

Thesis, FRANZEN Rachelle

Auteur : Messaoudi, Maryem

Promoteur(s) : Blomme, Arnaud

Faculté : Faculté de Médecine

Diplôme : Master en sciences biomédicales, à finalité spécialisée en biomédical data management

Année académique : 2025-2026

URI/URL : <http://hdl.handle.net/2268.2/25981>

Avertissement à l'attention des usagers :

Tous les documents placés en accès ouvert sur le site le site MatheO sont protégés par le droit d'auteur. Conformément aux principes énoncés par la "Budapest Open Access Initiative"(BOAI, 2002), l'utilisateur du site peut lire, télécharger, copier, transmettre, imprimer, chercher ou faire un lien vers le texte intégral de ces documents, les disséquer pour les indexer, s'en servir de données pour un logiciel, ou s'en servir à toute autre fin légale (ou prévue par la réglementation relative au droit d'auteur). Toute utilisation du document à des fins commerciales est strictement interdite.

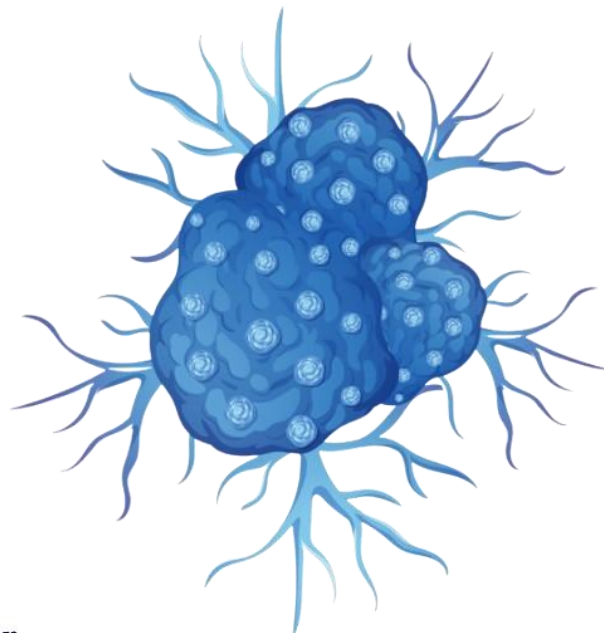
Par ailleurs, l'utilisateur s'engage à respecter les droits moraux de l'auteur, principalement le droit à l'intégrité de l'oeuvre et le droit de paternité et ce dans toute utilisation que l'utilisateur entreprend. Ainsi, à titre d'exemple, lorsqu'il reproduira un document par extrait ou dans son intégralité, l'utilisateur citera de manière complète les sources telles que mentionnées ci-dessus. Toute utilisation non explicitement autorisée ci-avant (telle que par exemple, la modification du document ou son résumé) nécessite l'autorisation préalable et expresse des auteurs ou de leurs ayants droit.

Faculté de Médecine
Mémoire

**Analyse par scRNA-seq du remodelage
immunitaire induit par la surexpression de
RICTOR dans un modèle murin de cancer
du poumon**

*Mémoire présenté en vue de l'obtention du diplôme de Master en Sciences
Biomédicales à finalité spécialisée en
Data Management*

Présenté par
Messaoudi Maryem



Promoteur
Blomme Arnaud

Co-promoteur
Lavergne Arnaud

Année académique 2025-2026

Remerciements

Je tiens tout d'abord à remercier Arnaud Blomme pour m'avoir offert l'opportunité de travailler sur ce projet. Ce travail, particulièrement enrichissant, m'a permis de découvrir une véritable passion pour la bioinformatique et d'explorer une nouvelle facette de la science que j'ai énormément appréciée. Je lui suis également reconnaissante pour la confiance qu'il m'a accordée tout au long de ce mémoire.

Je remercie également Arnaud Lavergne, mon promoteur en bioinformatique, pour sa disponibilité, ses réponses à mes questions et pour m'avoir ouvert les portes d'un laboratoire bienveillant. Grâce à ce projet, j'ai énormément appris sur le monde académique et sur ce que j'en attendais réellement.

Je souhaite adresser un immense merci à Elodie Renaude pour son encadrement tout au long de ce mémoire. Sa gentillesse et sa présence m'ont aidée à garder un repère et à continuer, même dans les moments plus compliqués.

Je remercie également mes amies, du fond du cœur, pour leur présence et leur soutien tout au long de ces années. Merci pour les moments partagés, les fous rires et l'énergie que vous m'avez apportée. Une mention particulière à Sarah, le yin de mon yang, le machine learning de ma bioinfo, dont les encouragements et l'ambition m'ont aidée à aller jusqu'au bout de ce mémoire. Merci aussi à Nada, mon acolyte de bioinformatique, mon doigt de la main, avec qui j'ai évolué tout au long de ce parcours, sa joie de vivre et sa bonne humeur ont été un véritable moteur. A cet appel passé à rire et à pleurer à vos côtés.

Je tiens également à remercier ma famille pour son soutien inconditionnel. Ma mère, pour m'avoir transmis la persévérance et la volonté d'aller au bout de ce que j'entreprends, et pour être une source d'inspiration au quotidien. Mes frères et ma sœur, pour leur présence et leur soutien, en particulier dans les moments plus difficiles, même quand mon code fonctionnait parfaitement... jusqu'à ce que je le relance une deuxième fois ;). Et ma nièce, mon rayon de soleil, qui m'a apporté de la joie tout au long de ce parcours.

Enfin, je souhaite avoir une pensée toute particulière pour ma Minnie d'amour, partie trop tôt cette année, mais dont la présence restera à jamais dans mon cœur.

Parce que ce parcours n'a pas toujours été simple, mais qu'il m'a menée bien plus loin que je ne l'aurais imaginé : *Per aspera ad astra*.

Utilisation de l'intelligence artificielle

Dans le cadre de la rédaction de ce mémoire, des outils d'intelligence artificielle ont été utilisés pour aider à la traduction de l'abstract, à la compréhension de certains articles scientifiques et à la création de la liste des abréviations. Ils ont également été utilisés pour la correction du texte, notamment pour vérifier l'orthographe, la grammaire et améliorer la clarté des phrases.

Résumé

Le carcinome non à petites cellules (NSCLC) constitue l'une des principales causes de mortalité par cancer dans le monde. Si l'immunothérapie néoadjuvante anti-PD-1 a modifié sa prise en charge, une proportion importante de patients ne répond pas au traitement, et les mécanismes de résistance restent mal compris. La surexpression de RICTOR, composante essentielle du complexe mTORC2, est observée dans une fraction des NSCLC et associée à un pronostic défavorable, mais son rôle dans le microenvironnement immunitaire tumoral demeure encore peu exploré.

Ce mémoire s'appuie sur l'analyse bioinformatique de données de séquençage d'ARN à cellule unique (scRNA-seq) issues d'un modèle murin génétique de NSCLC surexprimant RICTOR (KPR : **Tg-LSL-KRAS^{G12D}-TRP53^{lox/lox}-RICTOR^{OE}**) comparé au modèle contrôle KP, à deux stades de progression tumorale (Early : 1 mois ; Late : 3 mois post-induction). L'objectif était d'identifier une signature transcriptomique immunitaire associée au contexte KPR Late, puis d'évaluer son association avec la réponse à l'immunothérapie dans une cohorte humaine de NSCLC (Liu et al., 2025, Cell ; GSE243013 ; n = 243 patients, 1 239 631 cellules, 51 types cellulaires annotés).

Après contrôle qualité et intégration, 46 521 cellules immunitaires murines *CD45+* ont été retenues. Les macrophages représentaient la population majoritaire du compartiment immunitaire (environ 70 %). Un reclustering ciblé a permis d'identifier dix sous-populations macrophagiques distinctes. Les signatures candidates ont été construites à partir de marqueurs différentiels identifiés par FindMarkers, converties en orthologues humains, puis projetées sur 9 292 pseudobulks via trois approches complémentaires : scoring UCell, analyse différentielle gène par gène avec DESeq2, et test compétitif de gene set avec camera/limma-voom.

Les résultats suggèrent que la surexpression de RICTOR est associée à un remodelage transcriptomique du compartiment macrophagique murin. Certaines signatures montrent des associations partielles avec le statut de réponse pathologique dans les données humaines, bien que ces associations varient selon la signature et le sous-type macrophagique considérés. Ces résultats restent exploratoires, notamment parce que les échantillons humains proviennent de pièces de résection post-traitement.

Abstract

Non-small cell lung cancer (NSCLC) is one of the leading causes of cancer-related mortality worldwide. Although neoadjuvant anti-PD-1 immunotherapy has reshaped the clinical management of NSCLC, a substantial proportion of patients fail to respond, and the mechanisms underlying resistance remain incompletely understood. Overexpression of RICTOR, a key component of the mTORC2 complex, is observed in a subset of NSCLC and is associated with poor prognosis, yet its role within the tumor immune microenvironment remains largely unexplored.

This thesis is based on the bioinformatic analysis of scRNA-seq data derived from a genetically engineered mouse model of NSCLC overexpressing RICTOR (KPR : **Tg-LSL-KRAS^{G12D}-TRP53^{lox/lox}-RICTOR^{OE}**), compared to the control KP model, at two stages of tumor progression (Early: 1 month; Late: 3 months post-induction). The objective was to identify an immune-related transcriptomic signature associated with the Late KPR context, and subsequently assess its association with immunotherapy response in a human NSCLC cohort (Liu et al., 2025, Cell; GSE243013; n = 243 patients, 1,239,631 cells, 51 annotated cell types).

After quality control and integration, 46,521 murine CD45+ immune cells were retained. Macrophages constituted the dominant population within the immune compartment (approximately 70 %). Targeted reclustering identified ten distinct macrophage subpopulations. Candidate signatures were constructed from differentially expressed markers identified using FindMarkers, converted to human orthologs, and projected onto 9,292 pseudobulks using three complementary approaches: UCell scoring, gene-wise differential expression analysis with DESeq2, and competitive gene set testing using camera with limma-voom.

The results suggest that RICTOR overexpression is associated with transcriptomic remodeling of the murine macrophage compartment. Some signatures showed partial associations with pathological response status in the human dataset, although these associations vary depending on the signature and macrophage subtype considered. These findings remain exploratory, notably because the human samples were derived from post-treatment surgical resections.

Abréviations

AKT	protéine kinase B
ALK	anaplastic lymphoma kinase
anti-PD-(L)1	traitement ciblant PD-1 et/ou PD-L1
bulk RNA-seq	séquençage en vrac
C1Q	composant C1q du complément
CAFs	fibroblastes associés au cancer
DC	cellules dendritiques
DEG	gènes différentiellement exprimés
DE	Analyse différentielle
EGFR	récepteur du facteur de croissance épidermique
FCN1	ficolin 1
FGFR1	récepteur 1 du facteur de croissance des fibroblastes
FPI	fibrose pulmonaire idiopathique
HTO	hashtag oligonucléotides
ICI	inhibiteurs de points de contrôle immunitaire
ILC	cellules lymphoïdes innées
ImmGen	Immunological Genome Project
KP	modèle murin <i>KRAS</i>^{G12D}-<i>TP53</i>^{lox/lox}
KPR	modèle murin <i>Kras</i> ^{G12D} / <i>Trp53</i> ^{lox/lox} - <i>RICTOR</i> ^{OE} , avec surexpression conditionnelle de RICTOR
KRAS	Kirsten rat sarcoma viral oncogene homolog
LUAD	adénocarcinome pulmonaire
LUSC	carcinome épidermoïde pulmonaire
MDM	macrophages dérivés des monocytes
MPR	réponse pathologique majeure

mLST8	mammalian lethal with SEC13 protein 8
mTOR	mechanistic target of rapamycin
mTORC1	complexe 1 de mTOR / mechanistic target of rapamycin complex 1
mTORC2	complexe 2 de mTOR / mechanistic target of rapamycin complex 2
NK	natural killer
NR	non-répondeurs
NSCLC	carcinome non à petites cellules / non-small cell lung cancer
OMS	Organisation mondiale de la Santé
PCA	analyse en composantes principales
PCs	composantes principales
PD-1	programmed cell death protein 1
PD-L1	programmed death-ligand 1
pCR	réponse pathologique complète
PKCα	protéine kinase C alpha
R	répondeurs
RAPTOR	regulatory-associated protein of mTOR
RICTOR	rapamycin-insensitive companion of mTOR
RPCA	analyse en composantes principales réciproque
RVT	tumeur résiduelle viable
scRNA-seq	séquençage ARN unicellulaire / single-cell RNA sequencing
SGK1	serum/glucocorticoid-regulated kinase 1
LSL	Lox-STOP-Lox
SOX2	facteur de transcription SRY-box 2
SCLC	carcinome à petites cellules
SP-B	Surfactant Protein B
SPP1	secreted phosphoprotein 1

SV EC	cellules endothéliales veineuses systémiques
TAMs	macrophages associés aux tumeurs
TCGA	The Cancer Genome Atlas
TDM	tomodensitométrie
TEP-scan	tomographie par émission de positons couplée au scanner
TMB	charge mutationnelle tumorale
TME	microenvironnement tumoral
TNM-8	classification Tumor-Node-Metastasis, huitième édition
TP53	tumor protein p53
Tregs	lymphocytes T régulateurs
TREM2	triggering receptor expressed on myeloid cells 2
TRM	macrophages résidents
Trp53	transformation related protein 53
UMAP	Uniform Manifold Approximation and Projection

Table des matières

INTRODUCTION	1
1.1 CANCER DU POUMON : ÉPIDÉMIOLOGIE ET BIOLOGIE	1
1.1.1 Epidémiologie mondiale	1
1.1.2 Classification histologique et hétérogénéité moléculaire	1
1.1.3 Diagnostic, stadification et contexte de l'étude	2
1.1.4 La voie mTOR/RICTOR	4
1.2 IMMUNOTHÉRAPIE NÉOADJUVANTE ET HÉTÉROGÉNÉITÉ DE RÉPONSE	5
1.2.1 Points de contrôle immunitaires	5
1.2.2 Immunothérapie néoadjuvante dans le NSCLC résécable	5
1.2.3 Hétérogénéité de réponse : un problème non résolu	6
1.3. LE MICROENVIRONNEMENT TUMORAL DANS LE NSCLC	7
1.3.1 Composition et rôle du TME dans la réponse aux ICI	7
1.3.2 Hétérogénéité du TME	8
1.4 LES MACROPHAGES ASSOCIÉS AUX TUMEURS (TAMs)	9
1.4.1. Biologie et plasticité macrophagique	9
1.4.2 TAMs, progression tumorale et résistance aux ICI dans le NSCLC	10
1.4.3 Hétérogénéité des TAMs par le scRNA-seq	10
1.5 OUTILS BIOINFORMATIQUES	12
1.5.1 Pourquoi le <i>single-cell</i>	12
1.5.2 Problématique de projection inter-espèces	13
1.5.3 <i>Scoring</i> des signatures et robustesse clinique	14
1.5.4 Contexte expérimental et données préliminaires	14
2.OBJECTIFS ET HYPOTHESES	16
2.1 STRATÉGIE EXPÉRIMENTALE	16
3.Matériel et Méthodes	17
3.1 SOURCE DES DONNEES MURINES	17
3.1.1 Données murines : modèles et <i>design expérimental</i>	17
3.1.2 Données humaines	18
3.2 LOGICIELS ET PACKAGES	18
3.3 CONTRÔLE QUALITÉ ET FILTRAGE CELLULAIRE	19
3.4 DÉTECTION ET EXCLUSION DES DOUBLETS	20
3.5 NORMALISATION	20
3.6 INTÉGRATION DES DONNÉES MURINES	20

3.7 RÉDUCTION DE DIMENSION ET CLUSTERING	21
3.8 ANNOTATION CELLULAIRE	21
3.9 ANALYSE DIFFÉRENTIELLE GLOBALE	22
3.9.1 Approche pseudo-bulk (DESeq2)	22
3.9.2 Approche FindMarkers Wilcoxon	22
3.10 ANALYSES FOCALISÉES SUR LE COMPARTIMENT MACROPHAGIQUE.....	23
3.10.1 Extraction et construction du sous-objet macrophagique	23
3.10.2 Intégration par Harmony et reclustering macrophagique	23
3.10.3 Construction des signatures murines candidates	25
3.10.4 Conversion des gènes murins en orthologues humains	25
3.11 PRÉTRAITEMENTS DES DONNÉES HUMAINES	26
3.11.1 Chargement et contrôle qualité	26
3.11.2 Stratégie de traitement par lots.....	27
3.11.3 Agrégation pseudo-bulk.....	27
3.11.4 Classification des répondeurs	27
3.12 PROJECTION ET <i>SCORING</i> DES SIGNATURES SUR LES DONNÉES HUMAINES	27
3.12.1 UCell	27
3.12.2 Score patient par sous-type macrophagique.....	27
3.12.3 Analyse différentielle au niveau pseudobulk (DESeq2)	28
3.12.4 Test compétitif de gene sets (camera, limma-voom).....	28
4. Résultats	30
4.1 ANALYSE DU MICROENVIRONNEMENT IMMUNITAIRE MURIN	30
4.1.1 Contrôle qualité, filtrage et détection des doublets	30
4.1.2 Intégration et réduction de dimension	30
4.1.3 Annotation cellulaire et composition du TME murin.....	30
4.1.4 Analyse différentielle pseudobulk globale	31
4.2 ANALYSE DU COMPARTIMENT MACROPHAGIQUE MURIN	33
4.2.1 Réintégration par Harmony et reclustering	33
4.2.2 Caractérisation des clusters macrophagiques à partir de marqueurs de littérature ...	34
4.2.3 Construction des signatures murines candidates	35
4.3 PROJECTION SUR LES DONNÉES HUMAINES DE NSCLC	36
4.3.1 Cohorte humaine analysable et pseudobulks.....	36
4.3.2 <i>Scoring</i> UCell des signatures dans les pseudobulks macrophagiques	36
4.3.3 Analyse différentielle gène par gène dans les pseudobulks macrophagiques	37

4.3.4 Test compétitif des signatures par camera	38
5. Discussion	40
5.1 MISE EN PERSPECTIVE.....	40
5.2 REMODELAGE DU COMPARTIMENT MACROPHAGIQUE MURIN ET PORTÉE RÉELLE DE L'EFFET RICTOR	41
5.3 CE QUE LES SIGNATURES CAPTURENT RÉELLEMENT	42
5.4 PROJECTION HUMAINE : UN SIGNAL PARTIEL ET DÉPENDANT DU CONTEXTE	43
5.5 LIMITES, PERSPECTIVES ET CONCLUSION	45
6. Bibliographie.....	48

INTRODUCTION

1.1 CANCER DU POUMON : ÉPIDÉMIOLOGIE ET BIOLOGIE

1.1.1 Épidémiologie mondiale

Le cancer du poumon est la principale cause de mortalité par cancer dans le monde et représente donc un enjeu majeur de santé publique à l'heure actuelle. En 2022, il a été responsable d'environ 2.5 millions de nouveaux cas et de 1.8 million de décès à l'échelle mondiale, faisant de lui le cancer le plus meurtrier devant le cancer du sein et le cancer colorectal¹. Si cette tendance reste constante, les projections suggèrent que le nombre de cas et de décès pourrait encore augmenter fortement d'ici 2050².

À l'échelle belge, 9 487 patients ont été diagnostiqués avec un cancer du poumon en 2023, avec une majorité d'hommes atteints. Le pronostic reste mauvais, avec une survie nette à 5 ans de 27 % chez les hommes et de 35.7 % chez les femmes en Belgique sur la période 2019-2023³. Cette faible survie s'explique en partie par des diagnostics souvent tardifs, liés à l'absence de dépistage systématique et au fait que les symptômes précoces sont peu spécifiques. Le tabac reste le principal facteur de risque, mais d'autres facteurs interviennent aussi, notamment la pollution atmosphérique, en particulier les particules fines, certaines expositions professionnelles comme l'amiante ou le radon, ainsi que certaines maladies pulmonaires chroniques et prédispositions génétiques^{4,5}.

Ainsi, malgré les progrès thérapeutiques récents, le pronostic est parmi les plus défavorables en oncologie⁶. Cette situation souligne l'importance de mieux comprendre les mécanismes biologiques impliqués dans la progression tumorale et à la résistance aux traitements. Pour mieux comprendre ces mécanismes, des approches bioinformatiques permettent d'analyser des données de séquençage (génomique, transcriptomique, single-cell transcriptomique,...) réalisées sur des échantillons provenant de cohortes de patients et des métadonnées cliniques associées à la réponse thérapeutique. C'est dans cette perspective que s'inscrit ce mémoire.

1.1.2 Classification histologique et hétérogénéité moléculaire

Le cancer du poumon est classiquement divisé en deux grandes catégories histologiques : le carcinome non à petites cellules (NSCLC), qui représente environ 80 à 85 % des cas, et le carcinome à petites cellules, ou SCLC, qui représente les 15 à 20 % restants⁷. Dans ce travail, nous nous intéressons exclusivement au NSCLC, qui est le point central de notre étude. Selon la classification OMS 2021, il est divisé en trois sous-types principaux : l'adénocarcinome pulmonaire (LUAD), le carcinome épidermoïde pulmonaire (LUSC) et le carcinome à grandes cellules⁸.

Le sous-type le plus fréquent est le LUAD, qui est aussi le plus présent chez les non-fumeurs et chez les femmes, et se développe généralement en périphérie du poumon⁹. Au niveau moléculaire, il porte fréquemment des mutations *drivers*, en particulier des altérations du récepteur du facteur de croissance (*EGFR*) et Kirsten rat sarcoma viral oncogene homolog (*KRAS*), mais aussi des réarrangements de anaplastic lymphoma kinase (*ALK*). La répartition de ces mutations n'est pas la même dans le monde : les mutations

EGFR sont plus fréquentes en Asie de l'Est, tandis que les mutations *KRAS* prédominent en Occident et sont fortement associées au tabagisme⁹. Le LUSC est le deuxième sous-type le plus fréquent, il se développe préférentiellement au centre du poumon, à partir des bronches principales, et reste le sous-type le plus fortement associé au tabagisme¹⁰. Sur le plan moléculaire, il est surtout caractérisé par des altérations du gène tumor protein p53 (*TP53*), mais également par des amplifications du récepteur 1 du facteur de croissance des fibroblastes (*FGFR1*) et du facteur de transcription SRY-box 2 (*SOX2*). Enfin, le carcinome à grandes cellules occupe une place un peu particulière. Il s'agit en réalité d'une catégorie diagnostique d'exclusion, définie quand la tumeur ne montre pas de différenciation glandulaire, épidermoïde ou neuroendocrine identifiable à l'histologie et à l'immunohistochimie⁹.

L'une des principales difficultés dans le traitement du NSCLC est son importante hétérogénéité moléculaire. Il est composé de sous-groupes définis par différentes mutations *drivers*, tels que *KRAS*, *EGFR*, *TP53* et *ALK*, dont la fréquence varie selon le type histologique, l'origine géographique, l'ethnie et le statut tabagique du patient. Chaque mutation définit donc des sous-groupes tumoraux avec des microenvironnements tumoraux, c'est-à-dire tout l'ensemble des cellules non tumorales, des cellules immunitaires, des cellules stromales, des vaisseaux et des signaux moléculaires qui entourent et influencent la tumeur, et des réponses thérapeutiques différentes, qui ne sont pas équivalents sur le plan clinique. Seuls 25 à 30 % des cas de NSCLC ont une mutation ciblable par un traitement médicamenteux spécifique. Pour les autres, l'immunothérapie, qui est une stratégie thérapeutique qui a pour but de réactiver ou renforcer la réponse immunitaire antitumorale du patient, représente une option importante, mais sa réponse reste limitée, avec un taux de réponse global d'environ 20 % lorsqu'elle est utilisée seule¹¹.

L'effet des co-mutations illustre bien cette complexité. La combinaison *KRAS/TP53* est associée à une expression élevée du *programmed death-ligand 1* (PD-L1), ligand du *programmed cell death protein 1* (PD-1), formant l'axe de point de contrôle immunitaire PD-1/PD-L1, et à un microenvironnement plus inflammatoire, généralement corrélé à une meilleure réponse aux inhibiteurs de points de contrôle immunitaire (ICI). A l'inverse, les co-mutations *KRAS/STK11* sont plutôt considérées comme des marqueurs prédictifs négatifs, car elles sont associées à une évaison immunitaire et à une résistance aux immunothérapies. C'est un point important, car deux patients portant une mutation *KRAS* peuvent avoir des microenvironnements immunologiques très différents selon leur co-mutation. Cette diversité moléculaire entre patients se reflète dans la composition et l'organisation de leur microenvironnement tumoral, contribuant à la variabilité des réponses à l'immunothérapie¹².

1.1.3 Diagnostic, stadification et contexte de l'étude

Le cancer du poumon est diagnostiqué dans 70 % des cas à un stade avancé et inopérable, en raison de l'absence de symptômes spécifiques aux stades précoces et de l'absence de dépistage systématique. Le diagnostic repose sur l'imagerie thoracique, principalement le scanner ou tomodensitométrie (TDM), et la tomographie par émission de positons couplée au scanner (TEP-scan), auxquelles s'ajoutent une

confirmation histologique et un bilan moléculaire, indispensables pour choisir le traitement¹⁰. La stadification repose sur la classification *Tumor-Node-Metastasis*, huitième édition (TNM-8), où le T désigne la tumeur primitive (taille et invasion locale), N la présence de cellules tumorales dans les ganglions lymphatiques régionaux, et M les métastases à distance. Ce classement permet de définir la taille de la tumeur, son extension locorégionale, l'atteinte ganglionnaire et la présence de métastases, et donc d'orienter la décision thérapeutique¹³. Les grands principes de classement sont résumés simplement dans le **Tableau 1**.

Tableau 1. Synthèse simplifiée des principaux stades du NSCLC selon la classification TNM-8.

Ce tableau présente les grands stades du cancer du poumon non à petites cellules en fonction de :

- l'extension de la tumeur primitive,
- l'atteinte des ganglions lymphatiques régionaux,
- la présence de métastases à distance.

Il ne reprend pas tous les sous-stades TNM détaillés, mais sert de repère général pour l'interprétation clinique.

Synthèse réalisée d'après les principes de la classification TNM et les informations cliniques du National Cancer Institute^{13,14}.

Stade	Caractéristiques générales
Stade 0	Carcinome in situ : cellules tumorales limitées à l'épithélium, sans invasion des tissus voisins, sans atteinte ganglionnaire ni métastase.
Stade I	Tumeur localisée au poumon, sans atteinte ganglionnaire régionale ni métastase à distance.
Stade II	Tumeur plus volumineuse et/ou associée à une atteinte ganglionnaire régionale limitée, sans métastase à distance.
Stade III	Maladie localement avancée, avec extension locorégionale importante et/ou atteinte ganglionnaire médiastinale, sans métastase à distance.
Stade IV	Maladie métastatique, définie par la présence de métastases à distance ou d'une dissémination pleurale/péricardique maligne.

Les données analysées dans ce projet proviennent de patients atteints de NSCLC résécables, traités par chimio-immunothérapie néoadjuvante puis opérés. Les tissus étudiés proviennent donc de tumeurs résécables après traitement néoadjuvant. Malgré l'efficacité démontrée des ICI, en situation néoadjuvante, une proportion importante de patients ne répond pas au traitement, et les mécanismes de résistance restent encore mal compris^{15,16}.

Cette hétérogénéité de réponse nous amène à une question centrale : quels mécanismes biologiques distinguent les patients qui répondent de ceux qui résistent aux ICI ? Pour parvenir à y répondre, il faut comprendre comment les ICI agissent et quelles voies de signalisation intracellulaires, dans les cellules tumorales et dans le microenvironnement immunitaire, peuvent moduler l'efficacité.

1.1.4 La voie mTOR/RICTOR

1.1.4.1 La signalisation mTOR

La voie *mechanistic target of rapamycin* (mTOR) repose sur une sérine/thréonine kinase centrale qui intègre différents signaux de l'environnement cellulaire, tels que les facteurs de croissance, les nutriments, l'état d'énergie, pour réguler la croissance, le métabolisme et la survie cellulaire¹⁷⁻¹⁹. Bien que codée par un gène unique chez les eucaryotes supérieurs, elle n'agit pas seule et fonctionne comme sous-unité catalytique de deux complexes multiprotéiques distincts, mTORC1 et mTORC2. Ces deux complexes ont la même activité kinase mais diffèrent dans leur composition, leurs substrats et leurs fonctions^{17,20,21}. En effet, mTORC1 s'organise autour de la protéine *scaffold* regulatory-associated protein of mTOR (RAPTOR), tandis que mTORC2 lui repose sur la protéine *scaffold* rapamycin-insensitive companion of mTOR (RICTOR)^{17,19}. L'activation de mTORC1 favorise principalement la synthèse protéique et la croissance cellulaire, tout en inhibant l'autophagie^{17,21}. Ce complexe est sensible à la rapamycine et ses analogues, ce qui en fait une cible thérapeutique importante en oncologie^{18,21}. Son hyperactivation est fréquemment observée dans les cancers et contribue à la prolifération, à la résistance à l'apoptose et à une reprogrammation métabolique vers un état anabolique, notamment dans le NSCLC¹⁷.

Contrairement à mTORC1, mTORC2 est insensible à un traitement aigu à la rapamycine²⁰. Son principal substrat est la protéine kinase B (AKT), qu'il phosphoryle sur la sérine 473, une étape nécessaire à son activation complète^{19,21}. Il régule également la protéine kinase C alpha (PKC α), participe au remodelage du cytosquelette d'actine, et la glucocorticoïd-regulated kinase 1 (SGK1), qui intervient dans l'homéostasie ionique. Ces fonctions expliquent son rôle dans la survie cellulaire, la migration et l'invasion, des caractéristiques qui jouent un rôle important dans l'évolution de la tumeur¹⁹.

Les deux complexes entretiennent aussi un rétrocontrôle réciproque, ce qui complique leur ciblage pharmacologique^{17,19,20}. L'inhibition de mTORC1 par les rapalogues lève donc ce frein et peut paradoxalement réactiver mTORC2 et AKT, ce qui soutient la survie des cellules tumorales malgré le traitement^{17,19,20}. C'est justement pour contourner ce mécanisme de compensation que des inhibiteurs capables de bloquer en même temps mTORC1 et mTORC2, ont été développés. La plupart de ces inhibiteurs restent actuellement en phase préclinique ou en phase clinique précoce^{17,19,21}.

