau 24. Analyse de survie par le modèle de Cox-Hazard dans le Cluster 0.

Variable	Coef	Exp (Coef)	Se (coef)	IC 95% Coef	IC 95% exp (Coef)	z	p	-log2(p)
PDCD1-E	-0.37	0.69	0.14	[-0.66, -0.09]	[0.52, 0.91]	-2.59	0.01	6.70
CD274-E	-0.42	0.65	0.22	[-0.85, 0.00]	[0.43, 1.00]	-1.95	0.05	4.30
TLR4-E	0.64	1.90	0.26	[0.14, 1.15]	[1.15, 3.16]	2.50	0.01	6.31
TLR9-E	0.33	1.40	0.17	[-0.00, 0.67]	[1.00, 1.96]	1.94	0.05	4.26
CTLA4-E	0.10	1.10	0.17	[-0.23, 0.43]	[0.79, 1.54]	0.58	0.56	0.84
NFKB1-E	-0.13	0.88	0.33	[-0.77, 0.52]	[0.46, 1.68]	-0.38	0.70	0.51
ESR1-E	-0.07	0.93	0.08	[-0.23, 0.09]	[0.79, 1.10]	-0.82	0.41	1.28
CYP19A1-E	-0.17	0.85	0.13	[-0.42, 0.08]	[0.66, 1.09]	-1.30	0.19	2.36
PGR-E	-0.03	0.97	0.08	[-0.18, 0.11]	[0.83, 1.12]	-0.45	0.65	0.61
ERBB2-E	-0.06	0.95	0.14	[-0.34, 0.22]	[0.71, 1.25]	-0.40	0.69	0.53
Age au diagnostic	0.05	1.05	0.02	[0.02, 0.08]	[1.02, 1.08]	3.31	<0.005	10.05
statut Ménopausique	0.17	1.19	0.35	[-0.51, 0.86]	[0.60, 2.36]	0.50	0.62	0.70

Le modèle de Cox pour le Cluster 0, présente une concordance de 0,75, indiquant une bonne capacité discriminante. Le test du rapport de vraisemblance est significatif ($p < 0,005$), avec un $-\log_2(p)$ de 8,49, soutenant la qualité de l'ajustement.

Parmi les facteurs étudiés :

- ✓ **PDCD1-E** est significativement associé à une meilleure survie (HR = 0,69 ; p = 0,01), suggérant qu'une expression plus élevée de PDCD1 est protectrice.
- ✓ **CD274-E** montre une tendance protectrice (HR = 0,65) mais n'atteint la significativité que marginalement (p = 0,05).
- ✓ **TLR4-E** est significativement associé à un risque accru de décès (HR = 1,90 ; p = 0,01).
- ✓ **TLR9-E** présente également une association modérée avec une survie diminuée (HR = 1,40 ; p = 0,05).
- ✓ **L'âge au diagnostic** est fortement associé à la survie (HR = 1,05 par année supplémentaire ; p < 0,005).

Les autres biomarqueurs, notamment CTLA4-E, NF-κB1-E, ESR1-E, CYP19A1-E, PGR-E, ERBB2-E et le statut de ménopause, n'ont pas montré d'association statistiquement significative avec la survie dans ce cluster.

Tableau 25. Analyse de survie par le modèle de Cox-Hazard dans le Cluster 1

Variables	Coef	Exp (Coef)	Se (Coef)	IC 95% Coef	IC 95% exp (Coef)	z	p	-log2(p)
PDCD1-E	-0.11	0.90	0.21	[-0.52, 0.30]	[0.59, 1.35]	-0.53	0.60	0.74
CD274-E	0.25	1.29	0.21	[-0.16, 0.67]	[0.85, 1.96]	1.20	0.23	2.11
TLR4-E	0.43	1.54	0.29	[-0.14, 1.01]	[0.87, 2.75]	1.48	0.14	2.83
TLR9-E	-0.01	0.99	0.24	[-0.49, 0.47]	[0.62, 1.61]	-0.02	0.98	0.03
CTLA4-E	-0.28	0.75	0.22	[-0.72, 0.15]	[0.49, 1.16]	-1.28	0.20	2.32
NFKB1-E	-0.51	0.60	0.51	[-1.50, 0.49]	[0.22, 1.63]	-1.00	0.32	1.65
ESR1-E	-0.11	0.90	0.10	[-0.31, 0.10]	[0.73, 1.10]	-1.02	0.31	1.70
CYP19A1-E	0.15	1.16	0.15	[-0.14, 0.43]	[0.87, 1.54]	1.01	0.31	1.67
PGR-E	0.01	1.01	0.09	[-0.17, 0.19]	[0.84, 1.21]	0.12	0.91	0.14
ERBB2-E	-0.20	0.82	0.17	[-0.53, 0.13]	[0.59, 1.14]	-1.18	0.24	2.08
Age au diagnostic	0.01	1.01	0.02	[-0.03, 0.04]	[0.97, 1.05]	0.38	0.71	0.50
statut Ménopausique	-0.47	0.63	0.36	[-1.17, 0.23]	[0.31, 1.26]	-1.31	0.19	2.38

Dans le Cluster 1, l'analyse a porté sur 266 patients avec 30 événements observés. Le modèle présente également une concordance de 0,75, mais le test du rapport de vraisemblance est moins significatif ($p = 0,01$; $-\log_2(p) = 2,54$), indiquant un pouvoir explicatif plus faible par rapport au Cluster 0.

Les autres facteurs n'ont pas montré une signification statistique dans ce groupe :

- ✓ **TLR4-E** montre une augmentation non significative du risque (HR = 1,54 ; $p = 0,14$).
- ✓ **CD274-E** est associé à un effet protecteur modéré (HR = 1,29 ; $p = 0,23$).
- ✓ **CTLA4-E** tend à être protecteur (HR = 0,75 ; $p = 0,20$).
- ✓ **Le statut ménopausique** montre une association non significative vers une meilleure survie pour les patientes ménopausées (HR = 0,63 ; $p = 0,19$).

Ainsi, contrairement au Cluster 0, aucune variable ne ressort de manière marquée dans le Cluster 1, ce qui pourrait refléter une hétérogénéité biologique plus importante ou un nombre d'événements (nombre de décès dans chaque cluster) insuffisant pour détecter des associations robustes.

2. MODELES DEEP LEARNING

2.1. Réseaux de Neurones Artificiels (ANN)

Un modèle de réseaux de neurones entièrement connectés (ANN) a été réalisé afin de prédire la survie des patients en fonction des biomarqueurs épigénétiques et transcriptionnels et des paramètres cliniques.

L'entraînement a été effectué sur 2000 époques, avec une optimisation basée sur la descente de gradient et une fonction de coût logarithmique (log loss).

2.1.1. Performance globale

L'évaluation finale du modèle sur l'ensemble de test donne les résultats suivants :

- **Accuracy finale** : 96,80%
- **Loss finale** : 0.1674
- **AUC-ROC** : 0.99

Ces performances indiquent que le modèle ANN est très efficace pour distinguer les classes. L'accuracy élevée (96,80%) et l'AUC-ROC proche de 1.0 (0.99) confirment sa capacité à bien discriminer les patients survivants et décédés.

2.1.2. Convergence du modèle

Le suivi de l'entraînement montre (**Figure 90**):

- Une diminution progressive de la fonction de perte, indiquant une bonne convergence du modèle.
- Une augmentation parallèle de l'accuracy, démontrant une amélioration progressive des prédictions.

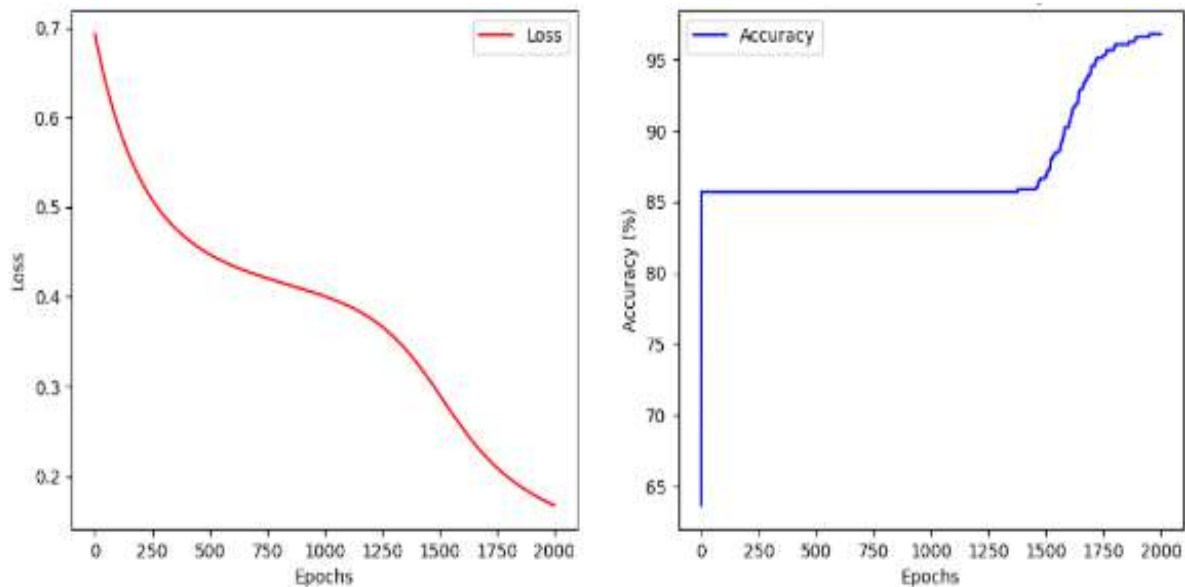


Figure 90. Courbes de perte et d'accuracy du modèle ANN.

2.1.3. Matrice de confusion

Analyse par Classe (Tableau 26) :

- **Classe 0 (Patients vivants) :**
 - **Précision :** 96,4 %
 - **Rappel :** 100 %
 - **F1-score :** 98,17 % Le modèle identifie parfaitement les patients survivants.
- **Classe 1 (Patients décédés) :**
 - **Précision :** 100 %
 - **Rappel :** 77,63 %, indiquant que 17 patients décédés ont été mal classés.
 - **F1-score :** 87,41 %

Tableau 26. Performance du modèle ANN par classe de survie.

Classe	Précision	Rappel	F1-score	Support
0 (Vivantes)	96.40%	100%	98.17%	455
1 (Décédés)	100%	77.63%	87.41%	76
Moyenne macro	98.20%	88.82%	92.79%	531
Moyenne pondérée	96.91%	96.80%	96.63%	531

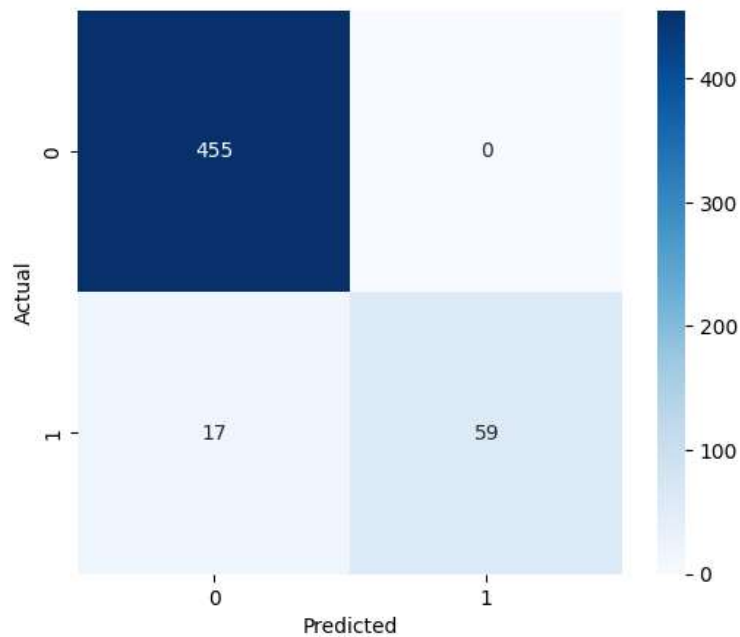


Figure 91. Matrice de confusion du modèle ANN.

Bien que la précision soit parfaite pour les patients décédés, le rappel indique que le modèle pourrait encore être amélioré pour détecter toutes les patientes à risque, décédées incorrectement et prédits comme vivantes (Faux Négatifs = 17) (Figure 91).

Le modèle ANN a atteint un AUC-ROC de 0,99 (Figure 92), indiquant sa capacité à distinguer très efficacement les patients vivants des décédés avec une performance très élevée.

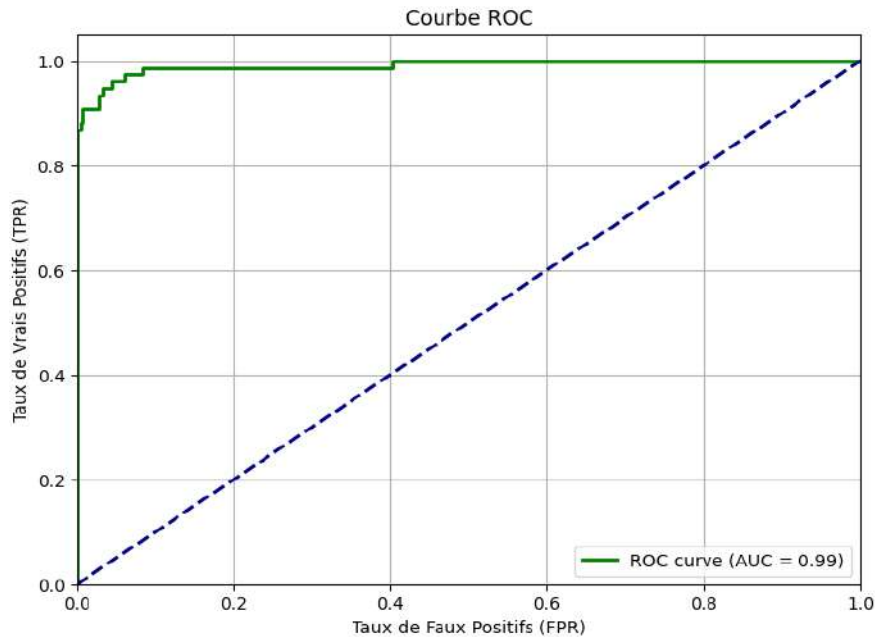


Figure 92. Courbe AUC-ROC du modèle ANN.

Le modèle ANN présente une excellente précision globale (96,80%) et un AUC-ROC très élevé (0.99), démontrant une grande capacité à distinguer les classes. Il identifie parfaitement la classe majoritaire (0 : les patients vivants) et ne génère aucun faux positif pour la classe minoritaire (1 : patientes décédés). Toutefois, un rappel de 77,63% pour la classe 1 indique une sous-estimation des patients décédés, suggérant la nécessité d'améliorations, notamment par le rééquilibrage des classes ou l'optimisation des hyperparamètres pour renforcer la détection des cas à risque.

2.2. Réseaux de Neurones Convolutifs (CNN)

Un modèle de réseaux de neurones convolutifs (CNN) a été mis en place afin de prédire la survie des patients en fonction de divers biomarqueurs et paramètres cliniques. L'entraînement a été effectué sur 1000 époques, avec une optimisation basée sur la descente de gradient et une fonction de coût logarithmique (log loss).

