Université de Lille – École Graduée Biologie – Santé

Institut de Recherches Interdisciplinaires en Cancérologie de Lille – ONCOLille – UMR 9020 CNRS / UMR 1277

Inserm CANTHER

THESE DE DOCTORAT

En vue de l'obtention du titre de

Docteur de l'Université de Lille

Discipline : Biologie cellulaire et moléculaire

Présentée par :

Geoffrey Pawlak

Biologie des systèmes et intelligence artificielle pour l'identification des réseaux de régulation de l'addiction oncogénique dans le cancer du poumon

Soutenue le 25 juin 2025 devant la commission d'examen :

Dr Charles Lecellier	Rapporteur
Dr Luca Grumolato	Rapporteur
Dr Laurence Calzone	Examinatrice
Pr Marie-Christine Copin	Présidente de jury
Pr. Mohamed Elati (Université de Lille)	Directeur de thèse
Dr. David Tulasne (Université de Lille)	Co-directeur de thèse

« The data are not as big as biology; they are a speckle compared with biology. But it is getting to that inflection point where they are big enough. And that's why it's so exciting for us. »

Aviv Regev

« La science consiste à passer d'un étonnement à un autre. »

Aristote

« Un jab à la fois, un pas à la fois, un round à la fois »

Résumé

La prise en charge du cancer du poumon repose notamment sur sa classification en sous-types histologiques et moléculaires, caractérisés par la présence ou non d'altérations moléculaires ciblables. Parmi les nombreuses altérations génétiques caractérisées, les mutations du récepteur MET, induisant un saut de l'exon 14 (METex14Del), sont retrouvés dans environ 3% des patients et représentent une nouvelle cible thérapeutique. Cependant, les mécanismes induits par METex14Del pour imposer une addiction oncogénique sont encore mal connus et les réponses aux thérapies ciblées sont perfectibles. Parallèlement, la multiplication des projets de séquençage de grande envergure génère d'importantes ressources « omiques » accessibles dans des bases publiques. Leur exploitation ouvre la voie à des approches de biologie computationnelle pouvant mener à des classifications plus fines à partir du profilage de l'expression génique des tumeurs pulmonaires. Néanmoins la diversité des techniques de séquençage rend difficile leur comparaison directe ou la fusion des données, ne permettant pas de classification unifiée à partie de large jeux de données.

Grâce aux approches de biologie des systèmes, et plus particulièrement à l'inférence de réseaux de régulation génique (GRN) appliqués au cancer du poumon, il est possible d'agréger de vastes jeux de données transcriptomiques. L'analyse de ces réseaux permet notamment d'identifier des régulateurs clés et de leur assigner une valeur représentant leur activité régulatrice, l'influence. Dans ce contexte, nous avons développé la plateforme web lung.cRegMap, qui intègre les profils d'expression de lignées cellulaires de cancer du poumon et de plusieurs cohortes comptant 5487 patients. Cette approche a permis de définir sept sous-groupes distincts de cancers du poumon, sur lesquels les données cliniques, moléculaires et histologiques ont été reportées. La classification obtenue est conforme à l'histologie, mais elle révèle également un sous type d'adénocarcinomes de mauvais pronostique et de nouveaux sous-types caractérisés par des signatures d'expression génique mésenchymateuse et neuroendocrine. Chaque groupe se distingue par un ensemble propre de régulateurs clés dont l'importance peut être estimée à partir d'essais de dépendance de ces gènes dans la prolifération et la survie de modèles cellulaires.

L'analyse en réseau de régulation et la notion d'influence des nœuds de régulation rend possible l'analyse dynamique du GRN. Nous avons tiré parti de cette originalité de notre approche pour étudier la dynamique de réponse à l'HGF des récepteurs METex14Del. Cette étude a mis en évidence l'activation d'un nœud de régulation majeur constitué de plusieurs facteurs de transcription interconnectés (ETS1, FOSL1, SMAD3) tous régulés par la voie signalisation RAS-ERK. Des validations expérimentales confirment l'implication de ces facteurs de transcriptions dans la régulation de l'expression des gènes induits par METex14Del et le rôle central de la voie RAS-ERK dans la migration et l'invasion cellulaire. Cette voie représente ainsi une cible thérapeutique prometteuse, notamment en cas de réponse incomplète aux inhibiteurs de MET.

Dans leur ensemble nos résultats valident la pertinence et la robustesse du modèle en réseau de régulation du cancer du poumon. L'approche comparative des tumeurs pulmonaires permet d'affiner les classifications avec l'identification de sous-types inédits. Par l'étude dynamique, nous avons montré que le réseau est efficace pour mettre en évidence des modules de nœuds majeurs, leur lien avec les voies de signalisation et leur implication dans le caractère transformant de METex14Del.

Abstract

Lung cancer management largely depends on classification into histological and molecular subtypes, defined by the presence or absence of actionable molecular alterations. Among the numerous genetic aberrations identified, mutations in the MET receptor that cause exon 14 skipping (METex14Del) are found in approximately 3% of patients and represent a new therapeutic target. However, the oncogenic mechanisms driven by METex14Del remain poorly understood, and responses to targeted therapies are still suboptimal. Meanwhile, the growing number of large-scale sequencing projects has generated substantial “omics” datasets, publicly accessible in various databases. These datasets enable computational biology approaches to achieve more refined classifications based on the gene expression profiles of lung tumors. Nevertheless, the diversity of sequencing technologies makes direct comparison or data merging challenging, preventing a unified classification built from large-scale datasets.

Using systems biology methods—particularly gene regulatory network (GRN) inference applied to lung cancer—it becomes possible to integrate vast transcriptomic datasets. Analyzing these networks facilitates the identification of key regulators, to which a regulatory activity metric (called influence) can be assigned. In this context, we developed the lung.cRegMap web platform, which incorporates expression profiles from lung cancer cell lines and multiple patient cohorts, totaling 5,487 samples. Through this approach, we established seven distinct subgroups of lung cancer, onto which we mapped clinical, molecular, and histological data. The resulting classification is consistent with histology yet also reveals a subgroup of adenocarcinomas with particularly poor prognosis, as well as new subtypes distinguished by mesenchymal and neuroendocrine gene expression signatures. Each group is characterized by a specific set of key regulators whose importance can be estimated via gene dependency assays in cell models.

Combining GRN analysis with the notion of regulator influence enables dynamic exploration of regulatory networks. We capitalized on this characteristic to investigate the dynamic response to HGF in cells harboring METex14Del. This study uncovered the activation of a major regulatory node composed of several interconnected transcription factors (ETS1, FOSL1, SMAD3), all regulated by the RAS-ERK signaling pathway. Experimental validations confirmed the involvement of these transcription factors in controlling the expression of genes induced by METex14Del, and underscored the central role of the RAS-ERK pathway in cell migration and invasion. This pathway thus constitutes a promising therapeutic target, particularly when responses to MET inhibitors are incomplete.

Overall, our findings validate the relevance and robustness of this gene regulatory network model for lung cancer. By comparing different lung tumor datasets, we refined classifications to reveal previously unrecognized subtypes. Through dynamic analyses, we demonstrated that the network effectively highlights major regulatory modules, their links to signaling pathways, and their involvement in the oncogenic properties of METex14Del.

Remerciements

Au Dr. Laurence CALZONE, au Dr. Marie-Christine COPIN de me faire l'honneur de participer à mon jury de thèse et de l'intérêt qu'ils portent à mon travail. Je remercie sincèrement Dr. Charles LECELLIER et Dr. Luca GRUMOLATO d'avoir accepté d'être mes rapporteurs.

Je remercie particulièrement mon directeur de thèse Mohamed Elati pour avoir su déceler chez moi les capacités pour effectuer cette thèse. Son expertise scientifique et sa vision critique ont été des guides précieux à chaque étape de ce travail. Pour avoir mis en place un environnement de travail propice à l'évolution intellectuelle et pour les nombreux échanges constructifs qui ont enrichi ma réflexion.

Je tiens à remercier chaleureusement mon co-directeur de thèse David Tulasne pour son encadrement exemplaire, son soutient, ses conseils et sa patience. Son regard avisé et sa rigueur scientifique ont su entretenir ma soif d'apprendre et ma curiosité scientifique.

Merci aux Dr Flavio Maina et Christophe Battail pour leur écoute et leurs recommandations toujours pertinentes lors de mes comités de suivi individuels.

À Wajdi pour ton expertise sur de nombreux sujets, ton exemplarité en matière d'enseignements, nos discussions inspirantes et stimulantes ont enrichi ma réflexion scientifique sur de nombreux points.

À Marie-José, merci pour ta détermination, ta bonne humeur et ta générosité. J'espère que notre collaboration continuera au-delà du projet METex14Del.

À Martin Figeac dont l'esprit scientifique et la rigueur font partie de l'ensemble des exemples que je garde en tête comme objectifs.

Je suis redevable à Nicolas Jonckheere et à Arnaud Janin pour leur confiance et nos discussions enrichissantes qui m'ont toujours motivé à aller chercher plus loin et à ne pas se contenter de l'acquis. Je suis convaincu que notre travail commun établi en synergie sera toujours productif.

À l'ensemble des collègues de l'équipe DiSCO, aux anciens membres ayant participé à la vie de l'équipe, thank you Konstantinos for you hapiness and your scientific spirit, Jules pour ta sagacité et ta capacité à relativiser, Emma pour ta curiosité et sa force de caractère. Un grand merci à l'ensemble de l'équipe Plasticité du laboratoire CANTHER, les déménagements successifs m'ont fait passer de votre open-space à l'autre bout du couloir et ensuite à l'étage du dessous, j'espère que nos collaborations scientifiques sauront être productives. Je remercie aussi l'ensemble de l'équipe de David pour leur regard essentiel et leurs questions qui m'ont permis de garder les pieds sur terre et de conserver la biologie et la clinique comme cap dans mon travail.

Je remercie notre directrice d'unité Isabelle Van Seuningen ainsi que l'ensemble de l'unité CANTHER.

Merci de tout cœur à l'ensemble de mes amis pour leur soutien sans faille durant ces années éprouvantes, pour ces nombreux moments partagés . Cela fait plusieurs années que nous nous suivons, depuis la fac de pharma nous n'avons cesser de nous soutenir mutuellement dans les projets qui nous tiennent à cœur. Merci à vous de faire partie de cette famille que l'on choisit.

Je tiens à souligner le rôle essentiel de Benjamin, notre appétit commun pour la Science et la recherche nous a mené tous les deux vers une thèse de science, ton esprit vif, exigeant et passionné est un modèle pour moi. Les discussions abordés lors de nos cafés quasi quotidiens ont fortement stimulé ma curiosité intellectuelle sur de nombreux sujets. J'ai aucun doute sur la qualité de nos futures collaborations.

Ma gratitude infinie va à l'ensemble de ma famille, frères, sœur, oncles, tantes. Vous avez eu sens cesse confiance en moi et en mes capacités à réussir ce que j'entreprends. Cette confiance que vous avez en moi, qui me fais parfois défaut, est une force qui m'a accompagné dans cette aventure. Je remercie aussi l'ensemble de ma belle-famille, merci pour votre gentillesse, vos encouragements, et pour m'avoir toujours fait sentir comme un membre à part entière de la famille.

À toi, Blandine, qui a vécu la thèse presque autant que moi, merci d'avoir survécu à mes crises existentielles, à mes monologues sur des sujets obscurs, et à mes soirées passés devant un écran plutôt qu'avec toi. Tu as su me porter dans les moments de doute et célébrer chaque petite victoire avec moi, cette thèse est un peu la tienne aussi.

À ma mère, mon héroïne, modèle de résilience, de combativité et de ténacité. Même si le sujet de mes recherches reste obscur, tu as toujours cru en moi. Je te remercie pour toutes les valeurs que tu m'as transmises. Cette thèse est aussi le fruit de tes sacrifices et de ton exemple.

À mon père, tu es un modèle d'ouverture, d'empathie et de dialogue avec autrui. La vision quotidienne du CHU Claude-Huriez me rappelle ton souvenir et me remémore les motivations profondes qui m'habitent. Tu m'as transmis la force d'avancer, même dans l'épreuve, et c'est aussi grâce à toi que j'ai pu aller au bout de cette thèse. Elle t'est dédiée.

Publications

1. Articles scientifiques

A coregulatory influence map of glioblastoma heterogeneity and plasticity

Chloé Bernhard, Konstantinos Geles, Geoffrey Pawlak, Wajdi Dhifli, Aurélien Dispot, Jules Dusol, Maria Kondratova, Sophie Martin, Mélissa Messé, Damien Reita, David Tulasne, Isabelle Van Seuningen, Natacha Entz-Werle, Silvia Anna Ciafrè, Monique Dontenwill & Mohamed Elati

npj Precision Oncology (2025)

Comprehensive map of the regulatory network triggered by MET exon 14 skipping reveals important involvement of the RAS-ERK signaling pathway

Marie-José Truong, Geoffrey Pawlak, Jean-Pascal Meneboo, Shéhérazade Sebda, Marie Fernandes, Martin Figeac, Mohamed Elati, David Tulasne

Cell Death & Disease (2025)

Gene regulatory network analysis identifies prognostic subgroups of lung adenocarcinoma driven by distinct master regulators.

Geoffrey Pawlak, David Tulasne, Mohamed Elati

(En préparation)

Network-based identification of master regulators in pancreatic neuroendocrine tumors through RNA-sequencing analysis

Arnaud Jannin, Geoffrey Pawlak, Sébastien Aubert, Martin Figeac, Mohamed Elati, Emmanuelle Leteurtre, Stéphanie Espiard, Catherine-Cardot-Bauters, Benjamin Chevalier, François Pattou, Camille Marcinia, Medhi El Amrani, Farid El Hajbi, Isabelle Van Seuningen, Marie-Christine Vantyghem, Christine Do Cao, Nicolas Jonckheere, Lucie Coppin

(En préparation)

2. Communications orales

- 2025 : Journée scientifique du Pancréas du Cancéropôle Nord-Ouest : PAAD-cRegMap: Une vue en réseau de co-régulation de l'hétérogénéité du cancer du pancréas
- 2025 : DMCM 2025 : lung.cRegMap: a web platform for unraveling the lung cancer coregulatory network and identifying potential key regulators.
- 2025 : Journées scientifiques du Cancéropôle Nord-Ouest : Lung.cregmap: Une application web pour explorer l'hétérogénéité du cancer du poumon au travers de son réseau de corégulation.
- 2023 : PhD-Days - Lung.cregmap : a co-regulatory influence map of lung cancer heterogeneity

- 2022 : PhD-Days - System biology and regulatory network for oncogenic addiction elucidation in MET 14 skipping mutation in lung cancer

3. Posters

- 2025 : ESMO ELCC2025 – Poster : 371P - Lung.cregmap: a web application to explore lung cancer heterogeneity through its coregulation network.
- 2023 : Journées scientifiques du Cancéropôle Nord-Ouest - lung.cregmap : a co-regulatory influence map of lung cancer heterogeneity
- 2022 : IBCN 2022 - BLCA-RegMap portal: a co-regulatory influence network view of bladder cancer heterogeneity and plasticity
- 2022 : Journées scientifiques du Cancéropôle Nord-Ouest - Inférence et interrogation du réseau de co-régulation dans le cancer du poumon. Etude de l'addiction oncogénique à la mutation METex14

4. Diffusion du savoir

- 2025 : Forum Industrie-Internat de la faculté de Pharmacie de Lille - Conférence "Intelligence artificielle en santé, apports, perspectives et place du pharmacien"
- 2025 : Forum Industrie-Internat de la faculté de Pharmacie de Lille - table ronde sur le devenir des pharmaciens

Table des matières

RESUME **4**

ABSTRACT **5**

REMERCIEMENTS **6**

PUBLICATIONS **8**

1. Articles scientifiques 8
2. Communications orales 8
3. Posters 9
4. Diffusion du savoir 9

ABREVIATIONS **11**

INDEX DES FIGURES ET TABLES **13**

INTRODUCTION **15**

1. Cancers : problematique de sante majeure 16
2. Focalisation sur le recepteur MET 34
3. Biologie des systemes 53

OBJECTIFS DE THESE **71**

RESULTATS **73**

1. Decloisonner la recherche et diffuser les modeles au moyen d'une application interrogeable 74
2. Gene regulatory network analysis identifies prognostic subgroups of lung adenocarcinoma driven by distinct master regulators 88
3. Transcriptional-program-based deciphering of the MET exon 14 skipping regulation network 116
4. A coregulatory influence map of glioblastoma heterogeneity and plasticity – From cells in vitro to tumors and back again 150