1.1.4.2 RICTOR dans le NSCLC

Dans le NSCLC, l'amplification du gène de RICTOR est un mécanisme fréquent d'activation de mTORC2. Cette altération génomique constitue une nouvelle cible thérapeutique pour l'inhibition de mTORC2²²⁻²⁴. Cette amplification, observée dans environ 12 % des cas de NSCLC selon les données *The Cancer Genome Atlas* (TCGA), avec une fréquence légèrement plus élevée dans le LUSC, est surtout retrouvée chez les patients ne présentant pas de mutation driver classique comme *EGFR* ou *KRAS* dans ce sous-groupe^{11,22,23}. L'amplification de *RICTOR* s'accompagne d'une surexpression de la protéine RICTOR^{22,25}. Au-delà de cet effet, cette amplification favorise la formation du complexe mTORC2, en recrutant les sous-unités

partagées comme mTOR et mLST8²². En aval, cela entraîne une augmentation de la phosphorylation d'AKT sur Ser473 et l'activation de PKC α , ce qui favorise la prolifération, renforce la résistance à l'apoptose, facilite la transition épithélio-mésenchymateuse et augmente le potentiel métastatique^{22,23,26}. La surexpression de RICTOR est alors associée à un pronostic défavorable dans plusieurs cohortes de NSCLC^{23,26}. L'ensemble de ces mécanismes est synthétisé dans la **Figure 1**.

1.2 Immunothérapie néoadjuvante et hétérogénéité de réponse

1.2.1 Points de contrôle immunitaires

Le système immunitaire possède des mécanismes physiologiques de tolérance, qui agissent comme des garde-fous (*gate-keepers*) en régulant l'activation des lymphocytes T pour éviter les réponses auto-immunes. Ces voies sont exploitées par les cellules tumorales pour échapper à la surveillance immunitaire²⁷.

Dans le NSCLC, l'axe PD-1/PD-L1 joue un rôle central dans le microenvironnement tumoral. PD-L1, exprimé à la surface des cellules tumorales et de certaines cellules immunitaires comme les macrophages et les cellules dendritiques (DC), se lie au récepteur PD-1 des lymphocytes T. Cette interaction induit une signalisation inhibitrice qui diminue la production de cytokines, la prolifération et l'activité cytotoxique des lymphocytes T²⁷. Dans ce contexte, la stimulation antigénique continue conduit progressivement à un état d'épuisement fonctionnel des lymphocytes T, caractérisé par une diminution de leur capacité proliférative et de leur fonction cytotoxique. Cet état s'accompagne d'une expression accrue de plusieurs récepteurs inhibiteurs, dont PD-1, TIM-3, LAG-3 et TIGIT, qui marquent un état d'épuisement avancé des lymphocytes²⁸. Les ICI agissent en bloquant ces interactions, notamment celle entre PD-1 et ses ligands, afin de restaurer une activité antitumorale endogène. Dans le NSCLC, les traitements les plus utilisés ciblent principalement PD-1 ou PD-L1, les anti-PD-1 tels que le nivolumab, le pembrolizumab et le cemiplimab, ou les anti-PD-L1 comme l'atézolizumab et le durvalumab. Ils peuvent être utilisés seuls, associés à la chimiothérapie, ou en combinaison avec un autre ICI ciblant un point de contrôle immunitaire différent²⁹.

1.2.2 Immunothérapie néoadjuvante dans le NSCLC résécable

L'administration des ICI avant la chirurgie, en situation néoadjuvante (traitement préalable avant la chirurgie), est particulièrement intéressante parce que la tumeur est encore en place et constitue une source d'antigènes tumoraux. Ce qui permet une meilleure activation des lymphocytes T contre les cellules cancéreuses. À l'inverse, après chirurgie, une grande partie de cette source antigénique est retirée, ce qui peut limiter la réponse immunitaire. Le but est alors de stimuler une réponse immunitaire systémique avant l'opération, afin d'éliminer la tumeur primitive mais aussi les micrométastases déjà disséminées, qui peuvent être à l'origine de récives²⁹. Ce mécanisme d'activation systémique en contexte néoadjuvant est illustré dans la **Figure 2**.

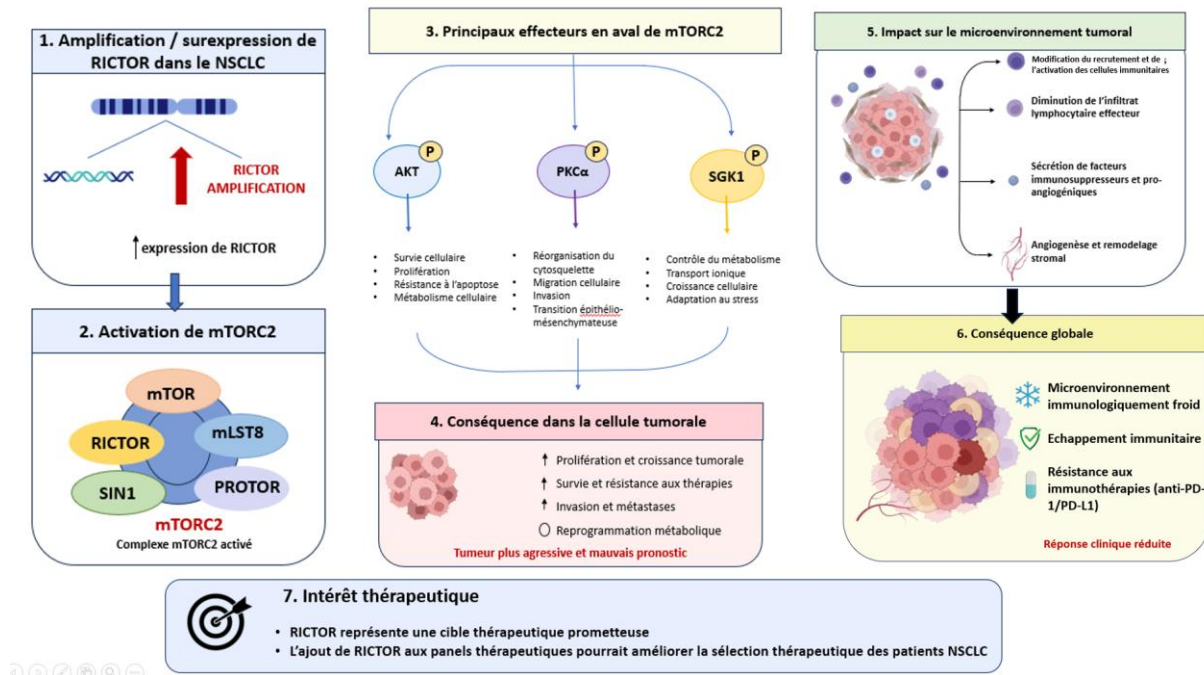


Figure 1. Conséquences de l'amplification de RICTOR dans le NSCLC.

L'amplification du gène RICTOR peut entraîner une surexpression de la protéine RICTOR, favorisant une activation accrue du complexe mTORC2. En aval, mTORC2 phosphoryle plusieurs effecteurs majeurs, notamment AKT, PKCα et SGK1. Ces protéines participent à la survie, à la prolifération, à la migration, à l'invasion et à la reprogrammation métabolique des cellules tumorales. Dans le microenvironnement tumoral, l'activation de cette voie peut également modifier le recrutement et l'activation des cellules immunitaires, réduire l'infiltration de lymphocytes effecteurs et favoriser la production de facteurs immunosuppresseurs et pro-angiogéniques. Ces mécanismes peuvent contribuer à l'établissement d'un microenvironnement immunologiquement froid, favorisant l'échappement immunitaire et une réponse aux immunothérapies anti-PD-(L)1 réduite.

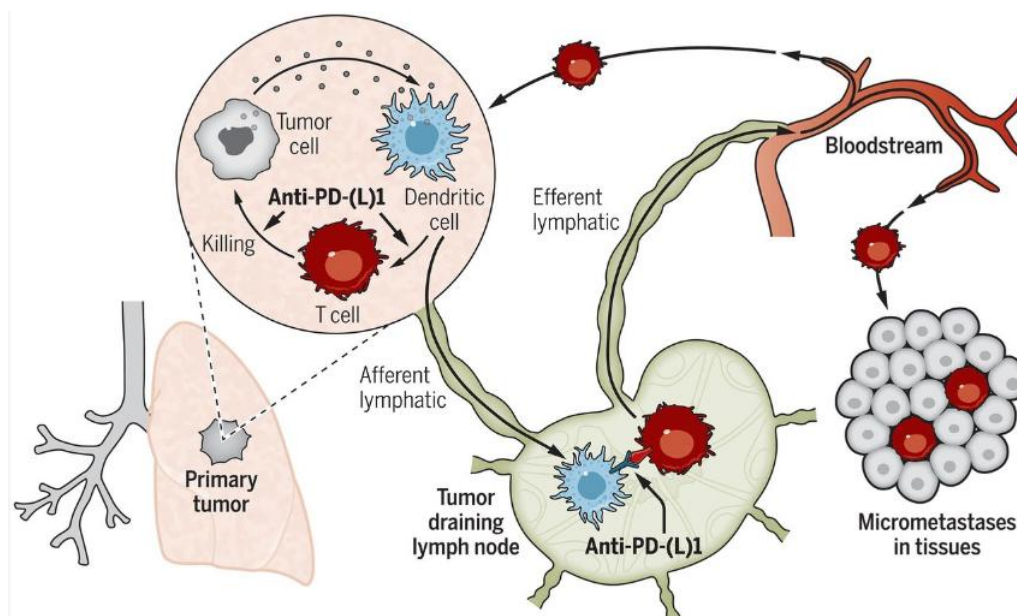


Figure 2. Mécanisme d'action systémique des ICI anti-PD-(L)1 en situation néoadjuvante.

Les cellules dendritiques transportent les antigènes tumoraux depuis la tumeur primaire vers le ganglion lymphatique drainant, où les lymphocytes T sont activés. Les anti-PD-(L)1 bloquent l'interaction inhibitrice entre PD-1 (présent sur les lymphocytes T) et ses ligands PD-L1/PD-L2 (exprimés notamment par les cellules tumorales ou certaines cellules immunitaires). En empêchant cette interaction, ils restaurent l'activité cytotoxique des lymphocytes T. Ces lymphocytes T activés peuvent ensuite rejoindre la circulation sanguine et cibler aussi bien la tumeur primaire que les micrométastases à distance. Adapté de Topalian et al.³⁰

La réponse au traitement est alors évaluée sur la pièce de résection par quantification du pourcentage de tumeur résiduelle viable (RVT). Nous parlons de réponse pathologique complète (pCR) lorsqu'aucune cellule tumorale viable n'est détectée, et de réponse pathologique majeure (MPR) lorsque le pourcentage de tumeur viable résiduelle est inférieur ou égal à 10 %²⁹.

L'essai NADIM a évalué l'association de nivolumab-chimiothérapie en néoadjuvant chez des patients atteints de NSCLC de stade IIIA résécable, et a rapporté 83 % de réponse pathologique majeure et 63 % de réponse pathologique complète chez les patients opérés. L'essai CheckMate 816 a ensuite confirmé l'intérêt de cette approche chez des patients de stade IB à IIIA, avec une amélioration de la survie sans événement et une augmentation du taux de réponse pathologique complète de 24 % contre 2,2 % sous chimiothérapie seule. Ces résultats ont conduit à l'approbation de cette stratégie néoadjuvante aux États-Unis^{16,31}. L'essai KEYNOTE-671 a ensuite étendu ces résultats avec le pembrolizumab en approche peropératoire, confirmant le bénéfice sur la survie et sur les taux de réponse pathologique¹⁵.

1.2.3 Hétérogénéité de réponse : un problème non résolu

Malgré ces résultats, seule une fraction des patients atteint une pCR ou une MPR sous un protocole identique, comme nous le voyons dans les données de CheckMate 816. Les biomarqueurs actuels ne permettent pas de prédire cette réponse de façon fiable. Le niveau d'expression de PD-L1 sur les cellules

tumorales est le marqueur le plus utilisé en clinique, mais il reste imparfait : dans l'essai NADIM, 58 % des patients avec un faible score de PD-L1 atteignaient malgré tout une MPR ou une pCR³¹.

Un autre aspect est la charge mutationnelle tumorale, ou le TMB, qui quantifie le nombre de mutations somatiques dans la tumeur. Cependant, le TMB n'a pas non plus permis de prédire la réponse pathologique dans CheckMate 816³². Cela suggère que la réponse aux ICI ne dépend pas uniquement des caractéristiques intrinsèques de la tumeur, mais aussi de l'ensemble des cellules qui constituent son microenvironnement. La caractérisation fine des populations immunitaires associées à la réponse ou à la résistance aux ICI est donc un enjeu important^{33,34}. Cette hétérogénéité cellulaire entre répondeurs et non-répondeurs est illustrée en **Figure 3**.

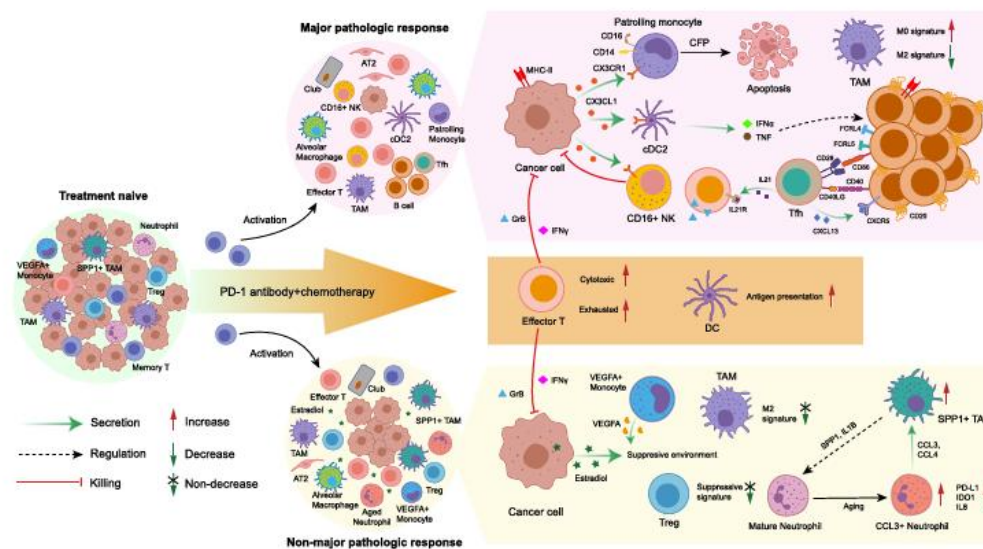


Figure 3. Dynamique du microenvironnement tumoral dans le NSCLC en réponse à la chimioimmunothérapie anti-PD-1.

Les patients avec une réponse pathologique majeure présentent une activation plus marquée des lymphocytes T effecteurs avec en parallèle une diminution des macrophages associés aux tumeurs (TAMs) à signature M2 et des lymphocytes T régulateurs (Tregs). À l'inverse, les non-répondeurs conservent un environnement immunosuppresseur dominé par les SPP1+ TAMs, les monocytes et les neutrophiles. Figure adaptée de Hu et al³⁵.

Cela nous amène à porter notre attention sur le microenvironnement tumoral, dont la composition cellulaire joue un rôle considérable dans la réponse aux ICI.

1.3. Le microenvironnement tumoral dans le NSCLC

1.3.1 Composition et rôle du TME dans la réponse aux ICI

La tumeur constitue un tissu hétérogène où les cellules cancéreuses coexistent avec un ensemble de cellules immunitaires, stromales et vasculaires formant le microenvironnement tumoral, ou TME. Ce microenvironnement n'est pas passif, les cellules tumorales le remodelent activement via des programmes oncogéniques, comme les mutations *KRAS* ou *BRAF*, pour favoriser leur survie et échapper à la surveillance immunitaire³⁶.

Le compartiment immunitaire est le plus étudié, il comprend les lymphocytes T cytotoxiques CD8+, principaux effecteurs de la réponse antitumorale, mais aussi des populations immunosuppressives comme les lymphocytes T régulateurs (Tregs) et les macrophages associés aux tumeurs (TAMs). D'autres

populations immunitaires, telles que les lymphocytes B, les cellules *natural killers (NK)*, les monocytes, les macrophages et les DC participent aussi à la composition du microenvironnement tumoral, avec des rôles variables selon leur état d'activation^{36,37}. Le compartiment stromal regroupe quant à lui des cellules non immunitaires, tels que les fibroblastes associés au cancer (CAFs), les cellules endothéliales et lymphatiques, qui structurent la tumeur en participant à la fibrose, au remodelage de la matrice extracellulaire et à la régulation du recrutement des cellules immunitaires au site tumoral^{36,38}.

La composition du TME conditionne directement la capacité d'un patient à répondre aux ICI. Trois profils distincts selon l'infiltration cellulaire et leur impact immunitaire ont été identifiés dans le NSCLC³⁹. Ces profils peuvent être rapprochés des catégories plus générales *immune-inflamed*, *immune-desert* et *immune excluded*³⁶. Le profil « *immuno-activite* », ou phénotype chaud, se caractérise par une forte infiltration en lymphocytes T CD8+, T CD4+ et B CD20+. Il est généralement associé à la meilleure survie sans maladie et à une bonne réponse à l'immunothérapie en monothérapie. Le profil « *immuno-exempt* », ou tumeur froide, présente peu de cellules immunitaires effectrices et est dominé par les neutrophiles et les Tregs, qui créent un environnement immunosuppresseur via des cytokines comme le TGF- β , limitant la réponse antitumorale. Ce profil est souvent associé à des stades cliniques avancés. Le profil « *immuno-defectueux* » est riche en cellules souches cancéreuses et en macrophages pro-tumoraux de type M2, favorisant l'immunosuppression, l'angiogenèse et est associé au pire pronostic chez les patients³⁹.

Au-delà de la simple présence des lymphocytes T, leur localisation et leur état fonctionnel sont donc essentiels. Une tumeur peut contenir des lymphocytes T sans répondre efficacement aux ICI si ceux-ci sont épuisés, exclus du contact avec les cellules tumorales ou inhibés par des populations immunosuppressives. Parmi ces populations, les Tregs, certains macrophages associés aux tumeurs selon leur polarisation et les cellules myéloïdes suppressives peuvent produire des cytokines inhibitrices comme IL-10 et TGF- β et limiter l'activité cytotoxique des lymphocytes T effecteurs³⁶.

1.3.2 Hétérogénéité du TME

Dans le NSCLC, le TME présente une hétérogénéité importante, confirmée à la fois par le séquençage unicellulaire et la transcriptomique spatiale. Contrairement à une idée simplifiée, les LUAD et les LUSC ne se distinguent pas seulement par leur histologie, mais aussi par leurs réseaux d'interactions cellule-cellule et par l'expression de différents couples ligand-récepteur⁴⁰.

L'analyse spatiale montre que la tumeur est organisée de manière très structurée, et non de façon aléatoire. Elle contient des régions distinctes au sein d'un même échantillon, avec des fonctions et des réseaux d'interaction moléculaire qui leur sont propres. Cette organisation a un impact sur la réponse thérapeutique, car l'environnement tumoral peut limiter l'abondance des lymphocytes T et NK, mais aussi freiner leur activité. Même en présence d'ICI, les lymphocytes T peuvent être réduits en nombre et inhibés fonctionnellement par leur environnement⁴⁰.

Les macrophages sont un autre type cellulaire occupant une place centrale dans le TME. Dans le contexte de la résistance aux ICI, l'épuisement fonctionnel des lymphocytes T peut être favorisé et entretenu par les signaux immunosuppresseurs émis par les cellules myéloïdes environnantes, surtout les macrophages³⁶. Une relation inverse a été observée entre le nombre de macrophages anti-inflammatoires et celui des lymphocytes T et NK cytotoxiques dans le tissu tumoral, ce qui souligne leur rôle dans la limitation de l'infiltration immunitaire au sein de la tumeur⁴¹. Par leur plasticité fonctionnelle et leur capacité à réduire les signaux nécessaires pour le recrutement des lymphocytes T, ils constituent un facteur clé de résistance aux ICI^{36,40}. Cette plasticité des macrophages les rend particulièrement intéressants pour comprendre les mécanismes de résistance primaire aux ICI, même si leur reprogrammation fonctionnelle reste complexe lorsque les macrophages sont déjà engagés dans un état polarisé³⁶. Les macrophages, par leur abondance, leur rôle régulateur et leur plasticité fonctionnelle, méritent un examen approfondi.

1.4 Les macrophages associés aux tumeurs (TAMs)

1.4.1. Biologie et plasticité macrophagique

Les macrophages sont des cellules ayant plusieurs fonctions qui jouent un rôle essentiel dans l'immunité innée⁴¹. Ils assurent par exemple les rôles de sentinelles en phagocytant les débris et les pathogènes, mais également la présentation d'antigènes et la régulation de l'inflammation pour maintenir l'homéostasie tissulaire⁴². Une propriété centrale marque leur biologie : la plasticité phénotypique. Cette plasticité leur permet de modifier leurs fonctions selon les signaux présents dans leur microenvironnement^{41,43}. Ils peuvent ainsi adopter des états fonctionnels très différents selon le contexte tissulaire et tumoral, allant de phénotypes pro-inflammatoires associés à des fonctions antitumorales à des phénotypes immunosuppresseurs favorisant la progression tumorale^{43,44}.

Deux types de macrophages coexistent dans le poumon, ils viennent de deux lignées distinctes : les macrophages résidents (TRM) d'origine embryonnaire, qui maintiennent leur population par auto-renouvellement local, et les macrophages dérivés des monocytes (MDM) qui sont recrutés depuis la circulation sanguine par des signaux inflammatoires comme CCL2^{44,45}. Cette dualité d'origine est importante car elle s'accompagne d'une distribution temporelle et spatiale différente au cours du développement tumoral : les TRM sont les premiers à répondre et s'accumulent près des cellules tumorales dès le début de la lésion, alors que les tumeurs plus avancées sont surtout dominées par les MDM⁴⁵.

Dans un contexte tumoral, cette plasticité est détournée par les cellules cancéreuses et leur environnement. Les signaux du TME, comme les cytokines IL-10 et TGF- β , l'hypoxie ou le lactate, orientent progressivement les macrophages vers les phénotypes pro-tumoraux^{42,43}. Ce détournement des macrophages fait des TAMs des acteurs actifs de la croissance tumorale, du remodelage tissulaire et de l'évasion immunitaire dans le NSCLC⁴⁴.

1.4.2 TAMs, progression tumorale et résistance aux ICI dans le NSCLC

Dans le NSCLC, les macrophages représentent une population immunitaire majeure du TME⁴³. Cette abondance est généralement associée à un microenvironnement immunosuppresseur plutôt qu'à une réponse antitumorale efficace⁴¹. Les TAMs favorisent souvent la progression tumorale par plusieurs mécanismes. Ils sécrètent des facteurs pro-angiogéniques comme le VEGF, des cytokines immunosuppressives telles que IL-10 et le TGF- β , qui inhibent l'activité des lymphocytes T effecteurs, et participent aussi en remodelant la matrice extracellulaire⁴⁴.

Le compartiment macrophagique, lorsqu'il est abondant, est associé à une moins bonne évolution clinique, notamment quand ces cellules ont un profil immunosuppresseur qui exprime CD163⁴⁴. Une forte abondance de TAMs peut être corrélée à une infiltration réduite en lymphocytes T cytotoxiques. Cela suggère qu'ils participent à la limitation de la réponse immune dans la tumeur⁴³. Les macrophages ont aussi un rôle dans l'angiogenèse en plus du rôle dans l'invasion tumorale et dans la survie des cellules cancéreuses par l'axe IL-6/STAT3 qui stimule directement la prolifération des cellules malignes et renforce leur résistance aux thérapies⁴⁴. D'un autre côté, pour l'évasion immunitaire, ils contribuent à un environnement riche en Tregs et exprimant eux-mêmes des molécules inhibitrices comme PD-L1⁴². Ce qui montre qu'ils participent activement à la progression du NSCLC⁴⁴.

L'influence des TAMs est aujourd'hui reconnue comme un déterminant majeur dans la réponse aux immunothérapies par ICI dans le NSCLC⁴⁴. Une des explications est leur capacité à exprimer PD-L1 à leur surface de manière dominante, et souvent de façon indépendante des cellules tumorales, ce qui permet d'inhiber directement l'activité des lymphocytes T cytotoxiques⁴². C'est d'ailleurs une des raisons pour lesquelles le niveau de PD-L1 tumoral reste un biomarqueur imparfait pour prédire la réponse des patients⁴⁴. En particulier, le TGF- β favorise la fibrose et le dépôt d'une matrice extracellulaire dense : ce qui constitue une barrière physique et limite l'accès des cellules immunitaires au cœur de la tumeur^{43,44}.

En plus de cela, certaines sous-populations de TAMs sont directement associées à l'échec thérapeutique comme les macrophages *TREM2+* (*triggering receptor expressed on myeloid cells 2*) qui s'accumulent préférentiellement dans les niches hypoxiques où ils diminuent l'activité des cellules NK. Ils favorisent également l'expansion des lymphocytes T CD8+ épuisés, ce qui crée un microenvironnement immunologique immunosuppresseur associé à un mauvais pronostic. Des études comme celle de Molgora et al. et Zhang et al., confirment que l'infiltration massive des TAMs *TREM2+* est corrélée à une progression tumorale avancée et que leur ciblage ou leur déplétion permet de restaurer la sensibilité des tumeurs aux traitements anti-PD1⁴³.

1.4.3 Hétérogénéité des TAMs par le scRNA-seq

Le modèle classique M1/M2 pour les macrophages est utile pour donner un cadre général mais il ne décrit pas correctement la diversité réelle des macrophages intratumoraux^{41,43}. En effet, les TAMs présentent des états mixtes, dynamiques et très dépendants du contexte tumoral⁴¹. Les études récentes proposent de

considérer l'activation des macrophages non pas comme des états fixes mais plutôt comme un continuum d'états fonctionnels⁴³. Ce continuum d'états macrophagiques et son impact sur la progression tumorale sont illustrés en **Figure 4**.

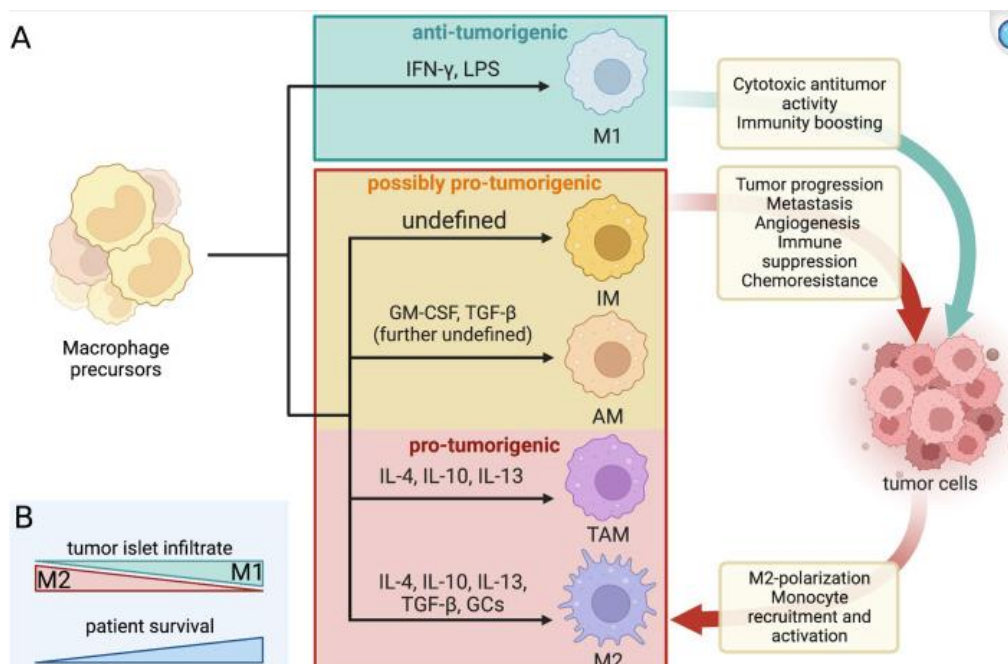


Figure 4. Représentation schématique du continuum d'activation des macrophages dans le microenvironnement tumoral.

Le modèle classique M1/M2 distingue des macrophages antitumoraux (proches du profil M1) et des macrophages pro-tumoraux (proches du profil M2). Cependant, les macrophages associés aux tumeurs présentent en réalité des états intermédiaires et mixtes, qui dépendent du contexte tumoral. Ces états peuvent influencer la progression tumorale, l'angiogenèse, l'immunosuppression et la résistance aux traitements. Figure tirée de Balážová et al.⁴¹.

Le séquençage ARN unicellulaire (scRNA-seq) a permis d'identifier plusieurs sous-populations macrophagiques dans le NSCLC, dont les macrophages *FCN1*⁺ (ficolin 1), *C1Q*⁺ (composant *c1q* du complément), *SPP1*⁺ (secreted phosphoprotein 1) et *TREM2*⁺^{43,45}. Les macrophages *FCN1*⁺ ont un profil transcriptomique proche des monocytes circulants et sont associés aux états précoces de recrutement inflammatoire⁴⁵. Les macrophages *C1Q*⁺ eux sont plus hétérogènes et peuvent exprimer des ligands inhibiteurs tels que PD-L1, PD-L2 ou HAVCR2 selon leur état d'activation. Cela leur confère alors un rôle d'immunosuppresseur important dans le blocage des lymphocytes^{43,44}. Ensuite, les macrophages *SPP1*⁺ sont, de leur côté, directement associés à des fonctions pro-fibrotiques et au remodelage de la matrice et à l'exclusion immunitaire des lymphocytes T. Pour finir, les macrophages *TREM2*⁺ s'accumulent surtout dans les niches hypoxiques et lipidiques, des zones du microenvironnement où se trouve une forte concentration de lipides ou un métabolisme des graisses altéré, où ils favorisent l'épuisement des lymphocytes T CD8⁺ et le freinage des cellules NK. Cela contribue à la résistance aux ICI⁴³.

Cette hétérogénéité moléculaire fine est invisible avec des techniques telles que le séquençage en vrac (*bulk RNA-seq*), qui est une technique qui moyenne le signal de toutes les cellules et masque les sous-populations spécialisées ou les microniches⁴³. Le *scRNA-seq* est donc une approche particulièrement adaptée pour étudier cette diversité macrophagique et comprendre leur rôle réel dans la progression tumorale⁴⁵. Cette technique de cartographie cellulaire a permis d'identifier des mécanismes moléculaires

qui régulent la plasticité des macrophages. Parmi ces mécanismes, le complexe mTORC2, dont RICTOR constitue le composant structural obligatoire, qui joue un rôle important dans la régulation métabolique et fonctionnelle des TAMs^{42,44}.