2.2.1. Performance globale

L'évaluation finale du modèle sur l'ensemble de test donne les résultats suivants :

- **Accuracy finale** : 98,31%
- **Loss finale** : 0.0454
- **AUC-ROC** : 0.996

Ces performances indiquent que le modèle CNN est extrêmement efficace pour distinguer les classes.

L'accuracy élevée (**98,31%**) et l'AUC-ROC proche de 1.0 (**0,996**) confirment sa capacité à bien discriminer les patients survivants et non survivants.

2.2.2. Convergence du modèle

Le suivi de l'entraînement montre (**Figure 93**) :

Une diminution rapide de la fonction de perte, indiquant une très bonne convergence du modèle.

Une hausse stable de l'accuracy, montrant une amélioration constante des prédictions.

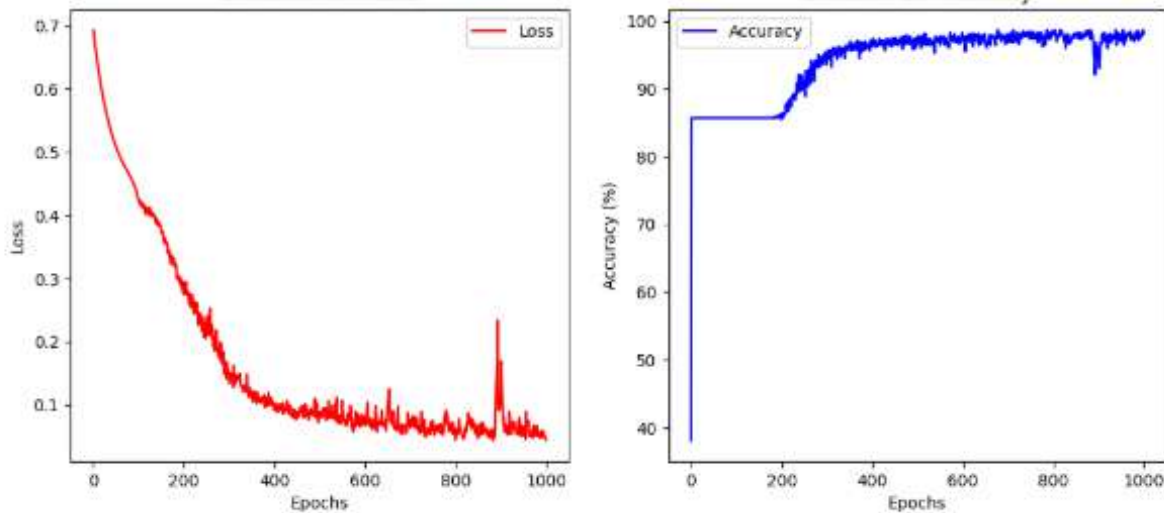


Figure 93. Courbes de perte et d'accuracy du modèle CNN.

2.2.3. Matrice de confusion

Analyse par Classe (**Tableau 27**) :

❖ **Classe 0 (Patients vivants) :**

✓ **Précision** : 98,00 %

✓ **Rappel** : 100,00 %

✓ **F1-score** : 99,00 %

Le modèle identifie parfaitement les patients vivants sans générer de faux négatifs.

❖ **Classe 1 (Patients décédés) :**

- ✓ **Précision :** 97,00 %
- ✓ **Rappel :** 90,79 %, ce qui indique que 7 patientes décédées ont été mal classées.
- ✓ **F1-score :** 94,00 %

Tableau 27. Performance du modèle ANN par classe de survie.

Classe	Précision	Rappel	F1-score	Support
0 (Vivantes)	98,00%	100,00%	99,00%	455
1 (Décédés)	97,00%	90,79%	94,00%	76
Moyenne macro	98,00%	95,40%	96,50%	531
Moyenne pondérée	98,00%	98,31%	98,00%	531

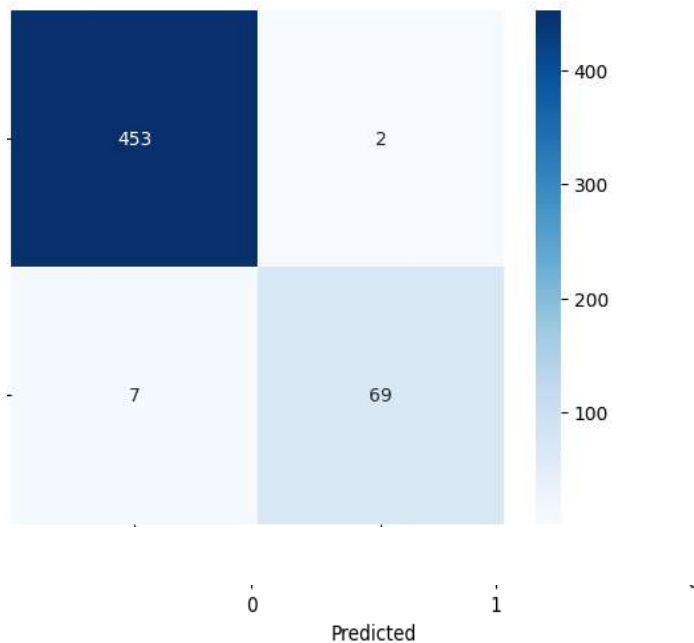


Figure 94. Matrice de confusion du modèle CNN.

AUC-ROC (Figure 95) : le modèle CNN a fourni un AUC-ROC de 0,99, traduisant ainsi une très bonne capacité de discrimination entre les patientes vivantes et décédées, et une performance robuste dans la prédiction de la survie.

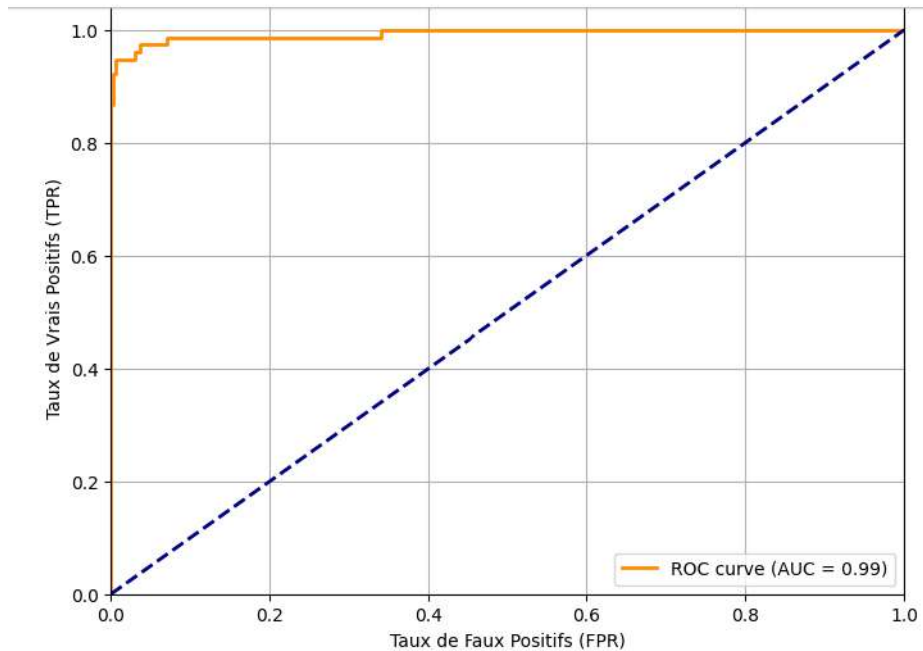


Figure 95. Courbe AUC-ROC du modèle CNN.

La comparaison des deux modèles de deep learning développés pour prédire la survie des patientes révèle des performances globalement excellentes pour chacun. Le modèle ANN démontre une très bonne précision (96,80 %) et une capacité remarquable de discrimination (AUC-ROC de 0,99), mais montre une légère faiblesse dans la détection des patientes décédées, avec un rappel de 77,63 %. En revanche, le modèle CNN présente des performances légèrement supérieures, avec une précision finale plus élevée (98,31 %) et un AUC-ROC atteignant 0,99, traduisant une capacité de prédiction robuste. Le CNN offre également un meilleur équilibre entre la précision et le rappel pour les patients décédés (rappel de 90,79 %), ce qui renforce sa pertinence clinique. Ainsi, bien que les deux modèles soient efficaces, le CNN se distingue par sa meilleure capacité à détecter les cas à risque et son excellente robustesse globale, en faisant un outil prometteur pour améliorer la prédiction de la survie des patientes atteintes de CS.

Le nombre d'époques d'entraînement a été adapté à chaque modèle afin d'assurer une convergence optimale. Le modèle ANN a été entraîné sur 2000 époques et le modèle CNN sur 1000 époques, selon leur dynamique d'apprentissage respective. La comparaison des performances repose donc sur leurs résultats finaux et non sur le nombre d'époques.



Discussion

Dans cette thèse, deux axes principaux ont été explorés afin de mieux comprendre la complexité du cancer du sein et d'améliorer la précision des diagnostics, du pronostic et des thérapies. Le premier axe porte sur la prédiction du statut HER2 à l'aide de quatre modèles d'apprentissage automatique, visant à optimiser l'identification de ce biomarqueur important pour les traitements ciblés. Le second axe concerne une analyse multiomique, qui explore l'impact de la méthylation de l'ADN et de l'expression génique sur la survie des patientes atteintes de cancer du sein, en intégrant les paramètres cliniques et biologiques. Ces deux recherches permettent d'approfondir la compréhension des mécanismes moléculaires du cancer du sein et d'améliorer les outils diagnostiques et pronostiques.

Dans la première étude, les résultats obtenus avec le modèle CatBoost ont montré des performances supérieures en termes de précision et de fiabilité, ce qui offre de nouvelles perspectives pour l'intégration de l'IA dans la personnalisation des traitements oncologiques. Les cancers HER2 positifs sont, dans la majorité des cas, associés à une croissance plus rapide et sont susceptibles de répondre à des thérapies ciblées, tandis que les cancers HER2 négatifs croissent plus lentement et sont moins susceptibles de récidiver ou de se propager. Environ 20 % des patientes atteintes d'un cancer du sein métastatique présentent une surexpression du gène HER2 (**Horimoto et al., 2022**). Le statut HER2 est évalué par immunohistochimie (IHC), qui quantifie la protéine HER2 sur les cellules cancéreuses, ou par hybridation in situ en fluorescence (FISH) pour détecter des copies supplémentaires du gène HER2 (**Che et al., 2023**). Dans le cancer du sein métastatique, la surexpression de la protéine HER2 est associée à des caractéristiques histologiques sévères, une mortalité élevée, un risque accru de récurrence et une réponse médiocre aux interventions thérapeutiques (**Slamon et al., 1987 ; Tandon et al., 1989 ; Moasser et al., 2007**). Les recherches récentes se sont concentrées sur la création de cadres prédictifs basés sur l'apprentissage automatique pour évaluer le statut HER2, dans le but d'améliorer la fiabilité et la compréhension des diagnostics pour les professionnels de santé. L'intelligence artificielle et les techniques de machine learning transforment ainsi le domaine de l'oncologie, notamment dans le diagnostic et le traitement du cancer du sein (**Conde-Sousa et al., 2022 ; Cordova et al., 2023**). Notre étude a mené une analyse comparative approfondie de plusieurs modèles de classification, incluant la régression logistique (LR), les forêts aléatoires (RF), LightGBM (LGBM) et CatBoost (CB), afin de prédire le statut HER2 dans le diagnostic du cancer du sein. Les résultats ont révélé que le modèle CatBoost surpassait les autres en termes de précision, d'exactitude et de performance globale, grâce à sa gestion avancée des

caractéristiques catégorielles. Son mécanisme de gradient boosting a permis de réduire de manière significative le surajustement, un défi fréquent dans les modèles d'apprentissage automatique (**Hancock et al., 2020**). De manière notable, le modèle CatBoost a présenté un équilibre exceptionnel entre la précision et le rappel, un critère clé pour minimiser les faux négatifs dans les diagnostics cliniques. Cela garantit que les patientes reçoivent un traitement adapté, réduisant ainsi les risques de sous-diagnostic. Une évaluation précise du statut HER2 est en effet essentielle pour identifier les patientes susceptibles de bénéficier d'une thérapie ciblée anti-HER2, seule ou en association avec une chimiothérapie, afin de favoriser une approche de médecine de précision (**Slamon et al., 2001 ; Liu et al., 2018**). L'efficacité de CatBoost dans notre étude corrobore les travaux existants, qui mettent en lumière le potentiel des modèles de gradient boosting dans le diagnostic médical. En exploitant un ensemble complexe d'arbres de décision, CatBoost capture efficacement les schémas subtils et complexes des données cliniques et pathologiques, souvent négligés par des modèles plus simples (**Zhang et al., 2019 ; Park et al., 2021**). Notre étude a utilisé plusieurs métriques de performance, notamment l'exactitude, le score F1 et l'AUC-ROC, conformément aux meilleures pratiques en machine learning. Cette approche garantit une évaluation exhaustive des performances des modèles. Nous avons également identifié plusieurs facteurs influençant la prédiction du statut HER2 chez les patientes atteintes de CS, tels que l'IMC, l'âge, le sous-type hormonal, l'implication nodale et la taille de la tumeur. En particulier, un IMC élevé était significativement associé à un statut HER2 négatif. L'obésité et le surpoids ont été corrélés à une réduction des chances d'obtenir une réponse pathologique complète (pCR) dans les cancers du sein luminal HER2 positif, notamment dans le cadre de la thérapie néoadjuvante (**Ebinç et al., 2022**). Des méta-analyses ont également confirmé l'impact négatif de l'IMC sur les taux de pCR sous traitement néoadjuvant à base d'anti-HER2. En intégrant des données cliniques et pathologiques complètes, les cliniciens peuvent prendre des décisions thérapeutiques plus personnalisées et mieux adaptées, influençant positivement la prise en charge du cancer du sein (**Ligorio et al., 2022**). Les principaux sous-types de cancer du sein, notamment Luminal A, Luminal B, HER2 positif et triple négatif (TNBC), présentent des caractéristiques distinctes selon le statut des récepteurs hormonaux et HER2. Ces différences influencent directement les stratégies thérapeutiques et les résultats cliniques des patientes. Par ailleurs, l'âge, la taille de la tumeur et l'implication des ganglions lymphatiques jouent un rôle critique dans la détermination du stade du cancer et dans l'orientation des décisions thérapeutiques (**O'Shaughnessy et al., 2023**). En combinant ces facteurs, les modèles prédictifs permettent aux professionnels de santé de prendre

des décisions éclairées sur les stratégies de traitement et la gestion des patientes. Notre étude confirme que les modèles avancés de machine learning, en particulier CatBoost, peuvent améliorer considérablement la précision de la prédiction du statut HER2 à partir de données cliniques et pathologiques. Ces résultats comblent une lacune dans la prédiction précise du statut HER2 et établissent une base solide pour l'application d'algorithmes avancés dans la personnalisation des stratégies thérapeutiques. Un modèle capable d'identifier avec précision les prédicteurs associés à la réponse thérapeutique pourrait influencer les décisions médicales dans un futur proche. Cependant, certaines limitations doivent être prises en compte. L'évaluation des performances des modèles dépend d'un jeu de données spécifique, ce qui peut restreindre la généralisation des résultats à d'autres populations ou contextes cliniques. Par exemple, la taille de l'échantillon et la diversité des données cliniques peuvent affecter la robustesse des prédictions. De plus, la sélection des variables et des biomarqueurs dans notre étude pourrait introduire des biais, notamment en raison de l'absence de certains facteurs cliniques potentiellement pertinents, comme les traitements antérieurs ou les variations de suivi. Enfin, l'absence de données longitudinales dans l'analyse limite la capacité à évaluer la dynamique temporelle de la progression du cancer et la réponse aux traitements au fil du temps. Des études supplémentaires, idéalement avec des jeux de données plus vastes et diversifiés, ainsi que l'inclusion de données longitudinales, sont nécessaires pour valider et étendre ces conclusions. L'évaluation des performances des modèles dépend d'un jeu de données spécifique, ce qui peut restreindre la généralisation des résultats. Des études supplémentaires sont nécessaires pour valider et étendre ces conclusions. En termes de performances, le classement des modèles selon nos critères est le suivant : 1. CatBoost (CB) : Meilleur modèle global, offrant une précision exceptionnelle et une perte minimale. 2. Régression Logistique (LR) : Bonne discrimination et rappel parfait, mais avec une perte plus élevée. 3. LightGBM (LGBM) : Compromis satisfaisant entre AUC-ROC et perte, bien que légèrement inférieur à CB et LR. 4. Random Forest (RF) : Moins performant en termes de précision et de perte, mais reste compétitif pour discriminer les cas HER2 positifs et négatifs.