DISCUSSION **170**

1. Reseau de regulation genique 171
2. Portail cRegMap 176
3. Étude du cancer du poumon 179

CONCLUSION **183**

REFERENCES **185**

Abréviations

ADAM : A Disintegrin and Metalloproteinase Domain

ANSM : Agence nationale de sécurité du médicament et des produits de santé

APC : Adenomatous polyposis coli

ARID2 : AT-Rich Interaction Domain 2

ASCL1 : Achaete-scute homolog 1

BRCA1 : Breast Cancer 1

C1GALT1 : Core 1 β1,3-galactosyltransférase

CAR-T : Chimeric Antigenic Receptor – T

coTF : cofacteurs de transcription

SCLC : Cancer bronchique à petites cellules

NSCLC : Cancer bronchique non à petites cellules

CCLE : Cancer Cell Line Encyclopedia

CPTAC : National Cancer Institute's Clinical Proteomic Tumor Analysis Consortium

EGFR : Epidermal Growth Factor Receptor

ESMO : European Society for Medical Oncology

ETS : E26 transformation-specific transcription factor

FDA : Food and Drug Administration

FGF : Facteur de croissance des fibroblastes

GAB1 : GRB2-associated binder-1

GBM : Glioblastome

GCP : Google Cloud Platform

G-CSF : Granulocyte-Colony Stimulating Factor

GDC : Genomic Data Commons

GDP : Guanosine diphosphate

GEF : Guanine nucleotide Exchange Factors

GEO : Gene Expression Omnibus

GRB2 : Growth factor receptor-bound protein 2

GRN : Gene Regulatory Network

GTP : Guanosine triphosphate

HES1 : Hes Family bHLH Transcription Factor 1

HGFR : Hepatocyte Growth Factor Receptor

ICI : Immune Checkpoint Inhibitors

IHC : Immunohistochimie

INSM1 : Insulinoma-associated protein 1

KEGG : Kyoto Encyclopedia of Genes and Genomes

KRAS : Kirsten Rat Sarcoma Viral Oncogene Homolog

LRRK2 : Leucine-Rich Repeat Kinase 2

MAM : Mitochondria-associated ER Membranes

MAPK : Mitogen-activated protein kinases

mTOR : mammalian target of rapamycin

NCCN : National Comprehensive Cancer Network

NCBI : National Center for Biotechnology Information

NEUROD1 : Neurogenic differentiation 1

NFIB : Nuclear Factor I B

NGS : Next-Generation Sequencing

NIH : National Institutes of Health

NSCLC : Non-Small Cell Lung Cancer

OMS : Organisation Mondiale de la Santé

PDGF : Facteur de croissance dérivé des plaquettes

PD-L1 : Programmed death-ligand 1

POU2F3 : POU Class 2 Homeobox 3

PPI : Protein-protein interaction

PROTAC : PROteolysis-TArgeting Chimera

PS : Performance Status

PTCH1 : Protein Patched Homolog 1

PTEN : Phosphatase and TENsin homolog

PTP : Protéine Tyrosine Phosphatase

RAS : Rat sarcoma virus

RB1 : Retinoblastoma-Associated Protein

RE : Réticulum Endoplasmique

RSK : Ribosomal S6 Kinase

RTK : Récepteur à tyrosine kinase

SCLC : Small-Cell Lung Caner

SHC : SRC homology 2 domain-containing

SOS1 : Son of Sevenless homolog 1

SRF : Serum Response Factor

TCGA : The Cancer Genome Atlas Program

TFBS : Transcripton factor binding site

TFI : Treatment Free Interval

TKI : Tyrosine Kinase Inhibitors

UMAP : Uniform Manifold Approximation and Projection

YAP1 : Yes-associated protein 1

Index des figures et tables

Figure 1 : Représentation simplifiée de la classification 2012 des tumeurs pulmonaires..... 18
Figure 2 : Évolution de la classification des SCLC 21
Figure 3 : Principales mutations drivers dans les NSCLC 23
Figure 4: Classes des récepteurs à activité tyrosine kinase. 24
Figure 5: Dimérisation des récepteurs tyrosine kinases 25
Figure 6 : Mécanismes d'activation et de signalisation des récepteurs à tyrosine kinase 26
Figure 7 : Résumé des principales mutations et les voies de signalisation impliquées..... 31
Figure 8 : Structure du récepteur MET 34
Figure 9: Production et activation du HGF. 35
Figure 10 : Modes de dimérisation du récepteur MET. 36
Figure 11 : Activation de MET et amarrage de molécules adaptatrices et transductrices du signal..... 37
Figure 12 : Médiateurs de la signalisation de MET. 41
Figure 13 : Mécanisme d'activation oncogénique de MET. 46
Figure 14 : Régulation de MET en condition physiologique et en présence du saut de l'exon 14 52
Figure 15: Le paysage épigénétique. 54
Figure 16 : Décomposition du paysage épigénétique par réseau de régulation génique. 55
Figure 17 : Construction d'un GRN 57
Figure 18 : Exemple d'une structure d'un modèle de Deep learning..... 61
Figure 19 : Principaux algorithmes d'apprentissage profonds appliqués aux graphes 62
Figure 20 : Mesures des performances du modèle SVM 66
Figure 21 : Performances du modèle h-LICORN 67
Figure 22 : Performances de h-LICORN face à d'autres algorithmes..... 67
Figure 23 : Calcul de la valeur d'influence d'un régulateur et visualisation 69
Figure 24 : Inférence d'un réseau de régulation avec le package R CoRegNet 70
Figure 25 : Vue générale du fonctionnement de l'application cRegMap 75
Figure 26 : Onglet CoRegNet 76
Figure 27 : Visualisation du GRN sur le module CoRegNet 76
Figure 28 : Étapes principales de l'algorithme UMAP 77
Figure 29 : Principe de l'algorithme de Louvain 79
Figure 30 : Illustration de la méthode de Louvain pour l'identification de clusters 79
Figure 31 : Onglet CoRegMap de lung.cregmap..... 80
Figure 32 : Visualisation de la suppression de l'effet de batch au sein de l'application cRegMap..... 81
Figure 33 : Projection de la dépendance à FOSL1 des lignées cellulaires cancéreuses pulmonaires..... 82
Figure 34 : Principe de fonctionnement d'un modèle SVM 83
Figure 35 : Illustration du kernel trick utilisé dans les SVM..... 84
Figure 36 : Évaluation comparative des performances de plusieurs modèles de classification 85

Figure 37: Onglet CoRegQuery de chargement de données locales 85
Figure 38 : Importation de données externes dans l'onglet CoRegQuery via l'option « GEO » 86
Figure 39 : Résumé de l'étude téléchargée après récupération des données du GEO 86
Figure 40 : Onglet CoRegMap où un GSE a été projeté 87
Figure 41 : Projection du jeu de données GSE249568. 175
Figure 42 : Carte de densité des tumeurs pulmonaires sur la projection 2D du modèle UMAP 180
Figure 43 : Sous-cluster NE riche en adénocarcinomes. 181

Introduction

1. Cancers : problématique de santé majeure

En dépit d'une prise en charge largement améliorée et allant de plus en plus vers une médecine personnalisée, le cancer est une des principales causes de décès dans le monde avec près de 20 millions de nouveaux cas et 9,7 millions de décès estimés en 2022 dans le monde d'après le dernier rapport GLOBOSCAN et 433 136 nouveaux cas et 157 400 décès par cancer en France d'après le panorama des cancers publié par l'Institut National du Cancer en 2023. En 2024, une projection estime à 28,4 millions de cas de cancer en 2040, soit une hausse de 47% par rapport aux données de 2020 (1). L'augmentation de l'incidence et de la mortalité par cancer est fortement liée à l'augmentation de la population mondiale, à son vieillissement, à l'indice de développement humain et à la modification des facteurs de risques principaux. La prise en charge du cancer reste une priorité de santé publique en raison de sa charge sociale, clinique et économique, particulièrement lourde.

Le cancer du poumon est la principale cause de décès par cancer dans le monde, avec 1,8 million de décès (18,7 %) et environ 2,5 millions de nouveaux cas (12,4 %) en 2022. Il représente le deuxième cancer chez la femme, et le premier chez l'homme en termes de mortalité et d'incidence, avec un ratio homme-femme proche de 2. En France, avec 52 777 nouveaux cas diagnostiqués pour 33 100 décès en 2023, il est le deuxième cancer le plus fréquent chez les hommes et le troisième chez les femmes, avec un âge médian de 68 ans chez les hommes et 66 ans chez les femmes.

Le principal facteur de risque de cancer du poumon reste le tabac. Chez les hommes, après une augmentation durant la majorité du XXème siècle, principalement induite par l'épidémie de tabagisme parmi les hommes, le taux de mortalité par cancer a subi une inflexion au cours des années 1990 grâce à une diminution du tabagisme, une amélioration de la prise en charge du cancer du poumon ainsi qu'un meilleur dépistage, cependant avec un décalage de 20 à 25 ans. L'évolution est différente chez les femmes, dans les pays développés le taux d'incidence des cancers pulmonaires continue d'augmenter, allant jusqu'à surpasser le taux d'incidence chez les hommes. Cependant, on note dans certains pays un ralentissement voire une diminution de celle-ci (2). En France, c'est uniquement chez les femmes que le taux d'incidence a augmenté de 4,3% sur la période 2010-2023 alors qu'il a baissé chez les hommes de 0,5% sur la même période.

L'absence de symptômes spécifiques de la maladie rend difficile le dépistage précoce du cancer du poumon, cette symptomatologie aspécifique contribue au faible taux de survie puisque celle-ci reste inférieure à 20% dans la plupart des pays (1). En France le taux de survie net standardisé (prenant en compte l'effet de l'âge) à 5 ans pour les personnes diagnostiquées entre 2010 et 2015 est de 20% avec une disparité selon les sexes puisqu'il est de 24% pour les femmes et de 18% pour les hommes.

Le docteur René Laennec fut le premier à décrire des lésions pulmonaires dont l'aspect différait de celui des cas de tuberculose décrits classiquement au XIXème siècle, elles furent alors qualifiées d'encéphaloïdes de par leur ressemblance au tissu cérébral (3). La microscopie optique, couplée aux protocoles de fixation tissulaire et aux méthodologies d'immunomarquage cytochimique, a permis l'analyse histopathologique approfondie de

ces altérations néoplasiques, établissant leur caractère tumoral malin et permettant une discrimination de plus en plus précise de leurs sous-types.

1.1 Histologie et immunohistochimie de la tumeur

1.1.1 Description histologique

Le cancer bronchique à petites cellules fut décrit pour la première fois en 1926 par W. G. Barnard dans une étude sur les « oat-cell sarcomas », ainsi nommés en raison de leur morphologie : de petites cellules à cytoplasme peu abondant, aux contours mal définis, ressemblant à des grains d'avoine, et supposément issues des cellules épithéliales pulmonaires. Les cancers bronchiques à petites cellules représentent environ 20 % des cancers pulmonaires. Ils se caractérisent par leur agressivité et sont fortement liés au tabagisme. La médiane de survie se situe entre 16 et 24 mois, et le taux de survie à 5 ans est, lui, estimé entre 10 et 20 % (4). Historiquement, les cancers primitifs du poumon ont été classés histologiquement en deux sous-types majeurs : les cancers bronchiques non à petites cellules (NSCLC, pour *non-small cell lung cancer*) et les cancers bronchiques à petites cellules (SCLC, pour *small cell lung cancer*).

Une nouvelle classification des tumeurs pulmonaires, publiée par l'OMS en 2021, remplace la précédente datant de 2015. Les principes de classification demeurent similaires, avec une approche fondée d'abord sur la morphologie, puis appuyée par l'immunohistochimie et les techniques moléculaires (5). La classification de 2021 est composée de cinq catégories (Figure 1), La première catégorie regroupe les tumeurs épithéliales : les papillomes, les adénomes, les lésions précurseurs glandulaires, les adénocarcinomes, les lésions précurseurs épidermoïdes, les carcinomes à grandes cellules, les carcinomes adénosquameux et les carcinomes sarcomatoïdes. La deuxième catégorie rassemble des groupes de tumeurs très différentes du point de vue de leur caractéristiques histologiques, de leur pronostic et de leur traitements sous le nom de tumeurs neuroendocrines (NE). On y retrouve les tumeurs carcinoïdes typiques (TC) et atypiques (AC), ainsi que les carcinomes neuroendocrines de haut grade, comprenant le carcinome neuroendocrine à grandes cellules (LCNEC) et le carcinome neuroendocrine à petites cellules (SCLC). Les trois derniers groupes sont le groupe des tumeurs hématolymphoïdes (lymphome MALT, lymphome diffus à cellules B et granulomatose lymphoïde), le groupe des tumeurs mésenchymateuses spécifiques du poumon et le groupe des tumeurs de tissus ectopiques (mélanome et méningiome).

Afin de rester cohérent avec la littérature scientifique, j'utiliserai les acronymes NSCLC pour désigner les carcinomes pulmonaires non à petites cellules, et SCLC exclusivement pour les carcinomes neuroendocrines à petites cellules.

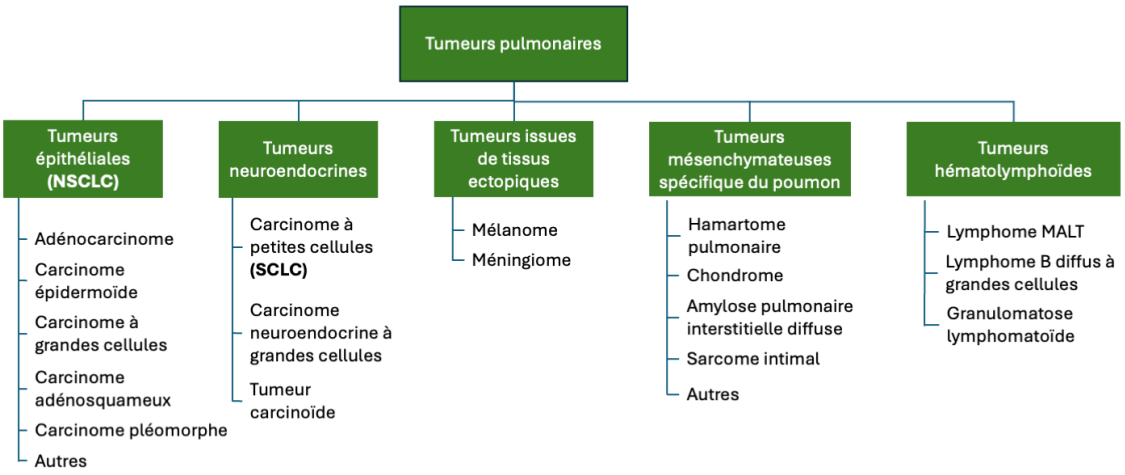


Figure 1 : Représentation simplifiée de la classification 2012 des tumeurs pulmonaires

Les tumeurs pulmonaires sont divisées en 5 groupes, les tumeurs épithéliales, les tumeurs neuroendocrines du poumon, les tumeurs issues de tissus ectopiques, les tumeurs mésenchymateuses et les tumeurs hématolymphoïdes.

Les NSCLC représentent près de 80 % des cancers pulmonaires. La majorité (60 %) d'entre eux sont des adénocarcinomes, se présentant généralement sous la forme d'une masse périphérique, à architecture glandulaire, avec une mucosécrétion intracellulaire et/ou l'expression en immunohistochimie (IHC) de TTF1, et/ou l'absence d'expression de p40 ou p63. Au sein des NSCLC, 20% des tumeurs sont des carcinomes épidermoïdes, apparaissant généralement le long des principales voies respiratoires, et pouvant former des cavités lorsqu'ils atteignent une taille importante. Ces cancers, associés au tabagisme, se développent à partir des cellules basales, qui n'expriment pas le facteur de transcription TTF1, un marqueur des pneumocytes et des adénocarcinomes bronchopulmonaires. En IHC, ils se caractérisent par l'expression de p40 ou p63 (6).

Bien que représentant une catégorie dans la classification des tumeurs thoraciques de l'OMS, la classe des tumeurs neuroendocrines regroupe des cancers très différents sur les plans clinique, épidémiologique, histologique, immunohistochimique et génétique. On peut distinguer deux catégories de tumeurs au pronostic différent. La première comprend les tumeurs carcinoïdes typiques et atypiques qui sont des tumeurs bien différenciées et de grade faible ou intermédiaire. La seconde admet les tumeurs faiblement différenciées et de haut grade qui sont les cancers neuroendocrines à petites cellules et les carcinomes neuroendocrines à grandes cellules. Le groupe des tumeurs carcinoïdes montre une évolution lente, de bon pronostic avec une survie à 5 ans de 87 à 100% pour les TC et compris entre 44 et 78% pour les AC (7), avec un plus grand risque d'extension nodulaire et métastasique et de récurrence chez les patients atteints d'un AC (8). Les LCNEC et SCLC représentent 3 et 13-15% des tumeurs pulmonaires. Bien qu'histologiquement différents, ils partagent une agressivité commune. Ils se caractérisent par un index mitotique et une nécrose plus élevés que ceux observés dans les TC et AC, et par un taux de survie à 5 ans faible : respectivement 15–20 % pour les LCNEC et 5 % pour les SCLC à un stade avancé (9). Le diagnostic de LCNEC nécessite une confirmation de différenciation neuroendocrine par immunohistochimie, contrairement au SCLC souvent identifiable en simple microscopie (10). Initialement, les SCLC répondent favorablement à la chimiothérapie, mais rechutent rapidement, tandis que les LCNEC y sont moins réceptifs (11).

Dans la suite de ce manuscrit, nous nous concentrerons sur les sous-types majoritaires de cancers du poumon, à savoir l'ensemble des NSCLC et, parmi les tumeurs neuroendocrines, uniquement les SCLC.

1.2 De l'histologie à l'approche moléculaire

Les progrès des techniques de séquençage et d'immunohistochimie ont ouvert la voie à une caractérisation moléculaire de plus en plus fine des cancers pulmonaires, dans l'objectif d'en comprendre les mécanismes à l'œuvre dans leur développement et leur maintien. Une meilleure compréhension des multiples voies de signalisation impliquées dans la pathogenèse moléculaire du cancer du poumon est cruciale pour le développement de stratégies thérapeutiques ciblant les aberrations moléculaires et les voies de signalisation activées en aval.

Les avancées spectaculaires en biologie moléculaire ont révolutionné notre compréhension et notre capacité à caractériser les cancers du poumon avec une précision sans précédent. L'émergence des technologies NGS, couplée aux analyses transcriptomiques, protéomiques et métabolomiques à haut débit, permet désormais d'établir une cartographie moléculaire détaillée de chaque tumeur, révélant non seulement les mutations spécifiques, mais aussi les signatures moléculaires complexes, le microenvironnement tumoral et les mécanismes de résistance thérapeutique.

1.2.1 Carcinogenèse multi-étape et mutations driver

Traditionnellement, l'émergence et l'évolution du cancer ont été vues selon un mécanisme darwinien où l'accumulation progressive de mutations génétiques dans un contexte de pression de sélection expliquerait la survenue et l'adaptation du cancer, suivant un mode d'évolution ramifié. Néanmoins cette vision d'une microévolution graduelle est insuffisante pour expliquer la complexité de l'évolution tumorale. Ce modèle est remis en question par l'existence de tumeurs relativement homogènes, par la résistance polyclonale aux traitements et par une macroévolution par événements catastrophiques où certains cancers peuvent évoluer par "sauts" brutaux via des événements massifs tels qu'un doublement du génome entier, ou un phénomène de chromothripsis (12). De récentes revues de la littérature ont résumé l'importance du microenvironnement tumoral. Ainsi, les cellules cancéreuses interagissent activement avec leur environnement, modifiant la matrice extracellulaire via les fibroblastes associés au cancer, détournant les cellules immunitaires de leur fonction protectrice, et stimulant la formation de nouveaux vaisseaux sanguins pour soutenir leur croissance (13,14).

Il est maintenant admis qu'une poignée de gènes serait responsables de la tumorigénèse (15). Les mutations qui affectent ces gènes peuvent être appelées mutations *drivers*, elles sont classées selon leur rôle dans le développement du cancer, généralement elles ciblent des gènes clés impliqués dans des voies de signalisation critiques. Elles seraient directement responsables de la transformation maligne des cellules, conférant un avantage sélectif qui favorise la prolifération, la survie ou la capacité invasive des cellules tumorales. Les gènes en causes sont soit des proto-oncogènes, c'est-à-dire des gènes qui requièrent une mutation de type « gain de fonction » pour activer leur potentiel oncogénique, ou des gènes suppresseurs de tumeur, dont la mutation de type « perte de fonction » leur confère un potentiel tumorigène. À l'inverse, de nombreuses autres

mutations, dites « *passengers* », sont considérées comme neutres et sont interprétées comme étant des mutations accumulées au cours de l'évolution clonale et n'affectent que marginalement le phénotype tumoral. La complexité du cancer, à la fois sur le plan biologique et clinique, est mise en évidence par la diversité des mutations driver et passenger au sein des tumeurs. Que ce soit par l'instabilité génomique, par les pressions exercées par l'environnement de la tumeur ou par le contexte clonal, il est difficile de dégager des signatures universellement pertinentes. Et, plus encore, de construire des classifications cliniquement pertinentes, utiles et robustes. Un ensemble de mutations drivers semble apparaître très tôt dans l'histoire naturelle des cancers spécifiques des tissus sans pour autant déclencher un cancer avancé. L'accumulation d'autres mutations drivers permettrait alors de franchir progressivement les étapes nécessaires à la tumorigénèse (16) et participerait à l'hétérogénéité intratumorale (17).

1.2.2 Vers une caractérisation plus fine des cancers pulmonaires

1.2.2.a SCLC

Les cancers bronchiques à petites cellules sont moins caractérisés moléculairement que les NSCLC. Les principaux marqueurs immunohistochimiques des SCLC sont la synaptophysine, une glycoprotéine vésiculaire synaptique impliquée dans la transmission synaptique des neurones, la chromogranine A, une protéine sécrétée par les cellules neuroendocrines, et le CD56, une protéine exprimée à la surface des cellules nerveuses.