1.5 Outils bioinformatiques

1.5.1 Pourquoi le *single-cell*

Le *scRNA-seq* est une technologie qui est capable de mesurer le profil transcriptomique de chaque cellule individuellement. D'un point de vue technique, les cellules sont encapsulées dans des gouttelettes, chacune associée à un code-barre unique, ce qui permet de rattacher chaque ARN séquencé à sa cellule d'origine. Le résultat de cette technique est une matrice de comptage où chaque colonne représente une cellule et chaque ligne un gène, représentant ainsi l'état transcriptomique de chaque cellule analysée. Cette résolution unicellulaire vient cependant avec une profondeur de séquençage plus faible par cellule que le bulk RNA-seq, ce qui peut limiter la détection des transcrits faiblement exprimés et causer des dropouts.

Le *bulk RNA-seq*, à l'inverse, repose sur un calcul de la moyenne des signaux d'expression de l'ensemble des cellules d'un échantillon, ce qui rend difficile l'attribution d'un profil à un type cellulaire précis⁴⁶⁻⁴⁹. Dans le TME, cette approche masque la diversité des populations immunitaires, surtout pour les macrophages, qui constituent une population très diversifiée, avec des sous-populations opposées, comme des macrophages pro-inflammatoires et immunosuppresseurs décrites précédemment^{46,47,49-51}. Le signal bulk, qui correspond à une moyenne pondérée selon la proportion de chaque type cellulaire, ne permet pas de distinguer si une variation reflète un changement d'état transcriptionnel ou un changement de proportion entre population^{46,51}. C'est une information que seul le *scRNA-seq* peut apporter^{47-49,51}.

Cet aspect est important pour notre question. Identifier un programme transcriptomique associé à une condition expérimentale spécifique suppose que ce programme corresponde à un état cellulaire particulier, présent dans une sous-population du microenvironnement immunitaire, et non à une variation globale du tissu tumoral ou du compartiment immunitaire global. En bulk, un signal pareil serait dilué dans la moyenne du tissu et probablement non détecté^{46,50}. Dans le NSCLC, le *scRNA-seq* a d'ailleurs déjà permis d'identifier des sous-types cellulaires avec des profils distincts, notamment dans le compartiment macrophagique, comme les macrophages *FABP4*+ dominants dans le LUAD et les macrophages *SPP1*+ dominants dans le LUSC^{50,51}.

Cette résolution est aussi nécessaire pour évaluer le lien entre un état cellulaire du TME et la réponse clinique. Plus généralement, plusieurs travaux montrent que certaines sous-populations identifiées par *scRNA-seq* peuvent être associées à la réponse ou à la résistance aux immunothérapies^{47,49-51}. En bulk, cette association aurait nécessité des algorithmes de déconvolution (méthode pour estimer les proportions de chaque type cellulaire dans un échantillon mélangé) basés sur des connaissances préalables, ce qui introduit une incertitude supplémentaire^{46,47}.

L'objectif de ce travail est d'identifier, au sein des populations de cellules immunitaires une signature transcriptomique spécifique à la condition expérimentale où RICTOR est surexprimé et d'évaluer son association avec la réponse clinique dans les cohortes humaines de NSCLC. Cette question est surtout abordée par le scRNA-seq, plutôt que par l'analyse en bulk^{46,47,50,51}.

1.5.2 Problématique de projection inter-espèces

Nous avons vu que le scRNA-seq permet d'identifier des états cellulaires spécifiques dans le microenvironnement. L'utiliser dans un contexte translationnel, c'est-à-dire en voulant projeter une signature venant d'un modèle murin sur des données humaines, amène une problématique : les différences entre souris et humain⁵²⁻⁵⁵.

Pour le faire, une étape de *mapping* des orthologues est nécessaire. Il faut trouver pour chaque gène murin son équivalent dans le génome humain⁵⁴. Cette étape n'est pas simple, car les relations d'orthologues ne sont pas toujours claires. Un gène murin peut correspondre à plusieurs gènes humains (*relation one-to-many*) ou n'avoir aucun correspondant^{52,54}. Ce qui peut réduire la taille d'une signature que l'on peut projeter. Par exemple, les souris ont une famille de récepteurs NK (Ly49) absente chez les humains, qui eux ont à la place les récepteurs KIR^{52,54,55}.

La limitation principale se trouve ailleurs : la conservation de séquence entre deux gènes orthologues ne garantit pas leur conservation fonctionnelle^{52,54,56}. En effet, même si les souris et les humains partagent environ 80 % de leurs gènes, seuls 40 % ont des correspondances complètes au niveau fonctionnel⁵². Une mutation d'un gène présent chez les souris et les humains peut donner des effets différents selon l'espèce. Par exemple, une mutation *IL7R* bloque seulement le développement des cellules T chez les humains, mais chez les souris elle affecte les cellules T et B. De nombreux cas sont connus dans le système immunitaire, de nombreux écarts sont relevés entre espèces^{52,54,56}.

Ces limites ne rendent pas la projection impossible, mais elles demandent une interprétation prudente^{54,55}. De nombreuses découvertes murines ont déjà été projetées avec succès sur les données humaines de poumon. C'est le cas de la population de cellules endothéliales veineuses systémiques (SV EC) exprimant *Col15a1*. Elle a d'abord été découverte chez la souris par *scRNA-seq* après une lésion pulmonaire, cette population ectopique a été retrouvée de manière presque identique chez les patients atteints de fibrose pulmonaire idiopathique (FPI)⁵⁵. Un signal qui est retrouvé chez les humains est une piste biologique potentielle, mais cela ne fait pas d'elle une preuve que le mécanisme est entièrement conservé⁵³⁻⁵⁵.

L'existence d'une signature projetable ne suffit pas ; il est nécessaire de démontrer sa pertinence biologique et sa robustesse avant d'établir sa valeur clinique.

1.5.3 *Scoring* des signatures et robustesse clinique

La taille et la composition d'une signature jouent un rôle sur sa capacité à être retrouvée et comprise dans un nouveau contexte⁵⁷. Les signatures comportant un grand nombre de gènes, surtout lorsqu'elles contiennent plusieurs dizaines à plusieurs centaines de gènes peu coexprimés ou peu spécifiques, prennent plus de bruit biologique et technique, ce qui les rend moins stables entre cohortes indépendantes et plus difficiles à interpréter biologiquement⁵⁷⁻⁶³. A l'inverse, les signatures courtes, avec des gènes coexprimés de façon cohérente, sont plus robustes et plus faciles à transposer⁵⁷. Un exemple très utilisé est le PAM50, dans le cancer du sein. Il est composé de 50 gènes et ils suffisent à définir des sous-types moléculaires utiles en clinique⁵⁹. Un autre exemple, qui est encore plus court, est l'Oncotype DX fait de 21 gènes⁶⁰. Cette signature a montré une valeur prédictive dans les essais randomisés et guide aujourd'hui les décisions thérapeutiques en pratique courante⁶⁰. Dans le NSCLC, des signatures de moins de dix gènes ont montré leur capacité à prédire la survie et l'infiltration immunitaire indépendamment du stade TNM, avec de meilleurs résultats quand elles sont combinées aux données cliniques^{61,63}.

Ce besoin d'avoir une signature courte est important dans notre approche surtout quand on veut passer de la recherche à la clinique⁵⁹. Les méthodes de *scoring* qui dépendent d'un groupe de référence ne sont pas adaptées à un test diagnostique chez un seul patient, parce que le score d'un même patient peut changer si la composition du groupe d'analyse change^{59,62}. Les approches du type *single-sample scoring*, qui calculent un score séparément pour chaque échantillon, sont donc plus adaptées à la médecine de précision et à une éventuelle validation clinique⁶².

Dans notre projet, identifier les gènes les plus discriminants de la signature KPR Late, plutôt que de projeter une longue liste de gènes, s'inscrit exactement dans cette logique : produire un signal simple, robuste, compréhensible et potentiellement testable en clinique.

1.5.4 Contexte expérimental et données préliminaires

L'ensemble des éléments de cette introduction montre l'importance de mieux comprendre le rôle de RICTOR dans le microenvironnement immunitaire tumoral. Des données préliminaires du laboratoire de *Cancer Signaling* (GIGA) permettent d'aborder cette question.

Une étude en cours au laboratoire suggère aussi que RICTOR pourrait favoriser la progression tumorale grâce à des voies métaboliques en aval, surtout celles liées au métabolisme lipidique et aux sphingolipides. Cela renforce l'idée que RICTOR agit comme un accélérateur tumoral plutôt que comme un oncogène classique. Un élément cliniquement important fait surface dans les données récentes concernant le carcinome épidermoïde : l'amplification de *RICTOR* et l'expression de PD-L1 sont inversement corrélés. Des tumeurs PD-L1 négatives, peuvent avoir une amplification de RICTOR, ce qui suggère que certains d'entre eux, normalement non éligibles aux immunothérapies, pourraient malgré tout bénéficier d'un ciblage de mTORC2. Cet argument renforce l'idée que le statut PD-L1 seul est insuffisant pour classer finement les

patients NSCLC. L'ajout de *RICTOR* aux panels moléculaires pourrait améliorer la sélection thérapeutique

25

Au-delà de l'aspect thérapeutique, des études récentes en protéomique montrent que l'activation de la voie mTOR peut aussi correspondre à un état tumoral « immunologiquement froid ». Dans l'étude de Lehtio et al., un sous-type de NSCLC peu infiltré par les cellules immunitaires était lié à une altération de l'axe *STK11*/AMPK/mTOR. Cette altération peut favoriser l'activation de mTOR et participer à l'échappement immunitaire. Ainsi, RICTOR est impliqué dans une voie qui pourrait connecter la progression de la tumeur aux changements immunitaires dans le microenvironnement tumoral⁶⁴.

Pour tester cette hypothèse, notre laboratoire a développé un modèle murin NSCLC (notamment par le principal investigator Blomme Arnaud), détaillé dans le matériel et méthode. Une surexpression conditionnelle de *RICTOR* dans le modèle de référence KP (***Tg-LSL- KRAS^{G12D}-TP53^{lox/lox}***), activant la voie mTORC2 dans l'épithélium pulmonaire. Dans ce modèle, quand on surexprime RICTOR, les souris survivent moins bien et leurs tumeurs se développent plus vite. Cela rend le modèle tumoral plus agressif, ce qui correspond au mauvais pronostic observé chez les patients NSCLC avec une forte expression de RICTOR. L'analyse scRNA-seq du compartiment immunitaire permet d'analyser comment la surexpression de RICTOR module le microenvironnement immunitaire tumoral chez la souris, en analysant les différentes populations immunitaires et en recherchant notamment l'apparition d'un cluster ou d'une signature macrophagique spécifique du modèle avec la surexpression de RICTOR. Ce travail vise d'abord à trouver et caractériser cette signature dans les données des souris, puis à évaluer si elle est conservée dans les tumeurs humaines et associée, voire prédictive, de la réponse aux ICI chez les patients NSCLC. Tester cette hypothèse nécessite des approches bioinformatiques adaptées à la complexité et à la taille des données de séquençage unicellulaire.

L'ensemble de ces idées développées pose le cadre de ce travail. Les données de *scRNA-seq*, provenant d'une grande cohorte de patients atteints de NSCLC traités par immunothérapie néoadjuvante, permettent de tester une hypothèse : un programme transcriptomique dans un modèle murin de cancer du poumon agressif induit par l'activation de la voie mTORC2, est-il aussi présent dans les tumeurs humaines, et est-il lié à une résistance à l'immunothérapie ? Pour y répondre, nous devons projeter cette signature murine sur les données humaines, évaluer sa présence, et vérifier son association avec la réponse clinique. Tout cela doit se faire en tenant compte des limites propres à chaque étape de cette démarche.

2.OBJECTIFS ET HYPOTHESES

Le cancer du poumon non à petites cellules (NSCLC) représente l'une des principales causes de mortalité par cancer dans le monde. La surexpression de RICTOR, composante essentielle du complexe mTORC2, dans les cellules épithéliales pourrait impacter le microenvironnement immunitaire tumoral, bien que son rôle précis reste peu étudié dans ce contexte²⁴. Ce complexe est associé à une progression tumorale accélérée du cancer du poumon et pourrait représenter une cible pour améliorer la prise en charge des patients.

Nous émettons l'hypothèse que la surexpression de RICTOR dans les cellules épithéliales tumorales induit un remodelage du microenvironnement immunitaire tumoral, caractérisé par l'apparition d'états cellulaires spécifiques au stade Late. Une signature transcriptomique associée à ce contexte KPR Late pourrait être retrouvée dans des données humaines de NSCLC et être associée à la réponse au traitement néoadjuvant.

Dans ce projet, nous avons d'abord cherché à caractériser le remodelage immunitaire du microenvironnement tumoral dans des modèles génétiques murins de cancer du poumon surexprimant ou non la protéine RICTOR (KP et KPR). Pour ce faire, nous avons réalisé une analyse longitudinale par scRNA-seq sur les tumeurs pulmonaires issues des deux modèles à différents moments de la progression tumorale (Early et Late). Le premier objectif de ce mémoire était donc d'évaluer, au travers de ces analyses, l'influence de la surexpression de RICTOR, et donc de l'activation du complexe mTORC2, sur le microenvironnement immunitaire dans le NSCLC et d'ainsi identifier les populations immunitaires potentiellement altérées en présence de cette activation. Dans un second temps, nous avons cherché à identifier une signature transcriptomique immunitaire associée à l'activation de mTORC2 et à évaluer le potentiel de cette signature à prédire la réponse aux immunothérapies dans une cohorte de patients atteints de NSCLC.

2.1 Stratégie expérimentale

Pour répondre à ces objectifs, une stratégie en plusieurs étapes a été mise en œuvre. Tout d'abord, une recherche bibliographique a permis d'identifier les approches bioinformatiques standards pour l'analyse de données scRNA-seq et de sélectionner un jeu de données public de patients NSCLC traités par immunothérapie répondant à des critères de qualité et de pertinence biologique. Ensuite, les données scRNA-seq murines, générées au laboratoire, ont été analysées selon un *pipeline* comprenant le contrôle qualité des données, l'intégration des quatre conditions expérimentales, l'annotation des populations immunitaires et l'analyse différentielle par approche pseudo-bulk, afin d'identifier une signature transcriptomique associée à la condition KPR Late dans le compartiment macrophagique. Cette signature a enfin été projetée sur les données humaines sélectionnées, afin d'identifier des sous-populations immunitaires correspondantes et d'évaluer leur association avec le statut de réponse pathologique au traitement néoadjuvant.

3. Matériel et Méthodes

3.1 Source des données murines

3.1.1 Données murines : modèles et *design expérimental*

Ce mémoire repose sur l'analyse bioinformatique de matrices de comptage issues scRNA-seq, générées au laboratoire en collaboration avec le Lab of Dendritic Cell Biology and Cancer Immunotherapy (Pr. Van Ginderachter and Pr. Laoui) de la Vrije Universiteit Brussel.

Les données murines ont été générées en amont du projet à partir de seize souris réparties en quatre groupes expérimentaux de quatre souris selon leur génotype et leur stade tumoral : KP Early, KP Late, KPR Early, KPR Late (n=4 par groupe). Les groupes Early (KP Early et KPR Early) sont composés de deux femelles et deux mâles, tandis que les groupes Late (KP Late et KPR Late) comprennent trois mâles et une femelle.

Le modèle KP (***Tg-LSL- KRAS^{G12D}-TRP53^{lox/lox}***) est un modèle murin de cancer du poumon non à petites cellules qui associe l'activation conditionnelle de l'oncogène *KRAS^{G12D}* (Kirsten rat sarcoma viral oncogene homolog) et l'inactivation biallélique du gène *TRP53* (Transformation related protein 53)^{65,66}. La mutation G12D correspond au remplacement de la glycine par un acide aspartique en position 12, entraînant une activation constitutive de *KRAS*. L'inactivation de *Trp53* repose sur la floxation des deux allèles^{65,66}. L'expression de la recombinaise Cre, obtenue par injection intratrachéale d'un vecteur adéno-associé AAV6.2FF-Cre (sérotypage à tropisme spécifique pour l'épithélium respiratoire pulmonaire) et sous le contrôle du promoteur Surfactant Protein B (SP-B), permet simultanément l'excision de la cassette LSL (Lox-STOP-Lox) qui active l'oncogène *KRAS*, et la suppression des deux allèles *TRP53^{lox/lox}*, spécifiquement dans les cellules épithéliales alvéolaires du poumon⁶⁵⁻⁶⁷. Cette double modification induit le développement de tumeurs pulmonaires agressives^{65,66}.

Le modèle KPR (***Tg-LSL-KRAS^{G12D}-TRP53^{lox/lox}-RICTOR^{OE}***), combine les modifications du modèle KP avec la surexpression conditionnelle de RICTOR dans l'épithélium pulmonaire^{68,69}. Lors de l'expression de la recombinaise Cre, la cassette LSL est excisée du transgène (*Tg-LSL-FLAG-RICTOR-IRES-EGFP-luciférase*), permettant l'expression de FLAG-RICTOR ainsi que des gènes rapporteurs (GFP et luciférase, co-exprimés via le site IRES) pour l'identification et le suivi des cellules transgéniques⁶⁶. Ce modèle entraîne l'activation constitutive de la voie mTORC2, qui accélère la progression tumorale⁶⁹.

Les stades Early et Late correspondent respectivement à 1 mois et 3 mois post-infection par le vecteur AAV-Cre. Les étapes expérimentales de prélèvement tissulaire, de dissociation cellulaire, d'enrichissement des cellules immunitaires et de génération des données biologiques n'ont pas été réalisées dans le cadre de ce mémoire. L'analyse présentée repose sur l'exploitation des données déjà produites.

Antérieurement au séquençage, un enrichissement a été réalisé par la sélection positive sur billes magnétiques couplées à l'anticorps anti-CD45 (CD45 MicroBeads mouse, Miltenyi Biotec, réf. 130-052-

301). Ainsi, seules les cellules immunitaires ont été conservées pour l’encapsulation sur puce 10x Genomics. De plus, les cellules d’un même groupe de souris ont été *poolées* dans un tube unique avant séquençage, sans marquage hashtag oligonucléotides (HTO), ce qui empêche la reconnaissance des souris individuelles au sein d’un groupe⁷⁰.

Les échantillons Early (KP Early et KPR Early) ont été séquencés en chimie 5’ (*gene capture 5’*), tandis que les échantillons Late l’ont été en chimie 3’ (*gene capture 3’*). Cette différence de protocole introduit un effet de lot technique lié à la capture différentielle des extrémités des transcrits : la technique 5’ privilégie les ARNm avec cap 5’ intact, tandis que la 3’ privilégie les queues poly(A). Ce biais a été corrigé lors de l’intégration des données (section 3.6). Les données brutes ont été alignées sur le génome de référence murin, aboutissant à la production d’une matrice de comptage par échantillon, qui constitue le point de départ de notre analyse bioinformatique^{70,71}.

3.1.2 Données humaines

Les données humaines proviennent de l’étude de Liu et al. (2025, Cell, GSE243013), qui représente l’une des plus larges cohortes de scRNA-seq disponibles pour le NSCLC traité par immunothérapie anti-PD-1, avec 243 patients séquencés après résection chirurgicale post-traitement néoadjuvant sous chimiothérapie combinée à un anti-PD-1⁷². Ce jeu de données a été sélectionné sur la base des critères suivants : publication dans un journal à fort impact (*Cell*), grande cohorte de patients, patients traités à l’immunothérapie anti-PD-1 (pembrolizumab, nivolumab, sintilimab, camrelizumab ou tislélizumab), données de NSCLC (180 patients avec un carcinome épidermoïde pulmonaire (LUSC) et 63 patients avec un adénocarcinome pulmonaire) analysées par scRNA-seq, accès aux données brutes et aux métadonnées cliniques incluant la réponse pathologique (pCR, MPR et non-MPR), nécessaire pour évaluer l’association entre le profil transcriptomique et la réponse au traitement néoadjuvant.

3.2 Logiciels et *packages*

L’ensemble de l’analyse bioinformatique a été réalisé sous R version 4.4.2. Les analyses ont été exécutées sur le cluster HPC du GIGA (Université de Liège). Les principaux *packages* utilisés sont détaillés dans le **Tableau 2** ci-dessous. Pour faciliter la compréhension de la stratégie analytique, une vue d’ensemble du *pipeline* bioinformatique est présentée en **Figure 5**.

Tableau 2. *Packages et logiciels utilisés dans le cadre de l’analyse bioinformatique.*

Package / Logiciel	Version
R	R version 4.4.2 (2024-10-31 ucrt)
Seurat ⁷³	5.3.0
SCTransform ⁷⁴ / glmGamPoi ⁷⁵	0.4.2/1.20.0
scDblFinder ⁷⁶	1.22.0
SingleR ⁷⁷	2.10.0
cellIdex ⁷⁷	1.18.0
DESeq2 ⁷⁸	1.48.1
ggplot2 ⁷⁹	4.0.0

dplyr ⁸⁰	1.1.4
pheatmap ⁸¹	1.0.13
patchwork ⁸²	1.3.2
Matrix	1.7.4
SingleCellExperiment ⁸³	1.30.1
SeuratDisk ⁸⁴	0.0.0.9021
UCell ⁸⁵	2.1
Limma ⁸⁶ /edger ⁸⁷	3.62.1/4.4.0
homologene ⁸⁸	1.4.68.19.3.27

3.3 Contrôle qualité et filtrage cellulaire

Les quatre échantillons ont été analysés individuellement au préalable afin de vérifier la qualité des données et de définir les paramètres de filtrage avant l'analyse intégrée. Pour chaque échantillon, la matrice de comptage a été importée via la fonction `Read10X()` et un objet Seurat distinct a été créé, en appliquant un filtre minimal de trois cellules par gène et de 200 gènes par cellule, permettant d'éliminer les gènes quasi-absents et les profils cellulaires de très faible qualité susceptibles de correspondre à des gouttelettes vides, des débris cellulaires ou des captures inefficaces. Ces seuils sont des paramètres empiriques couramment utilisés dans les *workflows standards* de Seurat, et leur adéquation au jeu de données a été vérifiée à partir de visualisations sur les échantillons individuels.

La qualité cellulaire a été évaluée à partir de deux métriques : le nombre de gènes détectés par cellule (*nFeature_RNA*) et le pourcentage de transcrits mitochondriaux (*percent.mt*), calculé à partir des gènes portant le préfixe *mt-*. Pour choisir les seuils de filtrage, la distribution de ces métriques a été visualisée sous forme de *violin plots* et de *scatter plots* pour chaque échantillon avant et après filtrage⁸⁹.

Les seuils retenus sont les suivants : nombre de gènes détectés compris entre 200 et 6 000 gènes, et pourcentage de transcrits mitochondriaux inférieur à 10 %. Ces seuils ont été définis sur la base de l'inspection visuelle des distributions par *violin plot* pour chaque échantillon. Ces valeurs sont cohérentes avec les pratiques courantes en scRNA-seq. Le seuil inférieur sur *nFeature_RNA* vise à exclure les débris cellulaires et les profils de faible qualité, tandis que le seuil supérieur permet d'écarter de potentiels doublets, qui sont des profils transcriptomiques mixtes résultant de la capture de plusieurs cellules dans une même gouttelette, et constitue une étape complémentaire à la détection algorithmique des doublets (section 3.4). Le seuil mitochondrial de 10 % permet d'écarter les cellules endommagées ou en cours d'apoptose^{89,90}.

Pipeline bioinformatique :
De l'identification d'une signature macrophagique murine à son évaluation dans une cohorte humaine de NSCLC

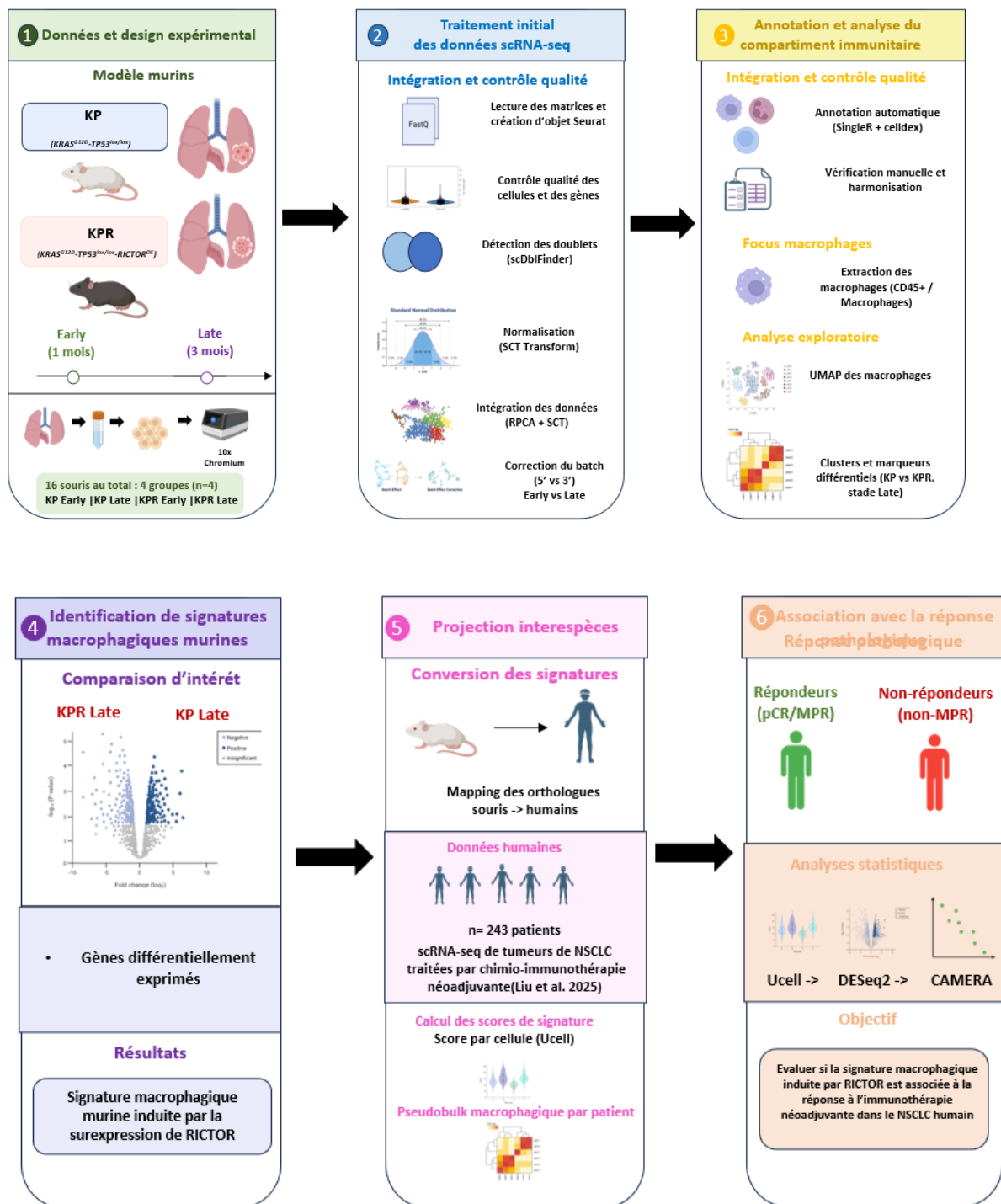


Figure 5. Vue d'ensemble du pipeline bioinformatique utilisé pour identifier et évaluer une signature macrophagique murine dans une cohorte humaine de NSCLC.

Le pipeline comprend six grandes étapes : le *design expérimental* murin KP/KPR, le contrôle qualité et l'intégration des données scRNA-seq murines, l'annotation cellulaire et l'analyse du compartiment macrophagique, l'identification de signatures macrophagiques candidates associées au contexte KPR Late, la conversion des gènes murins en orthologues humains, puis la projection et l'évaluation de ces signatures dans une cohorte humaine de NSCLC selon le statut de réponse pathologique.

3.4 Détection et exclusion des doublets

La détection des doublets a été réalisée sur chaque échantillon individuellement, après filtrage et avant intégration, à l'aide du *package* *scDblFinder*. Chaque objet Seurat a été converti au format *SingleCellExperiment*, ce format est attendu par *scDblFinder*. *scDblFinder* a été appliqué séparément à chaque échantillon afin de tenir compte des spécificités et d'éviter que les doublets ne biaisent l'annotation cellulaire et l'analyse différentielle en aval. Ce *package* a été retenu pour sa robustesse et sa faible dépendance aux paramètres.

3.5 Normalisation

La normalisation a été effectuée individuellement pour chaque échantillon à l'aide de *SCTransform* v2, implémentée dans Seurat. Contrairement à la méthode *LogNormalize*, *SCTransform* repose sur un modèle statistique qui corrige mieux le bruit technique et les variations liées à la fois à la profondeur de séquençage, tout en préservant la variabilité biologique d'intérêt.

La régression sur le pourcentage de transcrits mitochondriaux (*percent.mt*) a été appliquée afin de réduire l'influence de cet effet technique sur les analyses en aval. Les 3 000 gènes les plus variables ont été sélectionnés pour chaque échantillon, conformément aux paramètres par défaut de *SCTransform*. Toutefois, une normalisation de type *LogNormalize*, suivie d'une mise à l'échelle des données (*ScaleData*), a également été réalisée pour les représentations en *heatmaps*, qui requièrent des données mises à l'échelle.

3.6 Intégration des données murines

L'intégration des données murines a été réalisée avec la méthode *Reciprocal PCA (RPCA)* disponible dans Seurat v5. Cette méthode repose sur l'identification de 3 000 gènes hautement variables par la fonction *SelectIntegrationFeatures()*, et nécessite une analyse en composantes principales (PCA) réalisée séparément sur chaque objet avant l'intégration.

Une fois cette étape réalisée, les ancres d'intégration ont été identifiées à l'aide de la fonction *FindIntegrationAnchors()* en utilisant les 50 premières composantes principales (PCs). Ces ancres sont des paires de cellules issues de différents échantillons ayant des profils d'expression similaires. Elles servent de points de correspondance entre les jeux de données. Ce seuil de 50 PCs a été fixé de manière empirique, comme recommandé par les vignettes Seurat, qui recommandent un nombre plus élevé pour ne pas exclure le signal biologique potentiellement informatif. L'intégration a finalement été réalisée via *IntegrateData()*, qui aligne les données dans un espace commun en corrigeant les effets de lot⁹¹.

RPCA a été privilégiée pour conserver le signal transcriptomique propre à la condition KPR Late, tout en corrigeant l'effet de lot lié à la différence de capture chimique (3' versus 5') entre les échantillons Early et

Late⁹². Une seconde méthode d'intégration a été utilisée pour l'analyse focalisée sur le compartiment macrophagique, car cette approche était plus adaptée dans notre étude (section 3.10.3)⁹³.