Dans la seconde étude multiomique, nous avons exploré l'interaction complexe entre la méthylation de l'ADN, l'expression génique et les caractéristiques cliniques des patientes atteintes de cancer du sein, avec pour objectif d'identifier des biomarqueurs potentiels et des mécanismes moléculaires influençant la survie et la progression tumorale. En intégrant la méthylation et l'expression de gènes clés impliqués dans l'immunité et la régulation hormonale,

tels que *PDCD1*, *CD274*, *TLR4*, *TLR9*, *CTLA4*, *NFKB1*, *ESR1*, *PGR*, *ERBB2*, et *CYP19A1*, nous avons cherché à fournir une vision plus approfondie des dynamiques moléculaires sous-jacentes. Cette approche nous a permis de mieux comprendre la manière dont ces biomarqueurs, en interaction avec les caractéristiques cliniques des patientes. L'analyse de notre cohorte de 531 patientes atteintes de cancer du sein a révélé que le sous-type Luminal A représentait la majorité des cas (51,79 %), suivi des sous-types Luminal B et Basal. Ces résultats corroborent les observations antérieures, où le sous-type Luminal A est associé à un meilleur pronostic en raison d'une agressivité tumorale moindre et d'une meilleure réponse aux traitements hormonaux, par exemple, une étude marocaine rapporte 54,3 % de tumeurs Luminal A (**Fouad, A., et al., 2012**). En revanche, les sous-types HER2-enrichis et Basals sont associés à des formes plus agressives de cancer du sein, souvent caractérisées par une survie réduite, ce qui confirme l'importance de la classification moléculaire dans l'orientation des décisions thérapeutiques. Cette observation met en lumière le rôle essentiel de la classification moléculaire pour prédire la survie et la réponse aux traitements, notamment dans le cadre de traitements ciblés ou de thérapies combinées. L'âge moyen au diagnostic dans notre étude était de 57,56 ans, ce qui est similaire aux observations rapportées dans la littérature. En effet, plusieurs études épidémiologiques montrent que l'âge moyen au moment du diagnostic se situe généralement entre 55 et 60 ans dans les pays développés (**Global Cancer Observatory, 2023**). Cet âge avancé souligne la nécessité d'une meilleure sensibilisation pour le dépistage du cancer du sein chez les femmes âgées, un groupe de plus en plus touché par cette pathologie. Les analyses de la cohorte des profils de méthylation et d'expression génique dans le cancer du sein mettent en évidence des altérations épigénétiques et transcriptomiques spécifiques aux sous-types moléculaires, avec des conséquences cliniques majeures. Notre étude multiomique révèle des profils distincts des gènes immunitaires (*CD274*, *PDCD1*, *TLR9*) et hormonaux (*ESR1*, *PGR*, *ERBB2*, *CYP19A1*) qui varient selon ces sous-types. L'hypométhylation de ces gènes favorise leur activation, notamment une expression élevée de *PD-L1* et *PD-1* dans les sous-types HER2 et Basal-Like, liée à une évasion immunitaire accrue, tandis que les gènes hormonaux sont sur-exprimés dans les sous-types luminaux, témoignant de leur dépendance hormonale (**Saleh et al., 2020 ; Maet al., 2022 ; Chenet et al., 2024**). En revanche, le gène *CTLA4* présente une hyperméthylation, suggérant une suppression potentielle de son expression et une possible modulation négative de la réponse immunitaire (**Hoffmann et al., 2023 ; Jin et al., 2024**)

Grâce au clustering non supervisé (K-means et Agglomérative), deux groupes distincts de patientes ont été identifiés : un premier groupe, plus âgé, caractérisé par une forte méthylation

des gènes immunitaires et une expression hormonale élevée, et un second groupe, plus jeune, avec une faible méthylation et expression hormonale, probablement moins sensible aux traitements hormonaux mais plus susceptible de répondre à l'immunothérapie. Ces résultats soulignent l'importance d'une stratification multiomique fine permettant d'adapter les traitements du cancer du sein, en combinant ou ciblant spécifiquement les voies immunitaires et hormonales selon le profil moléculaire de la tumeur (Su et al., 2023). L'analyse des paramètres cliniques et démographiques selon les clusters issus du K-means révèle des différences notables suggérant des profils tumoraux distincts. Le cluster 0 regroupe majoritairement des patientes plus âgées, aux stades précoces (T1/T2, N0, M0) et présentant une meilleure survie globale, traduisant une maladie moins agressive, potentiellement plus sensible aux thérapies conventionnelles. En revanche, le cluster 1 se caractérise par des patientes légèrement plus jeunes, avec des stades plus avancés (notamment T3/T4, N1 et quelques M1) et une survie réduite, reflétant un profil tumoral plus agressif. Ce cluster est majoritairement constitué des sous-types luminal A et B, suggérant une forte dépendance hormonale, tandis que le cluster 0 est plus hétérogène, avec une proportion significative de tumeurs Basal-like et HER2. L'analyse conjointe des niveaux de méthylation et d'expression des gènes immunitaires et hormonaux met en évidence deux sous-groupes tumoraux distincts, porteurs de signatures épigénétiques et transcriptionnelles contrastées. Le Cluster 0 se caractérise par une hyperméthylation coordonnée de gènes clés tels que *PDCD1*, *CD274* ou *ESR1*, souvent associée à une expression paradoxalement maintenue, suggérant une régulation complexe et un état d'échappement immunitaire ou hormonal potentiellement stabilisé. À l'inverse, le Cluster 1 présente des profils de méthylation plus hétérogènes, traduisant une plus grande plasticité tumorale et possiblement une meilleure réactivité aux traitements ciblant les voies immunitaires ou endocrines (Kakushadze et Yu, 2017). L'analyse conjointe de la méthylation et de l'expression génique dans le cancer du sein a révélé des corrélations majoritairement négatives, notamment pour les gènes hormonaux *PGR*, *ESR1* et *ERBB2*, ce qui confirme le modèle épigénétique classique selon lequel une méthylation accrue réprime l'expression génique (Gorse et al., 2024). Cette régulation affecte particulièrement les récepteurs hormonaux, influençant potentiellement la réponse aux traitements hormonaux (Bertucci et al., 2008).

Les différences entre clusters illustrent la complexité biologique du cancer du sein. Le cluster 0, avec des corrélations négatives fortes, montre une forte dépendance hormonale. Par contre le cluster 1 présente des corrélations plus hétérogènes, notamment pour *TLR9*, traduisant une composante immunitaire plus marquée. Cela suggère des profils tumoraux distincts. Par ailleurs,

l'analyse des corrélations entre les expressions des gènes immunitaires et hormonaux a révélé un antagonisme transcriptionnel. Les marqueurs immunosuppresseurs CTLA4-E et PDCD1-E sont inversement corrélés à ESR1-E et PGR-E, indiquant que les tumeurs luminales présentent un microenvironnement moins immunosuppresseur, ce qui pourrait expliquer leur réponse limitée à l'immunothérapie (Peixoto et al., 2018). En revanche, une corrélation positive entre PDCD1-E et CYP19A1-E évoque un lien entre l'épuisement immunitaire et la production locale d'œstrogènes, justifiant l'étude de thérapies combinant inhibiteurs de l'aromatase et immunomodulateurs (Sidaway, 2017). Enfin, le statut ménopausique apparaît comme un facteur modulateur de la méthylation, en particulier pour les gènes *TLR4* et *ESR1*, suggérant que les changements hormonaux liés à la ménopause influencent l'épigénome tumoral et les interactions immuno-hormonales (Hilser et al., 2022). L'analyse conjointe de la méthylation et de l'expression génique des gènes immuno-hormonaux dans notre cohorte met en lumière des profils biologiques différenciés, aux implications pronostiques notables dans le cancer du sein. Contrairement à certaines études suggérant le rôle prépondérant de la méthylation comme biomarqueur isolé (Liyanage et al., 2019), nos résultats indiquent que la méthylation seule des loci étudiés (*PDCD1*, *CD274*, *TLR4*, etc...) ne distingue pas significativement les patientes selon leur survie. En revanche, l'expression génique, en particulier celle de *PDCD1*, *CTLA4* et *ERBB2*, s'avère plus discriminante, les patientes vivantes présentant une expression significativement plus élevée de ces gènes. Ce constat rejoint les travaux de Dieci et al. (2021), qui ont mis en évidence le rôle protecteur de l'expression de *PD-1/PD-L1* et d'autres régulateurs immunitaires dans la réponse anti-tumorale. La surexpression d'*ERBB2*, souvent associée à un mauvais pronostic dans les sous-types HER2-positifs, pourrait ici refléter une réponse favorable aux thérapies ciblées (trastuzumab, pertuzumab), comme le suggèrent les résultats de l'essai CLEOPATRA (Swain et al., 2015). Par ailleurs, les courbes de survie Kaplan-Meier et les tests de log-rank confirment que des profils combinés de faible méthylation et forte expression sont associés à une survie prolongée, tant pour les gènes immunitaires (*TLR4*, *CD274*, *CTLA4*) que hormonaux (*ESR1*, *CYP19A1*, *PGR*). Cela concorde avec les observations de Baylin et Jones (2016), qui ont montré que l'hypométhylation de promoteurs associés à des gènes régulateurs permet une transcription active favorable à la surveillance immunitaire ou à la réponse hormonale. L'analyse ROC soutient ces observations en soulignant la performance exceptionnelle de *TLR4* (AUC = 0,98) en tant que marqueur épigénétique, mettant en lumière son rôle dans l'activation de la réponse immunitaire innée et son potentiel comme biomarqueur diagnostique, en accord avec les données de Kawai et Akira (2010). Côté transcriptionnel,

ESR1-E (AUC = 0,85) se démarque également, confirmant son importance dans la biologie hormonale du sein et la pertinence de son expression dans le pronostic (Ali et Coombes, 2002). Enfin, l'hétérogénéité observée selon les clusters K-means suggère que certains biomarqueurs, tel *TLR9* (méthylé), sont particulièrement pertinents dans des sous-groupes spécifiques, illustrant la nécessité de stratégies personnalisées basées sur des profils multi-omiques (Curtis et al., 2012).

L'analyse par clustering fonctionnel des voies KEGG a permis d'identifier deux signatures moléculaires distinctes dans le cancer du sein. Le Cluster 0 est enrichi en gènes liés au développement mammaire et à la régulation hormonale, notamment via la signalisation œstrogénique impliquant *ESR1*, un facteur clé dans la différenciation des cellules mammaires et dans la réponse aux traitements endocriniens. Les termes significativement associés, tels que "*female gonad development*", "*response to estrogen stimulus*", ou encore "*mammary gland branching involved in pregnancy*", avec des p-values très faibles (ex. 0.000294), confirment l'implication de voies hormonales dans la tumorigenèse mammaire (Kechagias et al., 2021).

En revanche, le Cluster 1 met en évidence une signature immunitaire robuste, avec un enrichissement significatif des voies de régulation de la réponse immune, telles que *PD-1/PD-L1 checkpoint pathway*, *Toll-like receptor signaling*, et *cytokine-mediated signaling*. Des termes comme "*positive regulation of immune response*" ou "*interactions between immune cells and microRNAs*" montrent une très forte significativité ($p < 0.0001$, $-\log_{10}(p) > 4$), indiquant que ce cluster est dominé par des processus immuno-inflammatoires (Emens et al., 2019 ; Chen et al., 2021). Ces profils moléculaires évoquent des tumeurs immunogéniques, potentiellement sensibles à des immunothérapies ciblant les points de contrôle immunitaires, notamment les anticorps anti-PD-1/PD-L1. L'analyse de régression de Cox réalisée sur la population totale a mis en évidence que parmi les variables cliniques et moléculaires étudiées, l'âge au diagnostic pathologique et le stade métastatique (M) sont des facteurs pronostiques significatifs du risque de décès chez les patientes atteintes de cancer du sein. L'âge augmente le risque de 3 % par année supplémentaire, tandis que la présence de métastases accroît ce risque de 80 %. Sur le plan moléculaire, l'expression de *PDCDI* apparaît comme un facteur protecteur, réduisant le risque de décès de 15 %. En revanche, les autres marqueurs moléculaires (ESR1-E, CYP19A1-E, PGR-E) n'ont pas montré d'effet significatif. Malgré ces résultats, la performance prédictive globale du modèle reste faible (concordance = 0,24). L'analyse par modèles de Cox appliquée séparément aux clusters définis par la méthylation des biomarqueurs immuno-hormonaux n'a pas révélé de facteurs significatifs, mais plusieurs aspects importants émergent ; dans le Cluster

0, la méthylation d'*ESR1* et de *CYP19A1* est associée à une tendance protectrice (réduction du risque), probablement par inhibition des voies hormonales favorisant la croissance tumorale. Par contre, la méthylation du gène *PGR* semble liée à une augmentation du risque, bien que les intervalles de confiance larges invitent à la prudence. La méthylation des gènes immunitaires *PDCD1* et *CD274* montre une tendance non significative vers un pronostic défavorable. Dans le Cluster 1, certaines variables comme la méthylation de *CD274* (*PD-L1*) tendent à augmenter le risque (HR très élevé, mais non significatif), tandis que d'autres (*PDCD1*, *TLR4*) semblent protectrices. Ces résultats suggèrent une hétérogénéité biologique importante entre les clusters, soulignant la complexité de l'impact de la méthylation des gènes immuno-hormonaux sur la survie (Bouaziz et al., 2023). Deux approches de deep learning ont été mises en œuvre pour prédire la survie des patientes atteintes de cancer du sein à partir de données multiomiques combinant des biomarqueurs épigénétiques, transcriptomiques et des paramètres cliniques. Les modèles explorés incluent un réseau de neurones artificiels entièrement connecté (ANN) et un réseau de neurones convolutifs (CNN), chacun ayant fait l'objet d'un entraînement intensif et d'une évaluation rigoureuse.

La comparaison des deux modèles met en évidence que, bien que l'ANN présente déjà une performance solide, le CNN offre une précision plus élevée et un meilleur équilibre entre précision et rappel pour les deux classes. Ces résultats suggèrent que le CNN est mieux adapté à la complexité inhérente aux données multiomiques utilisées dans cette étude.