Histologiquement, les SCLC étaient initialement subdivisés en deux sous-types, classique et variant, sur la base de leur morphologie et de leurs caractéristiques de croissance (18).

Génétiquement, les analyses par séquençage de l'exome ou du génome des SCLC ont révélé des altérations des gènes *TP53* et *RB1*, retrouvées respectivement dans 60 à 90% des patients atteints de SCLC (19). La détermination de sous-types moléculaires neuroendocrines parmi les SCLC a commencé avec Poirier et al. en 2013. Ils ont étudié l'efficacité du SVV-001, un picornavirus oncolytique sélectif des cancers neuroendocrines, sur trois lignées de SCLC classique et de SCLC variant transplantées sur des modèles murins. Le profilage de l'expression génique a révélé que les lignées permissives sont caractérisées par une faible expression du facteur de transcription neuroendocrinien précoce ASCL1, et inversement, par une forte expression du facteur de transcription neuroendocrine tardif NEUROD1 (20). Des études ultérieures, in vivo et in vitro, ont permis de confirmer l'importance d'ASCL1 et NEUROD1 dans l'hétérogénéité des SCLC (21,22). Une analyse par classification non supervisée d'un panel de lignées cellulaires issues de SCLC a identifié deux sous-groupes caractérisés par l'expression d'un facteur de transcription neuroendocrine, INSM1 et d'un médiateur clé de la voie Hippo, YAP1. Dans cette étude, le groupe où YAP1 était fortement exprimé présentait une faible expression d'ASCL1 et NEUROD1. Inversement, le sous-groupe caractérisé par une expression d'INSM1 présentait des niveaux variables d'expression de ces facteurs de transcription. (23). La classification des SCLC s'est ensuite élargie par l'identification de POU2F3 comme facteur de transcription clé des tumeurs neuroendocrines, exprimé de manière exclusive et nécessaire au sous-type variant de SCLC dépourvu d'expression des autres marqueurs neuroendocrines. La classification alors obtenue comprend quatre groupes, SCLC-A, SCLC-N, SCLC-Y et SCLC-P

(24,25). Néanmoins, la signature YAP1 ne semble pas être spécifique d'un sous-type de SCLC. L'expression de YAP1 a été observée dans plusieurs sous-types et a été associée à des programmes de gènes liés à l'EMT. De plus, des analyses immunohistochimiques (IHC) ultérieures n'ont pas confirmé un sous-type YAP1 unique (26,27). La classification des SCLC a été raffinée par la suite grâce à l'analyse de données de séquençage ARN de tumeurs issues de résections chirurgicales (n=81) et de l'essai clinique IMpower133 (n=276), un essai de phase III randomisé visant à mesurer l'efficacité de l'ajout de l'anticorps anti-PD-L1 atezolizumab à une chimiothérapie classique à base de platine chez des patients atteints de SCLC au stade avancé (28), pour aboutir à quatre sous-types : ASCL1 (SCLC-A), NEUROD1 (SCLC-N), POU2F3 (SCLC-P) et un sous-type inflammé (SCLC-I) (Figure 2). Le sous-type SCLC-I présente un fort infiltrat immunitaire, notamment une surreprésentation des cellules T, des cellules NK et des macrophages, une transition épithélium-mésenchyme et un phénotype inflammatoire, avec une forte expression de gènes liés aux HLA, à l'activation de l'interféron-γ et aux points de contrôle immunitaires (29). Finalement, une analyse des données de l'essai IMpower133 par factorisation par matrice non négative — une technique d'analyse multivariée qui décompose une matrice de données complexes en un produit de plusieurs matrices de facteurs non négatifs, permettant d'identifier des patterns moléculaires sous-jacents — a abouti à l'identification de quatre sous-types moléculaires en concordance avec la classification précédente. Cette analyse a permis d'identifier deux sous-ensembles présentant des phénotypes neuroendocriniens et non neuroendocriniens distincts, caractérisés par une infiltration variable des cellules immunitaires. Les tumeurs neuroendocrines, avec peu de macrophages associés mais des signaux T-effecteurs élevés, présentent une survie globale plus longue sous immunothérapie anti-PD-L1 combinée à la chimiothérapie carboplatine-étoposide, par rapport à la chimiothérapie seule. Cette classification offre une approche cliniquement pertinente pour discriminer les patients atteints de SCLC les plus susceptibles de bénéficier des immunothérapies, et met en évidence les mécanismes complexes qui sous-tendent les réponses thérapeutiques (30,31).

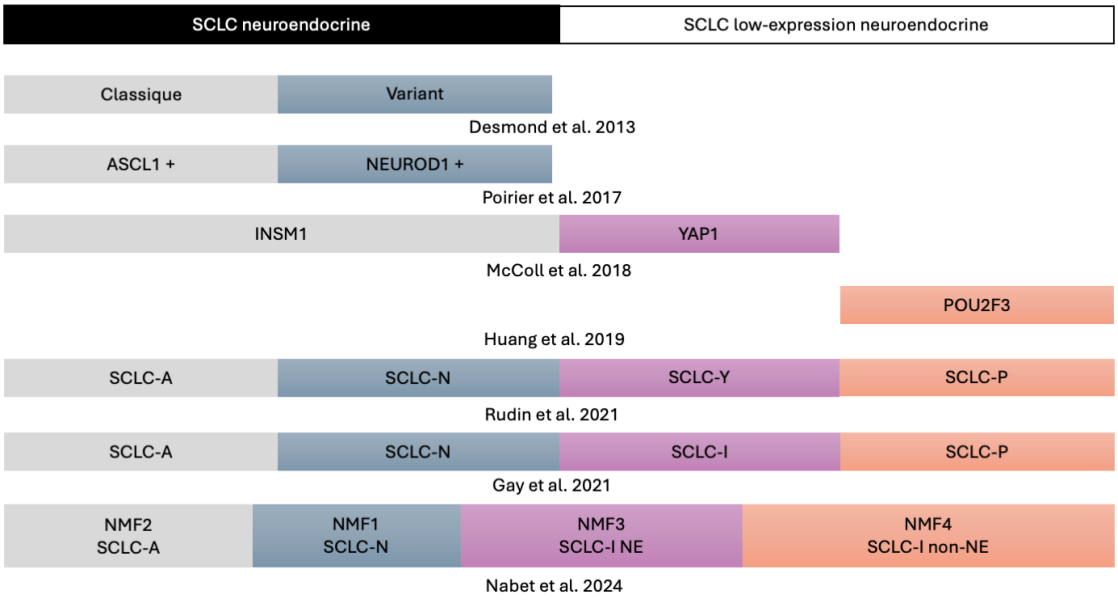


Figure 2 : Évolution de la classification des SCLC

L'évolution de la classification des SCLC est présentée de façon chronologique.

Même si de multiples thérapies moléculaires ciblées ont été testées dans des essais cliniques impliquant des patients atteints de SCLC, aucune n'a apporté de bénéfice durable (32). Très récemment, en mai 2024, la FDA a autorisé la mise sur le marché du tarlatamab (Imdelltra, Amgen, Inc.) pour les patients atteints d'un SCLC en rechute ou réfractaire aux thérapies à base de platine. Le tarlatamab est un anticorps bispécifique engageant les récepteurs DLL3 des cellules pulmonaires cancéreuses et les récepteurs CD3 des lymphocytes T sains, afin de favoriser la réponse immunitaire (33).

Le paysage mutationnel des SCLC est unique et très différent de celui des NSCLC. Comme dit plus tôt, les SCLC présentent une inactivation bi-allélique de TP53 et RB1, quasi universelle (34). D'autres gènes régulièrement altérés dans les SCLC incluent les gènes suppresseurs de tumeurs PTEN (7 %) et NF1 (4 %), des gènes impliqués dans la voie de signalisation NOTCH (10 %) (NOTCH1, NOTCH2, NOTCH3 et NOTCH4), ainsi que des régulateurs de l'intégrité de la chromatine (ARID1A (3 %), CREBBP (5 %), EP300 (2 %) et KMT2D (13 %)). De plus, des données en vie réelle rapportent une altération de STK11 (1,7 %), KEAP1 (3 %), PIK3CA (9,9 %) et RICTOR (5,6 %) (35,36).

1.2.2.b NSCLC

Les cancers bronchiques non à petites cellules sont les cancers bronchiques les plus caractérisés sur le plan moléculaire. L'identification de mutations rend possible le développement de médicaments ciblant la protéine responsable de l'émergence de la tumeur, améliorant ainsi les chances de survie du patient. Ainsi, de nombreux efforts ont été déployés pour cartographier le paysage mutationnel des NSCLC et identifier de potentielles cibles thérapeutiques. Les altérations moléculaires spécifiques qui favorisent la croissance tumorale et constituent des cibles thérapeutiques ont été bien définies dans les adénocarcinomes, tandis que le paysage moléculaire des carcinomes épidermoïdes suscite, quant à lui, un intérêt croissant. Une dépendance à l'égard des oncogènes s'établit lorsque la survie et la prolifération cellulaires dépendent de l'activation continue, de façon aberrante, des voies de signalisation, ce qui en fait des candidats idéaux pour les thérapies ciblées. Les voies de signalisation impliquées sont, pour la plupart, très fortement interconnectées, avec une prépondérance des voies engagées dans la carcinogenèse.

Les cancers pulmonaires à petites cellules et non à petites cellules présentent des paysages génomiques très différents. Ainsi, les NSCLC ont été largement étudiés, et de nombreuses mutations driver ont été identifiées, avec une prépondérance des mutations du gène KRAS, suivies de celles du gène EGFR. Les principales mutations driver rencontrées dans les NSCLC sont présentées dans la figure 3.

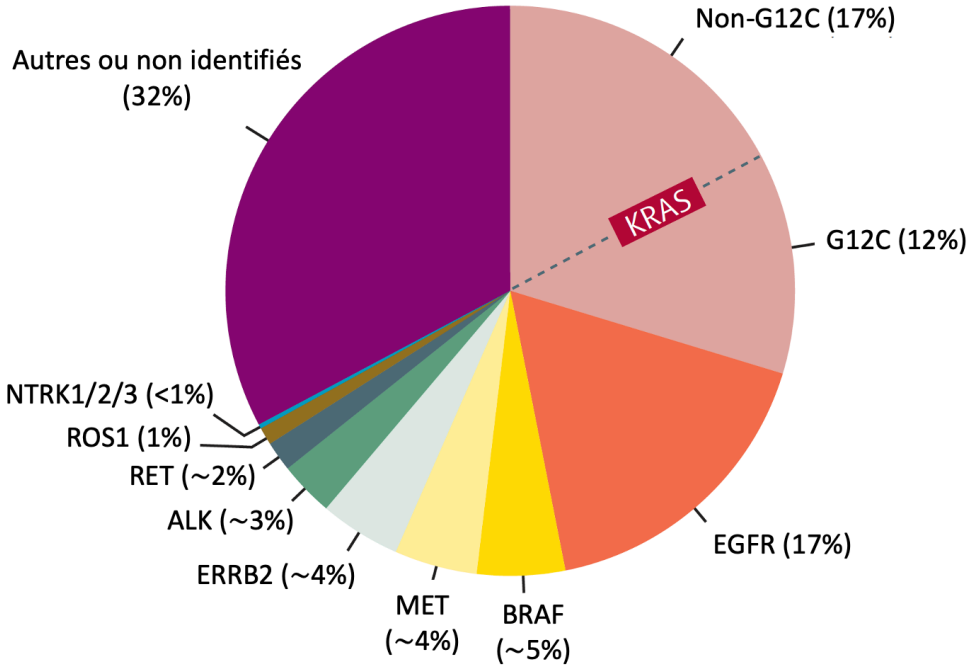


Figure 3 : Principales mutations drivers dans les NSCLC

D'après Thai et al. 2021.

Dans la suite de ce manuscrit, nous aborderons les principales mutations rencontrées dans les NSCLC, en commençant par celles ciblant les récepteurs à activité tyrosine kinase, puis celles affectant les acteurs situés en aval des voies de signalisation.

Récepteurs à activité tyrosine kinase

Les kinases constituent une vaste famille d'enzymes, avec 518 protéines kinases codées par le génome humain, dont 90 appartiennent au groupe des tyrosine kinases, une sous-classe majeure des kinases humaines. Ce groupe se divise en deux catégories : les récepteurs à activité tyrosine kinase et les tyrosine kinases non réceptrices (nRTK). Les RTK assurent la transduction des signaux extracellulaires vers l'intérieur de la cellule, tandis que les nRTK sont impliquées dans la signalisation intracellulaire (37). Les RTK présentent une forte similarité structurelle, comprenant un domaine extracellulaire capable de lier des ligands spécifiques, une unique hélice transmembranaire, et une région intracellulaire contenant un domaine catalytique à activité tyrosine kinase (Figure 4). Ce dernier, également appelé domaine kinase, adopte une structure bilobée composée d'un lobe N-terminal, d'un lobe C-terminal, et d'une cavité de liaison à l'ATP située entre ces deux lobes.

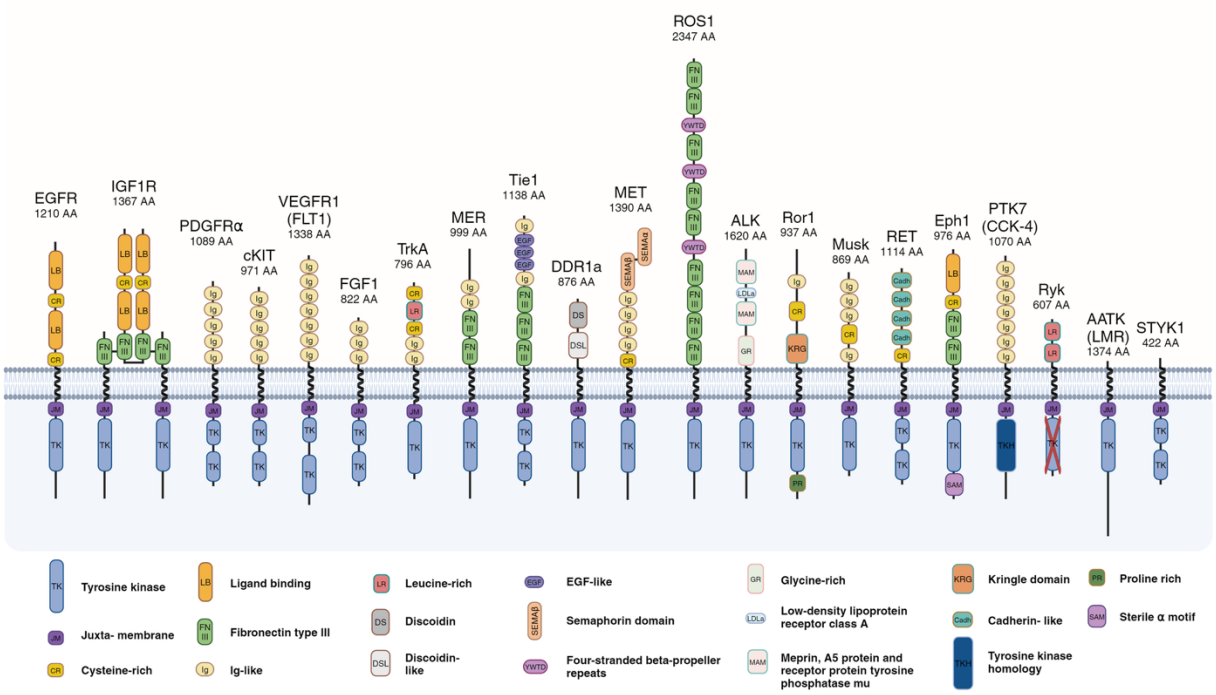


Figure 4: Classes des récepteurs à activité tyrosine kinase.

Les 20 classes de récepteurs à activité tyrosine kinase (RTK) présentent des structures distinctes, bien que partageant des caractéristiques communes. Les principales variations structurelles se situent au niveau des domaines extracellulaires et intracellulaires, influençant ainsi la reconnaissance des ligands et la transmission du signal. De gauche à droite, l'organisation des 20 classes de RTK met en évidence ces similarités et différences, illustrant la diversité fonctionnelle et structurelle de cette famille de récepteurs.
D'après Tomuleasa et al. (2024)

La fixation d'un ligand au domaine extracellulaire induit la dimérisation des récepteurs, permettant l'autophosphorylation des domaines intracellulaires et l'activation de l'activité tyrosine kinase. Les RTK catalysent ainsi le transfert d'un groupement phosphate issu de l'ATP vers le groupe hydroxyle des résidus tyrosine des protéines cibles. Cette famille regroupe notamment le récepteur de l'insuline, ainsi que ceux de divers facteurs de croissance, tels que VEGF, FGF, PDGF et l'EGF (38).

Les premières études de liaison de ligands à leurs récepteurs ont posé un modèle simple pour la dimérisation : un ligand bivalent interagit simultanément avec deux molécules réceptrices et les relie efficacement pour former un complexe dimérique. Ce modèle a été, plus tard, confirmé par études cristallographiques notamment pour le récepteur au VEGF ou au KIT. Le ligand, lui-même un dimère, réarticule les fragments de liaison au ligand de deux molécules réceptrices. Cependant, les structures récentes des régions extracellulaires plus complètes des RTK ont fourni des informations supplémentaires importantes sur l'éventail des mécanismes utilisés pour la dimérisation induite par les ligands. La dimérisation des récepteurs peut se présenter sous différents modes. Dans un scénario extrême, elle est exclusivement «médiée par le ligand », de sorte que les deux récepteurs n'entrent pas en contact direct (figure 5A). À l'opposé, la dimérisation est entièrement « médiée par le récepteur », le ligand n'intervenant pas directement dans l'interface du dimère (figure 5D). Entre ces deux extrêmes, des formes hybrides existent, combinant des interactions dépendantes du

ligand et du récepteur (figures 5B et 5C). La majorité des RTK semble adopter un mode de dimérisation s'apparentant à l'un de ces quatre modèles.