3.7 Réduction de dimension et clustering

Après l'intégration, une PCA a été appliquée sur *l'assay* intégré afin d'obtenir un espace réduit corrigé des effets de lot pour les analyses en aval. Pour les mêmes raisons que celles exposées en section 3.6, les 50 premières composantes principales ont été retenues. Ensuite, un graphe de voisinage a été construit avec la fonction `FindNeighbors()` dans l'espace PCA.

Le clustering a été réalisé avec `FindClusters()`, avec une résolution fixée à 1.4, ce qui a permis l'identification de 33 clusters. Ce paramètre de résolution a été retenu après inspection visuelle des UMAP à différentes résolutions, aboutissant à une granularité suffisante pour distinguer les sous-populations immunitaires sans fragmentation excessive des clusters. La visualisation a été réalisée par UMAP (*Uniform Manifold Approximation and Projection*)⁹⁴. Afin d'assurer la reproductibilité des analyses, une graine aléatoire (*seed*= 1234) a été utilisée.

3.8 Annotation cellulaire

Une étape essentielle de l'analyse consiste en l'annotation cellulaire, réalisée de manière automatisée à l'aide du *package* *SingleR*. La base de référence utilisée est *l'Immunological Genome Project (ImmGen)*, issue du *package* *celldex*. Cet ensemble de données bulk-RNA-seq a été constitué à partir de populations immunitaires murines triées^{77,95}. Cette référence a été choisie car elle est adaptée aux cellules CD45+, comme les données de ce projet.

L'annotation a été produite en deux étapes hiérarchiques. Dans un premier temps, une annotation large (*main labels*) a été effectuée sur l'ensemble des cellules, en les comparant aux neuf grandes familles immunitaires de référence : lymphocytes T, lymphocytes B, cellules NK, NKT, monocytes, macrophages, DC, neutrophiles et cellules lymphoïdes innées (ILC). Dans un second temps, une annotation fine (*fine labels*) a été réalisée indépendamment pour chaque famille afin d'identifier des sous-types cellulaires plus spécifiques.

Enfin, ces annotations fines ont été regroupées manuellement en une vingtaine de labels biologiques interprétables (Label20), à l'aide d'une fonction de *mapping* personnalisée, permettant d'assigner chaque combinaison de labels (main et fine) à une catégorie simplifiée. De plus, la cohérence biologique de l'annotation a été vérifiée par l'analyse d'expression de marqueurs connus pour chaque famille.

Après annotation, une analyse de la composition cellulaire a été réalisée en exploitant la distribution des types cellulaires (Label20) selon la condition (KP vs KPR), le stade tumoral (Early vs Late) et le groupe expérimental, avec une attention particulière portée aux macrophages, qui constituent la principale population immunitaire dans ces échantillons et le principal intérêt biologique de ce projet.

3.9 Analyse différentielle globale

3.9.1 Approche pseudo-bulk (DESeq2)

L'analyse différentielle (DE) a été réalisée selon une approche pseudo-bulk avec DESeq2, recommandée pour les données scRNA-seq car elle agrège les cellules en pseudo-échantillons avant l'analyse statistique, évitant ainsi la pseudo-réplication liée à l'analyse cellule par cellule^{96,97}.

Pour rappel, comme indiqué en section 3.1.1, l'absence de marquage HTO empêche de reconstruire de vrais réplicats biologiques. Des pseudo-réplicats ont donc été générés : pour chaque combinaison type cellulaire (Label20) x groupe expérimental (*exp_group*), les cellules ont été réparties aléatoirement en trois groupes de taille aussi équilibrée que possible^{96,97}. Les combinaisons comportant moins de trois cellules ont été exclues car elles ne permettent pas de générer des réplicats fiables. Nous avons ensuite agrégé les *counts* bruts de chaque pseudo-réplicat par sommation via *AggregateExpression()*, produisant une matrice pseudo-bulk. Chaque pseudo-bulk a reçu un identifiant unique appelé *pseudobulk_id* suivant un format de type : T_CD4_KPR_Late_rep2, assurant une traçabilité complète.

Un filtrage qualité a ensuite été réalisé : les gènes ayant moins de 10 *counts* totaux ont été exclus pour éviter de tester des gènes quasi absents et stabiliser les statistiques de DESeq2, seuil couramment utilisé dans les workflows Bioconductor. Pour l'analyse, chaque type cellulaire devait avoir au moins 6 pseudo-bulk *samples* et au moins 2 groupes expérimentaux distincts. Pour chaque type cellulaire, un objet *DESeqDataSet* a été construit à partir de la matrice de *counts* et des métadonnées associées (condition, *timepoint*, groupe expérimental), avec un *design* basé sur *exp_group*.

Cinq comparaisons ont été effectuées lorsque les effectifs le permettaient :

- KPR Late vs KP Late, afin d'évaluer l'effet KPR versus KP au stade Late.
- KPR Late vs KPR Early, afin d'étudier l'évolution temporelle dans le modèle KPR.
- KP Late vs KP Early, afin d'étudier l'évolution temporelle dans le modèle KP.
- KPR Early vs KP Early, afin d'évaluer l'effet KPR versus KP au stade Early
- KPR Late vs KP Early, pour évaluer l'effet combiné du génotype KPR et du stade tumoral

Les gènes ont été considérés comme différentiellement exprimés lorsqu'ils avaient une p-valeur ajustée < 0.05 et $|\log_2FC| > 1$, qui correspond à un changement d'expression d'un facteur 2 au minimum entre deux conditions.

3.9.2 Approche FindMarkers Wilcoxon

La seconde approche appliquée est *FindMarkers()* avec un test de Wilcoxon non paramétrique. Six comparaisons ont été réalisées pour chaque type cellulaire : KPR Early vs KP Early, KPR Late vs KP Late, KP Late vs KP Early, KPR Late vs KPR Early, KPR vs KP (tous stades confondus) et Early vs Late (toutes conditions confondues), utilisée en complément du pseudo-bulk. Un filtre de 10 cellules minimum par groupe a été

appliqué. Les gènes différentiellement exprimés (DEG) ont été définis en deux catégories : DEG larges (p-valeur ajustée < 0.05) et DEG robustes (p-valeur ajustée < 0.05 avec $|\log_2FC| > 1$).

Cette approche a été utilisée en complément du pseudo-bulk. Le pseudo-bulk est statistiquement plus robuste puisqu'il tient compte de la structure réelle de l'expérience, notamment la variabilité inter-souris, tandis que FindMarkers apporte une approche plus descriptive à l'échelle cellulaire, puisqu'il regarde cellule par cellule. C'est pourquoi les résultats de FindMarkers ont été utilisés comme un complément d'analyse pseudo-bulk.⁹⁶

3.10 Analyses focalisées sur le compartiment macrophagique

L'analyse globale des cellules immunitaires décrite ci-dessus a permis d'identifier les principales populations du microenvironnement tumoral, nous avons pu observer que les macrophages sont nombreux dans ces *datasets*. Cependant, une analyse à large échelle peut potentiellement limiter la résolution nécessaire pour détecter des différences transcriptionnelles fines chez les macrophages. Par ailleurs, Pauline Bardet avait déjà identifié un ensemble de gènes lors d'une première analyse réalisée par la VUB. L'objectif était également de pouvoir retrouver ces gènes.

3.10.1 Extraction et construction du sous-objet macrophagique

Pour l'analyse centrée sur le compartiment des macrophages, les cellules macrophagiques ont été extraites de l'objet global afin de construire un sous-objet dédié permettant une analyse plus fine. Toutes les cellules portant une annotation commençant par « Macrophages_ » ont été sélectionnées sur base de l'annotation Label20. Pour valider l'extraction, des vérifications sur le nombre de cellules extraites ont été réalisées. Ce sous-objet a ensuite été réintégré et reclusterisé, comme décrit dans la section suivante.

3.10.2 Intégration par Harmony et reclustering macrophagique

L'approche avec Harmony a été retenue car, contrairement à RPCA, elle ne repose pas sur l'identification d'ancres entre échantillons, ce qui la rend plus adaptée aux sous-ensembles déséquilibrés. Cette différence est importante dans notre sous-objet macrophagique, où certains sous-états peuvent être absents ou très peu représentés dans certaines conditions. Dans ce contexte, RPCA peut avoir des difficultés à identifier des ancres fiables pour des populations rares, ce qui est le cas de notre *set* de données. L'intégration appliquée cette fois au jeu complet plutôt qu'au seul sous-objet macrophagique a été réalisée par Harmony. Sur ce sous-objet, la chimie de capture et le stade tumoral sont confondus, donc la correction de l'effet de lot seule ne suffit pas. L'intégration globale qui contient un grand nombre de cellules et une grande diversité cellulaire, permet une estimation plus robuste de la composante. Après la normalisation avec SCTransform (section 3.5), les quatre objets individuels ont été fusionnés en un objet unique à l'aide de la fonction *merge()* de Seurat. Les 3000 gènes les plus variables ont été sélectionnés avec *SelectIntegrationFeatures()*, en excluant les gènes mitochondriaux (préfixes -mt), ribosomiaux et immunoglobuliniques, ce qui permet de limiter le bruit dû aux signaux techniques dans l'espace de

réduction de dimension. Une PCA a ensuite été calculée sur 50 composantes principales. Harmony a ensuite été appliqué sur les 50 premières composantes principales avec la fonction `RunHarmony()`, en corrigeant la variable de technique de capture (`group.by.vars= « technique »`). Ce choix est justifié car il correspond au principal effet de lot de *design expérimental* : les deux échantillons Early ont été séquencés en chimie 5' alors que les deux échantillons Late l'ont été en chimie 3'.

3.10.2.1 Méthode et choix du paramètre θ

Le paramètre θ , qui contrôle l'intensité de correction dans Harmony, a été testé pour plusieurs valeurs de 1 à 5. Les valeurs $\theta = 3, 4$ et 5 ont été écartées après inspection visuelle des UMAP. Pour départager $\theta = 1$ et $\theta = 2$, le mélange des lots dans l'embedding Harmony a été évalué quantitativement selon les variables échantillons, techniques, condition et *timepoint*. La valeur $\theta = 2$ a été retenue puisqu'elle offrait un meilleur compromis entre correction de l'effet technique et conservation de la structure biologique.

Le graphe de voisinage a ensuite été construit avec `FindNeighbors()` sur les 50 dimensions Harmony, avec $k=20$, une valeur adaptée aux jeux de données de taille moyenne à grande. Le clustering a été réalisé à une résolution de 1.4 sur l'objet global, choisie après avoir testé plusieurs résolutions allant de 0.3 à 1.5. Enfin, une représentation UMAP a été générée. L'annotation cellulaire a été réalisée selon la procédure décrite en section 3.8.

Les 32 610 cellules macrophagiques ont ensuite été extraites de l'objet intégré. Le recours à l'intégration globale plutôt qu'au seul sous-objet macrophagique est justifié empiriquement : sur ce sous-ensemble, la chimie de capture et le stade tumoral sont confondus et Harmony parvient mal à les dissocier, rendant la correction de l'effet de lot insuffisante (l'évaluation du mélange des lots décrite ci-dessus le montre). Le graphe de voisinage a été reconstruit avec `FindNeighbors()` (dimension 1:50, $k=20$) et clustering a été fait à la résolution de 0.5, retenue après test de plusieurs résolutions, donnant des clusters interprétables sans sur-fragmenter.

3.10.2.2 Analyse de composition et expression différentielle

Pour chaque cluster obtenu à la résolution 0.5 sur le sous-objet macrophagique extrait de l'intégration globale Harmony, une table de composition conditionnelle a été calculée, détaillant la proportion de cellules appartenant à chaque condition, à chaque échantillon, à chaque stade tumoral et à chaque technique de capture. Cette étape permet d'orienter l'analyse différentielle vers les clusters présentant un enrichissement marqué en cellules KPR Late et faible en KP Early.

Deux analyses différentielles ont été réalisées. La première vise à identifier, pour chaque cluster, les gènes significativement surexprimés par rapport au reste des cellules du sous-objet, à l'aide de `FindMarkers()` (Wilcoxon). Les 10 meilleurs marqueurs de chaque cluster, classés par $\log_2FC$ décroissant, ont été extraits et visualisés sous forme de *DotPlot* et *heatmap*.

La seconde approche compare les conditions expérimentales au sein des clusters d'intérêt. Les comparaisons incluaient notamment KPR_Late vs KP_Late, KPR_Late vs KPR_Early, KPR vs KP et KPR_Late vs KP_Early. L'analyse a été faite avec FindMarkers (Wilcoxon, min.pct= 0.1, logfc.threshold = 0). Cette approche avait pour but d'identifier des gènes robustes associés au profil expérimental recherché, plutôt qu'à une comparaison unique.

Enfin, un score de module a été calculé à l'aide de AddModuleScore() pour évaluer la capacité discriminante des gènes candidats entre les quatre conditions. Idéalement, une bonne signature devrait montrer un gradient maximal en KPR Late. Toutefois, une signature nettement plus élevée en KPR Late, même sans un gradient parfait, a aussi été considérée comme un bon candidat biologique.

3.10.3 Construction des signatures murines candidates

Deux clusters ont été retenus pour extraire une signature murine, sur base de leur composition conditionnelle. Le cluster 2, qui est le plus enrichi en cellules KPR et le cluster 0 avait une faible proportion de cellules KP Early, ce qui est cohérent avec l'hypothèse d'un modèle KPR accéléré. Pour chaque cluster, les gènes candidats ont été définis par l'intersection des gènes significatifs dans quatre comparaisons : KPR Late vs KP Late, KPR Late vs KPR Early, KPR vs KP et KPR Late vs KP Early (critères : $p_{adj} < 0.05$, $\log_2FC > 0.5$ et $pct.1 > 0.05$). Cette stratégie d'intersection permet de retenir les gènes robustes malgré les déséquilibres d'effectifs entre groupes^{40,98,99}. Cette analyse a conduit à deux signatures potentielles : la signature du cluster 0, composée de 24 gènes murins, et la signature du cluster 2, composée de 11 gènes murins.

Une troisième signature de quatre gènes, proposée par Elodie Renaude, a été gardée en se concentrant sur les clusters 0 et 2, identifiés comme les clusters les plus enrichis en KPR Late. Les marqueurs différentiels de ce cluster ont été identifiés par FindAllMarkers(), les gènes retenus sont ceux répondant aux critères suivants : p-valeur ajustée < 0.05 et un $\log_2FC > 0.5$, et triés selon un score combinant la significativité statistique et l'amplitude de l'effet ($-\log_{10}(p_{adj}) \times \log_2FC$), une approche courante en scRNA-seq.

3.10.4 Conversion des gènes murins en orthologues humains

Pour projeter la signature murine sur les données, chaque gène murin a été converti en son orthologue humain en deux étapes. Le *package* R homologene (v1.4.68.19.3.27, basé sur NCBI HomoloGene) a été utilisé en priorité avec les paramètres inTax = 10090 (souris) et outTax = 9606 (humain). Les gènes absents de HomoloGene ont été recherchés manuellement dans NCBI Gene (section « Orthologs »)¹⁰⁰. Cette approche à deux niveaux permet de conserver le plus d'orthologues identifiés possible.

Parmi les gènes murins des trois signatures, 31 orthologues humains ont été retenus avec deux orthologues humains pour C4b (C4A et C4B). Trois gènes ont été exclus, comme résumé dans le **Tableau 3** : *Zfp984*, absent de HomoloGene et de NCBI Gene chez *Homo sapiens*; *Tcaf1*, sans orthologue répertorié

et *Chil3* ayant une relation de type *one to many* avec les chitinases humaines (*CHIT1*, *CHIA*) selon OrthoDB^{101–104}. Ces gènes n’ont pas été conservés car ils n’ont pas de correspondance fonctionnelle directe. *ARG1* a été conservé dans la liste mais présentait une expression quasi nulle dans les pseudobulks macrophagiques humains. *APOBEC1* quant à lui possède bien un orthologue humain, mais il était absent de la matrice finale et n’a donc pas contribué au score. Il n’a pas été gardé pour l’analyse, nous avons donc 22 gènes pour le cluster 0 et 10 pour le cluster 2.

Tableau 3. Gènes humains pour la signature

Signature	Gènes murins	Gènes humains effectifs	Gènes humains
Cluster 0	24	22	<i>C4A</i> , <i>C4B</i> , <i>CD24</i> , <i>CD244</i> , <i>CLEC12A</i> , <i>CYP4A22</i> , <i>DHCR24</i> , <i>EMP1</i> , <i>FFAR2</i> , <i>HS3ST3B1</i> , <i>MFHAS1</i> , <i>MMP8</i> , <i>OSBPL1A</i> , <i>PHF20</i> , <i>PLA2G7</i> , <i>PLS3</i> , <i>PTPN7</i> , <i>SLC25A24</i> , <i>SPEG</i> , <i>SYNGR1</i> , <i>TMX4</i> , <i>TRIB1</i>
Cluster 2	11	10	<i>ADAM15</i> , <i>BASP1</i> , <i>CLEC12A</i> , <i>DUSP1</i> , <i>HCAR2</i> , <i>MMP8</i> , <i>PTPN18</i> , <i>RGS1</i> , <i>SYNGR1</i> , <i>TRIB1</i>
Signature Elodie	4	3 (<i>ARG1</i> conservé dans la table mais quasi-absent dans les données humaines)	<i>ARG1</i> , <i>FBP1</i> , <i>NPL</i>

3.11 Prétraitements des données humaines

3.11.1 Chargement et contrôle qualité

Les données humaines ont été chargées depuis une matrice MTX compressée (1 254 749 cellules x 31 831 gènes), transposée au format gènes x cellules attendu par Seurat. Un objet Seurat a été créé avec les mêmes paramètres minimaux que pour les données murines (min.cells = 3, min.features = 200). Le pourcentage de transcrits mitochondriaux a été calculé à partir du préfixe « MT- ». Les métadonnées cliniques (23 colonnes, incluant *sampleID*, *cancer_type*, statut de réponse pathologique, données démographiques et annotations cellulaires *sub_cell_type*) ont été alignées sur l’objet via la colonne *cellID*, avec une correspondance de 100 %. Un filtrage qualité a ensuite été appliqué : 200 < nFeature_RNA < 6 000 et percent.mt < 10 %. Les doublets n’ont pas été recalculés car Liu et al. avaient appliqué Scrublet en amont (taux de doublet attendu = 0.08, seuil de score 0.25 par séquençage), et les données mises à disposition étaient donc déjà filtrées⁷².

3.11.2 Stratégie de traitement par lots

En raison de la taille du *dataset* (> 1,2 million de cellules), les 243 patients ont été répartis en 24 lots (23 x 10 patients + 1 x 13 patients) via la fonction *sample()* avec une graine fixe pour la reproductibilité. Une vérification globale a été effectuée après traitement pour confirmer l'absence de doublons et le remplissage des fichiers produits.

3.11.3 Agrégation pseudo-bulk

Les cellules de chaque lot ont été agrégées en pseudo-bulk selon le schéma *sampleID x sub_cell_type*, où le *sub_cell_type* correspond aux 51 sous-types immunitaires annotés par Liu et al. Cette annotation a été réutilisée sans réannotation indépendante afin de garder la résolution fine validée par les auteurs et éviter d'ajouter de variabilité supplémentaire. Les *counts* bruts ont été extraits via *GetAssayData* (*assay* RNA, *layer counts*) et agrégés par sommation. Le seuil minimal de 10 cellules par pseudo-bulk a été appliqué, conformément aux recommandations de Crowell et al., aboutissant à 9 320 pseudobulks¹⁰⁵.

3.11.4 Classification des répondeurs

Treize patients présentaient des annotations en caractères chinois dans les colonnes *radiological_response*, *chemotherapy* et *pathological_response_rate*. Ces valeurs ont été traduites avec l'aide d'An Ning avant la construction des groupes. Une variable binaire de réponse a été construite à partir de la réponse pathologique : les patients pCR et MPR (tumeur résiduelle viable $\leq 10\%$) ont été classés répondeurs (R) et les patients non-MPR comme non-répondeurs (NR), conformément à la stratification de Liu et al. Un patient au statut inconnu a été exclu, aboutissant à 242 patients analysables (R = 130, NR = 112), avec 9 292 pseudobulks restant après exclusion du patient.

3.12 Projection et *scoring* des signatures sur les données humaines

Pour mesurer l'activité des signatures identifiées, trois méthodes ont été utilisées : UCell⁸⁵, qui repose sur le rang des gènes pour le *scoring* des signatures, DESeq2⁷⁸ pour l'analyse différentielle des gènes individuels et camera¹⁰⁶ pour tester l'enrichissement global des signatures. Elles ont été choisies parce qu'elles ne répondent pas au même objectif et ne reposent pas sur les mêmes principes.

3.12.1 UCell

UCell est une technique de *scoring* qui repose sur le rang des gènes dans chaque pseudobulk et non sur leurs valeurs d'expression brutes ou normalisées. Elle est adaptée aux petites signatures et robuste aux différences d'échelle entre pseudobulks. Elle a été utilisée avec la fonction *ScoreSignatures_UCell()* du *package* R UCell (v2.10.1, Bioconductor) sur la matrice de *counts* bruts des 9 292 pseudobulks R/NR. Les signatures ont été utilisées sous forme de listes de gènes.

3.12.2 Score patient par sous-type macrophagique

Pour éviter de surreprésenter les patients possédant plusieurs pseudobulks dans un même sous-type macrophagique, les scores UCell ont d'abord été résumés au niveau patient x sous-type macrophagique.

Pour chaque patient et chaque sous-type M ϕ *, le score retenu correspond à la médiane des scores des pseudobulks disponibles. Ce score est alors plus robuste qu'une moyenne aux valeurs extrêmes.

Un deuxième score par patient, le score macrophagique global, a aussi été calculé comme étant la médiane des scores sur tous les sous-types macrophagiques tumoraux disponibles par patient. Ce score a été comparé entre répondeurs versus non-répondeurs avec le test de Wilcoxon-Mann-Whitney. Pour le score macrophagique par sous-type, une comparaison R vs NR a été réalisée séparément pour chaque signature par test de Wilcoxon-Mann-Whitney, et les p-valeurs ont été corrigées avec la méthode de Benjamini-Hochberg sur tous les tests d'une même signature.

3.12.3 Analyse différentielle au niveau pseudobulk (DESeq2)

En plus de ces scores, une analyse différentielle avec DESeq2 au niveau pseudobulk a été réalisée pour identifier les gènes des signatures qui portent le signal R vs NR individuellement⁹⁶. Pour chaque sous-type macrophagique M ϕ * ayant au moins 5 pseudobulks par groupe, une analyse différentielle indépendante a été réalisée avec le *design ~ responder*, où la variable *responder* indiquait l'appartenance des échantillons au groupe R ou NR, avec NR utilisé comme groupe de référence. Avant l'analyse, les gènes faiblement exprimés ont été filtrés en conservant uniquement les gènes avec au moins 5 *counts* dans suffisamment de pseudobulks, c'est-à-dire dans un nombre égal à celui du groupe le moins représenté. Ce filtre réduit le bruit et améliore la puissance statistique. Le test de Wald a ensuite été utilisé pour évaluer l'association avec le statut de réponse.

Les résultats ont ensuite été filtrés pour ne retenir que les gènes des signatures murines converties. Le nombre de gènes testés varie selon le sous-type macrophagique. La correction de Benjamini-Hochberg a été utilisée pour obtenir les padj. Au total, 194 combinaisons gène x sous-type macrophagique ont été analysées avec 42 ayant un padj < 0.05.

3.12.4 Test compétitif de gene sets (camera, limma-voom)

Pour tester si les gènes de chaque signature variaient ensemble et dans un sens donné (R > NR ou NR > R), un test compétitif de *gene sets* a été réalisé avec *camera()* du *package* *limma*. Le paramètre *inter.gene.cor* = 0.01, qui est la valeur par défaut recommandée pour les données RNA-seq, a été utilisé pour tenir compte d'une dépendance statistique entre les gènes d'une même signature.

Les données ont été analysées avec *limma-voom*, adapté aux données de type bulk RNA-seq¹⁰⁷. Les gènes faiblement exprimés ont été filtrés avec *filterByExpr()*, ensuite, la normalisation a été faite avec la méthode TMM via *calcNormFactors()* du *package* *edgeR*¹⁰⁸. Ensuite, les *counts* ont subi une transformation avec *voom()* pour tenir compte du fait que leur variance augmente avec leur niveau d'expression. Le *design* utilisé est le même que celui de DESeq2 avec la comparaison R vs NR.

Ainsi, *camera* a été appliquée séparément pour chaque sous-type macrophagique ayant au moins 5 pseudobulks par groupe, et pour chacune des trois signatures si au moins 2 gènes de la signature étaient

présents après filtrage. Au total, 32 tests camera ont été réalisés après filtrage des signatures et des sous-types. Les résultats ont été analysés à partir de la FDR fournie par camera.

Une note importante est que camera est un test relativement conservateur, surtout quand le *set* de gènes est court. L'absence de significativité après correction FDR ne veut pas forcément dire qu'il n'y a pas d'effet, cela peut en effet être dû au petit nombre de gènes et à la prise en compte de la corrélation inter-gènes par camera¹⁰⁶. Ces résultats sont donc à interpréter en complément de ceux de DESeq2 et d'UCell.

4. Résultats

4.1 Analyse du microenvironnement immunitaire murin

4.1.1 Contrôle qualité, filtrage et détection des doublets

Après les étapes de contrôle qualité et la suppression des doublets, un total de 46 521 cellules a été gardé pour les analyses murines. Sur les 55 103 cellules de départ, 52 230 ont passé les critères de qualité, 2 873 ont été exclues. La détection des doublets a ensuite permis d'identifier et éliminer 5 709 cellules supplémentaires. Les quatre groupes expérimentaux présentaient encore des effectifs suffisants après filtrage : 8 996 cellules pour KP Early, 14 994 pour KP Late, 7 893 pour KPR Early et 14 638 pour KPR Late (**Tableau 4**). Ce nombre de cellules a permis de poursuivre les analyses, telles que l'intégration des échantillons, l'annotation des cellules et les comparaisons entre groupes.

Échantillon	Avant QC	Après QC	Cellules exclues QC	Après doublets
KP_Early	10 495	10 051	444	8 996
KP_Late	17 785	16 947	838	14 994
KPR_Early	9 327	8 714	613	7 893
KPR_Late	17 496	16 518	978	14 638
Total	55 103	52 230	2 873	46 521

Tableau 4. Nombre de cellules à chaque étape du filtrage et de la détection des doublets pour les quatre échantillons murins.

4.1.2 Intégration et réduction de dimension

L'intégration des quatre échantillons par RPCA a permis d'obtenir un objet commun regroupant 46 521 cellules, réparties en 33 clusters identifiés à une résolution de 1.4 (**Figure 6-A**). La projection UMAP montre que les cellules des quatre groupes expérimentaux se distribuent de manière homogène au sein des clusters, sans séparation nette selon l'échantillon, la condition ou la technique de séquençage (**Figure 6-B à E**). Ces résultats ont permis d'obtenir une vue globale du microenvironnement immunitaire avant l'annotation cellulaire.

La distribution homogène des cellules sur l'UMAP montre que l'intégration RPCA a bien corrigé l'effet du lot lié à la différence de chimie de capture. Cependant, la technique de séquençage est totalement liée au stade tumoral dans cette étude, et les deux facteurs ne peuvent pas être séparés. Cet aspect doit être pris en compte dans l'interprétation des analyses différentielles.

4.1.3 Annotation cellulaire et composition du TME murin

L'annotation cellulaire de l'objet intégré a permis d'identifier 22 populations immunitaires distinctes parmi les 46 521 cellules analysées (**Figure 7-A**). Les macrophages constituaient la population la plus abondante du jeu de données, avec 32 610 cellules sur 46 521, soit 70.1 % des cellules. Ils étaient principalement représentés par les sous-types Macrophages_MFAR (21 452 cellules, 46.1 %), et les Macrophages_autres (9 967 cellules, 21.4 %). Les neutrophiles représentaient 5 515 cellules, (11.9 %), tandis que les lymphocytes T CD4 comptaient 1 712 cellules (3,7 %), et les cellules B naïves/mémoires 1 264 cellules (2.7 %). Les compartiments DC, NK, NKT et ILC sont présents en proportions plus faibles.

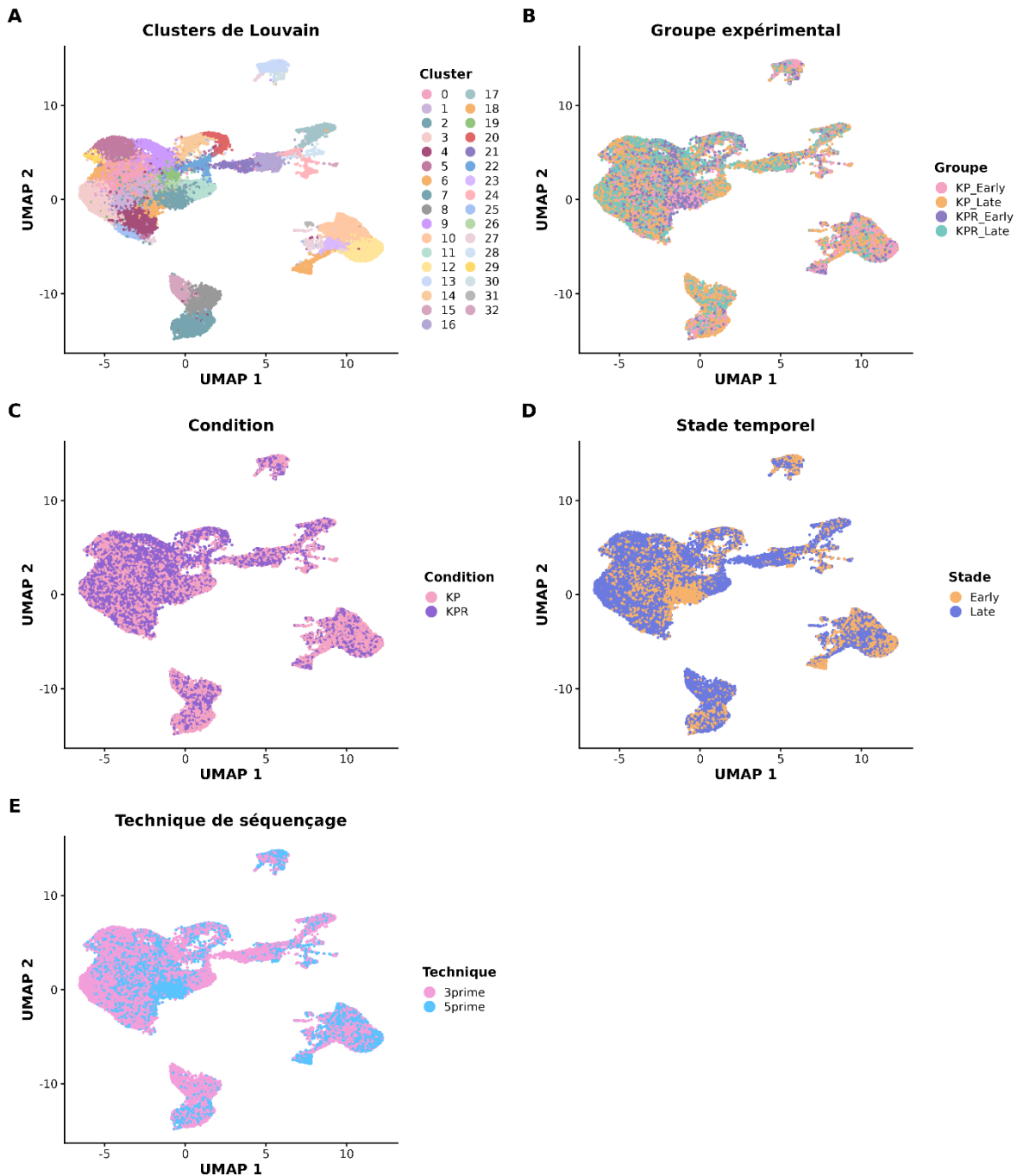


Figure 6. Projection UMAP des 46 521 cellules immunitaires murines après intégration par RPCA.