Le faible rappel observé pour les patientes décédées dans le modèle ANN souligne cependant certaines limites, en particulier dans la détection des cas à risque. Cela met en lumière la nécessité d'optimisations complémentaires, telles que : le rééquilibrage des classes (par oversampling des cas rares ou l'utilisation d'une fonction de perte pondérée) ; l'ajustement des hyperparamètres (par exemple, le taux d'apprentissage, la profondeur du réseau) ; l'intégration de techniques de régularisation (comme le dropout ou la normalisation) afin d'améliorer la généralisation du modèle.

À l'inverse, le CNN apparaît plus robuste, même avec un nombre d'époques d'apprentissage inférieur, et constitue une base prometteuse pour des approches plus avancées, telles que les réseaux hybrides, les mécanismes attentionnels ou l'intégration de réseaux récurrents (LSTM) en cas de données temporelles. Les deux modèles développés dans ce travail démontrent une capacité élevée à prédire la survie des patientes à partir de données multiomiques. Toutefois, le CNN se distingue par une meilleure capacité de généralisation et une précision accrue, notamment dans l'identification des patientes à risque. Ces résultats confirment la pertinence des

architectures avancées de deep learning dans le contexte de la médecine personnalisée et ouvrent des perspectives prometteuses pour le développement de systèmes d'aide à la décision médicale plus performants et mieux adaptés à la complexité des données cliniques et biologiques. Ces résultats confirment l'intérêt croissant des approches d'apprentissage profond pour l'analyse prédictive en oncologie, notamment en ce qui concerne la survie des patientes atteintes de cancer du sein, comme le montrent d'autres travaux basés sur l'imagerie histopathologique ou des approches multi-modales (Wetstein et al., 2022 ; Hamedi et al., 2024).

Ce travail de thèse a permis d'explorer deux approches complémentaires pour une meilleure compréhension du cancer du sein et une amélioration de sa prise en charge : la prédiction du statut HER2 par apprentissage automatique et l'analyse multiomique intégrant l'expression génique, la méthylation de l'ADN et les données cliniques.

La première partie a démontré l'efficacité du modèle CatBoost pour prédire le statut HER2 avec une précision notable, surpassant les autres modèles testés. Cette performance met en évidence le potentiel des techniques d'intelligence artificielle dans le soutien aux décisions cliniques, en particulier pour identifier les patientes éligibles à des thérapies ciblées. En intégrant des variables cliniques pertinentes, les modèles prédictifs peuvent contribuer à une médecine plus personnalisée, mieux adaptée aux profils individuels.

La seconde partie a permis de caractériser des profils moléculaires distincts à travers une approche multiomique, révélant des interactions complexes entre méthylation, expression génique et paramètres cliniques. Cette analyse a mis en évidence l'importance de certains biomarqueurs immuno-hormonaux dans la survie et la progression du cancer du sein, ouvrant la voie à des pistes prometteuses pour des thérapies combinées ou individualisées.

En somme, cette thèse souligne l'intérêt de combiner les approches computationnelles avancées et l'intégration multiomique pour affiner le diagnostic, le pronostic et les stratégies thérapeutiques dans le cancer du sein. Néanmoins, la généralisation de ces résultats nécessite des validations sur des cohortes plus larges et diversifiées, ainsi qu'une prise en compte des données longitudinales. Ces travaux s'inscrivent dans la dynamique actuelle de la médecine de précision et plaident en faveur d'une collaboration étroite entre cliniciens, biologistes et data scientists.



Conclusion

Cette thèse a permis d'approfondir la compréhension du cancer du sein à travers deux approches complémentaires : la prédiction du statut HER2 par apprentissage automatique et l'analyse multiomique intégrant méthylation de l'ADN, expression génique et données cliniques pour mieux cerner les interactions géniques influençant la survie des patientes.

La première partie a mis en lumière l'importance cruciale d'une évaluation précise du statut HER2 en soulignant l'impact de facteurs cliniques comme l'IMC, l'âge et le sous-type tumoral. Ce travail a démontré l'efficacité du modèle CatBoost pour prédire le statut HER2 avec une précision notable, surpassant les autres modèles testés. Cette performance met en évidence le potentiel des techniques d'intelligence artificielle dans le soutien aux décisions cliniques, en particulier pour identifier les patientes éligibles à des thérapies ciblées. En intégrant des variables cliniques pertinentes, les modèles prédictifs peuvent contribuer à une médecine plus personnalisée, mieux adaptée aux profils individuels.

La seconde partie a permis de caractériser des profils moléculaires distincts à travers une approche multiomique, révélant des profils moléculaires distincts selon les sous-types de cancer, avec des signatures épigénétiques et transcriptionnelles complexes qui conditionneraient la sensibilité aux traitements hormonaux ou immunothérapeutiques. La stratification fine des patientes en fonction de ces profils ouvre la voie à une médecine de précision plus efficace, adaptée à la diversité biologique des tumeurs. Cette analyse a mis en évidence l'importance de certains biomarqueurs immuno-hormonaux dans la survie et la progression du cancer du sein, ouvrant la voie à des pistes prometteuses pour des thérapies combinées ou individualisées.

Cette thèse met en lumière l'apport déterminant des approches multiomiques, notamment l'intégration des données épigénétiques et transcriptomiques, dans la compréhension fine des mécanismes du cancer du sein et l'amélioration de la prédiction de la survie des patientes. L'analyse conjointe de la méthylation de l'ADN et de l'expression génique a permis d'identifier des signatures moléculaires spécifiques aux sous-types tumoraux, révélant des profils distincts de régulation immunitaire et hormonale qui conditionnent à la fois l'agressivité tumorale et la réponse aux traitements. Les corrélations négatives observées entre la méthylation et l'expression des gènes hormonaux, ainsi que l'antagonisme transcriptionnel entre marqueurs immunitaires et hormonaux, illustrent la complexité des interactions moléculaires influençant la progression de la maladie et la survie. La stratification des patientes selon ces profils multiomiques a permis de distinguer des groupes à pronostic différencié, certains caractérisés

par une dépendance hormonale et une meilleure survie, d'autres par une plasticité épigénétique accrue et une composante immunitaire dominante, potentiellement plus sensibles à l'immunothérapie. Ces résultats soulignent l'importance d'une prise en charge personnalisée, fondée sur la caractérisation fine de la biologie tumorale, et ouvrent la voie à l'intégration des signatures épigénétiques et transcriptomiques dans les outils pronostiques et les choix thérapeutiques.

En somme, cette thèse souligne l'intérêt de combiner les approches computationnelles avancées et l'intégration multiomique et contribue significativement à l'amélioration des outils diagnostiques et pronostiques ainsi que les stratégies thérapeutiques dans le cancer du sein en intégrant les avancées de l'intelligence artificielle et de la biologie moléculaire.

Néanmoins, la généralisation de ces approches nécessite des validations sur des cohortes plus larges et diversifiées, ainsi qu'une prise en compte des données longitudinales pour valider la robustesse des signatures identifiées et affiner encore la prédiction de la survie, dans une perspective de médecine de précision toujours plus adaptée à la diversité des cancers du sein.. Ces travaux s'inscrivent dans la dynamique actuelle de la médecine de précision et plaident en faveur d'une collaboration étroite entre cliniciens, biologistes et data scientists.



Références Bibliographiques

A

Adetola V, Guay M. Integration of real-time optimization and model predictive control. *J Process Control*. 2010, 20(2):125–133. <https://doi.org/10.1016/j.jprocont.2009.09.001>

Admoun C, Mayrovitz H. Choosing Mastectomy vs. Lumpectomy-With-Radiation: Experiences of Breast Cancer Survivors. *Cureus*. 2021, 13(10):e18433. <https://doi.org/10.7759/cureus.18433>

Andrahennadi S, Sami A, Manna M, Pauls M, Ahmed S. Current Landscape of Targeted Therapy in Hormone Receptor-Positive and HER2-Negative Breast Cancer. *Current Oncology*. 2021, 28(3):1803–1822. <https://doi.org/10.3390/curroncol28030168>

Ali S, Coombes RC. Estrogen receptor alpha in human breast cancer: occurrence and significance. *J mammary gland biology neoplasia*. 2000, 5(3):271–281. <https://doi.org/10.1023/a:1009594727358>

Alleman I, Couvert-Mullot H, Béranger C, Gisserot O. Prise en charge du cancer du sein en cas de récepteurs hormonaux négatifs. *Le Pharmacien Hospitalier*. 2010, 45(1):25–32. <https://doi.org/10.1016/j.phhp.2010.01.005>

Alsaab HO, Sau S, Alzhrani R, Tatiparti K, Bhise K, Kashaw SK, Iyer AK. PD-1 and PD-L1 checkpoint signaling inhibition for cancer immunotherapy: Mechanism, combinations, and clinical outcome. *Frontiers in Pharmacol*. 2017, 8:561. <https://doi.org/10.3389/fphar.2017.00561>

Aranda CJ, Gonzalez-Kozlova E, Saunders SP, Fernandes-Braga W, Ota M, Narayanan S, He JS, Del Duca E, Swaroop B, Gnjjatic S, Shattner G, Reibman J, Soter NA, Guttman-Yassky E, Curotto de Lafaille MA. IgG memory B cells expressing IL4R and FCER2 are associated with atopic diseases. *Allergy*. 2023, 78(3):752–766. <https://doi.org/10.1111/all.15601>

Asaoka M, Gandhi S, Ishikawa T, Takabe, K. Neoadjuvant chemotherapy for breast cancer: Past, present, and future. *Breast Cancer*. 2020, 14:1178223420980377. <https://doi.org/10.1177/1178223420980377>

Audeh W, Blumencranz L, Kling H, Trivedi H, Srkalovic G. Prospective Validation of a Genomic Assay in Breast Cancer: The 70-gene MammaPrint Assay and the MINDACT Trial. *Acta medica academica*. 2019, 48(1):18–34. <https://doi.org/10.5644/ama2006-124.239>

B

Badowska-Kozakiewicz AM, Patera J, Sobol M, Przybylski J. The role of oestrogen and progesterone receptors in breast cancer - immunohistochemical evaluation of oestrogen and progesterone receptor expression in invasive breast cancer in women. *Contemp Oncol (Pozn)*. 2015, 19(3):220–225. <https://doi.org/10.5114/wo.2015.51826>

Barzaman K, Karami J, Zarei Z, Hosseinzadeh A, Kazemi MH, Moradi-Kalbolandi S, Safari E, Farahmand L. Breast cancer: Biology, biomarkers, and treatments. *Int Immunopharmacol*. 2020, 84:106535. <https://doi.org/10.1016/j.intimp.2020.106535>

Basse C, Arock M. The increasing roles of epigenetics in breast cancer: Implications for pathogenicity, biomarkers, prevention and treatment. *Int J Cancer*. 2015, 137(12):2785–2794. <https://doi.org/10.1002/ijc.29347>

Baylin SB, Jones PA. Epigenetic determinants of cancer. *Cold Spring Harb Perspect Biol*. 2016, 8(9):a019505. <https://doi.org/10.1101/cshperspect.a019505>

Bazira PJ, Ellis H, Mahadevan V. Anatomy and physiology of the breast. *Surgery (Oxford)*. 2022, 40(2):79–83. <https://doi.org/10.1016/j.mpsur.2021.11.015>

Bertucci F, Finetti P, Cervera N, Birnbaum D. Classification pronostique du cancer du sein et profils d'expression génique sur puces à ADN. *Med Sci (Paris)*. 2008, 24(6-7):599–606. <https://doi.org/10.1051/medsci/20082467599>

Björnström L, et Sjöberg M. Mechanisms of estrogen receptor signaling: convergence of genomic and nongenomic actions on target genes. *Mol Endocrinol*. 2005, 19(4):833–842. <https://doi.org/10.1210/me.2004-0486>

Boron WF, Boulpaep EL. *Medical Physiology*. 3rd ed. Philadelphia: Elsevier; 2016. ISBN: 9781455743773

Boundedjar A, Melzi MA, Idir H, Heba N. General oncology care in Algeria. In: Al-Shamsi HO, Abu-Gheida IH, Iqbal F, Al-Awadhi A, editors. *Cancer in the Arab World*. Springer; 2022. P.15–30. <https://doi.org/10.1007/978-981-16-7945-2>

Breiman L. Random forests. *Mach Learn*. 2001, 45(1):5–32. <https://doi.org/10.1023/A:1010933404324>

Briskin C. Progesterone signalling in breast cancer: A neglected hormone coming into the limelight. *Nat Rev Cancer*. 2013, 13(6):385–396. <https://doi.org/10.1038/nrc3518>

Brown M, Cohen J, Arun P, Chen Z, Van Waes C. NF-kappaB in carcinoma therapy and prevention. *Expert Opin Ther Targets*. 2008, 12(9):1109–1122. <https://doi.org/10.1517/14728222.12.9.1109>

C

Caciolla J, Bisi A, Belluti F, Rampa A, Gobbi S. Reconsidering aromatase for breast cancer treatment: New roles for an old target. *Molecules*. 2020, 25(22):5351. <https://doi.org/10.3390/molecules25225351>

Candelaria RP, Hwang L, Bouchard RR, Whitman GJ. Breast ultrasound: current concepts. *Seminars in Ultrasound, CT, and MR*. 2013, 34(3):213–225. <https://doi.org/10.1053/j.sult.2012.11.013>

Cardoso F, van't Veer LJ, Bogaerts J, Slaets L, Viale G, Delaloge S, et al. 70-Gene Signature as an Aid to Treatment Decisions in Early-Stage Breast Cancer. *New England J Med*. 2016, 375(8):717–729. <https://doi.org/10.1056/NEJMoa1602253>

Cartron PF, Pacaud R, Salbert G. Méthylation/déméthylation de l'ADN. *Rev Francoph Lab*. 2015, (473):37-48. [https://doi.org/10.1016/S1773-035X\(15\)30158-1](https://doi.org/10.1016/S1773-035X(15)30158-1)

Casalino, L. Verde, P. Multifaceted Roles of DNA Methylation in Neoplastic Transformation, from Tumor Suppressors to EMT and Metastasis. *Genes*, 2020, 11, Article 922. <https://doi.org/10.3390/genes11080922>

Chakravorty S, Manisha. Data analysis of malnutrition in India: a review of numerous factors. *Internl J Community Med Public Health*. 2023, 10(7):2629–2636. <https://doi.org/10.18203/2394-6040.ijcmph20232064>

Chaudhary SK, Kumar A, Singh R, Kumar M. Machine learning algorithms for soil moisture estimation using Sentinel-1: Model development and implementation. *Advances in Space Research*. 2022, 69(6):1799-1812. <https://doi.org/10.1016/j.asr.2021.08.022>

Che Y, Ren F, Zhang X, Cui L, Wu H, Zhao Z. Immunohistochemical HER2 Recognition and Analysis of Breast Cancer Based on Deep Learning. *Diagnostics*. 2023, 13:263.