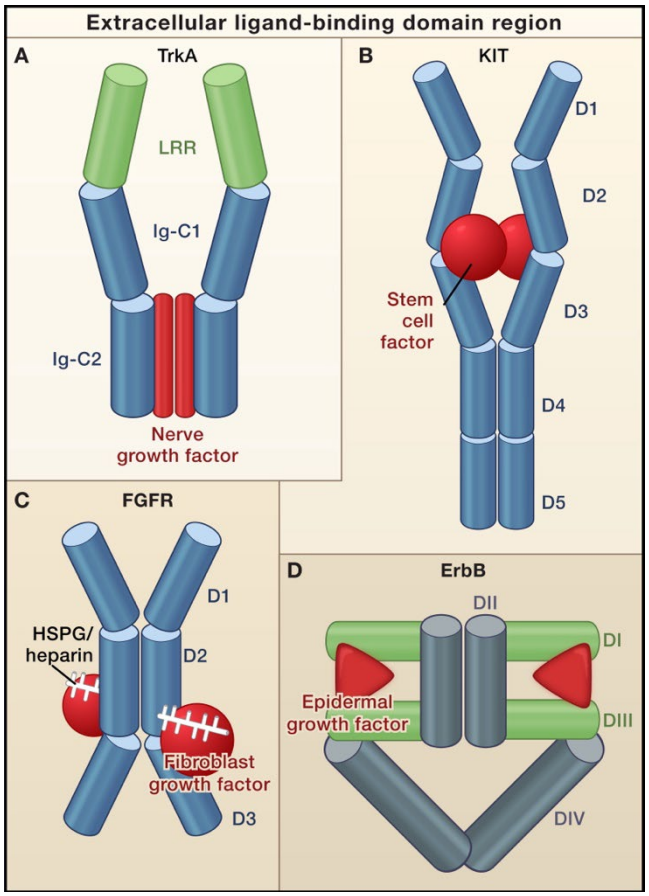


Figure 5: *Dimérisation des récepteurs tyrosine kinases*
D'après Lemmon, M. A., & Schlessinger, J. (2010)

En parallèle de l'activation des domaines catalytiques, la phosphorylation des régions non catalytiques génère des sites d'ancrage pour des effecteurs cytoplasmiques en aval, qui se lient aux résidus phosphotyrosines du récepteur via des séquences consensus spécifiques. La transmission du signal repose sur une organisation modulaire, impliquant notamment des domaines SH2 (src-homology domain 2) et PTB (phosphotyrosine binding domain), ainsi que des domaines catalytiques ou des modules d'interaction permettant d'assembler des complexes multiprotéiques (Figure 6). Cette modularité assure l'intégration et la diversification des signaux, en formant des réseaux de signalisation complexes. Cette complexité implique que, pour inhiber efficacement les voies de signalisation activées par les RTK et induire la mort des cellules tumorales, il est souvent nécessaire d'utiliser des stratégies thérapeutiques combinées ciblant plusieurs niveaux de la cascade de signalisation (39).

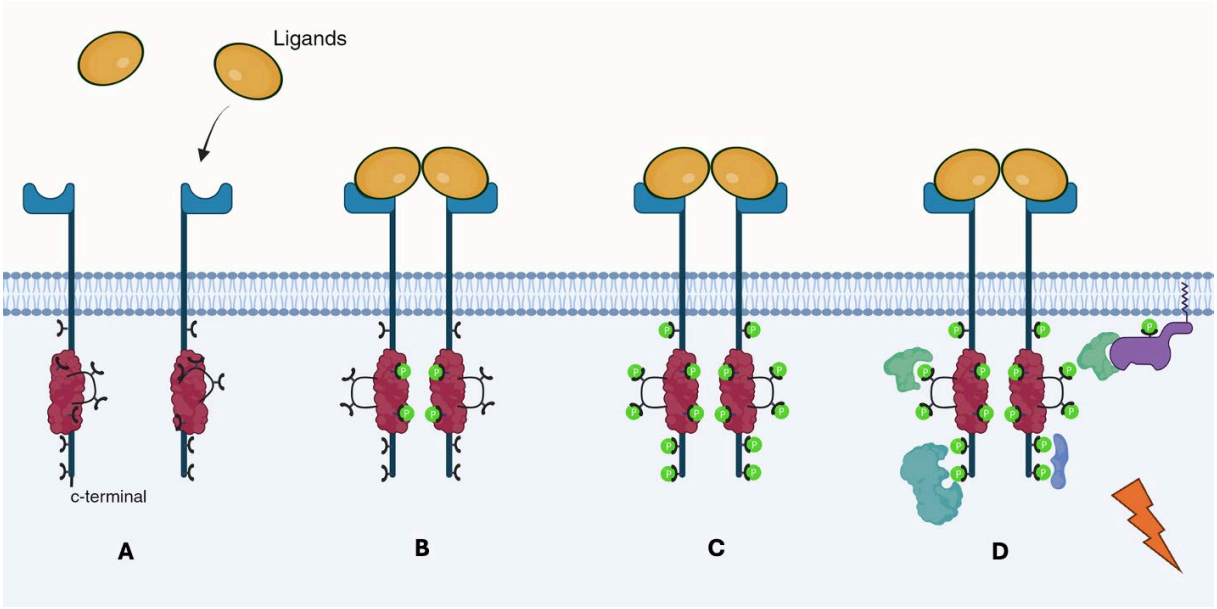


Figure 6 : Mécanismes d'activation et de signalisation des récepteurs à tyrosine kinase

Les ligands se lient aux RTK sous leur forme monomérique (A), induisant leur dimérisation et entraînant une phosphorylation croisée des domaines tyrosine kinase (B). Cette première étape est suivie d'une transphosphorylation secondaire impliquant non seulement les domaines tyrosine kinase, mais également les régions juxtamembranaires et C-terminales des récepteurs (C). Ces modifications post-traductionnelles génèrent des sites d'ancrage optimaux pour le recrutement de substrats intracellulaires, facilitant ainsi l'activation des protéines clés impliquées dans les voies de signalisation en aval (D).
D'après Tomuleasa et al. (2024)

Les récepteurs de la famille HER

La famille HER, pour *Human Epidermal Growth Factor Receptor* ou réceepteur humain au facteur de croissance épithélial (EGF), est composée de quatre récepteurs transmembranaires à activité tyrosine kinase : EGFR/ERBB1/HER1, ERBB2/HER2, ERBB3/HER3 et ERBB4/HER4. La protéine EGFR, codée par son gène en position 7p11.2, est constituée d'un domaine extracellulaire, d'un domaine transmembranaire et d'une région cytoplasmique. La région cytoplasmique représente environ 50 % de la protéine. Elle comporte un domaine juxtamembranaire, un domaine contenant l'activité tyrosine kinase et un domaine carboxy-terminal.

La fixation de l'EGF sur le domaine extracellulaire du récepteur entraîne une modification conformationnelle induisant une homodimérisation. Cette interaction favorise des réarrangements structuraux menant à la formation de dimères de domaines catalytiques juxtamembranaires asymétriques, à l'activation allostérique du domaine kinase du récepteur, et à la trans-autophosphorylation des résidus tyrosine de l'extrémité C-terminale, déclenchant ainsi la cascade de signalisation. EGFR peut aussi s'hétérodimériser avec un autre récepteur de la même famille ; néanmoins, les hétérodimères ont une affinité réduite pour leur ligand. Celui-ci se dissocie dans les endosomes, où il entraîne une fixation diminuée de Cbl, nécessaire à sa dégradation. La signalisation par les hétérodimères d'EGFR semblerait plus oncogène que celle médiée par ses homodimères (40). In fine, EGFR active de multiples voies de signalisation, y compris la voie Ras/MAPK, la voie PI3K/AKT et la cascade de signalisation de la phospholipase C/protéine kinase C (41).

Les mutations somatiques activatrices du domaine kinase d'EGFR sont observées chez 10% des patients caucasiens et jusqu'à 50% des patients asiatiques atteints d'un NSCLC avec une incidence plus élevée chez les femmes et les personnes qui n'ont jamais fumé ou qui sont d'anciens fumeurs (42). Les mutations sont responsables d'une activation constitutive d'EGFR indépendamment de son ligand. Ainsi, des thérapies ciblées ont été développées pour traiter les NSCLC sur la base de la présence de mutations driver. Plusieurs essais cliniques de phase III ont étudié le rôle des TKI EGFR de première (géfitinib et erlotinib) et de deuxième génération (afatinib et dacomitinib), qui ont montré des taux de réponse médians similaires (70 à 75 %) et des améliorations significatives de la survie sans progression allant de 10 à 14 mois, par rapport à la chimiothérapie à base de platine, dans les NSCLC avancés porteurs de mutations EGFR et naïfs de traitement. Plus récemment, l'osimertinib, un TKI de l'EGFR de troisième génération, a montré une efficacité supérieure, avec une amélioration de la SSP (18,9 mois contre 10,2 mois) et de la survie globale (38,6 mois contre 31,8 mois). Malgré ces avancées thérapeutiques majeures et une compréhension approfondie des déterminants génétiques, la résistance aux inhibiteurs d'EGFR survient inévitablement, entraînant une progression tumorale (43).

Les RTK, tels qu'EGFR, ALK, MET, ROS1 et RET, jouent un rôle central dans la signalisation cellulaire impliquée dans la prolifération, la survie et la formation de métastases. Les inhibiteurs de tyrosine kinase, comme l'erlotinib ou l'osimertinib pour EGFR, et l'alectinib ou le lorlatinib pour ALK, ont démontré une efficacité significative en induisant des réponses cliniques prolongées chez les patients porteurs de mutations activatrices (44). Toutefois, la réponse reste souvent limitée à une sous-population de patients, ce qui souligne l'importance d'une sélection moléculaire préalable (45). De nouvelles cibles, comme les récepteurs EphA2 et ROR1, sont également à l'étude afin de surmonter les mécanismes de résistance secondaire (46,47). Le développement de résistances, qu'elles soient dues à des mutations secondaires ou à l'activation de voies alternatives, demeure une limite majeure, nécessitant souvent une association avec d'autres inhibiteurs ou agents chimiothérapeutiques (48,49). Malgré ces défis, les RTK représentent une classe de cibles prometteuses, et leur inhibition continue d'être optimisée grâce à des stratégies de médecine de précision (38,50).

Les réarrangements chromosomiques

L'instabilité du génome dans les cancers du poumon peut mener au gain et à la perte de régions chromosomiques, voire de chromosomes entiers. Ces tumeurs présentent également des réarrangements structurels responsables d'altérations de la transcription et de l'expression des protéines, par fusion de gènes résultant d'insertions, d'inversions et de translocations.

La protéine ALK est codée par le gène du même nom, situé sur le chromosome 2 en position 2p23. Celui-ci appartient à la superfamille des récepteurs à l'insuline. C'est un RTK exprimé au cours de l'embryogenèse, dont l'expression diminue durant la période postnatale, entraînant une faible expression chez l'adulte. La plupart des mutations du gène ALK se présentent sous la forme d'une translocation avec un gène partenaire, conduisant à un gène de fusion oncogénique. Les réarrangements ALK ont été identifiés dans 5 à 6 % des NSCLC , en particulier chez des patients jeunes et non-fumeurs ou ayant peu fumé. Le transcrit de fusion EML4-ALK fut le premier produit d'un réarrangement ALK décrit (51), EML4 remplace alors les parties extracellulaires et

transmembranaires d'ALK par fusion avec son domaine juxtamembranaire, entraînant une activité kinase ALK constitutive, et représentant une voie de dépendance oncogénique dans le cancer du poumon. Son activation entraîne une stimulation des voies de signalisation PI3K/AKT/mTOR, MEK/ERK, PLCγ et JAK/STAT (52). D'autres gènes partenaires ont été décrits, comme KIF5B, KLC1, TFG, TPR, HIP1 et STRN (54), favorisant la dimérisation d'ALK.

Le gène ROS1 est un proto-oncogène situé sur le chromosome 6q22, qui code également pour un RTK partageant une forte homologie de séquence avec le domaine catalytique d'ALK (49 % d'acides aminés en commun) (53). ROS1 joue un rôle majeur dans l'activation de plusieurs voies de signalisation associées à la différenciation, à la prolifération, à la croissance cellulaire et à la survie (MEK/ERK, PI3K/AKT/mTOR, JAK/STAT). Les altérations de ROS1 de type réarrangement, conduisant à une protéine de fusion, sont présentes dans 0,9 à 2,6 % des NSCLC (54). Les fusions CD74-ROS1, EZR-ROS1, SDC4-ROS1 et SLC34A2-ROS1 représentent les réarrangements les plus fréquents (55) et sont préférentiellement retrouvées chez des patients jeunes, non-fumeurs, d'origine asiatique et atteints d'un adénocarcinome (56). Néanmoins le mécanisme d'activation constitutive de ROS1 reste peu connu et ne semblerait pas passer par une dimérisation de la protéine chimérique nécessaire à son activation (57–59).

Enfin, RET est situé sur le chromosome 10q11.2. La protéine correspondante est un RTK qui partage une homologie de séquence de 37 % avec le domaine catalytique du récepteur ALK. Son activation requiert l'intervention du complexe GDNF-GFRα, qui se lie au domaine extracellulaire. Cela entraîne sa dimérisation, l'autophosphorylation du domaine tyrosine kinase intracellulaire et, par conséquent, l'activation de plusieurs voies de signalisation, notamment MEK/ERK, PI3K/AKT/mTOR, JAK/STAT, PKA et PKC (60). Les réarrangements RET, identifiés dans 1 à 2 % des cas de NSCLC (61), conduisent à une dimérisation et à une activation des domaines catalytiques indépendamment de la fixation à leurs ligands. Les transcrits de fusion identifiés impliquent des gènes partenaires en amont tels que PTC, KIF5B, CCDC6, NCOA4, TRIM33, MYO5C, EPHA5, CLIP1, ERC1, PICALM, FRMD4A, RUFY2 et TRIM24 (62).

Les réarrangements chromosomiques impliquant des gènes de tyrosine kinases, tels que ALK, ROS1 et RET, ont ouvert la voie à des thérapies ciblées. Le réarrangement ALK, découvert en 2007, est présent dans environ 5 % des cas de NSCLC et entraîne une activation constitutive de voies de signalisation oncogènes. Le crizotinib, premier inhibiteur de ALK approuvé, a significativement amélioré les taux de réponse et de survie sans progression. Toutefois, la survenue rapide de résistances a conduit au développement de molécules de seconde génération, comme le céritinib et l'alectinib, aux profils de résistance plus larges et à une meilleure activité intracrânienne (63,64). De manière analogue, les réarrangements ROS1, détectés dans environ 1 à 2 % des patients, peuvent également être ciblés par le crizotinib, avec des réponses similaires à celles observées dans les cancers ALK-positifs (65). Quant aux cancers porteurs de réarrangements RET, bien que plus rares (environ 1 % des NSCLC), ils bénéficient désormais de traitements spécifiques. Les inhibiteurs multikinases, comme le vandétanib ou le cabozantinib, ont montré des réponses cliniques variables mais limitées, justifiant le

développement de molécules plus sélectives (66,67). La compréhension fine de ces altérations génomiques a profondément modifié la prise en charge du NSCLC, illustrant le potentiel de la médecine de précision.

KRAS

Le gène *KRAS* est un membre de la famille des oncogènes viraux du sarcome de rat (RAS), qui comprend deux autres isoformes chez l'homme : les oncogènes viraux du sarcome de rat de Harvey et du neuroblastome (*HRAS*, *NRAS*). Le gène KRAS, situé sur le chromosome 12 au locus 12p12.1, code pour deux isoformes protéiques très proches, KRAS-4B et KRAS-4A, composées respectivement de 188 et 189 acides aminés, en raison d'une coupure différente du quatrième exon (68). La protéine produite est une protéine G membranaire intracellulaire à activité guanosine triphosphatase, qui transforme le GDP en GTP. La transduction du signal depuis les RTK activés par leur ligand est médiée par un certain nombre d'acteurs clés. Les RTK activés par leur ligand s'homo- ou s'hétérodimérisent et se phosphorylent. Le RTK phosphorylé est lié par des protéines adaptatrices, dont GRB2, via leur domaine SH2 respectif, ce qui permet à leur tour de recruter des protéines GEF, parmi lesquelles SOS1, via son domaine SH3. SOS1 favorise la liaison du GTP à KRAS, le faisant basculer d'un état inactif à un état actif. La liaison du complexe GRB2/SOS1 peut être favorisée par la protéine à activité tyrosine phosphatase SHP2. Même si ce mécanisme d'activation de KRAS a lieu à la face intracellulaire de la membrane plasmique, Tulpule et al. ont mis en évidence une activation de KRAS indépendante de la localisation membranaire. Certaines protéines de fusion contenant des RTK, comme EML4-ALK, peuvent se combiner avec GRB2 et SOS1 pour former des granules protéiques cytoplasmiques sans membrane, afin d'activer KRAS et les voies de signalisation en aval (69).

Une fois activée, la protéine KRAS orchestre la transduction du signal à travers plusieurs voies de signalisation majeures. La première est la voie de signalisation RAF-MEK-ERK également appelée voie des MAPK. KRAS activé recrute la protéine kinase sérine/thréonine RAF par homo- ou hétérodimérisation, puis RAF se lie à MEK1/2 via son extrémité C-terminale et l'active par phosphorylation. MEK1/2 phosphorylé active ERK1/2 qui, à son tour, phosphoryle RSK, SRF, les protéines de la famille ETS et ELK1. L'activation de la transcription de leurs gènes cibles influence les processus de prolifération cellulaire, de différenciation et de migration (70). La seconde voie de signalisation est la voie PI3K-AKT-mTOR. KRAS activé se lie à PI3K, qui favorise la conversion du PIP2 en PIP3. PIP3, ainsi que le complexe mTOR, favorisent la phosphorylation d'AKT et son activation. AKT traverse la membrane nucléaire et active ou inhibe des processus impliqués dans le métabolisme cellulaire, l'apoptose et de régulation de la prolifération cellulaire (71). D'autres processus cellulaires peuvent être modulés par KRAS. On peut citer la migration cellulaire, via l'activation de RAL et de ses gènes cibles, ou encore de TIAM1 et RAC1, également impliqués dans l'adhésion cellulaire, la formation du cytosquelette d'actine et le trafic membranaire (70).

La conversion de la forme active de KRAS vers sa forme inactive est médiée par des protéines activatrices de GTPase (GAP), lorsqu'elles sont recrutées à la membrane plasmique et situées à proximité de KRAS. Parmi les gènes codant pour des protéines de la famille des GAP, on peut citer NF1, RASAL1-3, DAB2IP, SPRED1-3 et SYNGAP (72).

Le gène *KRAS* est l'oncogène le plus fréquemment muté parmi les protéines de la famille RAS. Il est à l'origine d'environ 90 % des adénocarcinomes canalaires pancréatiques, de 50 % des carcinomes colorectaux (CRC) et de 30 % des adénocarcinomes pulmonaires (73,74). Les mutations du gène KRAS sont majoritairement des mutations faux-sens, touchant dans 98 % des cas les codons 12 (G12), 13 (G13) ou 61 (G61) (75). Dans les NSCLC, la mutation G12C est la plus fréquemment observée, avec une fréquence plus élevée chez les fumeurs (25 à 35 % vs 5 %) (76). Les altérations de la protéine KRAS empêchent l'interaction de KRAS avec les GAP et l'hydrolyse du GTP lié, ce qui maintient KRAS dans son état constitutivement actif. Néanmoins, on note une hétérogénéité dans les conséquences des mutations KRAS. Ainsi, les protéines KRAS G13 semblent mieux répondre à l'hydrolyse par NF1, qui code la neurofibromine, une protéine appartenant à la famille des GAP (77). De plus, les mutations G12C et G12V semblent provoquer une augmentation des protéines liées à la voie de signalisation RalA/B et une diminution de p-AKT, par rapport aux autres altérations de KRAS ou au phénotype sauvage. En revanche, la mutation KRAS G12D semble préférentiellement activer la voie de signalisation PI3K-AKT (70).