- (A) Coloration selon les clusters (résolution = 1,4 ; 33 clusters).
- (B) Coloration selon le groupe expérimental (KP Early, KP Late, KPR Early, KPR Late). Le mélange des cellules au sein des clusters suggère une intégration efficace, sans ségrégation majeure liée aux variables expérimentales ou techniques.
- (C) Coloration selon la condition génétique (KP vs KPR).
- (D) Coloration selon le stade tumoral (Early vs Late).
- (E) Coloration selon la technique de séquençage (3' vs 5').

En concordance avec ces résultats, dans la projection UMAP, les différentes populations occupaient des régions distinctes de l'espace du graphique. Les macrophages formaient le cluster le plus grand, tandis que les neutrophiles, les lymphocytes T, les DC et les cellules NK se distribuaient dans les zones les plus restreintes (**Figure 7-A**).

L'analyse de la composition cellulaire par groupe expérimental montre que les macrophages MFAR sont la population la plus abondante de tous les groupes, avec une proportion qui augmente de KP Early à KPR Late (**Figure 7-B**). Les Macrophages_autres représentaient aussi une fraction importante dans chaque groupe. À l'inverse, les neutrophiles sont proportionnellement plus représentés dans les groupes KP, en particulier KP Late, que dans les groupes KPR. Ces résultats montrent la forte dominance du compartiment macrophagique dans le microenvironnement immunitaire murin étudié.

4.1.4 Analyse différentielle pseudobulk globale

L'analyse différentielle pseudobulk, réalisée sur les types cellulaires suffisamment représentés, a permis d'identifier 22 846 signaux différentiels significatifs, toutes comparaisons confondues ($p_{adj} < 0.05$; $|\log_2FC| > 1$). À l'échelle globale, les macrophages apparaissaient comme le compartiment présentant le plus grand nombre de gènes différentiellement exprimés (**Figure 8-A**).

Cette tendance se confirmait au sein des sous-types macrophagiques, où les populations Macrophages_MFAR et Macrophages_autres ont la majorité des signaux différentiels (**Figure 8-B**). Les comparaisons KPR_Early vs KP_Early et KPR_Late vs KP_Late permettent d'évaluer les différences associées au modèle KPR à temps post-infection AAV équivalent. Dans la comparaison KPR_Early vs KP_Early, les Macrophages_MFAR présentaient 217 DEG, contre 195 pour les Macrophages_autres. Dans la comparaison KPR_Late vs KP_Late, les Macrophages_MFAR présentaient 212 DEG significatifs, contre 146 pour les Macrophages_autres. Les autres sous-types présentaient des variations plus limitées, avec 20 DEG pour les Macrophages_MHCII+, 2 pour les Macrophages_CD103-CD11b+ et aucun pour les Macrophages_alveolaires (**Figure 8-B, Tableau 5**).

À l'inverse, les comparaisons impliquant une progression temporelle, à savoir KP_Late vs KP_Early et KPR_Late vs KPR_Early, montraient un nombre de DEG nettement plus élevé, pouvant atteindre 2 871 et 2 777 respectivement dans les Macrophages_MFAR.

La *heatmap* des gènes les plus différentiellement exprimés dans les sous-types macrophagiques entre KPR_Late et KP_Late (**Figure 8-C**) montre des profils d'expression distincts entre les pseudo-réplicats et les sous-types macrophagiques. Certains gènes présentant une expression plus marquée dans les pseudo-réplicats KPR_Late, notamment *Slpi*, *Saa3*, *Alox15* et *F5*, tandis que d'autres présentaient une expression plus élevée dans les pseudo-réplicats KP_Late comme *Zfp534*, *Camp*, *Cldn18* et *Clec1b*.

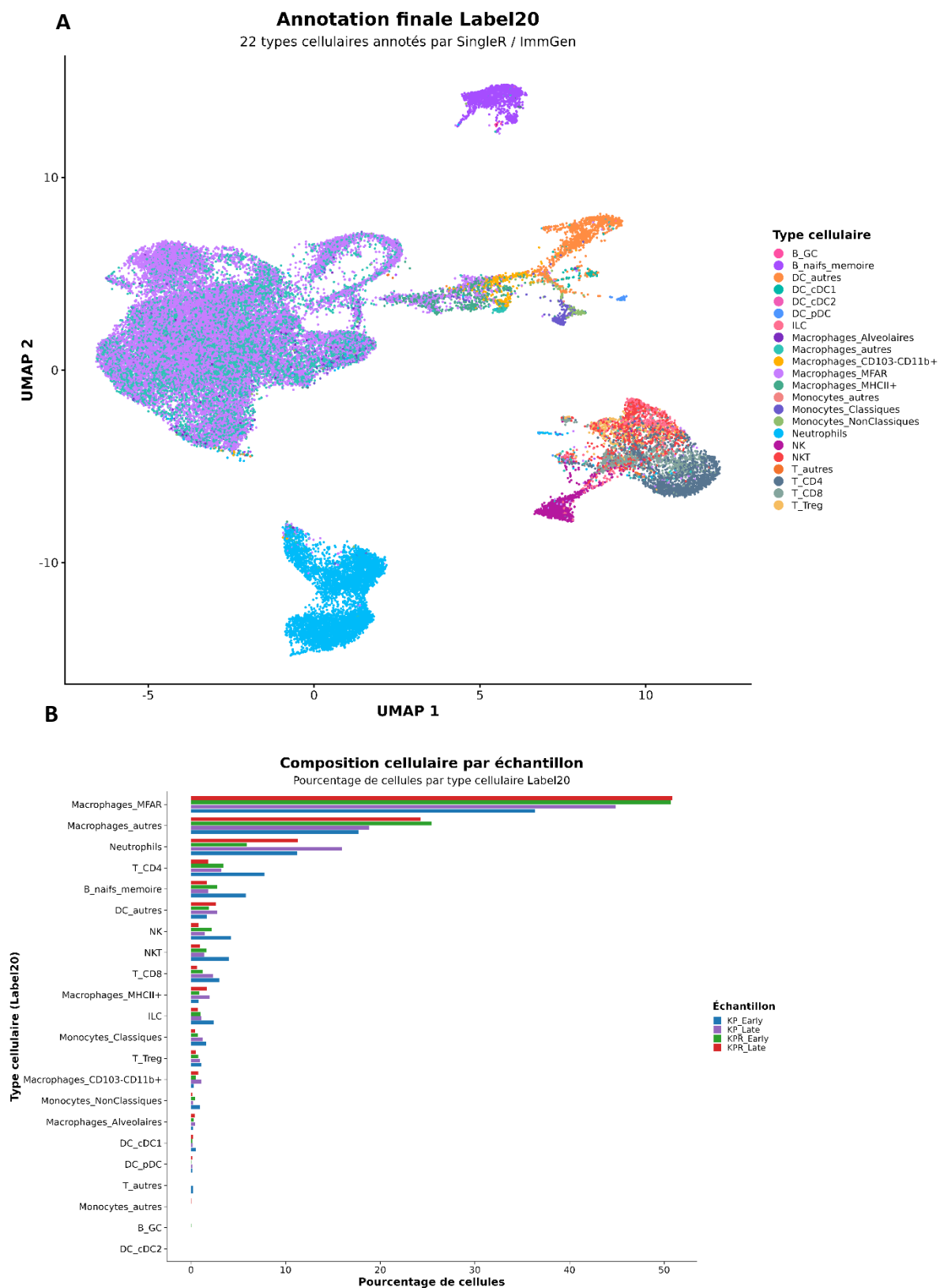


Figure 7. Annotation cellulaire du TME murin et composition par groupe expérimental.

- (A) Projection UMAP colorée selon les 22 types cellulaires de l'annotation finale Label20, obtenue par SingleR à partir de la référence ImmGen en deux étapes hiérarchiques
- (B) Composition cellulaire relative, exprimée en pourcentage, des différents types cellulaires Label20 selon le groupe expérimental (KP Early, KP Late, KPR Early, KPR Late).

Ainsi, les macrophages étant à la fois la population numériquement dominante du TME murin, avec environ 70 % des cellules comme indiqué dans la section 4.1.3, et le compartiment présentant le plus grand nombre de gènes différentiellement exprimés entre les conditions KP et KPR, l’analyse a ensuite été centrée sur ce compartiment dans la section suivante.

Tableau 5. Nombre de gènes différentiellement exprimés significatifs dans les sous-types macrophagiques selon les comparaisons pseudobulk testées.
 Les DEG significatifs sont définis par un padj < 0.05 et un |log2FC| > 1.

Comparaison	Macrophages_MFAR	Macrophages_autres	Macrophages_MHCII+	Macrophages_CD103-CD11b+	Macrophages_Alveolaires
KPR_Early vs KP_Early	217	195	16	4	0
KPR_Late vs KP_Late	212	146	20	2	0
KP_Late vs KP_Early	2871	2005	575	174	30
KPR_Late vs KPR_Early	2777	1899	374	193	40
KPR_Late vs KP_Early	2911	2057	459	184	30

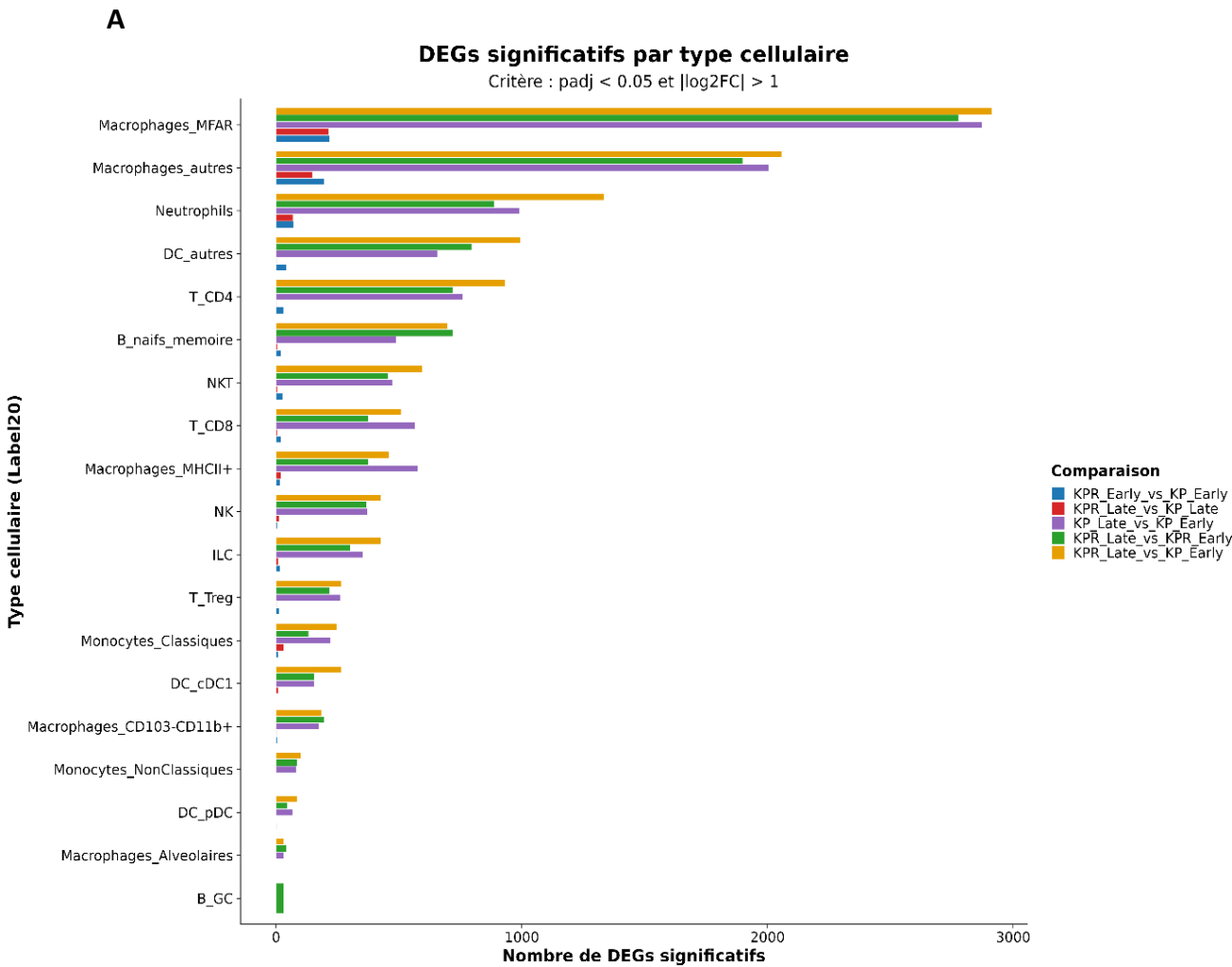


Figure 8. Analyse différentielle pseudobulk globale et focalisation sur les sous-types macrophagiques.
(A) Nombre de gènes différentiellement exprimés significatifs par type cellulaire Label20 et par comparaison biologique testée. Les DEG significatifs sont définis par un padj < 0.05 et un |log2FC| > 1.

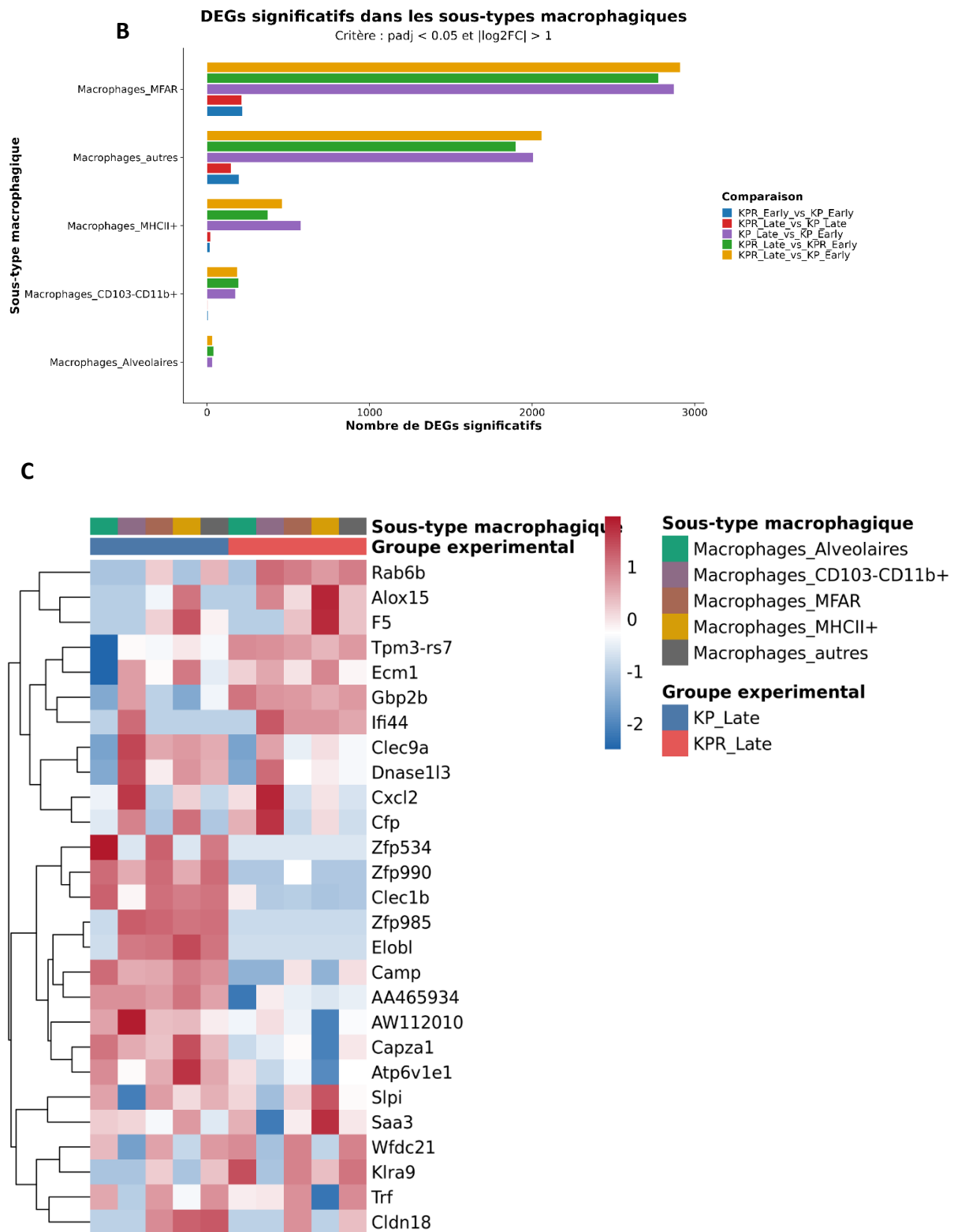


Figure 8, suite. Analyse différentielle pseudobulk globale et focalisation sur les sous-types macrophagiques.

(B) Nombre de gènes différentiellement exprimés significatifs dans les sous-types macrophagiques pour les mêmes comparaisons. Les macrophages MFAR et les Macrophages_autres présentent les nombres les plus élevés de DEG, en particulier dans les comparaisons impliquant le stade Late. (C) Heatmap des gènes différentiellement exprimés sélectionnés pour la comparaison KPR_Late vs KP_Late au niveau pseudobulk, selon la condition expérimentale et le sous-type macrophagique. Les colonnes représentent les pseudo-bulks et les lignes les gènes sélectionnés. Les valeurs correspondent à une expression normalisée et centrée-réduite.

4.2 Analyse du compartiment macrophagique murin

4.2.1 Réintégration par Harmony et reclustering

L'analyse globale du microenvironnement immunitaire murin ayant mis en évidence une forte dominance du compartiment macrophagique, celui-ci a été étudié plus en détail. Pour cette analyse ciblée, une seconde intégration a été réalisée à l'aide de Harmony sur l'ensemble des cellules immunitaires murines. Les cellules annotées comme macrophages ont ensuite été extraites de cet objet intégré. Le sous-ensemble macrophagique ainsi obtenu comprenait 32 610 cellules issues des quatre groupes expérimentaux : KP Early, KP Late, KPR Early et KPR Late.

Le reclustering de ce compartiment, basé sur l'intégration Harmony, a permis d'identifier 10 clusters macrophagiques distincts (**Figure 9-A**). Les clusters 0 et 1 étaient les plus représentés, avec respectivement 9 063 cellules (27.79 %) et 8 796 cellules (26.97 %). Les clusters 2 et 3 représentaient aussi des populations abondantes, avec respectivement 4 868 cellules (14.93 %) et 4 674 cellules (14.3 %). À l'inverse, les clusters 7, 8 et 9 étaient très peu abondants, avec 120, 109 et 31 cellules (**Tableau 6**).

La projection UMAP montrait plusieurs sous-populations macrophagiques distinctes (**Figure 9-A**). La coloration selon les groupes expérimentaux montrait une répartition non homogène des cellules KP Early, KP Late, KPR Early et KPR Late au sein des clusters (**Figure 9-B**). Cette hétérogénéité était aussi observable lorsque les cellules étaient regroupées selon la condition génétique KP/KPR (**Figure 9-C**). Par ailleurs, la coloration selon la technologie de séquençage ne montrait pas de séparation nette entre les cellules issues des chimies 3' et 5', contrairement à ce qui serait attendu en cas de structuration dominante par la technique (**Figure 9-D**).

L'analyse de composition confirme cette répartition non homogène (**Figure 10-A et B ; Tableau 6**). Le cluster 2 était majoritairement composé de cellules KPR (81.6 %), dont 60.4 % venant du groupe KPR Late. Il était aussi principalement composé de cellules Late (77.0 %). Le cluster 0 présentait également une forte proportion de cellules Late (71.9 % : 36.3 % KPR Late et 35.7 % KP Late), tandis que le cluster 1 était davantage représenté par des cellules KP (59.4 %) et le cluster 3 par des cellules Early (58.8 %).

Malgré leur enrichissement en cellules Late, les clusters 8 (109 cellules) et le cluster 9 (31 cellules) ont été exclus en raison de leurs faibles effectifs, limitant la fiabilité des analyses.

Au vu de leur composition et de leur représentation dans les groupes expérimentaux, les clusters 0 et 2 ont été retenus comme candidats pour la construction des signatures murines. Ces signatures peuvent donc être décrites comme des signatures macrophagiques murines associées au modèle KPR Late, c'est-à-dire à la surexpression de RICTOR à un temps post-infection tardif.

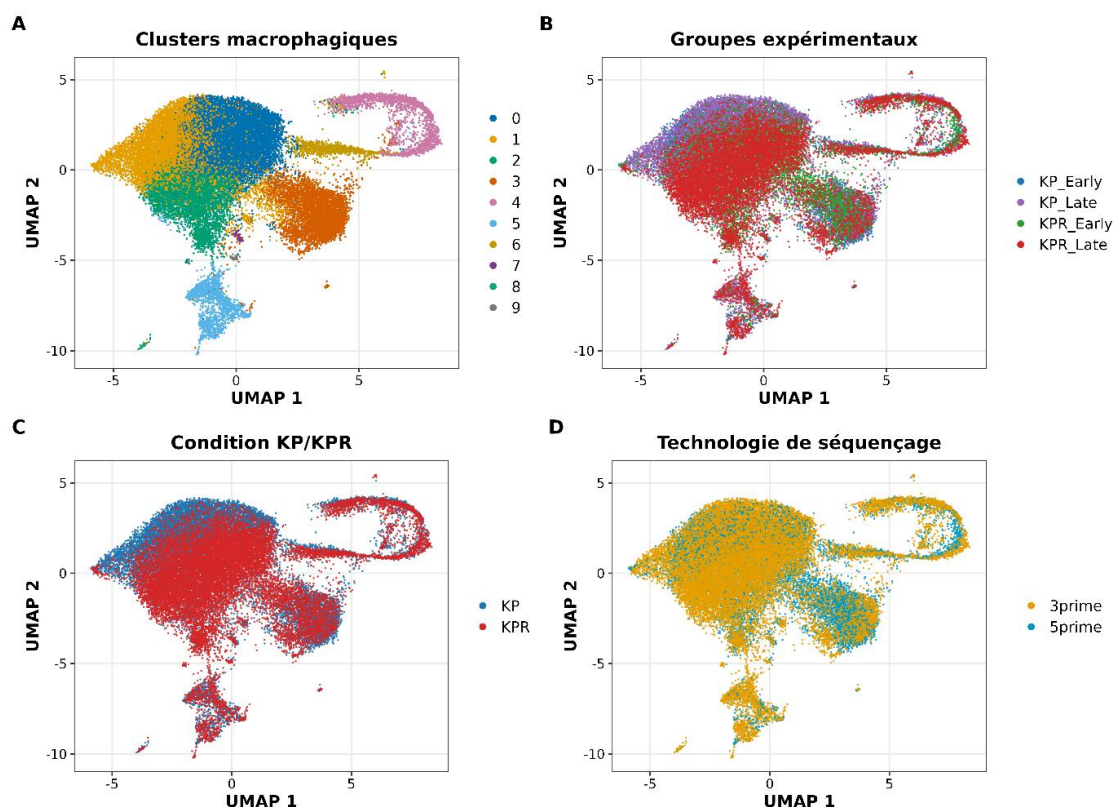


Figure 9. Extraction et reclustering du compartiment macrophagique murin à partir de l'intégration globale Harmony.

Projection UMAP des 32 610 cellules macrophagiques extraites de l'objet murin global intégré par Harmony, puis reclusterisées à une résolution de 0.5. **(A)** Coloration selon les clusters macrophagiques. **(B)** Coloration selon le groupe expérimental (KP Early, KP Late, KPR Early, KPR Late). **(C)** Coloration selon la condition génétique (KP vs KPR). **(D)** Coloration selon la technique de séquençage (3' vs 5').

Tableau 6. Effectifs cellulaires des clusters macrophagiques murins après reclustering.

Pour chaque cluster macrophagique, le tableau indique le nombre total de cellules, ainsi que les effectifs bruts et les proportions des groupes expérimentaux KP Early, KP Late, KPR Early et KPR Late. Les proportions selon la condition KP/KPR et selon le stade tumoral Early/Late sont également indiquées. Les pourcentages sont calculés au sein de chaque cluster.

Cluster	n	%	n KP	% KP	n KP	% KP	n KPR	% KPR	n KPR	% KPR
	cellules	total	Early	Early	Late	Late	Early	Early	Late	Late
0	9063	27.79	757	8.35	3232	35.66	1788	19.73	3286	36.26
1	8796	26.97	1832	20.83	3391	38.55	1172	13.32	2401	27.30
2	4868	14.93	87	1.79	808	16.60	1032	21.20	2941	60.41
3	4674	14.33	1552	33.20	1044	22.34	1194	25.55	884	18.91
4	2278	6.99	340	14.93	592	25.99	493	21.64	853	37.45
5	1650	5.06	148	8.97	705	42.73	156	9.45	641	38.85
6	1021	3.13	227	22.23	190	18.61	280	27.42	324	31.73
7	120	0.37	35	29.17	35	29.17	22	18.33	28	23.33
8	109	0.33	2	1.83	66	60.55	1	0.92	40	36.70
9	31	0.10	2	6.45	12	38.71	2	6.45	15	48.39

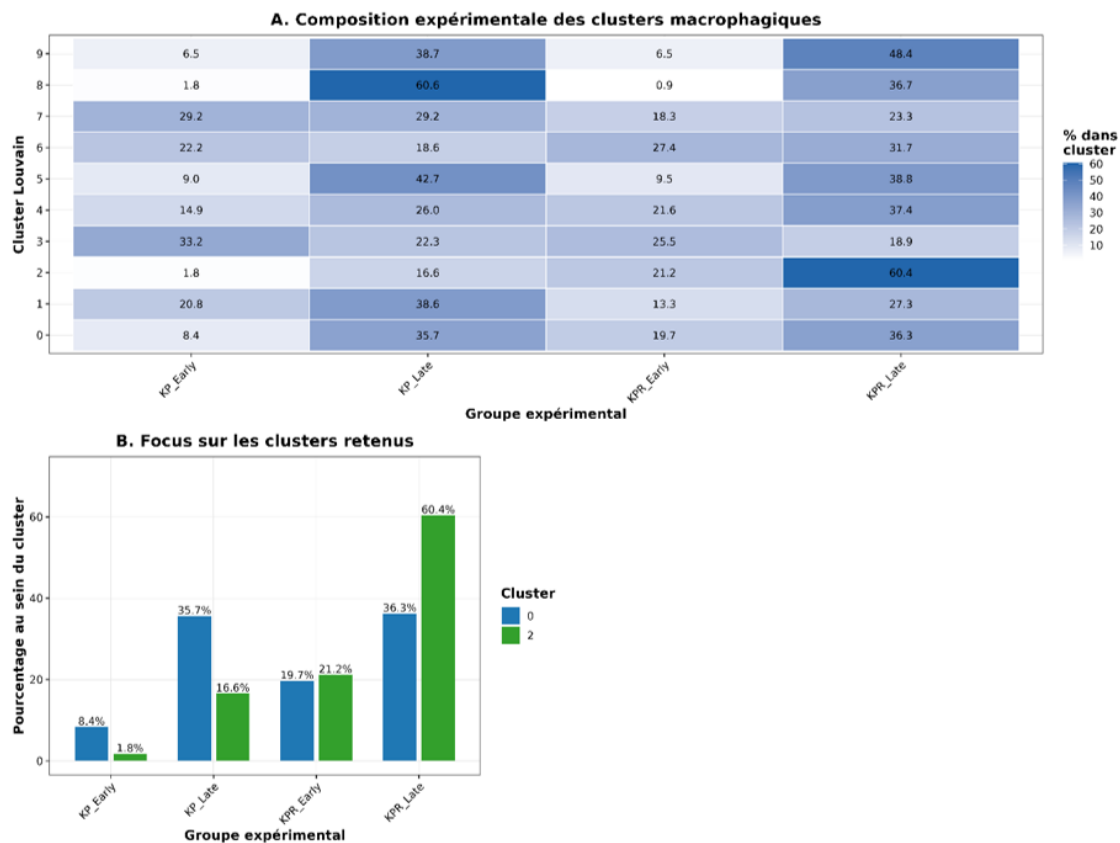


Figure 10. Composition expérimentale des clusters macrophagiques et sélection des clusters candidats.

(A) *Heatmap* représentant la proportion de cellules de chaque groupe expérimental au sein de chaque cluster macrophagique. Les pourcentages sont calculés à l'intérieur de chaque cluster. (B) Focus sur la composition des clusters 0 et 2 selon les quatre groupes expérimentaux.

4.2.2 Caractérisation des clusters macrophagiques à partir de marqueurs de littérature

Afin de caractériser les clusters macrophagiques indépendamment de la construction des signatures murines, l'expression de marqueurs macrophagiques décrits dans la littérature^{109,110} a été examinée dans les 10 clusters issus du reclustering. Ces marqueurs ont été sélectionnés à partir d'études sur l'hétérogénéité des populations myéloïdes et macrophagiques dans les cancers pulmonaires et murins. Ils couvrent des états macrophagiques généraux, alvéolaires, inflammatoires et associés aux tumeurs. La *heatmap* met en évidence des profils d'expression distincts entre clusters (**Figure 11**). Les clusters 0, 1 et 2 montraient une expression de marqueurs macrophagiques généraux et pulmonaires, notamment *Cd68*, *Adgre1*, *Itgax* et *Siglec f*. Le cluster 5 présentait une expression élevée de nombreux marqueurs associés à des macrophages résidents du tissu pulmonaire et à des macrophages associés aux tumeurs, dont *H2-Ab1*, *Cd80*, *Cd86*, *Arg1*, *Cd163*, *Timd4*, *Vsig4*, *Cd5l*, *Folr2*, *Spp1*, *Fn1*, *Trem2*, *Apoe*, *C1qa* et *C1qb*. Le cluster 7 présentait un profil plus restreint, notamment marqué par l'expression de *Cxcl10*. Le cluster 8 montrait également un profil activé, avec une expression élevée de *Nos2* et *Vegfa*. Les clusters 3, 4, 6 et 9 étaient moins clairement définis par cette sélection de marqueurs. Dans l'ensemble, cette analyse confirme que le reclustering permet de distinguer plusieurs états macrophagiques présentant des profils distincts, dont certains présentent des similarités avec des clusters macrophagiques déjà décrits dans la littérature.