- Chen DS, Mellman I.** Elements of cancer immunity and the cancer-immune set point. *Nature*. 2021, 541(7637):321–330. <https://doi.org/10.1038/nature21349>
- Chen X, Yu S, Chen J, Chen X.** Analysis of PD-L1 promoter methylation combined with immunogenic context in pancreatic ductal adenocarcinoma. *Cancer Immunology, Immunotherapy*. 2024, 73(8):149. <https://doi.org/10.1007/s00262-024-03745-y>
- Cheng Y, He C, Wang M, Ma X, Mo F, Yang S, Han J, Wei X.** Targeting epigenetic regulators for cancer therapy: mechanisms and advances in clinical trials. *Signal Transduction and Targeted Therapy*. 2019, 4:62. <https://doi.org/10.1038/s41392-019-0095-0>
- Cherif H, Bidoli E, Birri S, Mahnane A, Zaidi Z, Boukharouba H, et al.** Cancer estimation of incidence and survival in Algeria 2014. *J Cancer Research Therapy*. 2015, 3(9):100–104. <https://doi.org/10.14312/2052-4994.2015-14>
- Chuang YT, Lee YM, Chan MWY.** Combinatorial Epigenetic and Immunotherapy in Breast Cancer Management: A Literature Review. *Epigenomes*. 2020, 4(4):27. <https://doi.org/10.3390/epigenomes4040027>
- Chumsri S, Howes T, Bao T, Sabnis G, Brodie A.** Aromatase, aromatase inhibitors, and breast cancer. *J Steroid Biochemistry Molecular Biology*. 2011, 125(1–2):13–22. <https://doi.org/10.1016/j.jsbmb.2011.02.001>
- Cirolla VA.** Breast anatomy. *Virginia A. Cirolla's Lab*. 2017. <https://www.researchgate.net/publication/321011257>
- Coates AS, Winer EP, Goldhirsch A, Gelber RD, Gnant M, Piccart-Gebhart M, et al.** Tailoring therapies--improving the management of early breast cancer: St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2015. *Annals of Oncology*. 2015, 26(8):1533–1546. <https://doi.org/10.1093/annonc/mdv221>
- Concetti J, Wilson CL.** FKB1 and cancer: Friend or foe? *Cells*. 2018, 7(9):133. <https://doi.org/10.3390/cells7090133>
- Conde-Sousa E, Vale J, Feng M, Xu K, Wang Y, Della Mea V, La Barbera D, Montahaei E, Baghshah M, Turzynski A, Gildenblat J, Klaiman E, Hong Y, Aresta G, Araújo T, Aguiar P, Eloy C, Polónia A.** HEROHE Challenge: Predicting HER2 Status in Breast Cancer from Hematoxylin–Eosin Whole-Slide Imaging. *J Imaging*. 2022; 8(8):213. <https://doi.org/10.3390/jimaging8080213>

Cook DN, Pisetsky DS, Schwartz DA. Toll-like receptors in the pathogenesis of human disease. *Nature Immunology*. 2004, 5(10):975–979. <https://doi.org/10.1038/ni1116>

Cordova C, Muñoz R, Olivares R, Minonzio JG, Lozano C, Gonzalez P, Marchant I, González-Arriagada W, Olivero P. HER2 classification in breast cancer cells: A new explainable machine learning application for immunohistochemistry. *Oncol Lett*. 2022; 25(2):44. <https://doi.org/10.3892/ol.2022.13630>

Costello J.F, Plass C. Methylation matters. *J Medical Genetics*. 2001, 38(5):285–303. <https://doi.org/10.1136/jmg.38.5.285>

Curtis C, Shah SP, Chin SF, Turashvili G, Rueda OM, Dunning MJ, et al. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. *Nature*. 2012, 486(7403):346–352. <https://doi.org/10.1038/nature10983>

D

Damis E, Gucciardo L, Berrefas L, Goyens P. L'allaitement maternel : de la physiologie à la pratique. *Revue Médicale de Bruxelles*. 2012, 33(4):318–327. <https://pubmed.ncbi.nlm.nih.gov/23091937/>

Das K, Gunasegaran B, Tan IB, Deng N, Lim KH, Tan P. Mutually exclusive FGFR2, HER2, and KRAS gene amplifications in gastric cancer revealed by multicolour FISH. *Cancer Lett*. 2014, 353(2):167–175. <https://doi.org/10.1016/j.canlet.2014.07.021>

Debien V, De Caluwé A, Wang X, Piccart-Gebhart M, Tuohy VK, Romano E, Buisseret L. Immunotherapy in breast cancer: an overview of current strategies and perspectives. *NPJ Breast Cancer*. 2023, 9(1):7. <https://doi.org/10.1038/s41523-023-00508-3>

Delneste Y, Beauvillain C, Jeannin, P. Immunité naturelle : Structure et fonction des Toll-like receptors. *Médecine/Sciences (Paris)*. 2007, 23(1):67–74. <https://doi.org/10.1051/medsci/200723167>

Deng X, Qing Y, Horne D, Huang H, Chen J. The roles and implications of RNA m6A modification in cancer. *Nature Reviews Clinical Oncology*. 2023, 20(8):507–526. <https://doi.org/10.1038/s41571-023-00774-x>

Deshmukh A, Kumar S, Arfuso F, Newsholme P, Dharmarajan A, Vinod M. Secreted Frizzled-related protein 4 (sFRP4) chemo-sensitizes cancer stem cells derived from human breast, prostate,

and ovary tumor cell lines *Scientific Reports* 2017 7 2256 <https://doi.org/10.1038/s41598-017-02256-4>

Dieci MV, Criscitiello C, Goubar A, Viale G, Conte P, Guarneri V, Ficarra G, Mathieu MC, Delalogue S, Curigliano G, Andre F. Prognostic value of tumor-infiltrating lymphocytes on residual disease after primary chemotherapy for triple-negative breast cancer: a retrospective multicenter study. *Annals of Oncology*. 2014, 25(3):611–618 <https://doi.org/10.1093/annonc/mdt556>

Digel W, Lübbert M. DNA methylation disturbances as novel therapeutic target in lung cancer: preclinical and clinical results. *Critical Reviews in Oncology/Hematology*. 2005, 55(1):1–11. <https://doi.org/10.1016/j.critrevonc.2005.02.002>

D'Olne E, Neo VW, Naylor PA. Speech enhancement in distributed microphone arrays using polynomial eigenvalue decomposition In 2022 30th European Signal Processing Conference (EUSIPCO) 2022, (pp 55–59) IEEE. <https://doi.org/10.23919/EUSIPCO55093.2022.9909555>

Dumbrava EEI, Balaji K, Raghav K, Hess K, Javle M, Blum-Murphy M, Ajani J, Kopetz S, Broaddus R, Routbort M, Demirhan M, Zheng X, Pant S, Tsimberidou AM, Subbiah V, Hong DS, Rodon J, Shaw KM, Piha-Paul SA, Meric-Bernstam F. Targeting ERBB2 (HER2) Amplification Identified by Next-Generation Sequencing in Patients With Advanced or Metastatic Solid Tumors Beyond Conventional Indications. *JCO Precision Oncology*. 2019, 3:PO.18.00345. <https://doi.org/10.1200/PO.18.00345>

E

Ebiñç S, Oruç Z, Sezgin Y, Karhan O, Bilen E, Yerlikaya H, Işikdoğan A. Factors Influencing HER2 Discordance in Nonmetastatic Breast Cancer and The Role of Neoadjuvant Therapy. *Future Oncology*. 2022, 18(30):3399–3408. <https://doi.org/10.2217/fon-2022-0198>

Ebrahimi N, Abdulwahid AHRR, Mansouri A, Karimi N, Jafardoust Bostani R, Beiranvand S, Adelian S, Khorram R, Vafadar R, Hamblin MR, Aref AR. Targeting the NF-κB pathway as a potential regulator of immune checkpoints in cancer immunotherapy. *Cellular Molecular Life Scien.* 2024, 81:106. <https://doi.org/10.1007/s00018-023-05098-8>

Edge SB, Compton CC. The American Joint Committee on Cancer The 7th edition of the AJCC cancer staging manual and the future of TNM. *Annals of Surgical Oncology*. 2010, 17(6):1471–1474. <https://doi.org/10.1245/s10434-010-0985-4>

Egger G, Liang G, Aparicio A, Jones PA. Epigenetics in human disease and prospects for epigenetic therapy. *Nature*. 2004, 429(6990):457–463. <https://doi.org/10.1038/nature02625>

Elhaddad M, Hamam S. AI-Driven Clinical Decision Support Systems An Ongoing Pursuit of Potential Cureus 2024, 16(4):e57728. <https://doi.org/10.7759/cureus.57728>

Emens LA, Loi S. Immunotherapy Approaches for Breast Cancer Patients in 2023. *Cold Spring Harbor perspectives in medicine* 2023, 13(4) :a041332. <https://doi.org/10.1101/cshperspect.a041332>

Emens LA, Adams S, Barrios CH, Duffy AG, Kalinsky K, McDonnell AM, Middleton MR, Nanda R, Sharma P, Loi S. A review of immunotherapy and immunotherapeutic combinations in breast cancer. *J Clinical Oncology*. 37(27):2611–2619. <https://doi.org/10.1200/JCO.19.00907>

Esteller M. Cancer epigenomics DNA methylomes and histone-modification maps. *Nature reviews Genetics*. 2007, 8(4):286–298. <https://doi.org/10.1038/nrg2005>

Esteller M. Epigenetics provides a new generation of oncogenes and tumour-suppressor genes *British J cancer*. 2006, 94(2):179–183. <https://doi.org/10.1038/sj.bjc.6602918>

Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, Bussaglia E, Prat J, Harkes IC, Repasky EA, Gabrielson E, Schutte M, Baylin SB, Herman JG. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. *J the Nat Cancer Institute*. 2000, 92(7):–569. <https://doi.org/10.1093/jnci/92.7.564>

Eze E, Eze J. Artificial intelligence support for 5G/6G-enabled Internet of Vehicles networks An overview ITU. *Journal on Future and Evolving Technologies*. 2023, 4:178–195. <https://doi.org/10.52953/IEZN8770>

F

Fabian CJ, Kimler BF, Hursting SD. Omega-3 fatty acids for breast cancer prevention and survivorship. *Breast Cancer Research*. 2015, 17:62. <https://doi.org/10.1186/s13058-015-0571-6>

Feinberg AP, Levchenko A. Epigenetics as a mediator of plasticity in cancer *Science (New York, N Y)* 2023, 379(6632):eaaw3835. <https://doi.org/10.1126/science.aaw3835>

Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature*. 1983, 301(5895):89–92. <https://doi.org/10.1038/301089a0>

Fitz-James MH, Cavalli G. Molecular mechanisms of transgenerational epigenetic inheritance *Nature reviews Genetics*. 2022, 23(6):325–341. <https://doi.org/10.1038/s41576-021-00438-5>

Fouad A, Yousra A, Kaoutar Z, Omar elM, Afaf A, Sanae B. Classification moléculaire du cancer du sein au Maroc Molecular classification of breast cancer in Morocco. *The Pan African medical journal*. 2012, 13:91.

Foulkes WD. The ten genes for breast (and ovarian) cancer susceptibility. *Nat Rev Clin Oncol*. 2021, 18:259–260. <https://doi.org/10.1038/s41571-021-00491-3>

Feng Y, Wang Y, Wang P, Huang Y, Wang F. Short-Chain Fatty Acids Manifest Stimulative and Protective Effects on Intestinal Barrier Function Through the Inhibition of NLRP3 Inflammasome and Autophagy Cellular. *Physiol Biochem*. 2018, 49(1):190–205. <https://doi.org/10.1159/000492853>

G

Galbraith K, Snuderl M. DNA methylation as a diagnostic tool. *Acta neuropathol commun*. 2022, 10(1):71. <https://doi.org/10.1186/s40478-022-01371-2>

Gao Y, Zhang H, Peng J, Li L, Xiao Y, Li L, Liu Y, Qiao Y, Chou SL. A 30-year overview of sodium-ion batteries. *Carbon Energy*. 2024, 6:e464. <https://doi.org/10.1002/cey2.464>

Gajria D, Chandarlapaty S. HER2-amplified breast cancer mechanisms of trastuzumab resistance and novel targeted therapies Expert, *Rev Anticancer Ther*, 2011, 11(2):263–275. <https://doi.org/10.1586/era.10.226>

Ghosh TS, Shanahan F, O'Toole PW. The gut microbiome as a modulator of healthy ageing. *Nat Rev Gastroenterol Hepatol*. 2022,19(9):565–584. <https://doi.org/10.1038/s41575-022-00605-x>

Girault Y, Merle H. L'enseignement de l'astronomie. *Aster, recherches en didactique des sciences expérimentales*. 2003, (36):1-192.

González-Reyes S, Marín L, González L, González LO, del Casar JM, Lamelas ML, González-Quintana JM, Vizoso FJ. Study of TLR3, TLR4 and TLR9 in breast carcinomas and their association with metastasis. *BMC cancer*. 2010, 10 :665. <https://doi.org/10.1186/1471-2407-10-665>

Gorse M, Bianchi C, Proudhon C. Épigénétique et cancer la méthylation dans tous ses états. *Med Sci (Paris)*. 2024, 40(12) :925–934. <https://doi.org/10.1051/medsci/2024180>

Goulati P, Singh C V. The Crucial Role of Molecular Biology in Cancer Therapy: A Comprehensive Review. *Cureus*. 2024, 16(1):e52246. <https://doi.org/10.7759/cureus.52246>

Gremmo-Féger G. Allaitement maternel Quoi de neuf. *Rev Med Périnatol.* 2016, 8(4):213–220. <https://doi.org/10.1007/s12611-016-0381-9>

Ghazi H, Gonzales FA, Jones PA. Methylation of CpG-island-containing genes in human sperm fetal and adult tissues. *Gene.* 1992, 114(2):203–210. [https://doi.org/10.1016/0378-1119\(92\)90575-a](https://doi.org/10.1016/0378-1119(92)90575-a)

[Global Cancer Observatory iarc.fr \(site web\)](http://GlobalCancerObservatory.iarc.fr)

Goldhirsch A, Winer EP, Coates AS, Gelber RD, Piccart-Gebhart M, Thürlimann B, Senn H J. Panel members Personalizing the treatment of women with early breast cancer highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013 *Annals of oncology official journal of the European Society for Medical Oncology.* 2013, 24(9):2206–2223. <https://doi.org/10.1093/annonc/mdt303>

H

Haan J.C, Bhaskaran R, Ellappalayam A, Bijl Y, Griffioen C.J, Lujinovic E, Audeh W.M, Penault-Llorca F, Mittempergher L, Glas A.M. MammaPrint and Blueprint comprehensively capture the cancer hallmarks in early-stage breast cancer patients. *Genes Chromosomes Cancer.* 2022, 61(3):148–160. <https://doi.org/10.1002/gcc.23014>

Hamedi SZ, Emami H, Khayamzadeh M, Farahmand M. Application of machine learning in breast cancer survival prediction using a multimethod approach. *Scientific Reports.* 2024, 14:30147. <https://doi.org/10.1038/s41598-024-81734-y>

Hanahan D, Weinberg RA. The hallmarks of cancer. *Cell.* 2000, 100(1):57–70. [https://doi.org/10.1016/s0092-8674\(00\)81683-9](https://doi.org/10.1016/s0092-8674(00)81683-9)

Hancock JT, Khoshgoftaar TM. CatBoost for big data: an interdisciplinary review. *J Big Data.* 2020, 7(1):94. <https://doi.org/10.1186/s40537-020-00369-8>

Hassiotou F, Geddes D. Anatomy of the human mammary gland: Current status of knowledge. *Clinical Anatomy.* 2013, 26(1):29–48. <https://doi.org/10.1002/ca.22165>

Herman JG, Baylin SB. Gene silencing in cancer in association with promoter hypermethylation. *New Engl J Med.* 2003, 349(21):2042–2054. <https://doi.org/10.1056/NEJMra023075>