Longtemps considérée comme une cible thérapeutique difficile à atteindre en raison de la structure du récepteur et de sa forte affinité pour le GTP, la mutation G12C de KRAS est désormais ciblable grâce à l'émergence d'inhibiteurs covalents spécifiques. Le sotorasib a été le premier inhibiteur de KRAS G12C approuvé par la FDA, montrant une efficacité clinique significative en monothérapie dans les formes avancées de NSCLC préalablement traitées (78,79). L'adagrasib, un autre inhibiteur ciblant la mutation G12C, a également démontré des réponses prometteuses, en particulier chez les patients présentant des métastases cérébrales (80). Néanmoins, la réponse à ces thérapies reste limitée par l'émergence fréquente de résistances thérapeutiques secondaires, impliquant des mécanismes tels que la réactivation de voies de signalisation annexes, telles que la voie MAPK, ou des mutations de contournement (81). Malgré les progrès, la diversité biologique des tumeurs KRAS-mutées et la présence fréquente de co-mutations, notamment STK11 ou TP53, imposent une individualisation des traitements et une prise en charge tenant compte de l'émergence future de phénomènes d'échappement thérapeutique

BRAF

Le gène BRAF est un oncogène situé sur le chromosome 7, impliqué dans la différenciation, la prolifération, la survie et la croissance cellulaires. Il encode une protéine intracellulaire comportant un domaine de fixation à la protéine RAS, un domaine de régulation et un domaine kinase. Les protéines BRAF jouent un rôle dans la transduction du signal depuis les RTK vers le noyau. Situé en aval de KRAS, BRAF est activé à la suite de la phosphorylation de ses résidus Thr598 et Ser601. Activé, BRAF va s'hétéro- ou s'homodimériser et phosphoryler MEK1/2 qui va, à son tour, phosphoryler ERK1/2 (82). BRAF est muté dans environ 3 % des adénocarcinomes. La majorité de ces mutations se produit dans la boucle d'activation près du codon V600 ou dans la boucle de liaison au phosphate, aux résidus 464-469. Selon les mutations, le gène muté produit des protéines activées constitutivement en tant que monomères (V600E, V600K, V600D), dimères (K601E, L597V/Q/R, G469V/S/R/E/A, G464V), ou nécessitant une activation préalable de la voie de signalisation en amont pour induire une hyperactivation de la voie des MAPK (G596R, D594Y/N/G/E, N581Y/S/I, G466V/L/E/A, D287Y) (83). Il semble

exister une différence de conséquences cliniques entre les mutations BRAF V600 et les autres. Les mutations BRAF V600E sont plus fréquentes chez les femmes n'ayant pas d'antécédents de tabagisme, tandis que les mutations non V600 sont davantage liées aux hommes fumeurs ou anciens fumeurs (84). Les mutations de BRAF V600E dans les NSCLC, comme dans le cancer colorectal et le carcinome papillaire de la thyroïde, sont associées à un pronostic défavorable par rapport aux tumeurs présentant des mutations de BRAF non V600E. Plus précisément, dans une étude rétrospective portant sur 1046 NSCLC mutés BRAF, Marchetti et al. ont mis en évidence une corrélation entre la mutation V600E et les statuts T, N, ainsi que le stade pathologique, avec une diminution significative de la survie sans progression et de la survie globale par rapport aux patients ne présentant pas cette mutation. En revanche, les mutations non-V600E de BRAF n'ont été associées à aucune caractéristique histologique spécifique ni à un impact pronostique particulier (85).

La mutation V600E de BRAF fait l'objet de thérapies ciblées, notamment via la combinaison de dabrafenib et trametinib, qui a démontré une efficacité clinique significative, avec un taux de réponse objective atteignant 64 % (86). Les résultats de la cohorte EURAF ont confirmé l'activité de ces inhibiteurs dans des contextes cliniques en dehors des essais, avec une survie médiane de 10,8 mois et un contrôle de la maladie dans 85 % des cas (87). Cependant, la moitié des mutations BRAF sont non V600E et présentent une hétérogénéité biologique importante, rendant difficile le ciblage spécifique de BRAF. Néanmoins, certaines mutations comme G469A ou K601E montrent une réponse variable aux thérapies ciblées (88). Chez les patients non V600, la prise en charge par inhibiteurs de MEK associés à un inhibiteur de BRAF a montré une efficacité occasionnelle (83). Quel que soit le statut mutationnel, on observe également l'émergence de variations secondaires de BRAF après traitement par thérapies ciblées (89).

Les principales mutations sont présentées dans la figure 7.

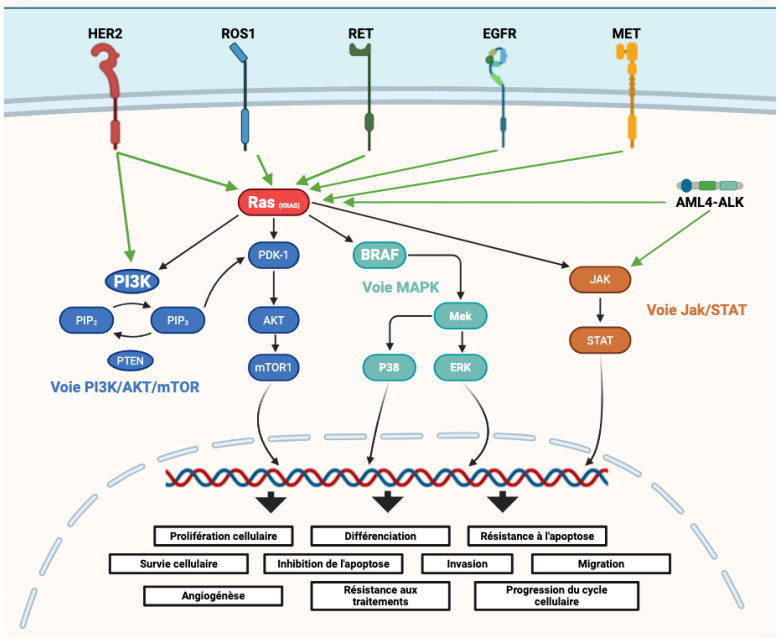


Figure 7 : Résumé des principales mutations et les voies de signalisation impliquées

1.3 Des cibles aux traitements : chimiothérapies, immunothérapies et thérapies ciblées

La prise en charge chirurgicale a été le premier traitement apporté aux tumeurs pulmonaires dès le XXe siècle, où le traitement chirurgical, puis la radiothérapie pour les patients inéligibles à la chirurgie, devinrent les approches couramment utilisées. Néanmoins, la mortalité liée au cancer du poumon restait élevée, atteignant 80 % (90). Dès 1919, les propriétés suppressives de l'hématopoïèse du sulfure de dichloroéthyle, utilisé comme gaz de combat durant la Première Guerre mondiale, ont été découvertes, et l'utilisation d'autres composés de la même famille a été étudiée, en particulier les moutardes à l'azote, qui se sont avérées moins toxiques et efficaces contre certains carcinomes pulmonaires (91). Cette période marque le début des découvertes en chimiothérapie. En 1969, Green met en évidence l'efficacité du cyclophosphamide chez des patients atteints d'un cancer du poumon, avec une efficacité supérieure à celle des moutardes à l'azote chez les patients atteints d'un cancer bronchique à petites cellules non différenciées. Ce rapport d'efficacité est inversé chez les patients atteints d'un carcinome épidermoïde (92). La différence histologique sous-entend une différence de sensibilité pharmacologique donc une différence de prise en charge.

La toxicité toujours élevée des dérivés des moutardes à l'azote a encouragé la recherche d'autres composés anticancéreux. La découverte des propriétés cytostatiques du cisplatine dans les années 1960 a ouvert la voie à son utilisation dans la prise en charge des cancers du poumon, dès son autorisation de mise sur le marché par la FDA en 1978 (93). Afin de réduire la néphrotoxicité, la neurotoxicité, l'ototoxicité et la myélosuppression provoquées par le cisplatine, d'autres dérivés du platine ont été développés, augmentant l'arsenal thérapeutique disponible, notamment le carboplatine (94) et permettant l'émergence d'associations thérapeutiques telles que l'association carboplatine-étoposide, combinant un agent liant l'ADN par création de liaisons intra- et interbrins, à un inhibiteur de la topoisomérase II, piégeant l'enzyme sur l'ADN fragmenté et amplifiant les cassures double brin. La décennie suivante inaugure la troisième génération : d'une part, les vinca semi-synthétiques, au premier rang desquelles la vinorelbine, qui se lie au site Vinca de la β-tubuline et empêche l'assemblage des microtubules lors de la mitose (95). Et d'autre part, les taxanes, paclitaxel puis docetaxel, qui, à l'inverse, stabilisent les microtubules polymérisés, figeant la cellule en métaphase. Parallèlement, la gemcitabine complète cet arsenal en tant qu'analogue de la déoxycytidine. Elle s'insère dans l'ADN et inhibe la ribonucléotide réductase, épuisant le stock de désoxyribonucléotides. Dans les années 2000, les antifolates, avec le pemetrexed, complètent les possibilités pharmacologiques en ciblant la thymidylate synthase. Enfin, au cours de la dernière décennie, les chimiothérapies classiques servent de traitements adjuvants et, associées aux immunothérapies, constituent un instrument de plus en plus précis au service de stratégies multimodales.

La protéine PD-1, pour *programmed death-1*, est une glycoprotéine transmembranaire exprimée par les lymphocytes T, qui a pour effet, lorsqu'elle se lie à son ligand, le PD-L1, d'engendrer des effets inhibiteurs sur la cellule, dont un arrêt du cycle cellulaire. Le PD-L1 exprimé par la tumeur joue un rôle important dans l'évasion immunitaire des cellules tumorales et peut favoriser leur croissance en encourageant l'apoptose des cellules T

situées au voisinage de la tumeur (96). C'est dès 2015, avec l'autorisation de mise sur le marché du nivolumab, qui inhibe le PD-1, empêchant ainsi son interaction avec PD-L1 et restaurant l'activité effectrice des lymphocytes, que les immunothérapies du cancer du poumon ont connu leur premier essor (97) suivi par le pembrolizumab (98). Dans leur sillage, les anti-PD-L1, dont l'atezolizumab, qui bloque le ligand PD-L1 porté par la tumeur ou les cellules présentatrices d'antigène, court-circuitant la synapse inhibitrice, ainsi que le durvalumab puis le cemiplimab, ont confirmé la valeur du ciblage du couple PD-1/PD-L1 dans des contextes de plus en plus précoces (99,100). D'autres immunothérapies ciblent également des glycoprotéines comme CTLA-4, exprimée à la surface des lymphocytes T régulateurs et des lymphocytes T cytotoxiques activés, qui induit un signal d'inhibition de l'activité des lymphocytes T lorsqu'ils se fixent aux récepteurs membranaires (CD80 et CD86), présents à la surface des cellules présentatrices d'antigène ou des cellules dendritiques. Son effet est de bloquer la production d'IL-2 et d'arrêter le cycle cellulaire. Ainsi, l'ajout de l'anti-CTLA-4 ipilimumab au nivolumab a montré un gain de survie durable dans l'essai clinique CheckMate-227 (101). Parallèlement, l'exploration de nouveaux freins immunologiques a conduit au blocage de LAG-3 par relatlimab, visant à lever également l'état d'épuisement lymphocytaire (102,103). Enfin, le tarlatamab, premier anticorps bispécifique indiqué dans les SCLC, contribue à rapprocher directement la molécule CD3 des cellules T du marqueur neuroendocrine DLL3 exprimé par les cellules tumorales, permettant ainsi de déclencher une cytolyse indépendante du HLA (33). Ainsi, l'arsenal s'est déployé selon une chronologie logique : du simple déblocage de PD-1 à la synergie avec CTLA-4, puis à la diversification des checkpoints émergents et, tout récemment, à la réorientation physique des effecteurs T vers la tumeur, témoignant d'une progression continue vers des stratégies de plus en plus précises et intégratives.

Enfin, l'identification des caractéristiques génétiques des tumeurs pulmonaires a contribué au développement de thérapies ciblées qui ont profondément transformé le traitement des cancers du poumon. Contrairement à la chimiothérapie, qui attaque indistinctement les cellules en division, ces traitements innovants ciblent des mutations jugées spécifiques de la tumeur et conférant un avantage sélectif aux cellules tumorales .

Depuis le développement du gefitinib, inhibiteur réversible de l'EGFR et ciblant la voie de signalisation RAS-RAF-MEK-ERK (104), les thérapies ciblées ont connu un développement ininterrompu au fur et à mesure que de nouveaux marqueurs oncogéniques ont été identifiés. Actuellement, selon les lignes directrices du *National Comprehensive Cancer Network* (NCCN), les thérapies ciblées sont utilisées dans les cancers bronchiques non à petites cellules de stade IV, après recherche des marqueurs oncogéniques. Ces tumeurs de stade IV sont des tumeurs au moment du diagnostic ou des cancers au stade précoce ou localement avancé qui ont évolué au stade métastasique. La thérapie systémique est le traitement le plus courant dans ce cas, et la recherche de biomarqueurs et de mutations spécifiques de la tumeur constitue la pierre angulaire de la stratégie thérapeutique. Elle diffère selon que la tumeur présente ou non un driver oncogénique. Les drivers oncogéniques (associés à leurs thérapies ciblées) suivants sont recherchés : mutations d'EGFR (ex19Del, L858R, T790M, insertions de l'exon 20, et autres mutations activatrices) traitées par osimertinib, gefitinib, erlotinib, afatinib et dacomitinib ; réarrangements d'ALK traités par alectinib, brigatinib, lorlatinib, crizotinib et ceritinib ; réarrangement de ROS1 traité par crizotinib, entrectinib et repotrectinib ; mutation BRAF V600E prise en charge par dabrafenib associé au trametinib ; réarrangement de RET traité par pralsetinib ou selpercatinib ; mutation

METex14Del et amplification de MET ciblées par le capmatinib ou le tepotinib ; mutations de HER2 prises en charge par trastuzumab associé au deruxtecan ; mutation KRAS G12C traitée par sotorasib ou adagrasib ; réarrangement de NTRK traité par entrectinib ou larotrectinib ; et enfin, fusions de NRG1 (105).

2. Focalisation sur le récepteur MET

Le récepteur MET est muté dans 3% des cancers pulmonaires (106). Il sera étudié en détail dans le chapitre qui suit, puisqu'il fait partie intégrante du sujet de thèse et que son réseau de régulation, dans un contexte de mutation METex14Del, a été étudié lors d'une collaboration avec l'équipe du Dr Tulasne.

2.1 Un récepteur et son ligand

Le proto-oncogène MET, nom donné en référence au N-methyl-N'-nitro-N-nitroso-guanidine, un agent carcinogène dont l'utilisation a permis d'identifier le gène responsable de la transformation de fibroblastes humains (107), est localisé sur le locus 7q21-q31 et encode le récepteur transmembranaire MET, également appelé récepteur au facteur de croissance hépatocytaire (HGFR). MET et son homologue étroit MST1R constituent une sous-famille spécifique au sein de la superfamille des récepteurs à activité tyrosine kinase (RTK) (41). D'un point de vue structural, le récepteur MET est constitué d'un domaine Sémaphorine (SEMA), d'un domaine plexine-sémaphorine-intégrine (PSI), de quatre domaines consécutifs immunoglobuline-plexine-facteurs de transcription (IPT1-4) dans la région extracellulaire, d'une hélice transmembranaire unique (TM) et d'un domaine kinase intracellulaire (KD) (Figure 8). La protéine MET est initialement exprimée sous forme de précurseur monocaténaire de 150 kDa. Elle subit ensuite un clivage protéolytique entre Arg307 et Ser308 par la protéase furine, menant à sa maturation. La forme mature est composée des sous-unités α (32 kDa) et β (120 kDa), reliées par au moins trois ponts disulfure (108). I faut noter que, par épissage alternatif d'un exon de 54 paires de bases, une isoforme longue de MET existe. Elle ne diffère que par 18 acides aminés de l'isoforme courte. Néanmoins, l'isoforme courte est la plus abondante et code pour l'hétérodimère p190 MET α/β (109). Dans la suite du manuscrit, nous ferons toujours référence à l'isoforme courte.

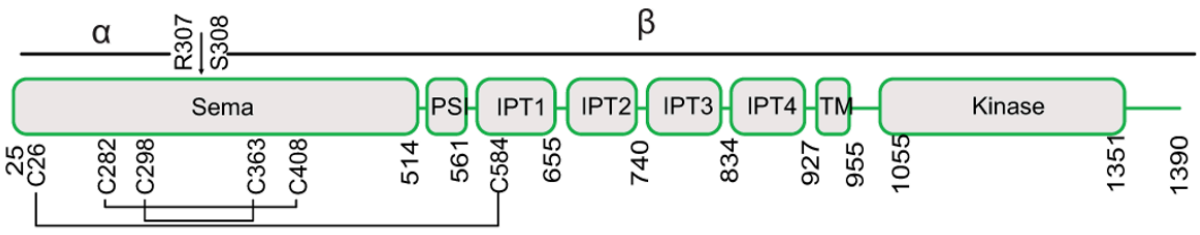


Figure 8 : Structure du récepteur MET

La structure du récepteur MET est décrite avec la présence des ponts disulfures représentés par les traits pleins liants deux acides aminés ainsi que le site de clivage entre les résidus Arg307 et Ser308.
D'après Uchikawa et al.

La même année, Nakamura *et al.* ont isolé un facteur mitogène hépatocytaire à partir du plasma de rats partiellement hépatectomisés, identifié sous le nom de *hepatocyte growth factor* (HGF) (110). En 1987, Stocker *et al.* ont découvert, dans le surnageant de cultures de fibroblastes, un facteur induisant la dispersion des cellules épithéliales, qu'ils ont nommé *scatter factor* (SF) (111). En 1991, il a été démontré que HGF et SF correspondent à une seule et même protéine, renommée HGF/SF, capable d'interagir avec le récepteur MET (112).

L'HGF est initialement sécrété sous la forme d'un précurseur pro-HGF de 728 acides aminés, qui subit ensuite un clivage protéolytique aboutissant à la forme mature de la protéine (Figure 9). La forme active de l'HGF est un hétérodimère composé d'une chaîne α et d'une chaîne β. Quatre domaines kringle forment la chaîne α (K1-4), tandis que des structures de type sérine-protéase constituent la chaîne β (Figure 9). Les activateurs de l'HGF comprennent des protéases sériques et des protéases membranaires, telles que l'activateur de l'HGF (HGF-A), l'activateur du plasminogène de type urokinase (uPA), la kallikréine plasmatique, les facteurs XII et XI de la coagulation, les métalloprotéinases et l'héparine. Le HGF-A est la protéase centrale responsable de l'activation du pro-HGF dans le sérum (113,114). Dans cette forme fonctionnelle, un pont disulfure interchaîne se forme entre Cys487 de la chaîne α et Cys604 de la chaîne β, contribuant ainsi à la stabilisation de la structure tertiaire de l'HGF (115).

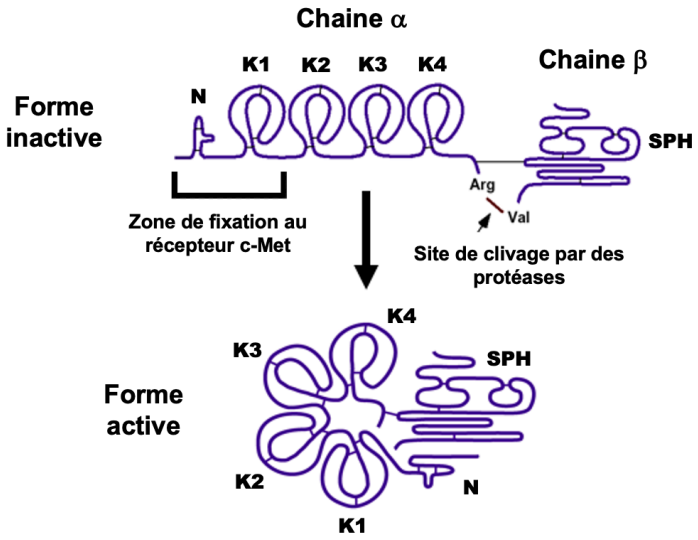


Figure 9: Production et activation du HGF.