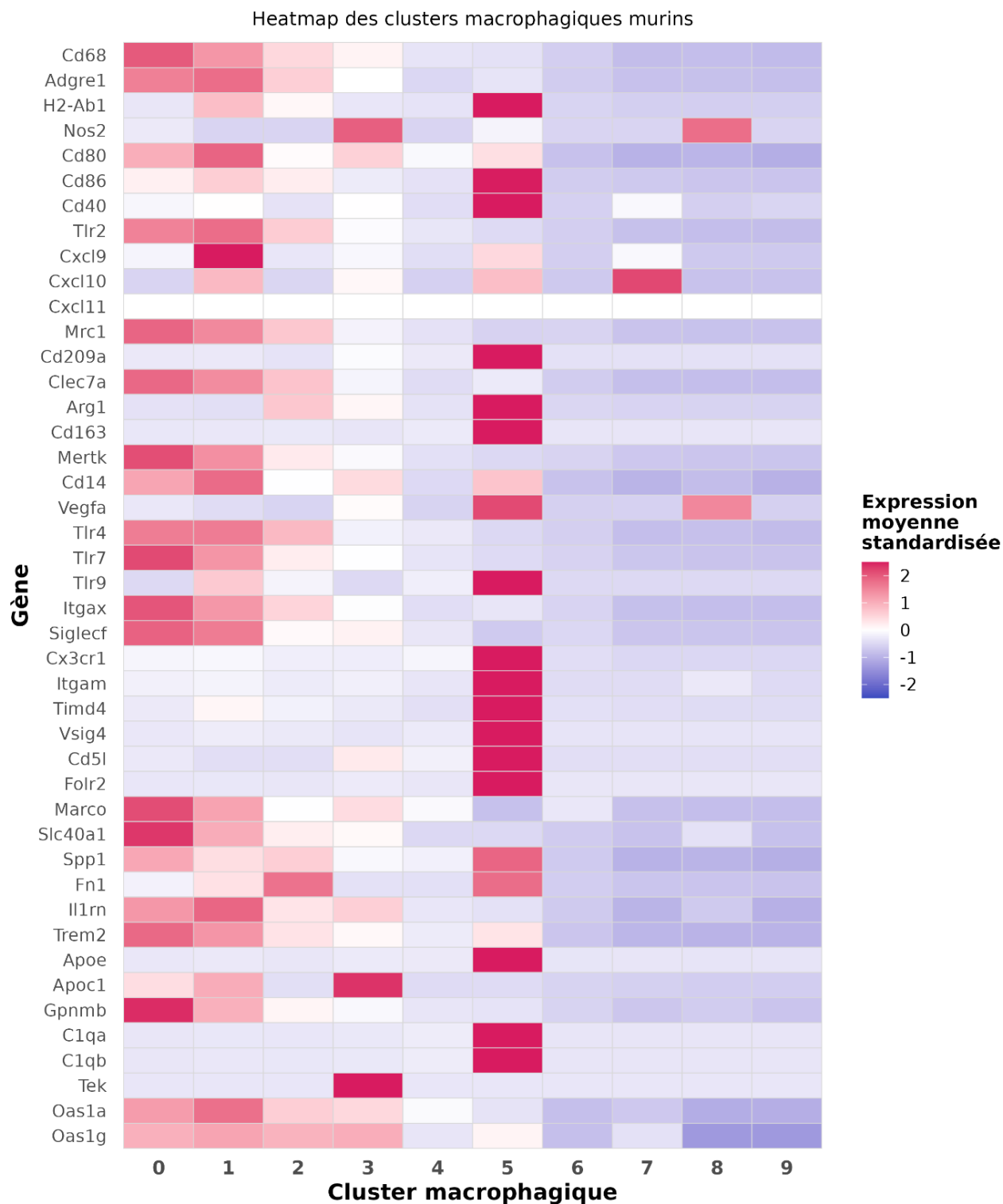


Figure 11. Expression moyenne de marqueurs macrophagiques issus de la littérature dans les clusters murins.

La *heatmap* montre l'expression moyenne standardisée de marqueurs macrophagiques sélectionnés dans la littérature, dans les 10 clusters obtenus après reclustering du compartiment macrophagique murin. Les lignes correspondent aux gènes et les colonnes aux clusters macrophagiques. Les couleurs indiquent le niveau relatif d'expression moyenne de chaque gène.

4.2.3 Construction des signatures murines candidates

Sur base de leur composition expérimentale, les clusters 0 et 2 ont été retenus pour la construction des signatures murines candidates. Le cluster 0 présentait une forte proportion de cellules Late, tandis que le cluster 2 présentait un enrichissement en cellules KPR, en particulier KPR_Late. Ces deux clusters ont donc été considérés comme les plus pertinents pour identifier les signatures liées à la progression tumorale (contexte Late), et plus spécifiquement à la surexpression de RICTOR (contexte KPR_Late) pour le cluster 2.

La signature dérivée du cluster 0 comprenait 24 gènes murins, dont 22 ont été conservés après conversion orthologique chez l'humain. La signature du cluster 2 comprenait 11 gènes murins, dont 10 ont été conservés chez l'humain. Une troisième signature complémentaire, appelée signature Elodie, composée de 4 gènes murins ayant montré un intérêt biologique au cours de précédentes expérimentations dans le cadre du projet, a également été retenue, avec 3 gènes conservés après orthologie (**Tableau 7**).

Pour mettre en évidence les signaux les plus robustes, un sous-ensemble de gènes communs aux quatre comparaisons différentielles liées au contexte KPR_Late a été sélectionné : KPR vs KP, KPR_Late vs KPR_Early, KPR_Late vs KP_Late et KPR_Late vs KP_Early. Neuf gènes répondaient à ce critère pour la signature du cluster 0, et deux pour celle du cluster 2. Dans les clusters dont sont issues les signatures, ces gènes présentaient une expression plus élevée dans le groupe KPR Late, avec une expression moyenne plus élevée et plus de cellules exprimant ces gènes (**Figure 12**). Cette observation suggère que les gènes retenus capturent principalement un programme associé au contexte KPR_Late et à la progression tumorale, surtout pour la signature du cluster 2.

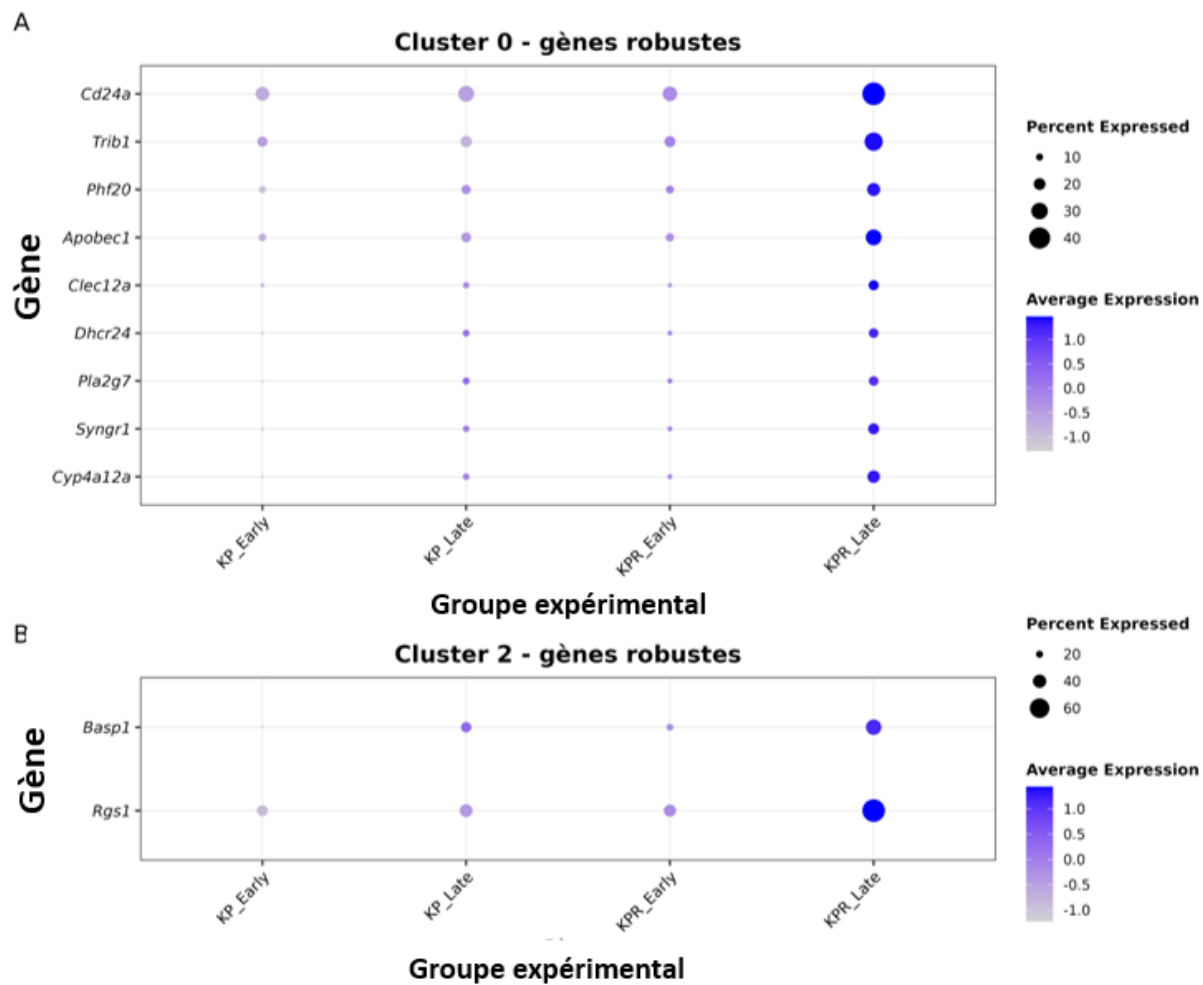
Tableau 7. Signatures murines candidates retenues pour la projection humaine.

Ce tableau présente les trois signatures murines utilisées pour la projection dans les données humaines de scRNA-seq. Les signatures associées aux clusters 0 et 2 ont été construites à partir de gènes surexprimés dans les clusters macrophagiques correspondants et retenus par intersection de plusieurs comparaisons différentielles liées au contexte KPR Late. La signature Elodie correspond à une signature complémentaire composée de gènes d'intérêt biologique. Pour chaque signature, le tableau indique l'origine de la signature, la justification du choix, le nombre de gènes murins retenus, le nombre de gènes humains conservés après conversion orthologique, ainsi que les gènes humains effectivement utilisés pour la projection.

Signature	Cluster d'origine	Justification du choix	N gènes murins	N gènes humains après orthologie	Gènes humains utilisés pour la projection
Signature Cluster 0	Cluster 0	Gènes surexprimés dans le cluster 0 et associés au contexte KPR Late, retenus par intersection de plusieurs comparaisons différentielles.	24	22	C4A, C4B, CD24, CD244, CLEC12A, CYP4A22, DHCR24, EMP1, FFAR2, HS3ST3B1, MFHAS1, MMP8, OSBPL1A, PHF20, PLA2G7, PLS3, PTPN7, SLC25A24, SPEG, SYNGR1, TMX4, TRIB1
Signature Cluster 2	Cluster 2	Gènes surexprimés dans le cluster 2, le plus enrichi en cellules KPR Late, retenus selon la même stratégie d'intersection.	11	10	ADAM15, BASP1, CLEC12A, DUSP1, HCAR2, MMP8, PTPN18, RGS1, SYNGR1, TRIB1
Signature Elodie	Signature complémentaire	Signature complémentaire composée de gènes d'intérêt biologique, utilisée comme comparateur.	4	3	ARG1, FBP1, NPL

Figure 12. Expression des gènes robustes des signatures murines selon les groupes expérimentaux.

Le DotPlot représente l'expression des gènes robustes retenus dans les signatures murines candidates, dans leur cluster d'origine. Le panneau A présente les neuf gènes robustes de la signature du cluster 0, tandis que le panneau B présente les deux gènes robustes de la signature du cluster 2. La taille des points indique la proportion de cellules exprimant le gène dans chaque groupe expérimental, et la couleur représente l'expression moyenne standardisée. Dans les deux clusters, les gènes robustes présentent une expression plus marquée dans le groupe KPR_Late, soutenant leur lien avec le contexte expérimental utilisé pour la construction des signatures.



4.3 Projection sur les données humaines de NSCLC

4.3.1 Cohorte humaine analysable et pseudobulks

La projection des signatures murines a été réalisée sur la cohorte humaine de NSCLC après contrôle qualité et construction des pseudobulks. La cohorte comprenait 243 patients, ce qui correspond à 1 239 631 cellules réparties en 51 types cellulaires annotés. Au total, 9 320 pseudobulks ont été générés, selon le schéma patient x type cellulaire.

Parmi les patients, 242 disposaient d'un statut de réponse pathologique exploitable. Un patient avec un statut inconnu a été exclu, ainsi que ses 28 pseudobulks associés, conduisant à un total de 9 292 pseudobulks pour les analyses comparatives. La cohorte analysable comprenait 130 répondeurs (85 pCR et 45 MPR) et 112 comme non-répondeurs.

La cohorte était majoritairement composée de LUSC (n = 180), contre 63 LUAD. Parmi les LUAD, 21 patients étaient répondeurs contre 42 non-répondeurs, alors que parmi les LUSC, 109 étaient répondeurs contre 70 non-répondeurs (**Tableau 8**).

Ainsi, la cohorte obtenue permettait d'évaluer les signatures murines projetées dans un contexte humain, à l'échelle du patient, de la réponse au traitement et des types cellulaires annotés. Pour rappel, les annotations cellulaires utilisées dans les analyses qui suivent sont celles de Liu et al., qui ont identifié 51 sous-types immunitaires dont 11 sous-types macrophagiques, nommés d'après leur gène marqueur principal (par exemple Mph_MARCO, Mph_FOLR2).

Tableau 8. Description de la cohorte humaine utilisée pour la projection des signatures murines.

Catégorie	Variable	Effectif
Cohorte	Patients inclus	243
Cohorte	Patients analysables R/NR	242
Réponse	Répondeurs (pCR + MPR)	130
Réponse	Non-répondeurs (non-MPR)	112
Réponse	pCR	85
Réponse	MPR	45
Données cellulaires	Cellules après QC	1 239 631
Données cellulaires	Types cellulaires annotés	51
Données cellulaires	Pseudobulks générés	9 320
Histologie	LUSC	180
Histologie	LUAD	63

4.3.2 Scoring UCell des signatures dans les pseudobulks macrophagiques

Les signatures murines converties en orthologues humains ont été scorées à l'aide de UCell sur les profils de pseudobulks macrophagiques humains. Les scores ont ensuite été agrégés au niveau patient afin de comparer les répondeurs et les non-répondeurs. Cette analyse portait sur 126 patients répondeurs et 101 patients non-répondeurs disposant de pseudobulks macrophagiques exploitables.

A l'échelle macrophagique globale, la signature du cluster 0 présentait un score significativement plus élevé chez les répondeurs que chez les non-répondeurs (médiane R = 0.0360 ; médiane NR = 0.0291 ; p-valeur ajustée < 0.001 ; **Figure 13**). A l'inverse, la signature Elodie présentait un score plus élevé chez les

non-répondeurs (médiane R= 0.1471 ; médiane NR = 0.2040 ; p-valeur ajustée = 0.0312). La signature du cluster 2 ne montrait pas de différence significative entre les deux groupes au niveau macrophagique global (médiane R= 0.2685 ; médiane NR = 0.2551 ; p-valeur ajustée = 0.46 ; **Figure 13**).

L'analyse par sous-type macrophagique a mis en évidence une projection hétérogène des signatures selon les états macrophagiques (**Figure 14**). La signature du cluster 0 était significativement plus élevée chez les répondeurs dans plusieurs sous-types, notamment Mph_DNAJB1, Mph_MARCO, Mph_CXCL2, Mph_FOLR2, Mph_S100A10 et Mph_VCAN (**Figure 14**). A l'inverse, la signature Elodie présentait des scores plus élevés chez les non-répondeurs dans certains sous-types, en particulier Mph_VCAN, Mph_S100A10 et Mph_DNAJB1. Le sous-type Mph_VCAN montrait l'effet le plus marqué pour cette signature ($\log_2FC = -1,757$; p-valeur ajustée = 0.0036). Enfin, la signature du cluster 2 présentait un profil plus hétérogène : elle était plus élevée chez les répondeurs dans le sous-type Mph_DNAJB1, mais plus élevée chez les non-répondeurs dans Mph_MMP9 et Mph_MARCO.

Les résultats de *scoring* UCell mettent en évidence des tendances distinctes selon la signature et les sous-types macrophagiques considérés. La signature du cluster 0 est globalement associée aux répondeurs, tant à l'échelle des macrophages que dans plusieurs sous-types spécifiques. A l'inverse, la signature Elodie est davantage enrichie chez les non-répondeurs, notamment dans les sous-types Mph_VCAN, Mph_S100A10 et Mph_DNAJB1. Enfin, la signature du cluster 2 ne présente pas de différence significative à l'échelle globale, mais montre des variations significatives dans certains sous-types macrophagiques.

4.3.3 Analyse différentielle gène par gène dans les pseudobulks macrophagiques

Afin d'identifier les gènes à l'origine des différences observées au niveau des signatures, une analyse différentielle gène par gène a été réalisée avec DESeq2 sur les pseudobulks macrophagiques. Cette analyse vient compléter le *scoring* UCell en évaluant individuellement chacun des gènes des signatures projetées dans les données humaines. L'analyse a été réalisée séparément pour chaque sous-type macrophagique, en comparant les patients répondeurs aux non-répondeurs. Le groupe NR a été utilisé comme référence, de sorte qu'un $\log_2FC$ positif indique une expression plus élevée chez les répondeurs, tandis qu'un $\log_2FC$ négatif correspond à une expression plus élevée chez les non-répondeurs.

Au total, 194 combinaisons gène x sous-type macrophagique ont été testées. Parmi celles-ci, 42 étaient significatives après correction de Benjamini-Hochberg ($p_{adj} < 0.05$; **Figure 15**). Plusieurs gènes étaient associés aux répondeurs dans différents sous-types macrophagiques, notamment *EMP1*, *TMX4*, *ADAM15*, *CLEC12A*, *PHF20* et *PTPN18*. Parmi eux, *TMX4* et *EMP1* apparaissaient comme des signaux récurrents, avec des $\log_2FC$ positifs dans plusieurs sous-types.

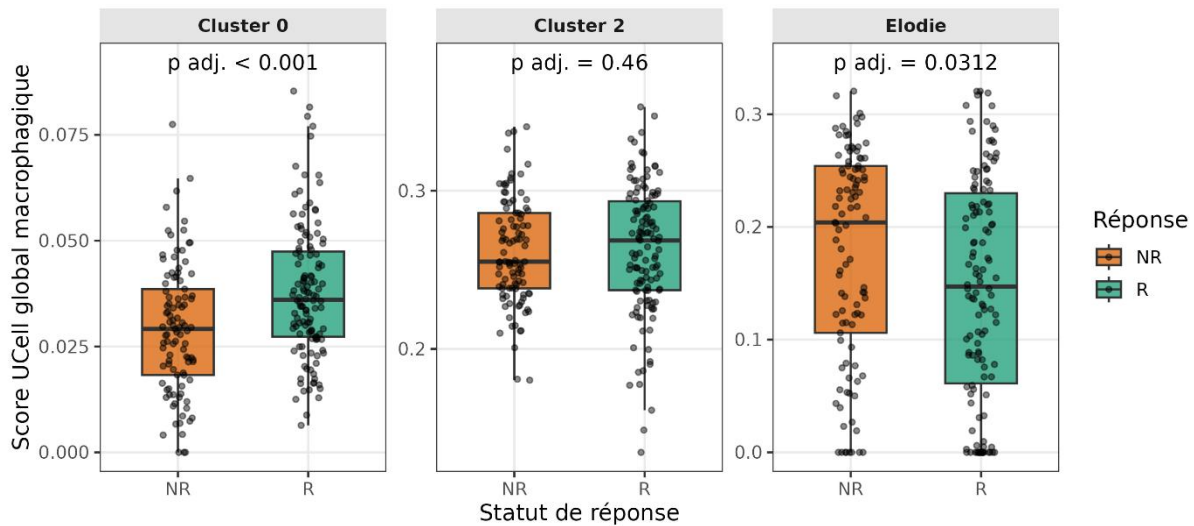


Figure 13. Scores UCell globaux des signatures murines projetées dans les pseudobulks macrophagiques humains.

Les scores UCell des signatures Cluster 0, Cluster 2 et Elodie ont été calculés à partir des profils pseudobulks macrophagiques, puis résumés au niveau patient par la médiane des sous-types macrophagiques disponibles. Les comparaisons entre répondeurs (R) et non-répondeurs (NR) ont été réalisées par test de Wilcoxon, avec correction de Benjamini-Hochberg pour les trois signatures testées. La signature du cluster 0 présente un score plus élevé chez les répondeurs, tandis que la signature Elodie est plus élevée chez les non-répondeurs. La signature du cluster 2 ne montre pas de différence significative au niveau macrophagique global.

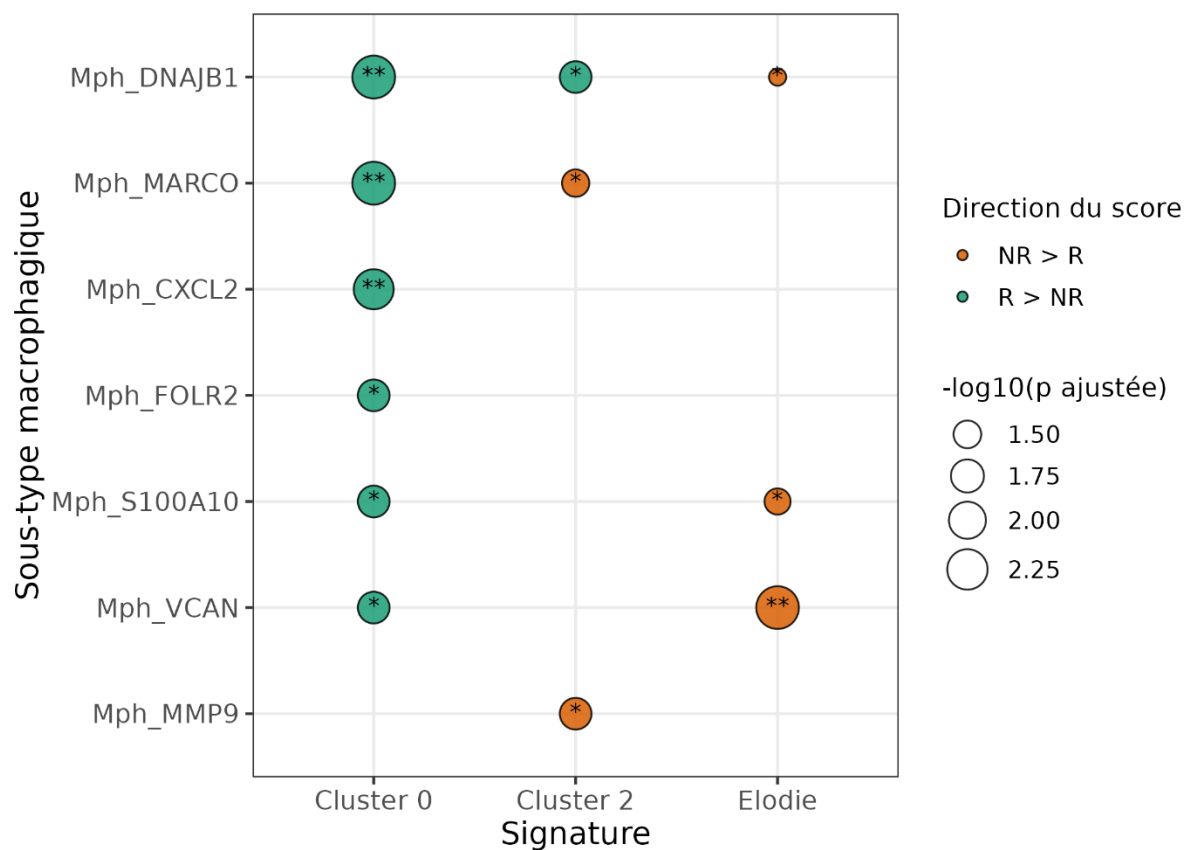


Figure 14. Sous-types macrophagiques présentant une différence significative de score UCell entre répondeurs et non-répondeurs.

Chaque point correspond à une combinaison signature × sous-type macrophagique significative après correction de Benjamini-Hochberg. La couleur indique la direction du score : vert pour un score plus élevé chez les répondeurs et orange pour un score plus élevé chez les non-répondeurs. La taille des points reflète la significativité statistique sous la forme $-\log_{10}(p\text{-valeur ajustée})$. Mph correspond aux sous-types macrophagiques annotés Mφ dans les données initiales.

A l'inverse, certains gènes étaient plus exprimés chez les non-répondeurs. *RGS1* présentait une direction NR > R dans plusieurs sous-types macrophagiques, notamment Mph_CXCL2, Mph_MARCO, Mph_VCAN, Mph_FCGR3A et Mph_ISG15. *FBP1* était également plus élevée chez les non-répondeurs, notamment dans Mph_VCAN. D'autres gènes comme *PTPN7*, *NPL* et *HS3ST3B1*, présentaient aussi des effets en faveur des non-répondeurs selon les sous-types. Certains gènes du cluster 0 (notamment *PTPN7* et *HS3ST3B1*) s'expriment dans la direction opposée aux autres. Cela explique le signal mixte de cette signature avec DESeq2 (**Figure 17**).

Ces résultats indiquent que les différences entre répondeurs et non-répondeurs reposent sur des gènes spécifiques, dont l'effet dépend du contexte macrophagique.

4.3.4 Test compétitif des signatures par camera

Afin d'évaluer les signatures projetées au niveau du *gene set* complet, un test compétitif a été réalisé avec camera sur les pseudobulks macrophagiques humains. Contrairement à l'analyse différentielle gène par gène avec DESeq2, cette approche considère chaque signature comme un ensemble et teste son enrichissement au sein de chaque sous-type macrophagique.

Au total, 32 combinaisons signatures x sous-types macrophagiques ont été testées. Parmi celles-ci, une seule combinaison était significative après correction FDR : la signature du cluster 2 dans le sous-type Mph_MMP9 (FDR = 0.0028 ; 33 pseudobulks ; **Figure 16**). La direction du test (NR > R), indique un enrichissement de cette signature chez les non-répondeurs dans ce sous-type.

Aucunes autres combinaisons n'atteignaient le seuil de significativité après correction multiple. Ces résultats suggèrent que, à l'échelle des signatures complètes, les effets observés restent globalement limités et dépendants du contexte macrophagique spécifique.

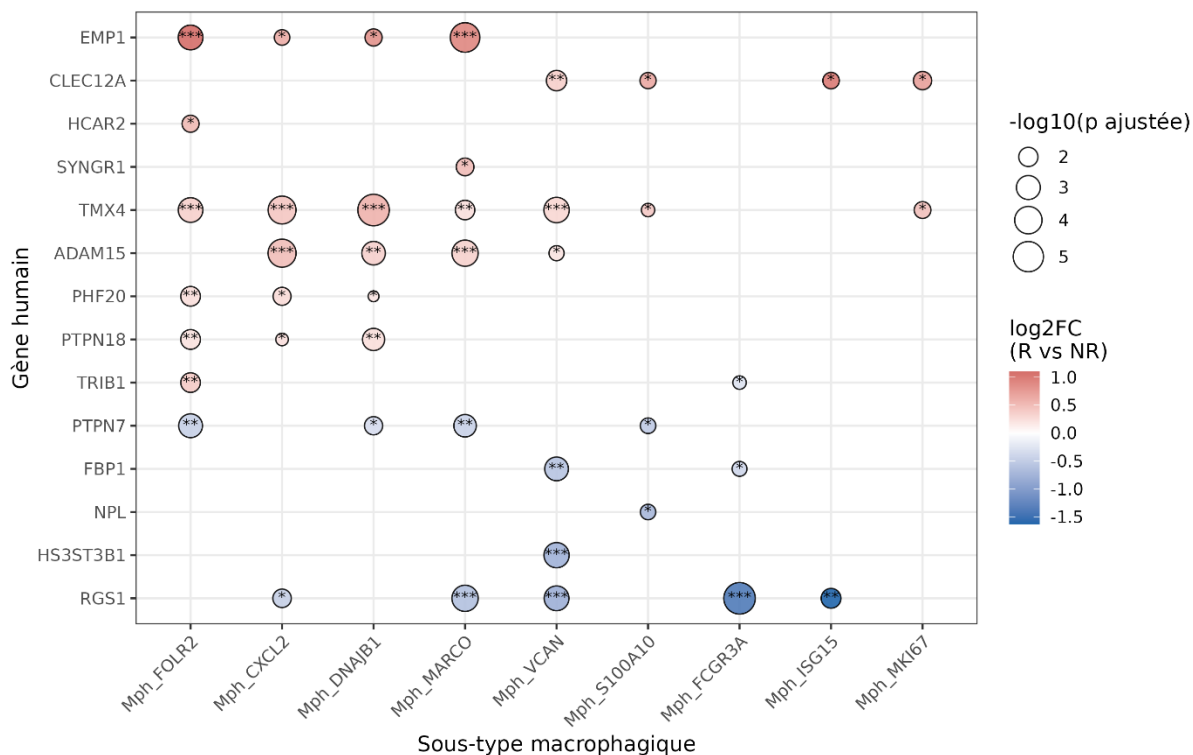


Figure 15. Analyse différentielle gène par gène des signatures murines projetées dans les pseudobulks macrophagiques humains.