Herzog SK, Fuqua SAW. ESR1 mutations and therapeutic resistance in metastatic breast cancer: progress and remaining challenges. *BritishJ Cancer.* 2022, 126(2):174–186. <https://doi.org/10.1038/s41416-021-01564-x>

Heisey R, Carroll JC. Identification et prise en charge des femmes ayant des antécédents familiaux de cancer du sein: Guide pratique à l'intention des médecins. *Canadian Family Physician*. 2016, 62(10):e572–e577. <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5063782/>

Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Perou CM. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biology*. 2007, 8(5):R76. <https://doi.org/10.1186/gb-2007-8-5-r76>

Hoffmann F, Franzen A, de Vos L, Wuest L, Kulcsár Z, Fietz S, Maas AP, Hollick S, Diop MY, Gabrielpillai J, Vogt T, Kuster P, Zarbl R, Dietrich J, Kristiansen G, Brossart P, Landsberg J, Strieth S, Dietrich D. CTLA4 DNA methylation is associated with CTLA-4 expression and predicts response to immunotherapy in head and neck squamous cell carcinoma. *Clinical Epigenetics*. 2023, 15(1):112. <https://doi.org/10.1186/s13148-023-01525-6>

Horimoto Y, Ishizuka Y, Ueki Y, Higuchi T, Arakawa A, Saito M. Comparison of tumors with HER2 overexpression versus HER2 amplification in HER2-positive breast cancer patients. *BMC Cancer*. 2022, 22(1):242. <https://doi.org/10.1186/s12885-022-09351-4>

Hsu JL, Hung MC. The role of HER2, EGFR, and other receptor tyrosine kinases in breast cancer. *Cancer Metastasis Reviews*. 2016, 35(4):575–588. <https://doi.org/10.1007/s10555-016-9649-6>

Huang J, Bai X, Xie X, Chen L, Lan X, Zhang Q, Song L, Hong P, Du C. A real-world study on oral vinorelbine for the treatment of metastatic breast cancer. *Oncologie*. 2022, 24(1):131–145. <https://doi.org/10.32604/oncologie.2022.019881>

Houdebine LM. Contrôle hormonal du développement et de l'activité de la glande mammaire. *Repr Nutr Develop*. 1986, 26(2B):523–541. <https://doi.org/10.1051/rnd:19860401>

Hu X, Wang L, Zhang J, Zhang Q, Ouyang Q, Wang X, Li W, Xie W, Tong Z, Xu F, Liu W, Wang S, Sun T, Chen Z, Wu J, Wang Y, Wang H, Yan M, Wang X, Wang J, Cao F, Du Y, Zhang Y, Chen L, Lu P, Sun S, Zhang R, Zang A, Nie X, Lei Y. ACE-Breast-02: A pivotal phase II/III trial of ARX788, a novel anti-HER2 antibody-drug conjugate (ADC), versus lapatinib plus capecitabine for HER2+ advanced breast cancer. *J Clin Oncol*. 2024, 42(16_suppl):1020. https://doi.org/10.1200/JCO.2024.42.16_suppl.1020

I

Iqbal N, et Iqbal N. Human epidermal growth factor receptor 2 (HER2) in cancers: Overexpression and therapeutic implications. *Mol Biol Intern*, 2014, 852748.

<https://doi.org/10.1155/2014/852748>

J

Jatoi I, Kaufmann M, editors. Management of breast diseases. 1st ed. Berlin: Springer; 2010.

<https://doi.org/10.1007/978-3-540-69743-5>

Javed F, Ahmed HB, Crespi R, Romanos GE. Role of primary stability for successful osseointegration of dental implants: factors of influence and evaluation. *Interv Med Appl Sci*. 2013, 5(4):162–167. <https://doi.org/10.1556/IMAS.5.2013.4.3>

Jin X, Mu P. Targeting breast cancer metastasis. *Breast Cancer (Auckl)*. 2015, 9(Suppl 1):23–34.

<https://doi.org/10.4137/BCBCR.S25460>

Jallah JK, Dweh TJ, Anjankar A, Palma OA. A review of the advancements in targeted therapies for breast cancer. *Cureus*. 2023, 15(10):e47847. <https://doi.org/10.7759/cureus.47847>

Jeong H, Park J, Kim HI, Lee M, Ko YJ, Lee S, Jun Y, Lee C. Mechanistic insight into the nucleus-vacuole junction based on the Vac8p-Nvj1p crystal structure. *Proc Natl Acad Sci USA*. 2017, 114(23):E4539–E4548. <https://doi.org/10.1073/pnas.1701030114>

Jin T, Park KS, Nam SE, Lim SH, Kim JH, Noh WC, Yoo YB, Park WS, Yun IJ. CTLA4 expression profiles and their association with clinical outcomes of breast cancer: a systemic review. *Ann Surg Treat Res*. 2024, 106(5):263–273. <https://doi.org/10.4174/ast.2024.106.5.263>

Joyon N, Penault-Llorca F, Lacroix-Triki M. Classification et signatures moléculaires des cancers du sein en 2017. *Oncologie*. 2017, 19:64–70. <https://doi.org/10.1007/s10269-017-2700-6>

K

Kakushadze Z, Yu W. K-means and cluster models for cancer signatures. *Biomol Detect Quantif*. 2017, 13:7–31. <https://doi.org/10.1016/j.bdq.2017.07.001>

Karapetyan L, Luke JJ, Davar D. Toll-like receptor 9 agonists in cancer. *Onco Targets Ther*. 2020, 13:10039–10060. <https://doi.org/10.2147/OTT.S247050>

Kawai T, Akira S. The role of pattern-recognition receptors in innate immunity: update on Toll-like receptors. *Nat Immunol.* 2010, 11(5):373–384. <https://doi.org/10.1038/ni.1863>

Khan M, Shah Z, Rooman M, Khan W, Vranceanu N, Alshehri M. Machine learning approach in breast cancer analysis. *Sci Rep.* 2023,13:21434. <https://doi.org/10.1038/s41598-023-48676-3>

Khan MM, Yalamarty SSK, Rajmalani BA, Filipczak N, Torchilin VP. Recent strategies to overcome breast cancer resistance. *Crit Rev Oncol Hematol.* 2024, 197:104351. <https://doi.org/10.1016/j.critrevonc.2024.104351>

Kim JY, Oh JM, Lee SK, Yu J, Lee JE, Kim SW, Nam SJ, Park YH, Ahn JS, Kim K, Im YH. Improved prediction of survival outcomes using residual cancer burden in combination with Ki-67 in breast cancer patients underwent neoadjuvant chemotherapy. *Front Oncol.* 2022, 12:903372. <https://doi.org/10.3389/fonc.2022.903372>

Knutson TP, Lange CA. Tracking progesterone receptor-mediated actions in breast cancer. *Pharmacol Ther.* 2014, 142(1):114–125. <https://doi.org/10.1016/j.pharmthera.2013.11.010>

Kulis M, Esteller M. DNA methylation and cancer. *Adv Genet.* 2010, 70:27–56. <https://doi.org/10.1016/B978-0-12-380866-0.60002-2>

L

Lauschke VM, Shafagh RZ, Hendriks DFG, Ingelman-Sundberg M. 3D primary hepatocyte culture systems for analyses of liver diseases, drug metabolism, and toxicity: emerging culture paradigms and applications. *Biotechnol J.* 2019, 14(7):e1800347. <https://doi.org/10.1002/biot.201800347>

Lei S, Zheng R, Zhang S, Wang S, Chen R, Sun K, Zeng H, Zhou J, Wei W. Global patterns of breast cancer incidence and mortality: A population-based cancer registry data analysis from 2000 to 2020. *Cancer Commun (Lond).* 2021, 41(11):1183–1194. <https://doi.org/10.1002/cac2.12207>

Le Romancer M, Poulard C, Cohen P, Sentis S, Renoir JM, Corbo L. Cracking the estrogen receptor's posttranslational code in breast tumors. *Endocr Rev.* 2011, 32(5):597–622. <https://doi.org/10.1210/er.2010-0016>

Li X, Yang J, Peng L, Sahin AA, Huo L, Ward KC, O'Regan R, Torres MA, Meisel JL. Triple-negative breast cancer has worse overall survival and cause-specific survival than non-triple-negative breast cancer. *Breast Cancer Res Treat.* 2017, 161(2):279–287.

<https://doi.org/10.1007/s10549-016-4059-6>

Li N, Song K, Chen H, Dai M. Advance and challenge of DNA methylation as cancer biomarkers for risk stratification, screening and early detection. *J Natl Cancer Cent.* 2025, 5(2):108–112.

<https://doi.org/10.1016/j.jncc.2024.12.007>

Li C, Fan Z, Lin X, Cao M, Song F, Song F. Parity and risk of developing breast cancer according to tumor subtype: A systematic review and meta-analysis. *Cancer Epidemiol.* 2021, 75:102050.

<https://doi.org/10.1016/j.canep.2021.102050>

Liang G, Weisenberger D. DNA methylation aberrancies as a guide for surveillance and treatment of human cancers. *Epigenetics.* 2017, 12(6):416–432.

<https://doi.org/10.1080/15592294.2017.1311434>

Ligorio F, Zambelli L, Fucà G, Lobefaro R, Santamaria M, Zattarin E, De Braud F, Vernieri C. Prognostic impact of body mass index (BMI) in HER2+ breast cancer treated with anti-HER2 therapies: from preclinical rationale to clinical implications. *Ther Adv Med Oncol.* 2022, 14:1–18.

<https://doi.org/10.1177/17588359221079123>

Liu ZH, Wang K, Lin DY, Xu J, Chen J, Long XY, Ge Y, Luo XL, Zhang KP, Liu YH, Xu FP. Impact of the updated 2018 ASCO/CAP guidelines on HER2 FISH testing in invasive breast cancer: a retrospective study of HER2 FISH results of 2233 cases. *Breast Cancer Res Treat.* 2019;175(1):51–57. <https://doi.org/10.1007/s10549-019-05148-5>

Lounis M, Belkessa S, Abdelhadi S, Bencherit D, Samara KA, Barqawi HJ. Breast cancer knowledge and practices amongst women in Algeria. *J Cancer Res Clin Oncol.* 2023, 149(11):8843–8852. <https://doi.org/10.1007/s00432-023-04786-z>

Lyko F. The DNA methyltransferase family: a versatile toolkit for epigenetic regulation. *Nat Rev Genet.* 2018, 19(2):81–92. <https://doi.org/10.1038/nrg.2017.80>

M

Ma F, Yan M, Li W, Ouyang Q, Tong Z, Teng Y, Wang Y, Wang S, Geng C, Luo T, Zhong J, Zhang Q, Liu Q, Zeng X, Sun T, Mo Q, Liu H, Cheng Y, Cheng J, Wang X, Nie J, Yang J, Wu X, Wang X, Li H, Ye C, Dong F, Wu S, Zhu X, Xu B; PHILA Investigators. Pyrotinib versus placebo in combination with trastuzumab and docetaxel as first line treatment in patients with

HER2 positive metastatic breast cancer (PHILA): randomised, double blind, multicentre, phase 3 trial. *BMJ*. 2023, 383:e076065. <https://doi.org/10.1136/bmj-2023-076065>

Ma X, Wu J, Wang B, Liu C, Liu L, Sun C. Epigenetic modifications: Critical participants of the PDL1 regulatory mechanism in solid tumors (Review). *Int J Oncol*. 2022, 61(5):134. <https://doi.org/10.3892/ijo.2022.5424>

Mailliez A. Ce que le radiologue doit savoir des nouvelles classifications moléculaires des cancers du sein. *Rev Sage Femme*. 2014, 13(6):313–318. <https://doi.org/10.1016/j.femme.2014.10.004>

Manouchehri E, Taghipour A, Ghavami V, Shandiz FH, Ebadi A, Roudsari RL. Menstrual and reproductive factors and risk of breast cancer in Iranian female population: A systematic review and meta-analysis. *Int J Prev Med*. 2022, 13:26. https://doi.org/10.4103/ijpvm.IJPVM_646_20

Martaindale SR. Breast MR Imaging: Atlas of Anatomy, Physiology, Pathophysiology, and Breast Imaging Reporting and Data Systems Lexicon. *Magn Reson Imaging Clin N Am*. 2018, 26(2):179–190. <https://doi.org/10.1016/j.mric.2017.12.001>

Menon A, O'Mahony M. Women's body image following mastectomy: Snapshots of their daily lives. *Appl Nurs Res*. 2019;47:36–41. <https://doi.org/10.1016/j.apnr.2019.03.002>

Mego M, Mani SA, Cristofanilli M. Molecular mechanisms of metastasis in breast cancer—clinical applications. *Nat Rev Clin Oncol*. 2010, 7(12):693–701. <https://doi.org/10.1038/nrclinonc.2010.171>

Mery B, Toussaint P, Heudel PE, Dufresne A, Carbonnaux M, Vanacker H, Bachelot T, Trédan O. Nouvelles stratégies thérapeutiques dans les cancers du sein HER2-surexprimé: New therapeutic strategies in HER2-positive breast cancer. *Bull Cancer*. 2021, 108(11S):11S8–11S18. [https://doi.org/10.1016/S0007-4551\(21\)00632-9](https://doi.org/10.1016/S0007-4551(21)00632-9)

Moasser MM. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene*. 2007;26:6469–6487. <https://doi.org/10.1038/sj.onc.1210477>

Moaz I, Fouad FA, Elmasry H, Tarek G, Elzoheiry A, Elgamal M, Ibrahim R, Hisham Y, Safwat G, Kamel MM, El-Batal HM, Fouda M. Associations Between Serum Soluble Toll-like Receptors 4 and 9 and Breast Cancer in Egyptian Patients. *Cancer Control*. 2023, 30:10732748231204755. <https://doi.org/10.1177/10732748231204755>

Morsberger L, Pallavajjala A, Long P, Hardy M, Park R, Parish R, Nozari A, Zou YS. HER2 amplification by next-generation sequencing to identify HER2-positive invasive breast cancer with negative HER2 immunohistochemistry. *Cancer Cell Int*. 2022;22(1):350.