Le HGF est produit et sécrété sous forme de pro-HGF par les cellules du stroma telles que les fibroblastes, les macrophages, le mésangium rénal, etc. Le pro-HGF sécrété est clivé en Arg494 et Val495 par les activateurs du HGF.La chaîne lourde du HGF se compose d'un domaine N-terminal et de quatre domaines kringle désignés K1-K4. Chaque domaine kringle comporte trois liaisons disulfure conservées, indiquées en noir. La chaîne légère du HGF contient un domaine de type protéase dépourvu d'activité enzymatique

La fixation de l'HGF peut s'effectuer sur le fragment β du domaine SEMA de MET via sa chaîne α. L'HGF peut aussi se lier à MET au niveau des domaines IPT 3-4, toujours via sa chaîne α (116,117). Son domaine SPH, situé sur la chaîne β, peut également se lier au domaine SEMA de MET, mais nécessite un clivage protéolytique pour permettre la libération d'une chaîne polypeptidique, qui vient ensuite se fixer au domaine SPH et stabiliser la poche d'activation allostérique Une étude récente, basée sur la cryo-microscopie électronique, a permis de

caractériser la structure tridimensionnelle du complexe formé entre le récepteur MET et l'HGF. Cette étude montre qu'une seule molécule de HGF établit un pont entre deux récepteurs MET en engageant à la fois les sites de liaison NK1 et SPH, une deuxième molécule de HGF et l'héparine stabilisant l'interaction (Figure 10).

Un variant d'épissage de l'HGF, NK1, peut également se fixer au récepteur MET. NK1 contient le domaine N et le premier domaine kringle de HGF. Il nécessite la présence d'héparane sulfate pour se dimériser, augmenter son affinité pour MET et ainsi assurer son activité agoniste (118–121).

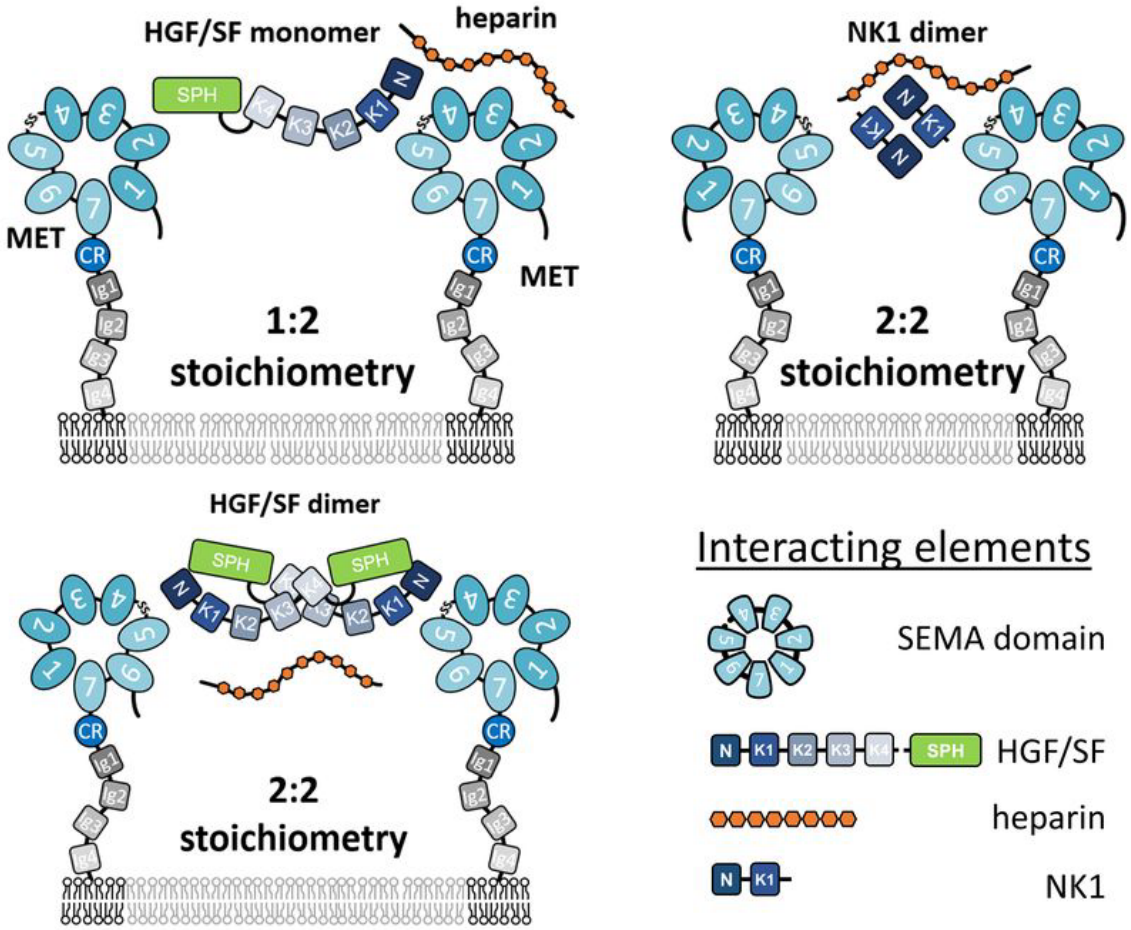


Figure 10 : Modes de dimérisation du récepteur MET.

Modes de liaison possibles aux récepteurs et stœchiométries pour HGF et NK1 en présence d'héparine selon l'hypothèse d'une stœchiométrie de liaison HGF/MET 1:2 et 2:2 ainsi qu'une liaison NK1 dimérisé / MET.
(122)

2.2 Mécanismes d'activation

2.2.1 Activation par son ligand

À la suite de son interaction avec son ligand, le récepteur MET subit une phosphorylation sur deux résidus du domaine catalytique, en position Y1234 et Y1235. D'autres résidus extra-catalytiques sont également phosphorylés, notamment Y1349 et Y1356, situés au sein de la queue C-terminale et constituant un site

d'amarrage multi-substrats (123). La phosphorylation des tyrosines de cette région permet le recrutement de divers effecteurs moléculaires et de molécules adaptatrices essentielles. GRB2 et STAT3 peuvent interagir directement avec les résidus carboxy-terminaux. La protéine adaptatrice GAB1 interagit indirectement (par l'intermédiaire de GRB2) ou directement, en fournissant une structure d'ancrage multisubstrat pour SHC, PI3K, SHP2, PLCγ1 et p120 (Figure 11), ce qui entraîne l'activation des voies de signalisation en aval (124,125). Ce site d'amarrage multi-substrats joue un rôle déterminant dans l'activation des cascades de signalisation intracellulaires médiées par MET.

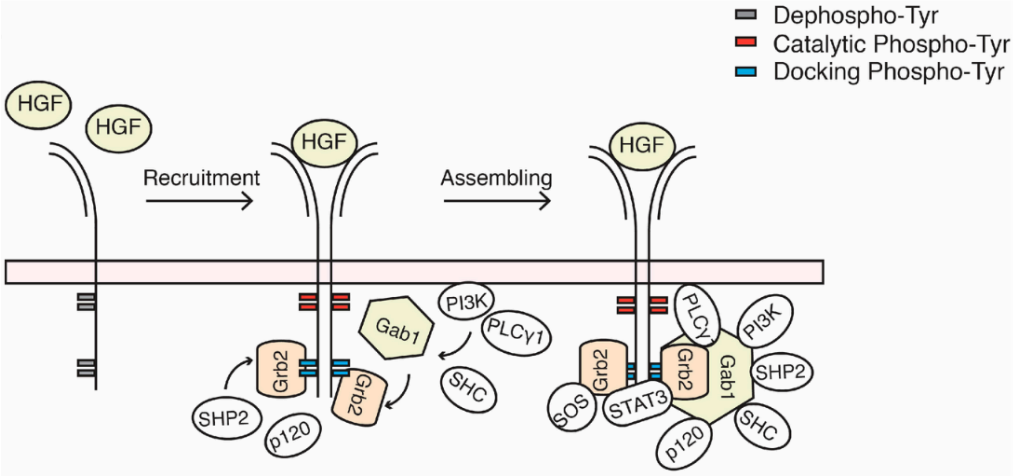


Figure 11 : Activation de MET et amarrage de molécules adaptatrices et transductrices du signal

L'activation de MET déclenche l'homodimérisation du récepteur, la transphosphorylation de ses résidus catalytique Y1234/1234 et extra-catalytique Y1349/Y1356, le recrutement de molécules adaptatrices et l'ancrage sur la plateforme multisubstrat de molécule de transduction du signal.
D'après Viticchiè & Muller, 2015.

2.2.2 Activation indépendant de l'HGF

Le récepteur MET peut être activé et participer à la réponse cellulaire indépendamment de son ligand HGF, notamment en interagissant avec d'autres partenaires, dont les intégrines, les sémaphorines, les plexines et d'autres RTK.

Les intégrines constituent des récepteurs membranaires hétérodimériques, orchestrant les interactions adhésives entre les cellules et leur microenvironnement, englobant la matrice extracellulaire ainsi que les cellules adjacentes. Outre leur rôle fondamental dans l'organisation architecturale des tissus et la modulation de leurs propriétés physico-mécaniques, les intégrines exercent une fonction essentielle en tant que transducteurs moléculaires, relayant les signaux extracellulaires par les voies de signalisation intracellulaires (126). Il a été démontré que l'ensemencement de cellules cancéreuses ovariennes, mammaires, pulmonaires ou prostatiques sur des substrats intégriniques tels que la fibronectine, le collagène ou la laminine induit une phosphorylation de MET dans divers modèles cellulaires (127,128). À l'inverse, la phosphorylation de MET diminue considérablement lorsque les cellules carcinomateuses sont cultivées en suspension, mais elle est restaurée lorsqu'elles sont réadhérées sur des substrats tels que la fibronectine ou le collagène (129,130). L'importance des intégrines dans cette activation a été mise en évidence par des approches d'interférence génétique et d'inhibition fonctionnelle.

L’extinction des intégrines α5, α5β1, β1 ou β5 via des siARN, ainsi que l’utilisation d’anticorps bloquants dirigés contre α5β1 ou α5, entraînent une diminution significative, voire une abolition complète, de la phosphorylation de MET, aussi bien in vitro que dans des xénogreffes murines (127,131). Les mécanismes exacts reliant la signalisation des intégrines à la phosphorylation de MET demeurent à élucider. Il est suggéré que le domaine cytoplasmique des intégrines, en particulier celui de l’intégrine β4, pourrait influencer la phosphorylation de MET (132). Les intégrines peuvent aussi agir en tant qu’adaptateurs moléculaires pour MET, indépendamment de leur fonction adhésive, et ainsi favoriser la survie cellulaire en l’absence d’ancrage (133,134). La formation d’un complexe MET/α6β4 intégrine induit la phosphorylation de la sous-unité β4, activant en aval les effecteurs Shc, PI3K et GAB1 (Grb2-associated binding protein 1). GAB1 et l’intégrine α6β4 agissent alors comme sites d’amarrage supplémentaires, recrutant des transducteurs qui interagissent avec ceux directement liés à MET, ce qui amplifie la signalisation (135).

Les sémaphorines constituent une famille protéique vaste et hétérogène, jouant des rôles essentiels dans la guidance axonale, les réponses immunitaires et l'angiogenèse (136). Elles possèdent un domaine extracellulaire SEMA, qui facilite les interactions avec leurs récepteurs, les plexines. Les plexines, également dotées d'un domaine SEMA extracellulaire, appartiennent à une famille diversifiée de récepteurs transmembranaires ayant des fonctions importantes dans les comportements cellulaires tumoraux (137). L'interaction entre MET et les plexines de classe B, notamment la Plexine B1, illustre la complexité de ces réseaux de signalisation. MET peut s'oligomériser avec la Plexine B1, entraînant l'activation du récepteur par la sémaphorine 4D (SEMA4D), indépendamment de HGF (138). Cette dynamique est modulée par l'équilibre entre différentes RTK. Dans les cellules de carcinome mammaire, la liaison de SEMA4D à la Plexine-B1 associée à ERBB2 favorise la migration cellulaire, tandis que cette même liaison avec la Plexine-B1 associée à MET l'inhibe (139). La Plexine B1 régule la signalisation de MET de manière directe, en inhibant son activation via l’association récepteur-récepteur, et de manière indirecte, en modulant les niveaux de SHP2 (140). La SEMA4D, fortement exprimée dans diverses tumeurs humaines, peut également induire l’angiogenèse via son interaction avec la Plexine B1 et MET dans les cellules endothéliales. Par ailleurs, la SEMA3C a été identifiée comme capable d’activer une variété de récepteurs à activité tyrosine kinase, y compris MET, formant un complexe de signalisation qui favorise la croissance invasive des cellules cancéreuses (141).

MET est également en étroite relation avec d’autres RTK tels que ROR1 et EGFR. MET phosphoryle ROR1, un récepteur surexprimé dans certains cancers. Cette phosphorylation, localisée dans le domaine riche en proline (PRD) de ROR1, recrute SRC, ce qui conduit à une phosphorylation supplémentaire dans le domaine kinase de ROR1 et à l’activation de la signalisation WNT. Dans l'adénocarcinome pulmonaire, ROR1 joue un rôle d'échafaudage moléculaire pour CAVIN1 et CAV1, deux protéines essentielles à la formation et au fonctionnement des cavéoles, des invaginations membranaires impliquées dans la transduction du signal. L'interaction entre CAVIN1 et CAV1 à la membrane plasmique protège CAV1 de la dégradation lysosomale, stabilisant ainsi les voies de survie médiées par AKT et stimulées par plusieurs RTK, dont MET (142).

L'interaction entre MET et EGFR a été particulièrement étudiée dans le contexte de la résistance aux traitements anticancéreux (143,144) MET et EGFR sont souvent coexprimés dans les tumeurs, et la co-immunoprécipitation d'un complexe MET-EGFR a été détectée dans des cellules de carcinome épidermoïde humain et de cancer du poumon (145,146). Bien que ces observations suggèrent la formation d'un complexe hétérodimérique entre les récepteurs MET et EGFR, permettant une transphosphorylation réciproque et leur activation, des études supplémentaires sont encore nécessaires pour prouver une interaction physique directe et non ambiguë entre eux. Une coopération fonctionnelle entre ces deux récepteurs peut également s'exercer de manière indirecte : MET est en effet capable d'activer la tyrosine kinase Src, laquelle peut à son tour phosphoryler EGFR, générant ainsi des sites de liaison pour divers effecteurs d'EGFR impliqués dans la transmission des signaux en aval (147,148). L'interaction MET-EGFR contribue significativement à la progression tumorale et à la résistance médicamenteuse dans le NSCLC. La signalisation d'EGFR favorise la phosphorylation de MET, qui, associée à la co-expression de HER3, amplifie l'activation des récepteurs. Cette interaction implique la régulation des niveaux de MET et la signalisation intermédiaire via les voies MAPK. Dans le NSCLC, qu'EGFR soit ou non muté, l'inhibition d'EGFR ou de MAPK réduit l'activation et les niveaux protéiques de MET (149). La double inhibition de MET et d'EGFR est nécessaire pour freiner efficacement la croissance des lignées cellulaires et des xénogreffes de NSCLC (147). La découverte que 22 % des échantillons de cancer du poumon résistants au gefitinib ou à l'erlotinib présentent une amplification de MET révèle un mécanisme majeur de résistance médicamenteuse (150).

2.3 Activateurs de voies de signalisation

Plusieurs voies de signalisation peuvent être activées à la suite de la liaison du récepteur MET. Parmi elles figurent la cascade des MAPK, la voie PI3K-AKT et l'activation du complexe NF-κB.

L'activation de la cascade de signalisation RAS-RAF-MEK-ERK implique la phosphorylation séquentielle de plusieurs kinases MAPK. Les sérine/thréonine kinases ERK, activées par l'axe RAS-RAF, constituent les principales kinases effectrices de cette voie en aval des RTK. Après l'autophosphorylation du domaine protéine kinase et l'activation par le complexe GRB2/SHC, RAS active successivement RAF et MEK1/2 (151). Une fois phosphorylé par MEK1/2, ERK se dimérise et transloque dans le noyau, où il phosphoryle et stabilise plusieurs facteurs de transcription impliqués dans la transition G1/S du cycle cellulaire, stimulant ainsi la prolifération et la survie cellulaire. Il a été démontré que l'activation de la voie MAPK par MET induit la phosphorylation activatrice du facteur de transcription ETS1 (152). L'expression de ce membre de la famille ETS potentialise la réponse à l'HGF, notamment en amplifiant les capacités cellulaires migratoires et invasives. ETS1 est particulièrement connu pour son rôle dans la régulation transcriptionnelle de plusieurs métalloprotéinases matricielles (MMP), incluant MMP1, MMP7 et MMP9. Les réponses invasives médiées par ETS1 peuvent ainsi être attribuées à la surexpression des MMP (153–155). Ainsi, les MAPK induisent également la dégradation des protéines et de la matrice extracellulaire, favorisent la migration cellulaire et soutiennent la prolifération tumorale (156). Par ailleurs, des travaux ont démontré que l'activation constitutive de MET induit la surexpression de facteurs de transcription de la sous-famille PEA3, appartenant à la famille ETS. Dans un contexte de surexpression du récepteur par amplification génique, les facteurs ETV1, ETV4 et ETV5 de la famille PEA3 s'avèrent être des régulateurs centraux

de la migration induite par MET. Ces facteurs régulent également l'expression de plusieurs MMPs, notamment MMP2 et MMP13 (157).

La liaison de l'HGF au récepteur MET induit son autophosphorylation, générant des sites d'ancrage pour la sous-unité p85 de la kinase PI3K (Figure 12). Une fois activée, PI3K phosphoryle plusieurs dérivés phosphatidylinositols et convertit le phosphatidylinositol-4,5-bisphosphate (PIP2) en phosphatidylinositol-3,4,5-trisphosphate (PIP3). Le PIP3 sert de site d'ancrage pour les protéines contenant un domaine PH, notamment AKT et PDK1. PDK1 phosphoryle AKT sur la thréonine 308, puis mTORC2 complète l'activation sur la sérine 473 (158). Une fois activé, AKT transloque vers le noyau, module des facteurs de transcription tels que FKHRL1, NF-κB et Bcl-2, et inhibe l'expression de gènes suppresseurs de tumeurs. La voie PI3K/AKT/mTOR peut également réguler l'expression de HIF-1α. HIF-1α module ensuite l'expression de protéines impliquées dans le métabolisme du glucose et l'angiogenèse, telles que le facteur de croissance endothélial vasculaire (VEGF) et l'érythropoïétine, favorisant ainsi l'adaptation à l'hypoxie (159). Cette voie peut être inhibée par PTEN (phosphatase and tensin homolog deleted on chromosome 10), qui facilite la déphosphorylation de PIP3 en PIP2, limitant ainsi l'activation des effecteurs en aval de PI3K, tels qu'AKT et mTOR. Dans les cellules tumorales, les mutations ou pertes d'expression de PTEN sont fréquentes, entraînant une activation constitutive et aberrante de la voie PI3K/AKT/mTOR (160).