Chaque point représente une combinaison gène × sous-type macrophagique significative après correction de Benjamini-Hochberg (padj < 0.05). La couleur indique la direction et l'amplitude du log2FC entre répondeurs et non-répondeurs : les valeurs positives, en rouge, correspondent à une expression plus élevée chez les répondeurs, tandis que les valeurs négatives, en bleu, correspondent à une expression plus élevée chez les non-répondeurs. La taille des points reflète le niveau de significativité statistique : plus le point est grand, plus la p-value ajustée est faible. Les étoiles indiquent le niveau de significativité (* padj < 0.05 ; ** padj < 0.01 ; *** padj < 0.001). Mph correspond aux sous-types macrophagiques annotés Mφ dans les données initiales.

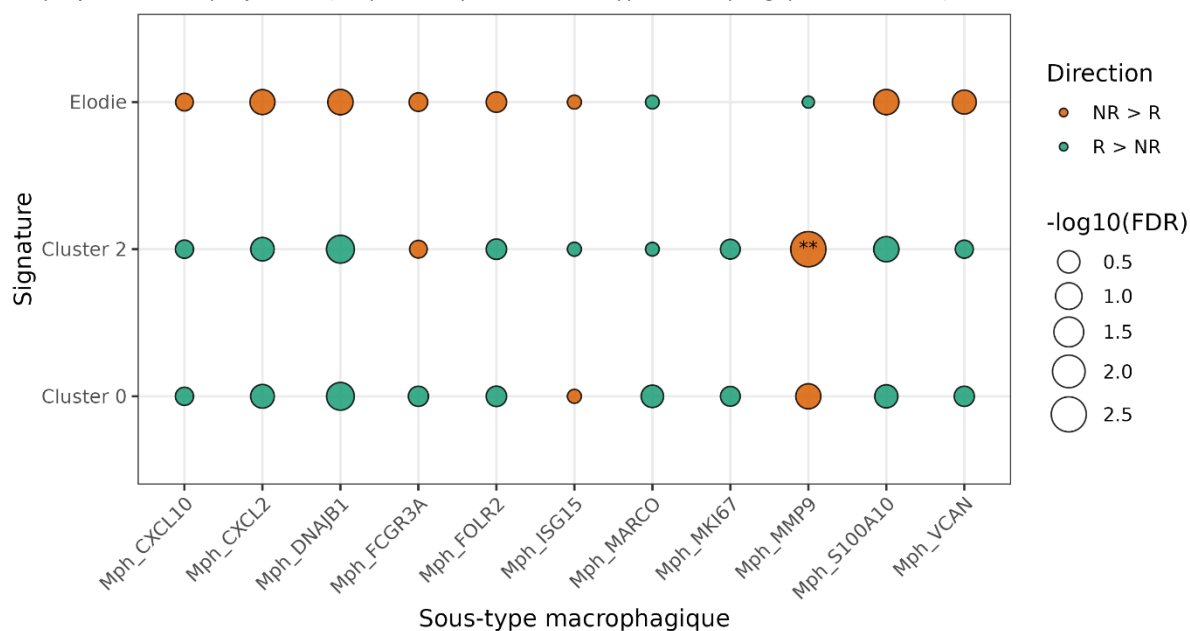


Figure 16. Test compétitif camera des signatures projetées dans les pseudobulks macrophagiques humains.

Chaque point représente une combinaison signature × sous-type macrophagique. La couleur indique la direction du test : vert pour R > NR et orange pour NR > R. La taille des points correspond à -log10(FDR), ce qui signifie que les points les plus grands ont les FDR les plus faibles. Les étoiles indiquent le niveau de significativité (* FDR < 0.05 ; ** FDR < 0.01 ; *** FDR < 0.001). Les sous-types macrophagiques annotés Mφ sont indiqués comme Mph dans la figure.

En résumé, les trois approches appliquées aux pseudobulks macrophagiques humains montrent des résultats qui varient selon la méthode utilisée (**Figure 17**). Le *scoring* UCell montre une association globale du cluster 0 avec les répondeurs et de la signature Elodie avec les non-répondeurs.

L'analyse DESeq2 gène par gène identifie plusieurs gènes différentiellement exprimés selon le statut de réponse, notamment *TMX4* et *EMP1* du côté des répondeurs, ainsi que *RGS1* et *FBP1* du côté des non-répondeurs.

Enfin, le test camera ne retrouve qu'une combinaison significative au niveau du *gene set* complet, correspondant à la signature cluster 2 dans le sous-type Mph_MMP9, orientée vers les non-répondeurs.

Ces résultats soulignent la complémentarité des approches utilisées et seront discutés afin d'évaluer leur cohérence et la pertinence biologique des signatures murines projetées dans la cohorte humaine de NSCLC.

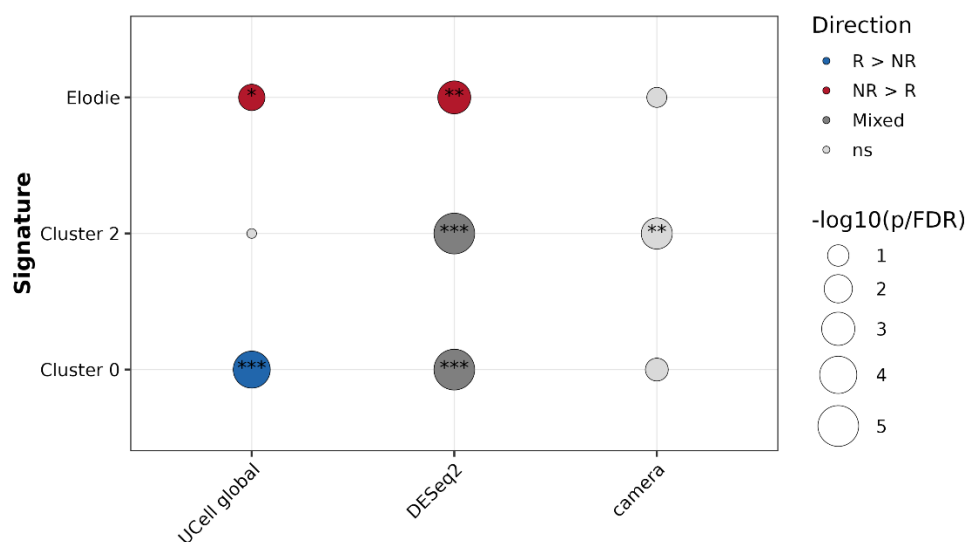


Figure 17. Synthèse de la projection humaine des signatures murines selon les trois approches analytiques. La figure résume les résultats obtenus pour les trois signatures murines projetées dans les pseudobulks macrophagiques humains. Chaque point représente le résultat global d'une signature selon une méthode donnée : *scoring* UCell global, analyse différentielle DESeq2 gène par gène, ou test compétitif camera. La couleur indique la direction du signal : bleu pour R > NR, rouge pour NR > R, gris foncé pour un signal mixte et gris clair pour un résultat non significatif. La taille des points représente le niveau de significativité sous la forme $-\log_{10}(\text{p-valeur ajustée})$ pour UCell et DESeq2, ou $-\log_{10}(\text{FDR})$ pour camera. Les étoiles indiquent le niveau de significativité statistique.

5. Discussion

5.1 Mise en perspective

Ce travail visait à étudier, par scRNA-seq, l'effet de la surexpression de *RICTOR* sur le microenvironnement immunitaire tumoral dans un modèle murin NSCLC, avec une attention portée au compartiment macrophagique. L'objectif était ensuite d'identifier une signature (macrophagique) enrichie dans le contexte KPR Late, puis d'évaluer sa projection dans une cohorte humaine de NSCLC et son association avec la réponse à l'immunothérapie néoadjuvante.

Les résultats mettent, dans l'ensemble, en évidence deux observations principales. Premièrement, le compartiment macrophagique murin apparaît fortement remodelé, avec plusieurs états transcriptomiques distincts, principalement associés au stade tumoral, tandis que l'effet du modèle KPR semble plus restreint et limité à certains sous-états macrophagiques. Deuxièmement, la projection de ces signatures dans les données humaines ne révèle pas une signature unique et homogène de résistance aux traitements par immunothérapie, mais plutôt des signaux partiels, dépendant de la signature considérée, du sous-type macrophagique et de l'approche analytique utilisée. Ces résultats doivent donc être interprétés comme exploratoires.

Il faut rappeler que le modèle KPR constitue avant tout une accélération de la progression tumorale induite par l'activation constitutive de mTORC2, et non un contexte biologique radicalement distinct du modèle KP. Retrouver des profils macrophagiques partiellement partagés entre les deux modèles n'est donc pas surprenant. Ce qui l'est en revanche, c'est l'ampleur du remodelage macrophagique observé : plus de 200 gènes différenciellement exprimés entre KPR et KP au stade Late suggèrent que l'activation de mTORC2 impacte l'état transcriptomique des macrophages tumoraux, au-delà du fait que la tumeur progresse plus vite.

Il est aussi important de préciser que le mot « prédictif » n'a pas lieu d'être utilisé dans ce contexte. Un biomarqueur prédictif correspond à un marqueur mesuré avant traitement, permettant d'estimer la probabilité de réponse d'un patient à une stratégie thérapeutique donnée, et donc ainsi d'orienter la décision thérapeutique. Cependant, la cohorte humaine utilisée, issue de Liu et al., repose sur des échantillons de résection obtenus après traitement néoadjuvant par chimiothérapie, et non sur des biopsies pré-thérapeutiques⁷². Les associations observées entre les scores de signature et le statut de réponse reflètent donc un état du microenvironnement tumoral après traitement, influencé à la fois par les caractéristiques initiales de la tumeur mais aussi par les effets thérapeutiques. L'interprétation doit ainsi rester associative : ces résultats ne permettent pas de conclure à une valeur prédictive, mais suggèrent un lien entre certains états macrophagiques post-traitement et la réponse pathologique^{15,16,111}.

5.2 Remodelage du compartiment macrophagique murin et portée réelle de l'effet RICTOR

Le premier résultat marquant de l'analyse murine est la forte dominance du compartiment macrophagique au sein du microenvironnement immunitaire. Dans l'objet global annoté, les macrophages représentent 32 610 des 46 521 cellules CD45+, soit environ 70 % des cellules, ce qui en fait la population immunitaire majoritaire. Cette observation est en accord avec la littérature du NSCLC, où les macrophages constituent un des types majeurs du compartiment myéloïde et jouent un rôle dans l'organisation du microenvironnement tumoral^{141,43–45}.

Cependant, cette population est loin d'être homogène. De nombreuses études en scRNA-seq, chez la souris comme chez l'humain, ont montré l'existence de plusieurs états macrophagiques aux fonctions distinctes. Dans ce contexte, il était pertinent de se concentrer spécifiquement sur ce compartiment^{37,45,48}. Le reclustering confirme cette hétérogénéité, avec l'identification de dix sous-populations macrophagiques dont la composition varie selon les groupes expérimentaux. Cela indique qu'une analyse globale aurait probablement masqué des signaux spécifiques à certains sous-états, et va dans le sens de l'approche consistant à construire des signatures à partir de clusters ciblés¹¹². Des études similaires sur des modèles murins de NSCLC ont aussi montré un remodelage macrophagique important au cours de la progression tumorale. Casanova-Acebes et al. ont vu, par scRNA-seq dans un modèle orthotopique KP, que les macrophages résidents tissulaires constituent une niche pro-tumorale dès les stades précoces, et que leur transcriptome se modifie rapidement pour favoriser le remodelage tissulaire et l'échappement immunitaire⁴⁵. Ces observations sont cohérentes avec nos résultats, qui montrent aussi une forte dominance macrophagique et des changements transcriptomiques marqués dès le stade Early, amplifiés au stade Late. Plus récemment, Karatza et al. ont fait une analyse scRNA-seq comparative sur trois modèles GEMM de LUAD (KL, KP et EGFR). Ils ont montré que les macrophages adoptent des états transcriptomiques distincts selon le génotype tumoral, ce qui correspond à notre approche centrée sur un modèle KPR distinct du modèle KP. Ces résultats, même s'ils viennent d'un preprint non encore évalué par les pairs, apportent un contexte comparatif pertinent pour notre démarche¹¹³.

L'impact du modèle KPR apparaît néanmoins plus limité que celui du stade tumoral. En effet, le nombre de DEG est nettement plus élevé dans les comparaisons Early vs Late que dans la comparaison KPR Late vs KP Late : 2871 DEG pour KP_Late versus KP_Early et 2777 pour KPR_Late versus KPR_Early, contre 217 DEG pour KPR_Early vs KP_Early et 212 pour KPR_Late vs KP_Late. Cela ne doit pas être interprété comme une mesure directe de l'intensité biologique de chaque effet, car le nombre de DEG dépend aussi des seuils statistiques, de la puissance de détection et des facteurs techniques⁷⁸. Il suggère toutefois que dans l'analyse différentielle pseudo-bulk faite dans les sous-types macrophagiques, que le signal associé à la progression tumorale est plus étendu que celui associé au modèle KPR. L'effet associé à *RICTOR*, bien que détectable, apparaît donc plus limité dans le cadre de cette analyse.

Cette interprétation doit toutefois être nuancée par une limite importante du jeu de données : le stade tumoral est confondu avec la chimie de séquençage (5' pour Early, 3' pour Late). Même si l'intégration par Harmony permet de réduire cet effet dans les analyses de reclustering, elle ne permet pas de le dissocier complètement du signal biologique. Ainsi, les différences observées entre Early et Late doivent être interprétées avec prudence^{70,71,93}. A l'inverse, la comparaison KPR_Early vs KP_Early et KPR Late vs KP Late, réalisée dans des conditions techniques identiques, fournit une estimation plus fiable de l'effet propre au modèle KPR, qui apparaît réel mais modéré.

5.3 Ce que les signatures capturent réellement

Les signatures macrophagiques retenues ne correspondent pas à un signal biologique unique et doivent être interprétées en fonction de leur origine. Les signatures des clusters 0 et 2 proviennent de sous-populations identifiées par reclustering, puis caractérisées par des analyses différentielles liées au contexte KPR_Late. Elles reflètent ainsi à la fois la composition des clusters et les gènes associés à ces états macrophagiques. Cette distinction explique pourquoi elles ne présentent pas toutes le même niveau de spécificité vis-à-vis du modèle KPR.

Le cluster 2 est celui qui se rapproche le plus de l'hypothèse initiale liée au modèle KPR, selon laquelle la surexpression de RICTOR dans l'épithélium pulmonaire accélère la tumorigenèse et pourrait remodeler le compartiment immunitaire. Il est majoritairement composé de cellules KPR (81.6 %), dont une grande proportion de cellules KPR Late. Sa signature reflète donc des gènes surexprimés dans ce cluster et associés, dans les comparaisons différentielles, au contexte KPR_Late. Elle est ainsi la meilleure candidate pour tester, dans les données humaines, la projection d'un signal lié au modèle KPR.

Le cluster 0 présente un profil différent. Il est également enrichi en cellules Late, qui représentent 71.9 % du cluster, mais cet enrichissement n'est pas spécifique au modèle KPR puisqu'il inclut à la fois des cellules KP_Late et KPR_Late. Sa signature a cependant été construite selon la même logique différentielle, en retenant des gènes associés au contexte KPR_Late dans ce cluster. Elle doit être interprétée comme signature issue d'un état macrophagique tardif, partagé entre KP et KPR, plutôt que comme une signature strictement spécifique du modèle KPR. Cela est cohérent avec les études de scRNA-seq dans le NSCLC, qui montrent l'existence de plusieurs états macrophagiques, certains partagés entre contextes tumoraux et d'autres plus spécifiques^{37,48,50}.

La signature d'Elodie occupe une position différente dans l'analyse. Contrairement aux signatures des clusters 0 et 2, elle ne provient pas directement des comparaisons différentielles utilisées pour construire ces signatures, mais d'une sélection raisonnée basée sur de précédentes expériences réalisées sur ce modèle et sur les résultats FindMarkers clusters versus cluster. Elle regroupe des gènes comme *Fbp1*, *Arg1*, *Npl* et *Chil3*, liés à des états macrophagiques enrichis dans les conditions KPR et/ou Late. Cette signature ne correspond donc pas à un cluster précis, mais à un signal ciblé, construit pour refléter un programme

macrophagique associé au modèle KPR, dont la projection permet d'évaluer sa présence et son lien avec la réponse chez les patients.

Enfin, la portée de ces signatures reste exploratoire, car leur construction repose sur des comparaisons différentielles murines pour lesquelles les vrais réplicats biologiques par souris ne pouvaient pas être reconstruits. Les signatures obtenues doivent donc être considérées comme des hypothèses issues du modèle murin, à projeter et tester dans les données humaines, plutôt que comme des programmes transcriptomiques définitivement validés^{96,97}.

5.4 Projection humaine : un signal partiel et dépendant du contexte

La projection des signatures murines dans la cohorte humaine de 243 patients, 9 320 pseudobulks et 51 types cellulaires, est une partie importante. Trois approches complémentaires ont été utilisées : le *scoring* UCell dans les pseudobulks macrophagiques, l'analyse différentielle gène par gène avec DESeq2, et le test compétitif camera par sous-types macrophagiques. Ces méthodes ne posent pas exactement la même question : Ucell donne le score global de la signature, DESeq2 teste les gènes individuellement, et camera regarde l'enrichissement d'un *gene set* complet. C'est probablement ce qui explique en partie les différences observées entre les résultats.

Le signal le plus clair concerne la signature du cluster 0. Les patients répondeurs ont des scores UCell plus élevés que les non-répondeurs dans plusieurs sous-types macrophagiques, et DESeq2 confirme que plusieurs gènes de cette signature sont plus exprimés chez les répondeurs, surtout *TMX4* (retrouvé dans six sous-types) et *EMP1*. Ces gènes doivent toutefois être interprétés comme des marqueurs exploratoires de cette association, et non comme la preuve d'un mécanisme fonctionnel direct. Ce résultat est intéressant car le cluster 0 n'est pas le signal le plus spécifique du modèle KPR : il correspond plutôt à un état macrophagique tardif, partagé entre KP Late et KPR Late. Le fait qu'il soit associé aux répondeurs suggère qu'un programme macrophagique lié à un contexte tumoral avancé dans le modèle murin ne correspond pas nécessairement à un signal de résistance chez l'humain.

Cette observation doit être interprétée avec prudence. La littérature montre que les macrophages du microenvironnement tumoral sont hétérogènes et que leur rôle dépend beaucoup du contexte biologique et thérapeutique¹¹⁴. Les études scRNA-seq menées dans le NSCLC traité en contexte néoadjuvant montrent également que la chimio-immunothérapie peut remodeler le microenvironnement tumoral et générer des profils immunitaires distincts selon la réponse pathologique^{35,114}. Dans cette cohorte post-traitement, l'association de la signature du cluster 0 avec les répondeurs pourrait refléter un état macrophagique compatible avec la réponse pathologique, sans permettre de déterminer si cet état était préexistant ou induit par le traitement.

A l'inverse, la signature du cluster 2, qui est pourtant la plus directement liée à l'hypothèse KPR, ne montre pas de signal global robuste entre répondeurs et non-répondeurs. Le test camera détecte un

enrichissement de cette signature dans le sous-type Mph_MMP9 du côté des non-répondeurs, mais ce résultat repose sur seulement 33 pseudobulks et doit donc être interprété avec prudence.

L'analyse gène par gène nuance toutefois cette absence de signal global. En effet, *RGS1*, qui appartient à la signature du cluster 2, est davantage exprimé chez les non-répondeurs dans plusieurs sous-types macrophagiques. Ce résultat est intéressant car il montre que certains gènes de la signature du cluster 2 se projettent bien dans la cohorte humaine, même si la signature complète ne présente pas d'association globale robuste. Ainsi, l'absence de signal au niveau du *gene set* complet ne signifie pas une absence totale de projection biologique, mais suggère plutôt un signal partiel, porté par certains gènes comme *RGS1* et dépendant du sous-type macrophagique considéré.

L'absence de signal global n'invalide donc pas l'hypothèse initiale. Elle suggère plutôt que le signal KPR murin ne se projette pas de manière homogène dans les macrophages humains, mais que certains composants de cette signature, notamment *RGS1*, pourraient conserver une pertinence fonctionnelle dans certains sous-types macrophagiques spécifiques. Des études récentes, notamment dans le cancer du sein triple négatif et le NSCLC, ont montré que *RGS1* est plus exprimé dans les TAMs de type M2. Son inhibition peut rendre les tumeurs plus sensibles à l'anti-PD-1 en restaurant l'infiltration et l'activité des lymphocytes T CD8+, ce qui suggère un rôle dans le maintien d'un microenvironnement immunosuppresseur¹¹⁵. Son enrichissement chez les non-répondeurs dans plusieurs sous-types macrophagiques de notre cohorte va dans le même sens, même si son rôle exact dans les macrophages tumoraux du NSCLC reste à confirmer.

La signature Elodie présente un profil opposé à celui du cluster 0. L'analyse UCell globale montre un enrichissement de cette signature chez les non-répondeurs, tandis que l'analyse gène par gène par DESeq2 met en évidence certains de ses composants dans des sous-types macrophagiques spécifiques, notamment *FBP1* dans Mph_VCAN et *NPL* dans Mph_S100A10. Ces résultats suggèrent qu'un signal macrophagique plus ciblé, associé aux états KPR et/ou Late dans le modèle murin, pourrait être retrouvé chez certains patients non-répondeurs humains.

Cette interprétation doit toutefois être nuancée par l'analyse des gènes individuels composant la signature. *FBP1* ressort dans Mph_VCAN, mais la littérature sur le NSCLC décrit principalement ce gène dans le compartiment tumoral, notamment dans des processus métaboliques et de progression tumorale^{116,117}. Cette littérature ne peut donc pas être directement transposée à notre résultat, qui concerne des pseudobulks macrophagiques. *NPL* ressort également dans Mph_S100A10. Cependant, les données disponibles sur *NPL* dans les macrophages du NSCLC restent limitées, ce qui en fait plutôt un signal exploratoire associé à un sous-type macrophagique précis. Enfin, *ARG1* mérite une interprétation particulière. Ce gène faisait partie de la signature murine initiale, ce qui est cohérent avec son utilisation fréquente comme marqueur d'états macrophagiques alternatifs/immunosuppresseurs dans les modèles murins. Cependant, dans le NSCLC humain, *ARG1* a été rapporté comme étant principalement associé aux

cellules de lignée neutrophilique plutôt qu'aux macrophages. Le fait qu'*ARG1* soit quasi absent des pseudobulks macrophagiques humains analysés ici n'est donc pas surprenant et suggère que cette composante murine de la signature ne se transpose pas directement au compartiment macrophagique humain¹¹⁸⁻¹²⁰. Ainsi, le signal observé chez les non-répondeurs semble principalement porté par les composants détectables de la signature, notamment *FBP1* et *NPL*, dans des sous-types macrophagiques spécifiques.

Au final, ces résultats mettent en évidence une divergence apparente : le cluster 0 et la signature d'Élodie sont tous deux liés à des états Late et/ou KPR dans le modèle murin, mais ils s'associent à des directions opposées dans la cohorte humaine, avec une association aux répondeurs pour le cluster 0 et aux non-répondeurs pour la signature d'Élodie. Cette observation indique que le contexte Late ne correspond pas à un programme macrophagique unique ou homogène. L'association observée chez l'humain ne semble donc pas dépendre uniquement de l'origine du cluster murin, mais aussi du contenu génique précis de chaque signature.

Un autre phénomène qui peut expliquer ces divergences est la différence entre les macrophages murins et humains. Même quand un gène est retrouvé dans les deux espèces, son rôle dans les macrophages peut être différent et donner des fonctions distinctes. Par exemple, *MARCO* est un marqueur spécifique des macrophages alvéolaires murins, qui sont des macrophages résidents, alors que chez l'humain son expression basale est relativement faible et il est plutôt associé à certaines populations de macrophages associés aux tumeurs. De même, *NOS2*, qui code iNOS, est fortement exprimé chez la souris mais souvent faiblement exprimé chez l'humain. Enfin, *Arginase-1*, *Ym1* et *Fizz1* sont des marqueurs M2 murins peu pertinents chez l'humain. Ces différences soulignent qu'il faut définir les macrophages humains avec des marqueurs adaptés à l'espèce humaine¹²¹.

Ainsi, certains composants du cluster 0, notamment *TMX4* et *EMP1*, sont associés aux répondeurs, tandis que certains composants de la signature d'Élodie, notamment *FBP1* et *NPL*, ressortent plutôt chez les non-répondeurs. *RGS1*, appartenant à la signature du cluster 2, constitue également un signal notable, puisqu'il est enrichi chez les non-répondeurs dans cinq sous-types macrophagiques distincts. Cette cohérence en fait une piste intéressante pour de futures analyses, même si son association avec la non-réponse reste descriptive à ce stade.

5.5 Limites, perspectives et conclusion

Les différences entre UCell, DESeq2 et camera ne sont pas une contradiction, elles montrent plutôt que le signal observé n'est pas porté de la même manière selon le niveau. UCell évalue l'activité globale d'une signature, DESeq2 teste les gènes un par un, et camera regarde l'enrichissement coordonné d'un *gene set* complet, en tenant compte de la corrélation entre gènes. Le fait que DESeq2 trouve plusieurs gènes significatifs, tandis que camera ne détecte qu'un signal localisé dans *Mph_MMP9* pour le cluster 2, suggère

que la projection humaine repose plus sur quelques gènes porteurs du signal que sur une activation homogène de toute la signature. Cela va dans le sens d'une réduction du nombre de gènes des signatures murines à un nombre plus limité de gènes robustes.

La problématique principale concerne l'analyse murine : les souris d'une même condition ont été poolées avant la préparation des librairies, sans marquage HTO ni autre stratégie de démultiplexage individuel. Il n'est donc pas possible de reconstruire l'identité de chaque souris après séquençage. Les comparaisons différentielles murines doivent alors être vues comme exploratoires, car elles comparent essentiellement des conditions poolées plutôt que de vrais réplicats biologiques indépendants. Une limite supplémentaire est que seuls deux clusters macrophagiques ont été retenus pour les signatures, sur la base de leur composition conditionnelle. D'autres états macrophagiques dans le dataset n'ont pas été explorés, et d'autres signatures auraient pu être faites avec d'autres clusters ou d'autres stratégies de sélection de gènes. Les signatures testées ne sont donc qu'un sous-ensemble des programmes transcriptomiques potentiellement liés au contexte KPR Late. Une deuxième limite importante concerne la cohorte humaine, faite de pièces de résection après chimio-immunothérapie néoadjuvante. Sans biopsies pré-traitement, les associations observées ne peuvent pas être interprétées comme des biomarqueurs prédictifs, mais plutôt comme des états du microenvironnement tumoral au moment de la résection, potentiellement remodelés par le traitement. Enfin, la projection inter-espèces reste une limite importante : la conversion orthologique entraîne des pertes et ne garantit pas une conservation fonctionnelle parfaite, comme on le voit avec *ARG1*, présent dans la signature murine mais quasi absent des pseudobulks macrophagiques humains. Par ailleurs, la cohorte a un déséquilibre histologique notable : les répondeurs sont majoritairement des LUSC (109/130, 83.8 %), tandis que les non-répondeurs ont plus de LUAD (42/112, 37.5 %). Ce déséquilibre peut biaiser les résultats, car le LUSC et le LUAD ont des microenvironnements tumoraux distincts et des profils macrophagiques différents. Les associations entre les signatures et le statut de réponse pourraient donc refléter en partie des différences histologiques, pas juste un signal lié à la réponse thérapeutique.

L'utilisation de trois approches analytiques complémentaires — UCell, DESeq2 et camera —, combinée à une cohorte humaine de 243 patients, permet quand même d'évaluer la robustesse du signal, même si les limites décrites avant en nuancent la portée. Il est par ailleurs important de souligner que ces résultats ne doivent pas être interprétés comme négatifs : identifier des changements transcriptomiques macrophagiques aussi marqués à partir d'une surexpression de RICTOR limitée à l'épithélium pulmonaire constitue en soi un résultat notable. Les perspectives principales seraient de valider ces signatures dans une cohorte indépendante, idéalement avec des biopsies pré-traitement et des données de survie, afin d'évaluer leur reproductibilité et leur éventuelle valeur prédictive. Il serait également pertinent de réduire les signatures à quelques gènes robustes, notamment *TMX4* et *EMP1* du côté des répondeurs, et *RGS1*, *FBP1* ou *NPL* du côté des non-répondeurs. Ces candidats pourraient ensuite faire l'objet d'une validation complémentaire à l'échelle tissulaire, par immunohistochimie ou transcriptomique spatiale, afin de

confirmer leur expression et de mieux situer les populations macrophagiques porteuses du signal dans le microenvironnement tumoral. Côté murin, une expérience avec réplicats biologiques individuels et démultiplexage par marquage HTOPermettrait de générer des signatures plus robustes.

En conclusion, ce travail montre que le compartiment macrophagique murin est remodelé dans le modèle KPR, mais que sa projection dans la cohorte humaine NSCLC ne révèle pas une signature unique de résistance. Les signaux observés sont partiels, dépendants du sous-type macrophagique et du contenu génique de chaque signature. Le lien entre RICTOR, macrophages et réponse à l'immunothérapie reste donc indirect à ce stade, et devra être validé dans des cohortes indépendantes.