<https://doi.org/10.1186/s12935-022-02761-1>

Mukhopadhyay K, Liu Z, Bandyopadhyay A, Kirma NB, Tekmal RR, Wang S, Sun LZ. Aromatase expression increases the survival and malignancy of estrogen receptor positive breast cancer cells. *PLoS ONE*. 2015, 10(4):e0121136. <https://doi.org/10.1371/journal.pone.0121136>

N

Najjar MK, Manore SG, Regua AT, Lo HW. Antibody-Drug Conjugates for the Treatment of HER2-Positive Breast Cancer. *Genes*. 2022, 13(11):2065. <https://doi.org/10.3390/genes13112065>

Navarrete-Bernal MGC., Cervantes-Badillo MG, Martínez-Herrera JF, Lara-Torres CO, Gerson-Cwilich R, Zentella-Dehesa A, Ibarra-Sánchez MJ, Esparza-López J, Montesinos JJ, Cortés-Morales VA, Osorio-Pérez D, Villegas-Osorno DA, Reyes-Sánchez E, Salazar-Sojo P, Tallabs-Utrilla LF, Romero-Córdoba S, Rocha-Zavaleta L. Biological landscape of triple negative breast cancers expressing CTLA-4. *Front Oncol*. 2020, 10:1206. <https://doi.org/10.3389/fonc.2020.01206>

Nagpal M, Singh S, Singh P, Chauhan P, Zaidi MA. Tumor markers: A diagnostic tool. *Nat J Maxillofac Surg*. 2016, 7(1):17–20. <https://doi.org/10.4103/0975-5950.196135>

Narayan A, Ji W, Zhang XY, Marrogi A, Graff JR, Baylin SB, Ehrlich M. Hypomethylation of pericentromeric DNA in breast adenocarcinomas. *Int J Cancer*. 1998, 77(6):833–838. [https://doi.org/10.1002/\(sici\)1097-0215\(19980911\)77:6](https://doi.org/10.1002/(sici)1097-0215(19980911)77:6)

National Cancer Institute. Breast cancer treatment (PDQ®)—Patient version. National Institutes of Health. Disponible sur : <https://www.cancer.gov/types/breast/patient/breast-treatment-pdq>

National Cancer Institute. CTLA-4. NCI Dictionary of Cancer Terms. Bethesda (MD): National Cancer Institute. Disponible sur : <https://www.cancer.gov/publications/dictionaries/cancer-terms/def/ctla-4>

Ngan TT, Nguyen NTQ, Van Minh H, Donnelly M, O'Neill C. Effectiveness of clinical breast examination as a 'stand-alone' screening modality: an overview of systematic reviews. *BMC Cancer*. 2020, 20(1):1070. <https://doi.org/10.1186/s12885-020-07521-w>

Nielsen TO., Leung SCY, Rimm DL, Dodson A, Acs B, Badve S, Denkert C, Ellis MJ, Fineberg S, Flowers M, Kreipe HH, Laenkholm AV, Pan H, Penault-Llorca FM, Polley MY, Salgado R, Smith, IE, Sugie T, Bartlett JMS, McShane LM, Hayes DF. Assessment of Ki67 in Breast

Cancer: Updated Recommendations From the International Ki67 in Breast Cancer Working Group. *J Natl Cancer Inst.* 2021, 113(7):808–819. <https://doi.org/10.1093/jnci/djaa201>

O

Oblak A, Jerala R. Toll-like receptor 4 activation in cancer progression and therapy. *J Immunol Res*, 2011, Article 609579. <https://doi.org/10.1155/2011/609579>

O'Shaughnessy J, Gradishar W, O'Regan R, Gadi V. Risk of Recurrence in Patients With HER2+ Early-Stage Breast Cancer: Literature Analysis of Patient and Disease Characteristics. *Clin breast cancer*, 2023, 23(4):350–362. <https://doi.org/10.1016/j.clbc.2023.03.007>

Obeagu EI, Obeagu GU. Breast cancer: A review of risk factors and diagnosis. *Medicine*, 2024, 103(3):e36905. <https://doi.org/10.1097/MD.00000000000036905>

Oliver CM, Warnakulasuriya S, McGuckin D, Singleton G, Martin P, Santos C, Bedford J, Wagstaff D, Sahni A, Gilhooly D, Wilson J, Edwards K, Baumber R, Vindrola-Padros C, Dorey J, Leeman I, Boyd-Carson H, Vohra R, Singh P, Bedford M, PQIP collaborative. Delivery of drinking, eating and mobilising (DrEaMing) and its association with length of hospital stay after major noncardiac surgery: observational cohort study. *British J anaesthesia*, 2022, 129(1) :114–126. <https://doi.org/10.1016/j.bja.2022.03.021>

P

Pal, S., Morgan, X., Dar, H. Y., Gacasan, C. A., Patil, S., Stoica, A., Hu, Y.-J., Weitzmann, M. N., Jones, R. M., & Pacifici, R. Gender-affirming hormone therapy preserves skeletal maturation in young mice via the gut microbiome. *J Clin Investig.* 2024, 134(6):e175410. <https://doi.org/10.1172/JCI175410>

Palanca-Ballester C, Rodriguez-Casanova A, Torres S, Calabuig-Fariñas S, Exposito F, Serrano, D, Redin E, Valencia K, Jantus-Lewintre E, Diaz-Lagares A, Montuenga L, Sandoval J, Calvo A. Cancer Epigenetic Biomarkers in Liquid Biopsy for High Incidence Malignancies. *Cancers*. 2021, 13(12):3016. <https://doi.org/10.3390/cancers13123016>

Pandey V, Mishra A, Singh D, Kumar P. A review on breast cancer: risk factors, diagnosis and management. *Zenodo*. 2023. <https://doi.org/10.5281/zenodo.7918734>

Park DJ, Park MW, Lee H, Kim YJ, Kim Y, Park YH. Development of machine learning for Diagnostic Disease Prediction Based on Laboratory Tests. *Research Square*. 2021. <https://doi.org/10.21203/rs.3.pex-1429/v1>

Parvez A, Choudhary F, Mudgal P, Khan R, Qureshi K, Farooqi H, Aspatwar A. PD-1 and PD-L1: Architects of immune symphony and immunotherapy breakthroughs in cancer treatment. *Frontiers in Immunology*. 2023,14:1296341. <https://doi.org/10.3389/fimmu.2023.1296341>

Peixoto P, Renaude E, Boyer-Guittaut M, Hervouet E. Epigenetics, a key player of immunotherapy resistance. *Cancer Drug Resistance*, 2018, 1:219–229. <https://doi.org/10.20517/cdr.2018.17>

Penault-Llorca F, Radosevic-Robin N. Ki67 assessment in breast cancer: an update. *Pathology*. 2017, 49(2):166–171. <https://doi.org/10.1016/j.pathol.2016.11.006>

Péqueux C, Raymond-Letron I, Blacher S, Boudou F, Adlanmerini M, Fouque MJ, Rochaix P, Noël A, Foidart JM, Krust A, Chambon P, Brouchet L, Arnal JF, Lenfant F. Stromal estrogen receptor- α promotes tumor growth by normalizing an increased angiogenesis. *Cancer Research*. 2012, 72(12):3010–3019. <https://doi.org/10.1158/0008-5472.CAN-11-3768>

Phung MT., Tin Tin S, Elwood JM. Prognostic models for breast cancer: a systematic review. *BMC cancer*. 2019, 19(1):230. <https://doi.org/10.1186/s12885-019-5442-6>

Prat A, Perou CM. Mammary development meets cancer genomics. *Nat Med*, 2009, 15(8):842–844. <https://doi.org/10.1038/nm0809-842>

Q

Qu G, Dubeau L, Narayan A, Yu MC, Ehrlich M. Satellite DNA hypomethylation vs. overall genomic hypomethylation in ovarian epithelial tumors of different malignant potential. *Mutation research*. 1999, 423(1-2):91–101. [https://doi.org/10.1016/s0027-5107\(98\)00229-2](https://doi.org/10.1016/s0027-5107(98)00229-2)

R

Ralsler M, Varma SJ, Notebaart RA. The evolution of the metabolic network over long timelines. *Curr Opin Syst Biol*. 2021, 28:100402. <https://doi.org/10.1016/j.coisb.2021.100402>

Rehman HT, Jawaid H, Tahir A, Imtiaz M, Zulfiqar T, Aziz T. Breast cancer knowledge among health professionals: A pre–post-knowledge-based intervention study. *J Family Med Prim Care*. 2022, 11(9):5649-5655. <https://doi.org/10.4103/jfmprc.jfmprc.1226.22>

Rochefort H, Rouëssé J, Ancelle-Park MRM, Hill C, Sancho-Garnier H, Stoppa-Lyonnet D, Tardivon A, Birnbaum D, Bouchard Ph, Estève J, Jeanteur Ph, Le Bouc Y, Léridon H, Maudelonde T, Schaison G, Tubiana M. Cancers du sein, incidence et prévention. *Bull Acad Natl Med.* 2008, 192(1):161–179. [https://doi.org/10.1016/S0001-4079\(19\)32856-0](https://doi.org/10.1016/S0001-4079(19)32856-0)

Rountree MR, Bachman KE, Herman JG, Baylin SB. DNA methylation, chromatin inheritance, and cancer. *Oncogene.* 2001, 20(24):3156–3165. <https://doi.org/10.1038/sj.onc.1204339>

Raghav KPS, Moasser MM. Molecular pathways and mechanisms of HER2 in cancer therapy. *Clin Cancer Res.* 2023, 29(13):2351–2361. <https://doi.org/10.1158/1078-0432.CCR-22-0283>

Razavi P, Chang MT, Xu G, Bandlamudi C, Ross DS, Vasan N, Cai Y, Bielski CM, Donoghue MTA, Jonsson P, Penson A, Shen R, Pareja F, Kundra R, Middha S, Cheng ML, Zehir A, Kandath C, Patel R, Huberman K, Smyth LM, Jhaveri K, Modi S, Traina TA, Dang C, Zhang W, Weigelt B, Li BT, Ladanyi M, Hyman DM, Schultz N, Robson ME, Hudis C, Brogi E, Viale A, Norton L, Dickler MN, Berger MF, Iacobuzio-Donahue CA, Chandarlapaty S, Scaltriti M, Reis-Filho JS, Solit DB, Taylor BS, Baselga J. The Genomic Landscape of Endocrine-Resistant Advanced Breast Cancers. *Cancer Cell.* 2018, 34(3):427–438.e6. <https://doi.org/10.1016/j.ccell.2018.08.008>

Renoir JM, Marsaud V, Lazennec G. Estrogen receptor signaling as a target for novel breast cancer therapeutics. *Biochem Pharmacol.* 2013, 85(4):449–465. <https://doi.org/10.1016/j.bcp.2012.10.018>

Robertson JF, Jaeger W, Syzmendera JJ, Selby C, Coleman R, Howell A, Winstanley J, Jonssen PE, Bombardieri E, Sainsbury JR, Gronberg H, Kumpulainen E, Blamey RW. The objective measurement of remission and progression in metastatic breast cancer by use of serum tumour markers. European Group for Serum Tumour Markers in Breast Cancer. *Eur J Cancer.* 1999, 35(1):47–53. [https://doi.org/10.1016/s0959-8049\(98\)00297-4](https://doi.org/10.1016/s0959-8049(98)00297-4)

Rolfo C, Giovannetti E, Martinez P, McCue S, Naing A. Applications and clinical trial landscape using Toll-like receptor agonists to reduce the toll of cancer. *NPJ Precis Oncol.* 2023, 7:26. <https://doi.org/10.1038/s41698-023-00364-1>

Romero Q, Bendahl PO, Fernö M, Grabau D, Borgquist S. A novel model for Ki67 assessment in breast cancer. *Diagn Pathol.* 2014, 9:118. <https://doi.org/10.1186/1746-1596-9-118>

Rosso R, D'Alonzo M, Bounous VE, Actis S, Cipullo I, Salerno E, Biglia N. Adherence to Adjuvant Endocrine Therapy in Breast Cancer Patients. *Curr Oncol.* 2023, 30(2):1461–1472. <https://doi.org/10.3390/curroncol30020112>

Rothenberger NJ, Somasundaram A, Stabile LP. The Role of the Estrogen Pathway in the Tumor Microenvironment. *Int J Mol Sci.* 2018, 19(2):611. <https://doi.org/10.3390/ijms19020611>

Rotte A. Combination of CTLA-4 and PD-1 blockers for treatment of cancer. *J Exp Clin Cancer Res.* 2019, 38:255. <https://doi.org/10.1186/s13046-019-1259-z>

Russo J, Russo IH. The role of estrogen in the initiation of breast cancer. *J Steroid Biochem Mol Biol.* 2006, 102(1-5):89–96. <https://doi.org/10.1016/j.jsbmb.2006.09.004>

S

Saber M, Boulmaiz T, Guermoui M, Abdrabo KI, Kantoush SA, Sumi T, Boutaghane H, Hori T, Binh DV, Nguyen BQ, Bui TT, Vo ND, Habib EF, Mabrouk E. Enhancing flood risk assessment through integration of ensemble learning approaches and physical-based hydrological modeling. *Geomatics Nat Haz Risk.* 2023, 14:2203798. <https://doi.org/10.1080/19475705.2023.2203798>

Saghafinia S, Mina M, Riggi N, Hanahan D, Ciriello G. Pan-Cancer Landscape of Aberrant DNA Methylation across Human Tumors. *Cell Rep.* 2018, 25(4):1066–1080.e8. <https://doi.org/10.1016/j.celrep.2018.09.082>

Sahoo A, Ghose DK. Imputation of missing precipitation data using KNN, SOM, RF, and FNN. *Soft Comput.* 2022, 26:5919–5936. <https://doi.org/10.1007/s00500-022-07029-4>

Senkus E, Kyriakides S, Ohno S, Penault-Llorca F, Poortmans P, Rutgers E, Zackrisson S, Cardoso F, ESMO Guidelines Committee. Primary breast cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Ann Oncol.* 2015, 26(Suppl)5:v8–v30. <https://doi.org/10.1093/annonc/mdv298>

Seo DO, O'Donnell D, Jain N, Ulrich JD, Herz J, Li Y, Lemieux M, Cheng J, et al. ApoE isoform- and microbiota-dependent progression of neurodegeneration in a mouse model of tauopathy. *Science.* 2023, 379(6628):eadd1236. <https://doi.org/10.1126/science.add1236>

Shi S, Xu C, Fang X, Zhang Y, Li H, Wen W, Yang G. Expression profile of Toll-like receptors in human breast cancer. *Mol Med Rep.* 2019, 21(2):786–794.

<https://doi.org/10.3892/mmr.2019.10853>

Sidaway P. Buparlisib is an effective second-line treatment. *Nat Rev Clin Oncol.* 2017, 14(4):198. <https://doi.org/10.1038/nrclinonc.2017.22>

Singh N, Baby D, Rajguru JP, Patil PB, Thakkannavar SS, Pujari VB. Inflammation and cancer. *Ann Afr Med.* 2019, 18(3):121–126. https://doi.org/10.4103/aam.aam_56_18

Slamon DJ, Clark GM, Wong SG, Levin WJ, Ullrich A, McGuire WL. Human breast. *Science.* 1987, 235:177–182. <https://doi.org/10.1126/science.3798106>

Slamon DJ, Leyland-Jones B, Shak S, et al. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med.* 2001, 344:783–792. <https://doi.org/10.1056/NEJM20010315344110>

Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B. Show more. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *J Natl Cancer Inst.* 2006, 98(4):262–272. <https://doi.org/10.1093/jnci/djj052>

Stankov A, Bargallo-Rocha JE, Ñamendys-Silva A, Ramirez MT, Stankova-Ninova K, Meneses-Garcia A. Prognostic factors and recurrence in breast cancer: Experience at the National Cancer Institute of Mexico. *ISRN Oncol.* 2012, 825258:1-7. <https://doi.org/10.5402/2012/825258>

Stierschneider A, Wiesner C. Shedding light on the molecular and regulatory mechanisms of TLR4 signaling in endothelial cells under physiological and inflamed conditions. *Front Immunol.* 2023, 14:1264889. <https://doi.org/10.3389/fimmu.2023.1264889>

Su D, Xiong Y, Wang S, Wei H, Ke J, Li H, Wang T, Zuo Y, Yang L. Structural deep clustering network for stratification of breast cancer patients through integration of somatic mutation profiles. *Comput Methods Programs Biomed.* 2023, 242:107808. <https://doi.org/10.1016/j.cmpb.2023.107808>

Sung H, Ferlay J, Siegel RL, Laversanne M, Soerjomataram I, Jemal A, Bray F. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. *CA Cancer J Clin.* 2021, 71(3):209–249. <https://doi.org/10.3322/caac.21660>

Surakasula A, Nagarjunapu GC, Raghavaiah KV. A comparative study of clinicopathological features of breast cancer in younger and older women. *J Clin Diagn Res.* 2014, 8(1):96–100.