De manière analogue, la protéine STAT3 peut être recrutée au niveau du site d'ancrage situé à l'extrémité carboxy-terminale du récepteur MET après sa phosphorylation (161,162). La phosphorylation de STAT3 induit alors sa dissociation du récepteur, suivie de sa dimérisation via ses domaines SH2. Ces dimères transloquent ensuite dans le noyau, où ils régulent la transcription de gènes impliqués dans la prolifération cellulaire et les processus de différenciation.

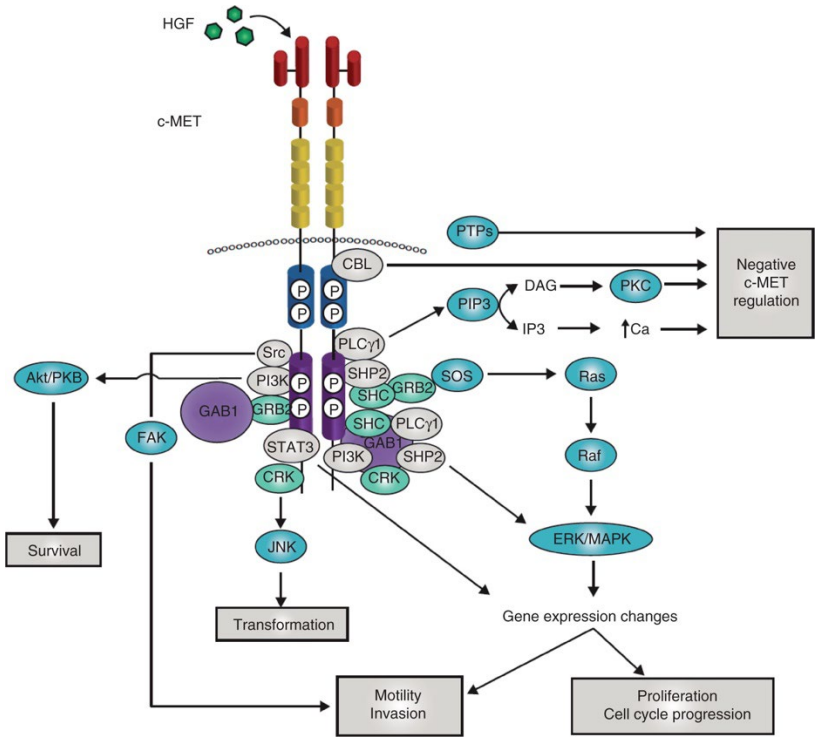


Figure 12 : Médiateurs de la signalisation de MET.

Extrait de "An overview of the MET signaling pathway" (Organ & Tsao, 2011).

Pour finir, la voie NF-κB fait partie des voies de signalisation majeures activées par le récepteur MET. Cette voie comprend des facteurs de transcription présents à l'état inactif dans le cytoplasme (p65/RelA, RelB, p100/p52 NF-κB2, p105/p50 NF-κB1) et qui, lorsqu'ils sont activés, favorisent l'expression de gènes impliqués dans l'inflammation, la prolifération cellulaire, la réponse immunitaire et l'apoptose (163). L'activation des voies Ras/MAPK et PI3K/AKT conduit à l'activation des kinases IKK (IκB kinase), qui sont phosphorylées et dimérisées, puis phosphorylent IκB, entraînant sa dégradation et la libération du complexe NF-κB NF-κB transloque alors dans le noyau et induit l'expression de gènes impliqués dans l'adhésion cellulaire, l'apoptose et la survie cellulaire, tels que cIAP1, cIAP2 et TRAF2 (164,165).

2.4 Une activité finement régulée

2.4.1 Action des phosphatases

L'activité des RTK est régulée négativement par les PTP, une famille de plus de 100 protéines cytoplasmiques ou membranaires. Plusieurs de ces phosphatases interviennent spécifiquement dans la régulation de MET. La phosphatase PTP-S peut se lier au domaine juxtamembranaire du récepteur activé après stimulation par le ligand, sur les résidus 995 et 1009 (166). Par ailleurs, PTP1B et TCPTP, quant à eux, déphosphorylent les résidus Y1234 et Y1235 du domaine kinase activé (167). Plusieurs rapports ont indiqué que PTP1B agit comme un régulateur négatif de l'activation de MET induite par le HGF (168,169). Huang et collaborateurs ont révélé un nouveau rôle oncogénique de THEMIS2 et son mécanisme sous-jacent, qui fonctionne en supprimant l'association de PTP1B avec MET, conduisant ainsi à son activation (170). D'autres

phosphatases jouent également un rôle important : PTPRB déphosphoryle spécifiquement le résidu Y1356, impliqué dans le recrutement des protéines adaptatrices, réduisant ainsi les capacités migratoires induites par l'HGF via l'inhibition de la voie ERK. PTPRF, dont l'expression et l'activité augmentent après stimulation par l'HGF, régule l'activité des voies MAPK et PI3K liées à l'inhibition de contact (171). PTPRJ induit une déphosphorylation préférentielle des résidus Y1349 et Y1356, ainsi que de la protéine GAB1 et de la sous-unité p120 de la β-caténine, impliquée dans la motilité liée à MET (172). Malgré l'importance de ces mécanismes de régulation négative, les modes précis de recrutement et d'action des phosphatases restent encore mal compris.

2.4.2 Clivage protéique

Le récepteur MET peut être soumis à un mécanisme communément désigné sous le terme de « ectodomain shedding », qui implique le clivage de son domaine extracellulaire par l'action de diverses métalloprotéinases matricielles (MMP) et de membres de la famille ADAM, notamment la métalloprotéinase sensible au TIMP-3, ADAM10 et ADAM17 (173). ADAM10/17 clivent MET au niveau de son domaine extracellulaire, sur un site encore mal défini, libérant un fragment soluble nommé soluble MET (sMET), d'une taille d'environ 90 kDa. Le domaine extramembranaire ainsi libéré n'est plus capable de lier HGF, neutralisant ainsi l'activité de MET. Le fragment résiduel transmembranaire et cytoplasmique, d'une taille de 55 kDa subsiste à la membrane. Cependant, ce fragment résiduel peut persister et subir un second clivage intramembranaire, catalysé par le complexe γ-sécrétase, dont la préséniline constitue la sous-unité enzymatique active. Cette étape libère un fragment intracellulaire labile de MET, désigné MET-ICD (Intracellular Domain) ou p50 MET, qui est ensuite dirigé vers le protéasome cytosolique pour y être dégradé (174,175). Le clivage du récepteur MET par ADAM10 agit comme un mécanisme de rétrocontrôle négatif qui régule l'activité de signalisation et maintient l'homéostasie cellulaire. Cependant, dans les cancers, des altérations de l'activité d'ADAM10 peuvent perturber cette régulation. Par exemple, une diminution de l'activité d'ADAM10 peut entraîner une accumulation de récepteurs tyrosine kinase non clivés à la surface cellulaire, amplifiant ainsi les signaux de prolifération et de survie qui contribuent à la progression tumorale et à la résistance aux thérapies ciblées (176).

2.4.3 Régulation par les caspases

Bien que l'activité canonique du récepteur MET soit classiquement associée à la promotion de la survie cellulaire, des données récentes ont mis en évidence une fonction pro-apoptotique contextuelle de ce récepteur. En effet, lors de l'induction de l'apoptose, un fragment intracellulaire d'environ 40 kDa, contenant le domaine kinase de MET, est généré ; il est désigné p40-MET caspase Ce fragment résulte d'un clivage séquentiel par les caspases : un premier clivage survient au niveau de la région carboxy-terminale, sur les résidus D1376 ou D1380 (site $DNAD^{1376}$/$DEVD^{1380}$), suivi d'un second clivage au sein du domaine juxtamembranaire, sur le résidu D1002 (site $ESVD^{1002}$), catalysé par la caspase 3 (177). Ce double clivage entraîne la dissociation du domaine extracellulaire, empêchant ainsi toute activation ultérieure des voies de signalisation pro-survie dépendantes de HGF. Le fragment p40 MET ne se contente pas d'être inactif : il exerce lui-même une activité pro-apoptotique autonome. Sa surexpression dans des cellules MDCK entraîne une augmentation de la mort cellulaire, corrélée à

une activation accrue de la caspase 3 (178). De manière concordante, ce fragment est également détecté in vivo dans un modèle d'hépatite fulminante, suggérant une implication physiopathologique (179). La capacité pro-apoptotique de p40 MET dépend strictement de la réalisation des deux clivages successifs. Des études chez des souris transgéniques knock-in exprimant une forme mutée de MET (p.D1374N), incapable de subir le clivage carboxy-terminal, ont montré une protection accrue des hépatocytes exposés à un agoniste de FAS, confirmant ainsi le rôle protecteur du fragment non clivé (180).

Les travaux de l'équipe du Dr Tulasne ont mis en évidence que le fragment p40 MET se localise préférentiellement au niveau des MAM, des structures spécialisées formant l'interface fonctionnelle entre le réticulum endoplasmique et la mitochondrie. Cette localisation stratégique semble jouer un rôle déterminant dans l'induction de la perméabilisation de la membrane mitochondriale, une étape clé de la cascade apoptotique (180). Les protéines de la famille BCL2, connues pour leur rôle central dans la régulation de l'apoptose, sont localisées à la fois à la membrane externe de la mitochondrie et au niveau du RE (Rong et al., 2008). Parmi elles, BAK apparaît comme un acteur essentiel dans la médiation des effets pro-apoptotiques de p40 MET. En effet, la répression de l'expression de BAK diminue significativement la mort cellulaire induite par p40 MET dans les cellules mammaires MCF10A, tandis qu'aucun effet comparable n'est observé lors de la répression de BAX (179,180).

Des expériences basées sur le transfert d'énergie par résonance de fluorescence (FRET) ont permis de démontrer une interaction physique entre p40 MET et BAK, à la fois dans les cellules mammaires MCF10A et dans les cellules hépatiques IHH. Une fois activé, BAK est capable de provoquer un relargage massif de calcium depuis le RE, conduisant à une accumulation calcique dans la mitochondrie, un phénomène connu pour favoriser l'activation des voies apoptotiques dépendantes de la mitochondrie (181). Ainsi, l'interaction de p40 MET avec BAK, localisée au sein des MAM, pourrait représenter un mécanisme clé par lequel ce fragment pro-apoptotique renforce l'engagement de la cellule vers la mort, en facilitant les transferts calciques entre le RE et la mitochondrie.

2.4.4 Glycosylation

La glycosylation est essentielle à la maturation et aux fonctions des protéines membranaires, notamment des RTK, en régulant leur acheminement, leur conformation et leur liaison aux ligands. La N-glycosylation constitue la modification la plus courante des RTK, consistant en un transfert d'oligosaccharide sur un résidu asparagine. Onze sites potentiels de N-glycosylation sont présents sur la protéine, confirmés par dé-glycosylation enzymatique à l'aide de la neuraminidase et de l'endoglycosidase F (182). Dans certains types cellulaires, la protéine MET peut également subir une O-glycosylation, impliquant le transfert de glycanes sur le groupement hydroxyle (OH) d'un résidu de sérine ou thréonine. Dans les lignées cellulaires de carcinome hépatocellulaire HA22T et PLC5, la modification de MET par des O-glycanes est médiée par C1GALT1, une O-glycosyltransférase de la famille des mucines. La déplétion de C1GALT1 par siARN diminue la phosphorylation et la dimérisation de MET, inhibant ainsi la migration et l'invasion cellulaires (183). Dans les cellules HCT116, la déplétion de la β-galactoside α2,6-sialyltransférase (ST6Gal-I) entraîne une réduction de l'α2,6-sialylation de la MET. Cela diminue

l'expression de la sous-unité β mature sans affecter l'ARNm de MET, suggérant que l'α2,6-sialylation est importante pour le clivage approprié de la pro-MET, et donc pour sa fonction (184). In vivo, l'absence de O-glycosylation semble perturber le clivage et la phosphorylation correcte de MET. De plus, elle induit une séquestration cytoplasmique de la protéine, accélère sa dégradation protéasomale et atténue la signalisation médiée par MET (185,186).

2.4.5 Internalisation et Intégration du signal

À la suite de la liaison du HGF à son récepteur, MET est internalisé et dirigé vers les endosomes par une endocytose dépendante de la clathrine et de la dynamine. En effet, l'inhibition de l'expression de ces deux protéines compromet significativement l'endocytose du récepteur (187,188). L'ubiquitine ligase CBL est également un acteur central de ce processus, pouvant être recrutée à MET selon deux modalités distinctes : directe et indirecte. D'une part, l'ubiquitination de MET par CBL nécessite une interaction directe entre le domaine de liaison aux tyrosines kinases de CBL et Y1003, localisé dans le domaine juxtamembranaire du récepteur (189). D'autre part, CBL peut s'associer indirectement à MET via une interaction entre son domaine SH3 et la protéine adaptatrice GRB2, elle-même liée à MET. De façon similaire, SOCS1 favorise la dégradation des récepteurs tyrosine kinases, dont MET. SOCS1, fréquemment réprimé dans le carcinome hépatocellulaire humain, est considéré comme un suppresseur de tumeur. Le domaine SH2 de SOCS1 se lie à la protéine MET, tandis que sa boîte SOCS recrute la machinerie d'ubiquitination E3, conduisant à une dégradation dépendante du protéasome. De plus, la dégradation de MET médiée par SOCS1 n'est pas dépendante des lysosomes, puisque la bafilomycine et la chloroquine, inhibiteurs lysosomaux, ne bloquent pas cette dégradation. De façon cohérente, la surexpression de SOCS1 diminue la phosphorylation de MET et atténue ainsi sa signalisation hépatique. Il est intéressant de noter que les modifications des lysines, notamment la polyubiquitination liée à K48, sont nécessaires à la dégradation médiée par SOCS1 (190,191).

À partir des endosomes, MET peut soit être recyclé vers la membrane plasmique, soit être acheminé vers les lysosomes pour sa dégradation. Des études récentes indiquent que la PKC joue un rôle crucial dans l'acheminement des vésicules vers le compartiment périnucléaire le long des microtubules. La déplétion de PKCε bloque la dégradation de MET dans les cellules de carcinome hépatocellulaire (192). Contrairement à Y1003, la phosphorylation de Y1313 joue un rôle dans la stabilisation de la protéine MET. Cette phosphorylation permet la liaison de TNS4 (Tensine 4) à MET via son domaine SH2 Une fois liée, TNS4 inhibe l'internalisation de MET vers les lysosomes. Une forte corrélation entre l'expression de TNS4 et de MET est observée dans les cancers colorectaux et ovariens (193).

À l'instar d'autres récepteurs membranaires, MET peut être recyclé vers la membrane plasmique après son internalisation, bien que les mécanismes moléculaires régulant ce processus demeurent partiellement élucidés. Des études menées sur les cellules HeLa ont démontré qu'une fraction du récepteur MET internalisé est soumise à un recyclage consécutif à la stimulation par l'HGF. Ce processus de recyclage nécessite une interaction spécifique entre MET et la protéine GGA3 (Golgi-localised gamma ear-containing Arf-binding protein 3), au niveau des compartiments endosomaux (194). Il est particulièrement notable que les mutations activatrices M1268T (ou

M1250T) et D1246N (ou D1228N), localisées dans le domaine kinase du récepteur, induisent un recyclage constitutif de MET vers la membrane plasmique, associé à une altération de sa dégradation (187).

2.5 Altérations activatrices de MET dans les cancers du poumon

2.5.1 Mécanismes multiples d'activation de MET dans le cancer

Les altérations génomiques affectant MET et entraînant une dérégulation de son activité peuvent jouer un rôle dans l'oncogenèse, faisant de MET un proto-oncogène. À ce jour, quatre grands types d'altérations génétiques ont été identifiés comme responsables de l'activation du proto-oncogène MET, contribuant à une tumorigenèse dépendante de MET (Figure 13). Le premier mécanisme est l'amplification du gène MET, ou gain de copies, qui entraîne une surexpression du récepteur et son activation indépendante du ligand. Ce mécanisme a été impliqué dans plusieurs types de cancers tels que le glioblastome, l'adénocarcinome pulmonaire, le cancer gastrique, le cancer colorectal et le médulloblastome, et a été initialement décrit dans le carcinome gastrique, où il induisait la transformation cellulaire via l'activation de MET (61,195–197). Le deuxième mécanisme repose sur la surexpression de son ligand, HGF, qui établit une boucle autocrine favorisant l'activation aberrante de MET, notamment dans les cancers du sein, de l'estomac, du côlon et du poumon (198). Le troisième mécanisme implique des réarrangements chromosomiques menant à la formation de gènes de fusion avec MET, tels que KIF5B-MET dans l'adénocarcinome pulmonaire, ou BAIAP2L1-MET et C8ORF34-MET dans les cancers du rein, entraînant une activation constitutive ou une surexpression du récepteur (199). Enfin, les mutations activatrices de MET constituent un mécanisme majeur d'activation, décrit pour la première fois en 1997 dans le carcinome papillaire rénal héréditaire. La majorité de ces mutations, localisées dans le domaine tyrosine kinase, ont été identifiées dans les cancers rénaux papillaires. Cependant, des mutations hors de ce domaine ont également été identifiées, notamment dans les domaines SEMA et juxtamembranaire (JM), ainsi que de nombreuses mutations provoquant l'exclusion de l'exon 14, particulièrement fréquentes dans les NSCLC. Ces altérations sont en constante expansion, avec la découverte régulière de nouvelles mutations de MET dans divers contextes (200).