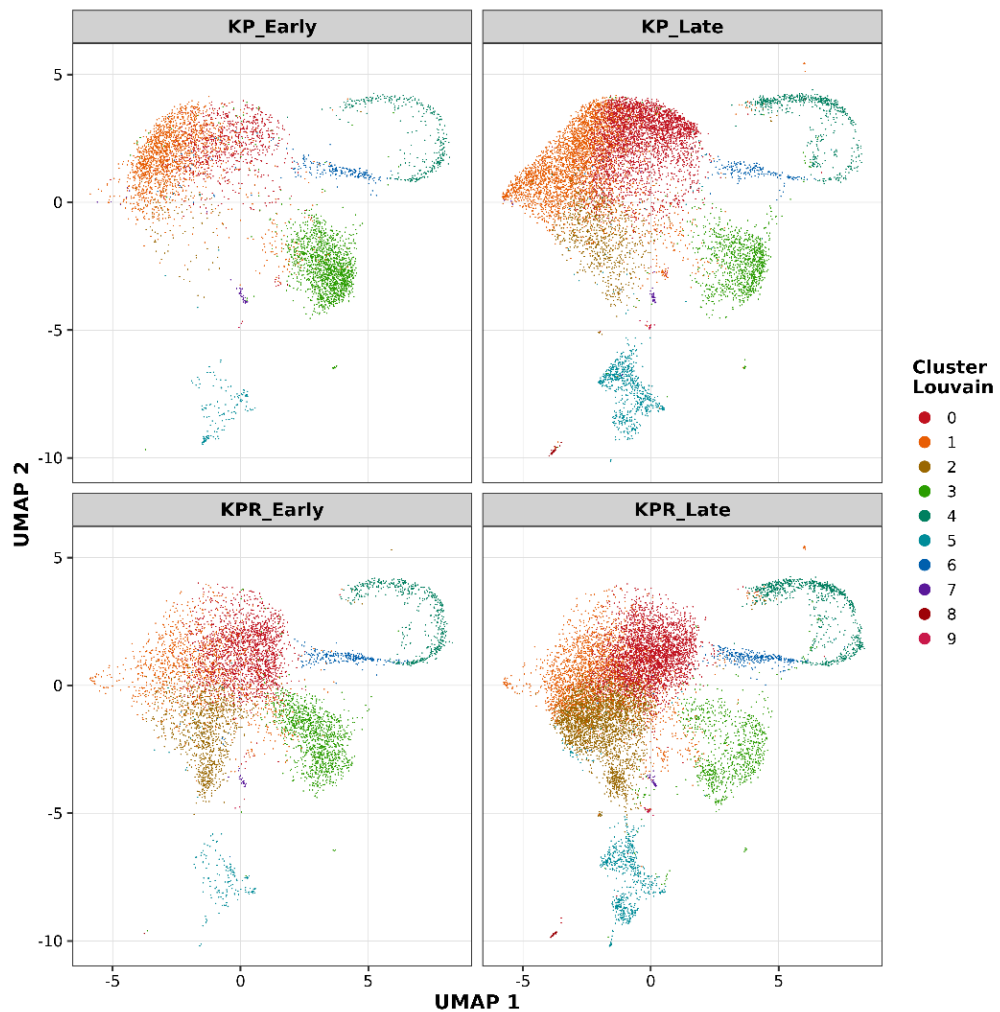
6. Bibliographie

- Bray, F. *et al.* Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. *CA. Cancer J. Clin.* **74**, 229–263 (2024).
- Zhou, J. *et al.* Global burden of lung cancer in 2022 and projections to 2050: Incidence and mortality estimates from GLOBOCAN. *Cancer Epidemiol.* **93**, 102693 (2024).
- Cancer Fact Sheet Lung Cancer, 2023 | Belgian Cancer Registry BELGE SITE. <https://kankerregister.org/en/publicaties/cancer-fact-sheet-lung-cancer-2023>.
- Cancer du poumon OMS. <https://www.who.int/fr/news-room/fact-sheets/detail/lung-cancer>.
- Cancer, I. N. D. Cancer du poumon site fr. <https://www.cancer.fr/professionnels-de-sante/statistiques-et-chiffres-sur-les-cancers/epidemiologie-des-cancers/cancer-du-poumon> (2025).
- Table:Survie à 5 ans de certains types de cancers. *Édition professionnelle du Manuel MSD* <https://www.msmanuals.com/fr/professional/multimedia/table/survie-a-5-ans-de-certains-types-de-cancers>.
- Cancer thoracique et pulmonaire | Institut Roi Albert II. <https://www.institutroi-albert.be/fr/cancer/cancer-thoracique-et-pulmonaire>.
- Nicholson, A. G. *et al.* The 2021 WHO Classification of Lung Tumors: Impact of Advances Since 2015. *J. Thorac. Oncol.* **17**, 362–387 (2022).
- Luo, G. *et al.* Estimated worldwide variation and trends in incidence of lung cancer by histological subtype in 2022 and over time: a population-based study. *Lancet Respir. Med.* **13**, 348–363 (2025).
- Sabbula, B. R., Gasalberti, D. P., Mukkamalla, S. K. R. & Anjum, F. Squamous Cell Lung Cancer. in *StatPearls* (StatPearls Publishing, Treasure Island (FL), 2026).
- Rocca, A. *et al.* Refining treatment strategies for non-small cell lung cancer lacking actionable mutations: insights from multi-omics studies. *Br. J. Cancer* **133**, 1405–1427 (2025).
- Budczies, J. *et al.* KRAS and TP53 co-mutation predicts benefit of immune checkpoint blockade in lung adenocarcinoma. *Br. J. Cancer* **131**, 524–533 (2024).
- Travis, W. D. *et al.* The IASLC Lung Cancer Staging Project: Proposals for Coding T Categories for Subsolid Nodules and Assessment of Tumor Size in Part-Solid Tumors in the Forthcoming Eighth Edition of the TNM Classification of Lung Cancer. *J. Thorac. Oncol.* **11**, 1204–1223 (2016).
- Traitement du cancer du poumon non à petites cellules - NCI. <https://www.cancer.gov/types/lung/patient/non-small-cell-lung-treatment-pdq> (2026).
- Wakelee, H. *et al.* Perioperative Pembrolizumab for Early-Stage Non-Small-Cell Lung Cancer. *N. Engl. J. Med.* **389**, 491–503 (2023).
- Forde, P. M. *et al.* Neoadjuvant Nivolumab plus Chemotherapy in Resectable Lung Cancer. *N. Engl. J. Med.* **386**, 1973–1985 (2022).
- Panwar, V. *et al.* Multifaceted role of mTOR (mammalian target of rapamycin) signaling pathway in human health and disease. *Signal Transduct. Target. Ther.* **8**, 375 (2023).
- Regulation and metabolic functions of mTORC1 and mTORC2 | Physiological Reviews | American Physiological Society. <https://journals.physiology.org/doi/full/10.1152/physrev.00026.2020>.
- Sztankovics, D. *et al.* mTOR hyperactivity and RICTOR amplification as targets for personalized treatments in malignancies. *Pathol. Oncol. Res.* **30**, (2024).
- Ragupathi, A., Kim, C. & Jacinto, E. The mTORC2 signaling network: targets and cross-talks. *Biochem. J.* **481**, 45–91 (2024).
- Battaglion, S., Benjamin, D., Wälchli, M., Maier, T. & Hall, M. N. mTOR substrate phosphorylation in growth control. *Cell* **185**, 1814–1836 (2022).
- Kim, L. C., Rhee, C. H. & Chen, J. Rictor amplification promotes NSCLC cell proliferation through formation and activation of mTORC2 at the expense of mTORC1. *Mol. Cancer Res. MCR* **18**, 1675–1684 (2020).
- Szalai, F. *et al.* Rictor-A Mediator of Progression and Metastasis in Lung Cancer. *Cancers* **16**, 543 (2024).
- Cheng, H. *et al.* RICTOR amplification defines a novel subset of lung cancer patients who may benefit from treatment with mTOR1/2 inhibitors. *Cancer Discov.* **5**, 1262–1270 (2015).
- Krencz, I. *et al.* RICTOR amplification is associated with Rictor membrane staining and does not correlate with PD-L1 expression in lung squamous cell carcinoma. *Pathol. Oncol. Res.* **30**, 1611593 (2024).
- Zhou, H. *et al.* Epithelial-mesenchymal reprogramming by KLF4-regulated Rictor expression contributes to metastasis of non-small cell lung cancer cells. *Int. J. Biol. Sci.* **18**, 4869–4883 (2022).
- He, X. & Xu, C. Immune checkpoint signaling and cancer immunotherapy. *Cell Res.* **30**, 660–669 (2020).
- Liu, X. *et al.* Targeting T cell exhaustion: emerging strategies in non-small cell lung cancer. *Front. Immunol.* **15**, (2024).
- Neoadjuvant checkpoint blockade for cancer immunotherapy | Science. <https://www.science.org/doi/10.1126/science.aax0182>.
- Topalian, S. L., Taube, J. M. & Pardoll, D. M. Neoadjuvant checkpoint blockade for cancer immunotherapy. *Science* **367**, eaax0182 (2020).
- Provencio, M. *et al.* Neoadjuvant chemotherapy and nivolumab in resectable non-small-cell lung cancer (NADIM): an open-label, multicentre, single-arm, phase 2 trial. *Lancet Oncol.* **21**, 1413–1422 (2020).
- Deutsch, J. S. *et al.* Association between pathologic response and survival after neoadjuvant therapy in lung cancer. *Nat. Med.* **30**, 218–228 (2024).
- Min, J. *et al.* The NSCLC immunotherapy response predicted by tumor-infiltrating T cells via a non-invasive radiomic approach. *Front. Immunol.* **15**, (2024).
- Shi, X. *et al.* Overcoming resistance to immune checkpoint inhibitor in non-small cell lung cancer: the promise of combination therapy. *Front. Immunol.* **16**, (2025).
- Tumor microenvironment remodeling after neoadjuvant immunotherapy in non-small cell lung cancer revealed by single-cell RNA sequencing | Genome Medicine | Springer Nature Link. <https://link.springer.com/article/10.1186/s13073-023-01164-9>.
- Binnewies, M. *et al.* Understanding the tumor immune microenvironment (TIME) for effective therapy. *Nat. Med.* **24**, 541–550 (2018).
- Zilionis, R. *et al.* Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity* **50**, 1317–1334.e10 (2019).
- Cords, L. *et al.* Cancer-associated fibroblast phenotypes are associated with patient outcome in non-small cell lung cancer. *Cancer Cell* **42**, 396–412.e5 (2024).

39. Peng, H. *et al.* Profiling Tumor Immune Microenvironment of Non-Small Cell Lung Cancer Using Multiplex Immunofluorescence. *Front. Immunol.* **12**, (2021).
40. De Zuani, M. *et al.* Single-cell and spatial transcriptomics analysis of non-small cell lung cancer. *Nat. Commun.* **15**, 4388 (2024).
41. Balázsová, K., Clevers, H. & Dost, A. F. The role of macrophages in non-small cell lung cancer and advancements in 3D co-cultures. *eLife* **12**, e82998 (2023).
42. Tao, Y., Xiao, J., Li, Y. & Zou, A. Tumor-Associated Macrophages in Lung Cancer: Origins, Functional Heterogeneity, and Therapeutic Implications. *J. Inflamm. Res.* **18**, 15257–15280 (2025).
43. Rahal, Z., Darzi, R. E., Moghaddam, S. J., Cascone, T. & Kadara, H. Tumour and microenvironment crosstalk in NSCLC progression and response to therapy. *Nat. Rev. Clin. Oncol.* **22**, 463–482 (2025).
44. Hao, D. & Chen, S. Targeting tumor-associated macrophages in non-small cell lung cancer: mechanisms, prognosis, and therapeutic opportunities. *Front. Immunol.* **16**, 1679537.
45. Casanova-Acebes, M. *et al.* Tissue-resident macrophages provide a pro-tumorigenic niche to early NSCLC cells. *Nature* **595**, 578–584 (2021).
46. Marquez-Galera, A., de la Prida, L. M. & Lopez-Atalaya, J. P. A protocol to extract cell-type-specific signatures from differentially expressed genes in bulk-tissue RNA-seq. *STAR Protoc.* **3**, 101121 (2022).
47. Kuksin, M. *et al.* Applications of single-cell and bulk RNA sequencing in onco-immunology. *Eur. J. Cancer* **149**, 193–210 (2021).
48. Wu, F. *et al.* Single-cell profiling of tumor heterogeneity and the microenvironment in advanced non-small cell lung cancer. *Nat. Commun.* **12**, 2540 (2021).
49. Cao, K. *et al.* Integrating bulk, single-cell, and spatial transcriptomics to identify and functionally validate novel targets to enhance immunotherapy in NSCLC. *Npj Precis. Oncol.* **9**, 112 (2025).
50. Leader, A. M. *et al.* Single-cell analysis of human non-small cell lung cancer lesions refines tumor classification and patient stratification. *Cancer Cell* **39**, 1594–1609.e12 (2021).
51. Bica, C. *et al.* Tumor heterogeneity assessment using single-cell RNA sequencing (scRNA-seq): applications in lung cancer for diagnosis and treatment. *Front. Immunol.* **16**, (2026).
52. Bin, Y. *et al.* Against all odds: The road to success in the development of human immune reconstitution mice. *Anim. Models Exp. Med.* **7**, 460–470 (2024).
53. Oswald, E. *et al.* Immune cell infiltration pattern in non-small cell lung cancer PDX models is a model immanent feature and correlates with a distinct molecular and phenotypic make-up. *J. Immunother. Cancer* **10**, e004412 (2022).
54. Gros, P. & Casanova, J.-L. Reconciling Mouse and Human Immunology at the Altar of Genetics. *Annu. Rev. Immunol.* **41**, 39–71 (2023).
55. Truchi, M. Le séquençage de transcrits sur cellule unique appliqué à l'étude des mécanismes de formation et de résolution de la fibrose pulmonaire. (Université Côte d'Azur, 2023).
56. Masopust, D., Sivula, C. P. & Jameson, S. C. Of Mice, Dirty Mice, and Men: Using Mice to Understand Human Immunology. *J. Immunol. Baltim. Md 1950* **199**, 383–388 (2017).
57. Tang, H. *et al.* A 12-Gene Set Predicts Survival Benefits from Adjuvant Chemotherapy in Non-Small-Cell Lung Cancer Patients. *Clin. Cancer Res. Off. J. Am. Assoc. Cancer Res.* **19**, 1577–1586 (2013).
58. Gene Signatures Revisited | JNCI: Journal of the National Cancer Institute | Oxford Academic.
<https://academic.oup.com/jnci/article/104/4/262/980404?guestAccessKey=#no-access-message>.
59. Stable gene expression for normalisation and single-sample scoring | Nucleic Acids Research | Oxford Academic.
<https://academic.oup.com/nar/article/48/19/e113/5913300?login=false>.
60. The impact of Oncotype DX breast cancer assay results on clinical practice: a UK experience - PMC.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC7103011/>.
61. Identification of immune-related gene signature for non-small cell lung cancer patients with immune checkpoint inhibitors - ScienceDirect.
<https://www.sciencedirect.com/science/article/pii/S2405844024030056>.
62. Benchmarking single-sample gene set scoring methods for application in precision medicine | Briefings in Bioinformatics | Oxford Academic.
<https://academic.oup.com/bib/article/26/6/bbaf684/8382580?login=false>.
63. A Robust 8-Gene Prognostic Signature for Early-Stage Non-small Cell Lung Cancer - PMC.
<https://pmc.ncbi.nlm.nih.gov/articles/PMC6684755/>.
64. Lehtiö, J. *et al.* Proteogenomics of non-small cell lung cancer reveals molecular subtypes associated with specific therapeutic targets and immune-evasion mechanisms. *Nat. Cancer* **2**, 1224–1242 (2021).
65. Jackson, E. L. *et al.* The Differential Effects of Mutant p53 Alleles on Advanced Murine Lung Cancer. *Cancer Res.* **65**, 10280–10288 (2005).
66. Rindler, T. N. *et al.* Efficient Transduction of Alveolar Type 2 Cells with Adeno-associated Virus for the Study of Lung Regeneration. *Am. J. Respir. Cell Mol. Biol.* **65**, 118–121.
67. DuPage, M., Dooley, A. L. & Jacks, T. Conditional mouse lung cancer models using adenoviral or lentiviral delivery of Cre recombinase. *Nat. Protoc.* **4**, 1064–1072 (2009).
68. Lin, S. *et al.* Surfactant protein B (SP-B) α -/- mice are rescued by restoration of SP-B expression in alveolar type II cells but not Clara cells. *J. Biol. Chem.* **274**, 19168–19174 (1999).
69. Sarbassov, D. D., Guertin, D. A., Ali, S. M. & Sabatini, D. M. Phosphorylation and regulation of Akt/PKB by the rictor-mTOR complex. *Science* **307**, 1098–1101 (2005).
70. Wallace, A. Finding the perfect fit: How to choose the right single cell assay. *10x Genomics*
<https://www.10xgenomics.com/blog/finding-the-perfect-fit-how-to-choose-the-right-single-cell-assay>.
71. What is the difference between Single Cell 3' and 5' Gene Expression libraries?
https://kb.10xgenomics.com/s/article/360000939852-What-is-the-difference-between-Single-Cell-3-and-5-Gene-Expression-libraries?utm_source=chatgpt.com.
72. Liu, Z. *et al.* A single-cell atlas reveals immune heterogeneity in anti-PD-1-treated non-small cell lung cancer. *Cell* **188**, 3081–3096.e19 (2025).
73. Hao, Y. *et al.* Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* **42**, 293–304 (2024).
74. Hafemeister, C. & Satija, R. Normalization and variance stabilization of single-cell RNA-seq data using regularized negative binomial regression. *Genome Biol.* **20**, 296 (2019).
75. Ahlmann-Eltze, C. & Huber, W. glmGamPoi: fitting Gamma-Poisson generalized linear models on single cell count data. *Bioinformatics* **36**, 5701–5702 (2021).
76. Germain, P.-L., Lun, A., Garcia Meixide, C., Macnair, W. & Robinson, M. D. Doublet identification in single-cell sequencing data using scDblFinder. *F1000Research* **10**, 979 (2021).

77. Aran, D. *et al.* Reference-based analysis of lung single-cell sequencing reveals a transitional profibrotic macrophage. *Nat. Immunol.* **20**, 163–172 (2019).
78. Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome Biol.* **15**, 550 (2014).
79. H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York, 2016.
80. dplyr: A Grammar of Data Manipulation. *R-universe* <https://tidyverse.r-universe.dev/dplyr>.
81. Kolde R (2025). pheatmap: Pretty Heatmaps. R package version 1.0.13, <https://github.com/raivokolde/pheatmap>.
82. Pedersen T (2025). patchwork: The Composer of Plots. R package version 1.3.2.9000, <https://patchwork.data-imaginist.com>.
83. Chapter 4 The SingleCellExperiment Class | Introduction to Single-Cell Analysis with Bioconductor.
84. Hoffman P (2023). SeuratDisk: Interfaces for HDF5-Based Single Cell File Formats. R package version 0.0.0.9021, <https://github.com/mojaveazure/seurat-disk>.
85. Andreatta, M. & Carmona, S. J. UCell: Robust and scalable single-cell gene signature scoring. *Comput. Struct. Biotechnol. J.* **19**, 3796–3798 (2021).
86. Ritchie, M. E. *et al.* limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47 (2015).
87. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**, 139–140 (2010).
88. homologene: Quick Access to Homologene and Gene Annotation Updates. *R-universe* <https://ogannm.r-universe.dev/homologene>.
89. Luecken, M. D. & Theis, F. J. Current best practices in single-cell RNA-seq analysis: a tutorial. *Mol. Syst. Biol.* **15**, e8746 (2019).
90. McGinnis, C. S., Murrow, L. M. & Gartner, Z. J. DoubletFinder: Doublet Detection in Single-Cell RNA Sequencing Data Using Artificial Nearest Neighbors. *Cell Syst.* **8**, 329–337.e4 (2019).
91. Stuart, T. *et al.* Comprehensive Integration of Single-Cell Data. *Cell* **177**, 1888–1902.e21 (2019).
92. Fast integration using reciprocal PCA (RPCA). https://satijalab.org/seurat/articles/integration_rpca.html?utm_source=chatgpt.com.
93. Korsunsky, I. *et al.* Fast, sensitive and accurate integration of single-cell data with Harmony. *Nat. Methods* **16**, 1289–1296 (2019).
94. Becht, E. *et al.* Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2019).
95. Heng, T. S. P., Painter, M. W., & Immunological Genome Project Consortium. The Immunological Genome Project: networks of gene expression in immune cells. *Nat. Immunol.* **9**, 1091–1094 (2008).
96. Squair, J. W. *et al.* Confronting false discoveries in single-cell differential expression. *Nat. Commun.* **12**, 5692 (2021).
97. Murphy, A. E. & Skene, N. G. A balanced measure shows superior performance of pseudobulk methods in single-cell RNA-sequencing analysis. *Nat. Commun.* **13**, 7851 (2022).
98. Shafi, A., Nguyen, T., Peyvandipour, A. & Draghici, S. GSMA: an approach to identify robust global and test Gene Signatures using Meta-Analysis. *Bioinformatics* **36**, 487–495 (2020).
99. Wu, T. *et al.* Tau polarizes an aging transcriptional signature to excitatory neurons and glia. *eLife* **12**, e85251 (2023).
100. Home - Gene - NCBI. <https://www.ncbi.nlm.nih.gov/gene/>.
101. Wang, Y. *et al.* Chitinase-like proteins de-N-glycosylating CD36 modify cholesterol metabolism in atherosclerotic macrophages. *Nat. Commun.* <https://doi.org/10.1038/s41467-026-71388-x> (2026) doi:10.1038/s41467-026-71388-x.
102. Ober, C. & Chupp, G. L. The Chitinase and Chitinase-Like Proteins: A Review of Genetic and Functional Studies in Asthma and Immune-Mediated Diseases. *Curr. Opin. Allergy Clin. Immunol.* **9**, 401–408 (2009).
103. Kanneganti, M., Kamba, A. & Mizoguchi, E. Role of chitotriosidase (chitinase 1) under normal and disease conditions. *J. Epithel. Biol. Pharmacol.* **5**, 1–9 (2012).
104. Tegenfeldt, F. *et al.* OrthoDB and BUSCO update: annotation of orthologs with wider sampling of genomes. *Nucleic Acids Res.* **53**, D516–D522 (2024).
105. Crowell, H. L. *et al.* muscat detects subpopulation-specific state transitions from multi-sample multi-condition single-cell transcriptomics data. *Nat. Commun.* **11**, 6077 (2020).
106. Wu, D. & Smyth, G. K. Camera: a competitive gene set test accounting for inter-gene correlation. *Nucleic Acids Res.* **40**, e133 (2012).
107. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
108. Robinson, M. D. & Oshlack, A. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol.* **11**, R25 (2010).
109. Guan, F. *et al.* Tissue macrophages: origin, heterogeneity, biological functions, diseases and therapeutic targets. *Signal Transduct. Target. Ther.* **10**, 93 (2025).
110. Zilionis, R. *et al.* Single-Cell Transcriptomics of Human and Mouse Lung Cancers Reveals Conserved Myeloid Populations across Individuals and Species. *Immunity* **50**, 1317–1334.e10 (2019).
111. La Thangue, N. B. & Kerr, D. J. Predictive biomarkers: a paradigm shift towards personalized cancer medicine. *Nat. Rev. Clin. Oncol.* **8**, 587–596 (2011).
112. Squires, M. & Qiu, P. Recursive Clustering of Cellular Diversity in scRNA-Seq Data. *J. Comput. Biol. J. Comput. Mol. Cell Biol.* **32**, 444–460 (2025).
113. Karatza, A. *et al.* Cross-species analysis identifies genotype-driven vulnerabilities in lung adenocarcinoma. 2025.08.24.671070 Preprint at <https://doi.org/10.1101/2025.08.24.671070> (2025).
114. Toledo, B. *et al.* Deciphering the performance of macrophages in tumour microenvironment: a call for precision immunotherapy. *J. Hematol. Oncol. J. Hematol Oncol* **17**, 44 (2024).
115. Targeting CREB1/p300-mediated RGS1 expression in tumor-associated macrophages improves the efficacy of anti-PD-1 therapy in triple-negative breast cancer - PubMed. <https://pubmed.ncbi.nlm.nih.gov/41047025/>.
116. Wang, Z., He, T., Lv, W. & Hu, J. Down-regulation of FBP1 in lung adenocarcinoma cells promotes proliferation and invasion through SLUG mediated epithelial mesenchymal transformation. *Transl. Cancer Res.* **12**, 236–246 (2023).
117. He, T. *et al.* FBP1 inhibits NSCLC stemness by promoting ubiquitination of Notch1 intracellular domain and accelerating degradation. *Cell. Mol. Life Sci. CMLS* **81**, 87 (2024).
118. Rondono, R. *et al.* IL-8 induces exocytosis of arginase 1 by neutrophil polymorphonuclears in nonsmall cell lung cancer. *Int. J. Cancer* **125**, 887–893 (2009).
119. Zhou, Y., Shen, G., Zhou, X. & Li, J. Therapeutic potential of tumor-associated neutrophils: dual role and phenotypic plasticity. *Signal Transduct. Target. Ther.* **10**, 178 (2025).
120. Colombo, G., Pessolano, E., Talmon, M., Genazzani, A. A. & Kunderfranco, P. Getting everyone to agree on gene signatures for murine macrophage polarization in vitro. *PLOS ONE* **19**, e0297872 (2024).
121. Mosser, D. M. The many faces of macrophage activation. *J. Leukoc. Biol.* **73**, 209–212 (2003).

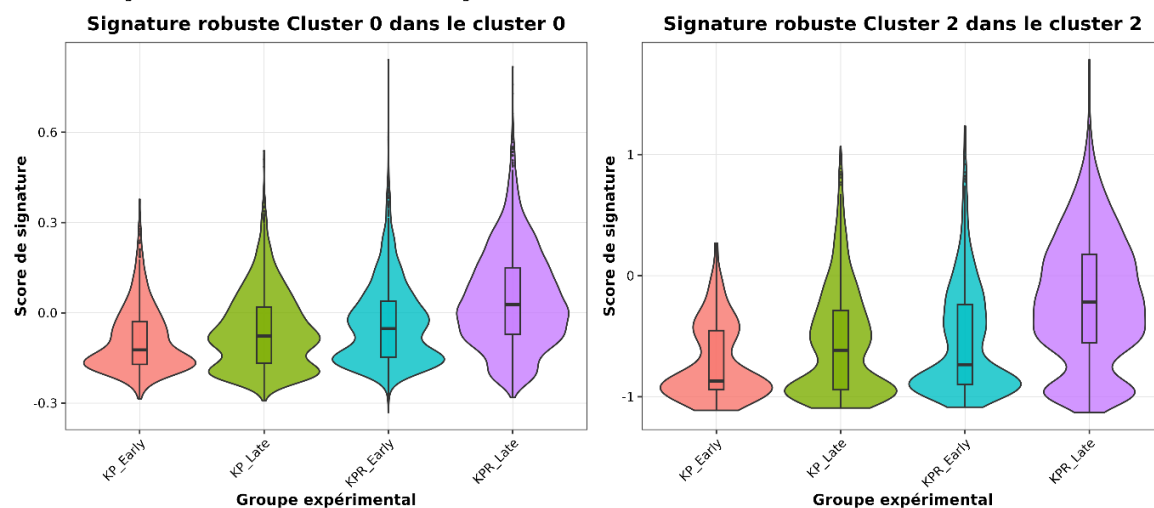
Annexe



Annexe 1. Répartition des 10 clusters macrophagiques selon les quatre groupes expérimentaux.

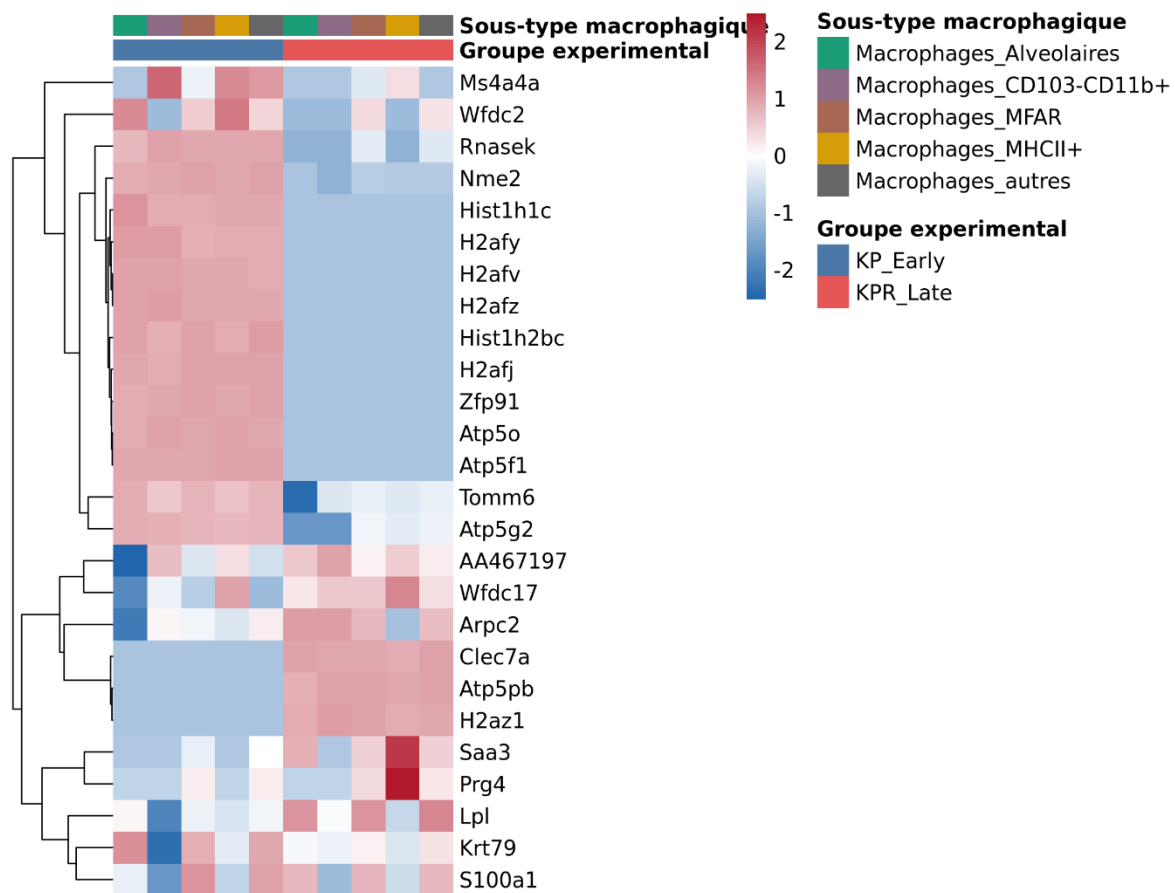
Projection UMAP des 32 610 macrophages murins issus de l'analyse Harmony, colorés selon les clusters obtenus à la résolution 0.5 (clusters 0 à 9). Les cellules sont présentées séparément pour chacun des quatre groupes expérimentaux (KP_Early, KP_Late, KPR_Early et KPR_Late) afin de visualiser la distribution relative des clusters selon la condition et le stade tumoral.

Scores des signatures robustes dans leur cluster d'origine



Annexe 2. Scores des signatures murines robustes dans leur cluster d'origine.

Les violonplots représentent les scores des signatures robustes du cluster 0 et du cluster 2 dans leur cluster macrophagique d'origine. Les groupes expérimentaux sont indiqués sur l'axe des x. Les deux signatures montrent une tendance à des scores plus élevés dans le groupe KPR Late, ce qui soutient leur association avec le contexte KPR Late avant leur projection dans les données humaines.



Annexe 3. Heatmap des gènes différemment exprimés sélectionnés pour la comparaison KPR_Late vs KP_Early.

Les colonnes représentent les pseudo-bulks macrophagiques ordonnés selon le groupe expérimental et le sous-type macrophagique, tandis que les lignes correspondent aux gènes sélectionnés. Les valeurs représentées correspondent à une expression normalisée et centrée-réduite. Le rouge indique une expression relative plus élevée et le bleu une expression relative plus faible. Cette comparaison doit être interprétée comme complémentaire, car elle combine un effet de condition génétique et un effet de stade tumoral.