<https://doi.org/10.7860/JCDR/2014/8579.4325>

Susini T, Biglia N, Bounous VE. Prognostic factors research in breast cancer patients: New paths. *Cancers (Basel)*. 2022, 14(4):971. <https://doi.org/10.3390/cancers14040971>

Swain SM, Baselga J, Kim SB, et al. Pertuzumab, trastuzumab, and docetaxel in HER2-positive metastatic breast cancer. *N Engl J Med*. 2015, 372(8):724–734. <https://doi.org/10.1056/NEJMoa1413513>

T

Tandon AK, Clark GM, Chamness GC, Ullrich A, McGuire WL. HER-2/neu oncogene protein and prognosis in breast cancer. *J Clin Oncol*. 1989, 7(8):1120-1128. <https://doi.org/10.1200/JCO.1989.7.8.1120>

Thompson D, Easton D. The genetic epidemiology of breast cancer genes. *J Mammary Gland Biol Neoplasia*. 2004, 9(3):221–236. <https://doi.org/10.1023/B:JOMG.0000048770.90334.3b>

Torrisani, Jerome et Lopez, F. DNA methylation and epigenetic regulation of cancers. *Hepato-Gastro*. 2003, 10(6):455-467.

Tortora GJ, Derrickson BH, Burkett B, Dye D, Cooke J, Diversi T. *Principles of Anatomy and Physiology*. 1st Asia-Pacific ed. John Wiley & Sons; 2016. <http://www.wileydirect.com.au/buy/principles-of-anatomy-physiology-1st-asia-pacific-edition>

Totura AL, Whitmore A, Agnihothram S, Schäfer A, Katze MG, Heise MT, Baric RS. Toll-like receptor 3 signaling via TRIF contributes to a protective innate immune response to severe acute respiratory syndrome coronavirus infection. *mBio*. 2015, 6(3):e00638-15. <https://doi.org/10.1128/mBio.00638-15>

Travis RC, Key TJ. Oestrogen exposure and breast cancer risk. *Breast Cancer Res*. 2003, 5(5):239–247. <https://doi.org/10.1186/bcr628>

Tuval A, Strandgren C, Heldin A, Palomar-Siles M, Wiman KG. Pharmacological reactivation of p53 in the era of precision anticancer medicine. *Nat Rev Clin Oncol*. 2024, 21(2):106–120. <https://doi.org/10.1038/s41571-023-00842-2>

V

Van Seuning I. Qu'est-ce que l'épigénétique ? Stratégies d'étude et application au cancer. *Ann Pathol.* 2009, 29(Spec No 1):S28–S30. <https://doi.org/10.1016/j.annpat.2009.07.009>

Varaganti P, Buddolla V, Anantha Lakshmi B, Kim Y-J. Recent advances in using folate receptor 1 (FOLR1) for cancer diagnosis and treatment. *Life Sci.* 2023, 326:121802. <https://doi.org/10.1016/j.lfs.2023.121802>

Vargo-Gogola T, Rosen JM. Modelling breast cancer: one size does not fit all. *Nat Rev Cancer.* 2007, 7(9):659–672. <https://doi.org/10.1038/nrc2193>

Vazquez-Sequeiros E, De Miquel DB. Indications de la ponction-biopsie à l'aiguille fine sous échographie (EUS-FNA) dans la pathologie ganglionnaire. *Acta Endosc.* 2005, 35:25–33. <https://doi.org/10.1007/BF03002644>

W

Wachsman JT. DNA methylation and the association between genetic and epigenetic changes: relation to carcinogenesis. *Mutat Res.* 1997, 375(1):1–8. [https://doi.org/10.1016/s0027-5107\(97\)00003-1](https://doi.org/10.1016/s0027-5107(97)00003-1)

Wang H, Mao X. Evaluation of the efficacy of neoadjuvant chemotherapy for breast cancer. *Drug Des Devel Ther.* 2020, 14:2423–2433. <https://doi.org/10.2147/DDDT.S253961>

Wang J, Singh P, Yin K, Schadt EE, Wang W, Zhang W, Zhang J. Disease spectrum of breast cancer susceptibility genes. *Front Oncol.* 2021, 11:663419. <https://doi.org/10.3389/fonc.2021.663419>

Lingyue W, , Hu P, , Hongwei Z, Ying L, Cao X, Olaf H, Tie L, Geping L, Anming B, Chen X. Integrative modeling of heterogeneous soil salinity using sparse ground samples and remote sensing images. *Geoderma.* 2023, 430. <https://doi.org/10.1016/j.geoderma.2022.116321>

Wang XQ, Danenberg E, Huang CS, Lim CJ, Choo KS, Lim JC, Tan TZ, Ong CW, Ngeow J, Lim TK, Lim B, Tai WM, Iqbal J, Rozen SG, Ginhoux F, Tan DSP. Spatial predictors of immunotherapy response in triple-negative breast cancer. *Nature.* 2023, 621(7980):868-76. <https://doi.org/10.1038/s41586-023-06498-3>

Warsy A, Almukaynizi F, Aldaihan S, Alam S, Dagastani M. Genetic polymorphisms in aromatase (CYP19) gene and cancer. In: IntechOpen; 2017. <https://doi.org/10.5772/intechopen.69208>

Wei SC, Duffy CR, Allison JP. Fundamental mechanisms of immune checkpoint blockade therapy. *Cancer Discov.* 2018, 8(9):1069-86. <https://doi.org/10.1158/2159-8290.CD-18-0367>

Wetstein SC, de Jong VMT, Stathonikos N, van Diest PJ. Deep learning-based breast cancer grading and survival analysis on whole-slide histopathology images. *Sci Rep.* 2022,12:15102. <https://doi.org/10.1038/s41598-022-19112-9>

Widschwendter M, Jones PA. DNA methylation and breast carcinogenesis. *Oncogene.* 2002, 21(35):5462-82. <https://doi.org/10.1038/sj.onc.1205606>

Wilson AS, Power BE, Molloy PL. DNA hypomethylation and human diseases. *Biochim Biophys Acta.* 2007, 1775(1):138-62. <https://doi.org/10.1016/j.bbcan.2006.08.007>

Wojtukiewicz MZ, Rek MM, Karpowicz K, Gluszuk K, Hrynkiewicz R, Rutkowski T, Politynska B, Tokajuk P, Kozłowski R, Wojtukiewicz AM. Inhibitors of immune checkpoints—PD-1, PD-L1, CTLA-4: new opportunities for cancer patients and a new challenge for internists and general practitioners. *Cancer Metastasis Rev.* 2021, 40(3):949-82. <https://doi.org/10.1007/s10555-021-09976-0>

Y

Yan, M., Ouyang, Q., Sun, T., Niu, L., Yang, J., Li, L., Song, Y., Hao, C., Chen, Z., Liu, Z., Lv, H., Zhang, M., Liu, L., Yang, X., Xiao, H., Gao, Z., Li, X., Dong, F., Zhang, L., Dong, D., ... Xia, F. Pyrotinib plus capecitabine for patients with HER2-positive metastatic breast cancer and brain metastases (PERMEATE trial): overall survival results from a multicenter, single-arm, two-cohort, phase 2 trial. *EClinicalMedicine*, 2024, 76, 102837. <https://doi.org/10.1016/j.eclinm.2024.102837>

Yamamoto, M., Sato, S., Hemmi, H., Hoshino, K., Kaisho, T., Sanjo, H., Takeuchi, O., Sugiyama, M., Okabe, M., Takeda, K., & Akira, S. Role of adaptor TRIF in the MyD88-independent toll-like receptor signaling pathway. *Science.* 2003;301(5633):640–643. <https://doi.org/10.1126/science.1087262>

Z

Zhang D. Artificial neural network. In: *Fundamentals of image data mining.* Springer; 2019. p. 193-213. https://doi.org/10.1007/978-3-030-17989-2_9

Zhang H, Katib I, Hasan H. Research on the psychological distribution delay of artificial neural network based on the analysis of differential equation by inequality expansion and contraction method. *Appl Math Nonlinear Sci.* 2021;7. Article 00001024782021200029. <https://doi.org/10.2478/amns.2021.2.00029>

Zhang J, Simpson CM, Berner J, Chong HB, Fang J, Ordulu Z, Weiss-Sadan T, Possemato AP, Harry S, Takahashi M, Yang TY, Richter M, Patel H, Smith AE, Carlin AD, Hubertus de Groot AF, Wolf K, Shi L, Wei TY, Dürr BR, Chen NJ, Vornbäumen T, Wichmann NO, Mahamdeh MS, Pooladanda V, Matoba Y, Kumar S, Kim E, Bouberhan S, Oliva E, Rueda BR, Soberman RJ, Bardeesy N, Liao BB, Lawrence M, Stokes MP, Beausoleil SA, Bar-Peled L. Systematic identification of anticancer drug targets reveals a nucleus-to-mitochondria ROS-sensing pathway. *Cell.* 2023;186(11):2361-79.e25. <https://doi.org/10.1016/j.cell.2023.04.026>

Zhang Z, Zhao Y, Canes A, Steinberg D, Lyashevskaya O; AME Big-Data Clinical Trial Collaborative Group. Predictive analytics with gradient boosting in clinical medicine. *Ann Transl Med.* 2019;7:1-7. <https://doi.org/10.21037/atm.2019.03.29>



Annexes

Type histologique:

Date récidive (local, métastase) :

Nombre de ganglions atteints :

ER : Positif - Négatif

PR : Positif - Négatif

HER2 : Positif - Négatif

Ki67 : Positif - Négatif

Antécédents personnels de :

- Cancer :

- Pathologies non cancéreuses (Diabète/HTA...Autre) :

Antécédents familiaux de cancer :

(Lien de parenté, nombre de cas dans la famille, les organes affectés par la tumeur) :

Prise en charge thérapeutique :

Chirurgie :

Atteinte Ganglionnaire : IHC :

Chimiothérapie :

Radiothérapie :

Hormonothérapie, thérapie ciblée :

Suivi et évolution :

o Durée du suivi :mois

o Rechute : Oui Non

* Délai de rechute :

Production Scientifique

1/ PUBLICATIONS INTERNATIONALES

1- **Boulmaiz A**, Berredjem H, Cheikchouk K, Boulkrah H, Aouras H, Djedi H. (2024). Predicting HER2 Status Associated with Breast Cancer Aggressiveness Using Four Machine Learning Models. *Asian Pacific Journal of Cancer Prevention*. 25(10):3609-3618.

<https://doi.org/10.31557/APJCP.2024.25.10.3609>

2- Merabet N, Ramoz N, **Boulmaiz A**, Bourefis A, Benabdelkrim M, Djeflal O, Moyse E, Tolle V, Berredjem H. (2023). SNPs-Panel Polymorphism Variations in GHRL and GHSR Genes Are Not Associated with Prostate Cancer. *Biomedicines*, 11(12), 3276.

<https://doi.org/10.3390/biomedicines11123276>

3- **Boulmaiz A**, Berredjem H, Cheikchouk K, Boulkrah H, Aouras H, Djedi H. (2025). Exploring the Multi-Omic Role of DNA Methylation and Gene Expression in Breast Cancer: Findings from TCGA Data (en cours de publication).

4- Atba W, Cherifi M, Hazourli S, **Boulmaiz A**, Lapicque F, Grid A. (2025). Parametric Study and Applications of Gradient Boosting Regressor for Optimization and Automation of Urea Removal from Wastewater by Electrocoagulation (en coursed publication).

2/ LISTE DES COMMUNICATIONS

❖ Communications internationales

1. **Boulmaiz A**, Berredjem H, Cheikchouk K, Aouras H, Djedi H. (2025). *Comparaison de la régression de Cox et des réseaux de neurones artificiels pour la prédiction de la survie dans le cancer du sein à partir des données TCGA*. E-poster présenté aux Premières Journées Internationales d'Oncologie Médicale de l'EHS Ahmed Ben Bella, 16–17 mai 2025, Hôtel Renaissance, Tlemcen, Algérie.

2. **Boulmaiz A**, Berredjem H, Cheikchouk K, Boulkrah A, Aouras H, Djedi H. (2024). *Immune Marker Methylation and Its Impact on Breast Cancer Progression*. ICHSMT'24: International Congress on Health Sciences and Medical Technologies, Tlemcen, Algeria, 3–5 December 2024. In: *Book of Posters International Congress on Health Science and Medical Technologies 2024*.

3. **Boulmaiz A**, Bourefis A, Berredjem H, Boulkrah A, Cheikchouk K, Aouras H, Djedi H. (2024). *Profiling immune and hormonal in breast cancer: a machine learning-based clustering study with clinical and biological correlations*. Présentée au 8^e Congrès de Biochimie et de Génétique Médicale, Société Algérienne de Biochimie et de Génétique Médicale, 21–22 septembre 2024, Hôtel Mercure, Bab Ezzouar, Alger, Algérie.
4. Bourefis A, **Boulmaiz A**, Berredjem H, Djeflal O, Rocchi P. (2024). Etude de l'expression sérique de la ménine chez les patients atteints par cancer de la prostate. Présentée au 8^e Congrès de Biochimie et de Génétique Médicale, Société Algérienne de Biochimie et de Génétique Médicale, 21–22 septembre 2024, Hôtel Mercure, Bab Ezzouar, Alger, Algérie.
5. Boulkrah A, Berredjem H, **Boulmaiz A**, Cheniki R, Benaziza J.R, Chadli N, Kermiche N, Aouras H. (2020). *Clinicopathological Characteristics of Breast Cancer in a Population of Eastern Algeria*. *Medical Technologies Journal*, Poster presented in June 2020.
6. **Boulmaiz A**, Boulkrah A, Touaibia S, Sedira W, Lankar A, Berredjem H. (2019). *Epidemiological Study of Breast Cancer in a Population of Eastern Algeria*. ICHSMT'19: International Congress on Health Sciences and Medical Technologies, December 2019.

3/ FORMATIONS

1. **Atelier de Biologie Moléculaire** 23-26 Mars **2019** organisé par l'Association Tunisienne de Psycho-Neuro-Endocrino-Immunologie (AT-PNEI) à l'institut Supérieur de Biotechnologie de Monastir (ISBM) (Tunisie)
2. **Certificate of Excellence – Immunoassay Technologies**, 10 juin **2020** Délivré par la *Biological Society of Iraq* – Cours en ligne sur les technologies immuno-analytiques. Résultat du quiz final : 80/100. Encadré par : Marwan M. Mohammed & Ali Mohammed (MSc Microbiology & Immunology).
3. **Webinaire – ABC Oncology Educational Course : Biosimilars Module – Session I**, 28 juillet **2020** Organisé par *ABC Oncology*- Participation reconnue avec 2 crédits CME. Sujet: *Guidance on the Use of Biosimilars in the Cancer Setting*. Présenté par Dr. Sami Khatib & Dr. Fadi Farha.

4. **Formation** Neural Networks and Deep Learning sur la Plateforme : *Coursera, proposé par DEEP LEARNING AI (fondée par Andrew Ng)*, Terminé le 2 septembre **2023**, Note obtenue : 95,50 %

5. **Formation** Teach English Now! Foundational Principles sur la Plateforme : *Coursera, proposé par Arizona state university* Terminé le 9 mai **2023**, Note obtenue : 88,45 %

6. **Formation** Introduction to Genomic Technologies sur la Plateforme : *Coursera, proposé par Johns Hopkins University*, Terminé le 2 mai **2024**, Note obtenue : 79,50 %