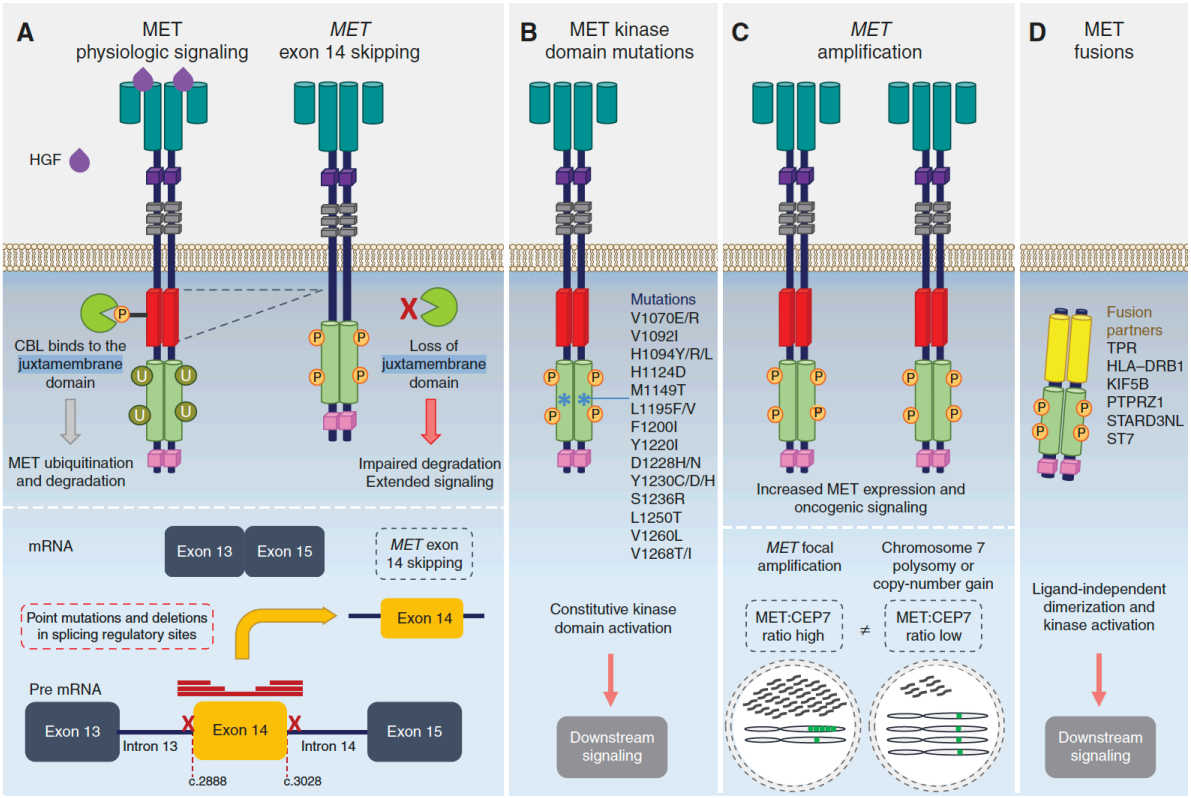


Figure 13 : Mécanisme d'activation oncogénique de MET.

A. L'activation du récepteur MET débute par la fixation du facteur de croissance hépatocytaire (HGF) à son sous-domaine SEMA extracellulaire, induisant sa dimérisation et la phosphorylation du domaine intracellulaire. L'ubiquitine ligase E3 CBL interagit avec le résidu Y1003 du domaine juxtamembranaire (codé en partie par l'exon 14), favorisant l'ubiquitination et la dégradation du récepteur. Les altérations de l'exon 14 (mutations ou délétions) entraînent un saut d'exon, une suppression partielle du domaine juxtamembranaire, une altération de la liaison de CBL et une activation constitutive des voies oncogéniques. B. Les mutations ponctuelles du domaine tyrosine kinase (TKD) de MET induisent une auto-phosphorylation indépendante du ligand, favorisant une signalisation oncogénique persistante. C. L'amplification focale de MET entraîne une surexpression du récepteur, distincte de la polysomie du chromosome 7 (rapport MET:CEP7 élevé vs. faible). Cette surexpression favorise une signalisation oncogénique constitutive via autodimérisation, oligomérisation et auto-phosphorylation du récepteur. D. Les réarrangements impliquant le domaine TKD de MET génèrent des protéines de fusion capables de s'auto-dimériser indépendamment du ligand, entraînant une activation persistante du récepteur.
Issu de Chen et al. 2024

2.5.2 Mutations du domaine kinase

Bien que de nombreuses altérations de la signalisation MET aient été décrites, aucun lien causal direct entre MET et la progression tumorale n'avait été établi jusqu'en 1997, date à laquelle des mutations activatrices de MET ont été identifiées dans le carcinome papillaire rénal héréditaire (HPRC), marquant la première démonstration d'une association génétique directe avec le développement tumoral (201). Cette relation causale a été établie grâce à une analyse de polymorphismes génétiques montrant une ségrégation familiale avec un locus situé sur le chromosome 7q, où se localise le proto-oncogène MET. Contrairement à la perte d'hétérozygotie classiquement observée dans les cancers, une duplication du chromosome 7 suggérait ici l'implication d'un oncogène, renforçant l'hypothèse d'un rôle oncogénique pour MET. Ce gène fut d'autant plus ciblé qu'il appartient à la même famille de récepteurs à activité tyrosine kinase que le proto-oncogène RET, déjà connu pour son implication dans d'autres syndromes tumoraux héréditaires (202). Bien que les mutations MET les plus connues parmi les NSCLC concernent l'exon 14, une nouvelle catégorie de mutations activatrices dans le domaine

tyrosine kinase de MET suscite un intérêt croissant. Dans une large cohorte, 0,5 à 0,9 % des échantillons de NSCLC présentaient des mutations du domaine kinase sans altérations concomitantes de METex14, incluant des variantes telles que H1094Y, D1228H/N et M1250T (203). Ces mutations définissent un nouveau sous-type moléculaire potentiellement sensible à des inhibiteurs de MET tels que l'elzovantinib, avec des réponses partielles observées chez certains patients ne présentant pas d'autres altérations oncogéniques. Une autre étude multicentrique a identifié des mutations du domaine kinase de MET chez des patients atteints principalement d'un adénocarcinome pulmonaire avancé, souvent en association avec des altérations de TP53, KRAS ou NRAS (204). Ces mutations étaient réparties sur les exons 15 à 19, avec H1112Y et H1097R parmi les plus fréquentes. Toutefois, aucun patient de cette cohorte n'avait été traité avec un inhibiteur spécifique de MET, laissant ouverte la question de leur potentiel thérapeutique. La classification des mutations de MET reste un défi. Enfin, l'analyse fonctionnelle à haut débit a permis d'identifier des régions structurales critiques dans la régulation du domaine kinase de MET, révélant des mutations activatrices jusque-là inconnues, susceptibles d'influencer l'oncogenèse et la résistance aux traitements ciblés (205).

2.5.3 Conséquences fonctionnelles des mutations du domaine kinase

À la suite de la mise en évidence de mutations activatrices du proto-oncogène MET, de nombreuses études fonctionnelles ont été entreprises afin de caractériser leur impact biochimique et tumorigène. Des essais menés sur des fibroblastes murins transfectés par des constructions d'ADNc de MET muté ont démontré que l'introduction de mutations spécifiques induit une augmentation significative de l'autophosphorylation du récepteur, révélant une activation constitutive de son activité tyrosine kinase en l'absence de ligand (106,206,207). L'intensité de cette activation kinase varie selon la nature de la mutation : la substitution M1268T présente le plus fort potentiel d'activation, suivie de Y1248H, L1213V et D1246H (208). Comme vu précédemment, certaines mutations entraînent une activation de MET indépendamment de son ligand, tandis que d'autres, à activité plus modeste (M1149T, V1206L, L1213V, V1238I), nécessitent une stimulation exogène par HGF pour induire une phosphorylation accrue de MET et un phénotype transformant (209,210). Ainsi, la dépendance au ligand et l'intensité du signal oncogénique semblent corrélées au niveau d'activation intrinsèque conféré par chaque mutation. Le caractère oncogénique intrinsèque de ces altérations a été validé par des modèles murins « knock-in » portant des mutations germinales de MET, qui développent spontanément des tumeurs, confirmant leur rôle direct dans l'initiation tumorale (211,212). Enfin, les données disponibles sur les voies de signalisation en aval de MET muté restent limitées. Toutefois, pour les mutations M1268T et Y1248H, une augmentation du recyclage membranaire du récepteur ainsi qu'une réduction de sa dégradation ont été observées, suggérant un mécanisme de stabilisation post-transductionnelle renforçant l'intensité et la durée du signal oncogénique (187).

2.5.4 Mutations du domaine juxtamembranaire

Alors que les mutations de MET associées aux cancers rénaux se concentrent essentiellement dans le domaine tyrosine kinase, celles identifiées dans les cancers pulmonaires affectent majoritairement le domaine

juxtamembranaire, codé par l'exon 14. Les premières mutations décrites dans cette région étaient des substitutions ponctuelles, telles que R988C, P1009S et T1010I, responsables de modifications d'acides aminés au sein de la région juxtamembranaire du récepteur (213,214). Parmi celles-ci, R988C est retrouvée chez environ 1 % des patients atteints de NSCLC (215). Bien que ces mutations ne conduisent pas à une activation classique de la kinase MET, plusieurs études ont montré qu'elles peuvent favoriser des comportements tumoraux, notamment en augmentant la migration cellulaire, la capacité à former des colonies en milieu semi-solide et la production intracellulaire d'espèces réactives de l'oxygène — autant de processus impliqués dans la transformation oncogénique (216,217). Le caractère réellement activant de ces mutations reste cependant débattu. En effet, elles n'induisent pas d'autophosphorylation du domaine kinase, ce qui rend leur mécanisme d'action moins évident que celui des mutations classiques d'activation (218). En 2017, Montagne et al. ont mis en évidence que la mutation R988C favorise le clivage protéolytique de MET par la calpaïne, une protéase activée par le calcium. Ce clivage produit un fragment intracellulaire de 45 kDa, désigné p45 MET, qui, en présence de stimulation par le HGF, induit la migration et l'invasion des cellules épithéliales (219). Contrairement au fragment p40 MET, généré lors de l'apoptose, le fragment p45 ne participe pas à la mort cellulaire, ce qui suggère que la protéolyse de MET est un processus contextuel, capable de générer des formes protéiques distinctes selon les conditions cellulaires, avec des fonctions divergentes (220).

En somme, même si les mutations juxtamembranaires comme R988C ne suivent pas les modèles classiques d'activation oncogénique, elles participent vraisemblablement à la progression tumorale via des mécanismes non canoniques, en particulier par la modulation du microenvironnement intracellulaire et le remodelage fonctionnel de la protéine MET.

2.5.5 Mutations du domaine extracellulaire

Longtemps considérées comme de simples variantes génétiques sans retentissement fonctionnel, les mutations affectant le domaine extracellulaire du récepteur MET n'ont suscité qu'un intérêt limité jusqu'à récemment. Parmi celles-ci, la substitution N375S, localisée dans le domaine Semaphorine, constitue le polymorphisme germinal le plus fréquemment identifié du gène MET, avec une prévalence particulièrement élevée dans les populations d'Asie de l'Est (215). Initialement détectée dans divers cancers, notamment pulmonaires, cette mutation fut longtemps interprétée comme une variation neutre sur le plan biologique (221). Toutefois, des travaux récents ont révélé que MET N375S confère une affinité de liaison anormalement élevée pour le récepteur HER2, favorisant la formation de dimères MET/HER2 indépendamment du ligand HGF et engendrant une signalisation aberrante via l'axe HER2. De manière paradoxale, ces tumeurs se révèlent sensibles aux inhibiteurs de HER2, mais non aux inhibiteurs de MET, ce qui suggère une dépendance fonctionnelle détournée vers l'axe HER2, malgré l'implication de MET dans la structure du dimère (222). Par ailleurs, des études fonctionnelles ont montré que la mutation N375S est associée à une résistance accrue aux inhibiteurs de MET, comme SU11274, traduisant une altération de la sensibilité thérapeutique en lien avec une perte partielle d'affinité pour HGF (223). De plus, cette mutation a été corrélée à une chimiorésistance à la cisplatine dans certains cancers, comme celui de l'endomètre, mettant en évidence son rôle dans la plasticité adaptative des cellules tumorales face aux traitements cytotoxiques (224).

2.5.6 Ciblage pharmacologique des mutations du domaine kinase

Le domaine tyrosine kinase du récepteur MET constitue une cible thérapeutique majeure dans plusieurs cancers, notamment les NSCLC, en particulier lorsqu'il présente des mutations activatrices indépendantes des altérations classiques de l'exon 14. Des mutations telles que H1094Y, L1195F, D1228H/N et M1250T ont été décrites comme induisant une activation constitutive du récepteur, favorisant ainsi une signalisation oncogénique soutenue et une dépendance tumorale à MET (203). Les inhibiteurs de tyrosine kinase dirigés contre MET se classent en deux grandes catégories : les inhibiteurs de type I, tels que le capmatinib et le tepotinib, qui ciblent la conformation active de la kinase, et les inhibiteurs de type II, comme le cabozantinib, qui se lient à la conformation inactive. Les mutations secondaires D1228X et Y1230X ont été associées à une résistance aux inhibiteurs de type I, nécessitant le recours à des composés de type II pour restaurer une efficacité thérapeutique (225). Des essais cliniques ont également mis en lumière l'intérêt du ciblage du domaine kinase de MET dans d'autres cancers, comme le carcinome rénal papillifère. Par exemple, le foretinib, un ITK multi-cible, a montré une réponse objective dans 50 % des patients porteurs de mutations germinales de MET, contre aucune réponse significative chez les patients sans mutation (226). De même, le crizotinib a permis un contrôle durable de la maladie (> 12 mois de survie sans progression) dans 75 % des cas MET-positifs (227). L'efficacité du savolitinib a aussi été rapportée dans l'essai SAVOIR, où son taux de réponse atteignait 27 %, contre 7 % avec le sunitinib (228). Enfin, les approches combinées, telles que l'association du savolitinib avec l'inhibiteur de PD-L1 durvalumab, ont montré une survie médiane sans progression de 12 mois et une survie globale de 27,4 mois dans les cancers métastatiques présentant une altération de MET (229).

2.5.7 Mutations secondaires de résistance

Les mutations secondaires du récepteur MET constituent l'un des principaux mécanismes de résistance acquise aux inhibiteurs ciblant cette tyrosine kinase. Sous pression thérapeutique, la tumeur développe des altérations génétiques spécifiques compromettant l'efficacité initiale des traitements. Les mutations les plus fréquemment impliquées dans ce processus sont localisées dans le domaine kinase de MET, en particulier sur les résidus D1228, Y1230, L1195 et F1200. Ces modifications structurales perturbent la conformation du site de liaison des inhibiteurs, réduisant leur affinité et, par conséquent, leur capacité à bloquer l'activité du récepteur. Par exemple, les substitutions D1228N/V et Y1230C/H confèrent une résistance marquée aux inhibiteurs de type I (tels que le capmatinib et le tepotinib), tout en maintenant, dans certains cas, une sensibilité partielle aux inhibiteurs de type II tels que le cabozantinib (230). Dans les formes de CPNPC combinant une mutation de EGFR avec une amplification de MET, des mutations secondaires telles que D1246N, D1228N ou Y1230H ont été identifiées après échec de traitements combinés EGFR/MET. Ces altérations illustrent la plasticité génétique des cellules tumorales, capables de contourner simultanément plusieurs cibles thérapeutiques. À cela s'ajoute la réactivation de voies de signalisation parallèles, notamment RAS/MAPK via des mutations secondaires de KRAS, ou encore des altérations concomitantes de HER2, contribuant à une résistance indépendante de MET (231–233). Sur le plan fonctionnel, ces mutations n'altèrent pas seulement la pharmacodynamie des inhibiteurs ; elles

peuvent également renforcer l'activation constitutive du récepteur MET de manière indépendante du ligand, amplifiant ainsi le signal oncogénique. Cette complexité moléculaire souligne la nécessité de stratégies thérapeutiques adaptatives, combinant divers agents ciblés ou intégrant de nouveaux inhibiteurs capables de surmonter les barrières posées par ces mutations de résistance. Les progrès récents en cartographie fonctionnelle des mutations offrent des pistes prometteuses pour anticiper ces résistances et guider la conception de traitements de nouvelle génération.

2.6 Conséquences de la mutation METex14Del

La fréquence globale du saut de l'exon 14 (METex14Del) dans les cancers pulmonaires non à petites cellules (NSCLC) est estimée à environ 2 %, et cette altération est généralement mutuellement exclusive des autres événements oncogéniques majeurs, tels que les mutations de BRAF (5–8 %), les réarrangements d'ALK (2–7 %) ou de ROS1 (1–2 %). Contrairement à d'autres altérations génétiques fréquemment associées à des profils cliniques distincts (telles que les mutations EGFR chez les femmes asiatiques non fumeuses), le saut de l'exon 14 de MET est plus fréquemment observé chez des patients âgés, souvent fumeurs, et est associé à un pronostic défavorable (234). Le saut de l'exon 14 de MET a initialement été décrit en 1994 comme un variant d'épissage alternatif de l'ADNc, isolé à partir de tissus normaux de rein, foie et cerveau murins, en l'absence de modification des séquences consensus d'épissage (235). Ce n'est qu'en 2005 que METex14Del a été identifié pour la première fois dans des tissus de NSCLC humains comme résultant d'une mutation somatique (236). Dans les NSCLC, METex14Del, correspondant à la délétion complète des acides aminés du domaine juxtamembranaire (Δaa963–1009), résulte de diverses aberrations moléculaires. Ces altérations affectent principalement les sites donneurs et accepteurs d'épissage de l'exon 14. Une analyse génomique portant sur 69 219 patients atteints de NSCLC a révélé que 1 592 d'entre eux (soit 2,3 %) présentaient une altération METex14, répartie sur plusieurs sites fonctionnels distincts. Les tumeurs porteuses de METex14Del présentaient un faible taux de mutations somatiques et une surexpression de PD-L1, éléments potentiellement prédictifs d'une réponse favorable aux immunothérapies. Parmi les altérations moléculaires associées, des coamplifications de MDM2 (34 %), CDK4 (19 %) et MET (11 %), ainsi que des mutations de TP53 (42 %), ont été fréquemment observées dans les échantillons tissulaires (237).

La mutation METex14Del est ciblée par plusieurs inhibiteurs de tyrosine kinase (TKI), notamment les inhibiteurs spécifiques tepotinib et capmatinib, ainsi que par l'inhibiteur multikinases crizotinib. Toutefois, les taux de réponse et les survies sans progression observés sous MET-TKI restent inférieurs à ceux obtenus avec les TKIs ciblant d'autres oncogènes des NSCLC, tels qu'EGFR ou ALK. Le crizotinib, le capmatinib et le tepotinib présentent des taux de réponse objective de 44 %, 41 % et 46 %, respectivement (238–240). Ces résultats soulignent la nécessité croissante de mieux comprendre les mécanismes moléculaires sous-jacents à l'oncogénicité de METex14Del, ainsi que les déterminants de la sensibilité ou de la résistance aux thérapies ciblant MET.

Une étude récente menée par l'équipe du Dr Tulasne a démontré, à l'aide d'un modèle de cellules pulmonaires humaines exprimant spécifiquement la mutation METex14Del, que celle-ci constitue bien une mutation driver, mais que son pouvoir transformant est strictement dépendant de la présence de HGF, tant in

vitro qu'in vivo. Lorsqu'HGF est exprimé à un niveau endogène et en l'absence d'altérations génétiques concomitantes, l'activation de METex14Del induit une transformation cellulaire sensible à l'inhibition par plusieurs inhibiteurs de tyrosine kinase ciblant MET (241).

L'impact pronostique des mutations de l'exon 14 de MET demeure encore insuffisamment caractérisé. Néanmoins, une analyse rétrospective portant sur une cohorte de 687 patients atteints de NSCLC a montré que les mutations de l'exon 14 de MET, ainsi que les amplifications du gène, étaient associées à un pronostic défavorable. Une étude complémentaire menée dans une population asiatique a montré que les patients présentant une mutation de l'exon 14 de MET avaient un taux de récidive significativement plus élevé après résection chirurgicale, comparativement à ceux porteurs de réarrangements d'ALK. Toutefois, leur survie globale reste comparable à celle des patients présentant d'autres altérations oncogéniques, telles que des mutations de EGFR ou des fusions impliquant ROS1, ALK ou RET (242). Enfin, d'autres études ont montré que, bien que le saut de l'exon 14 de MET soit associé à un pronostic défavorable, il n'est pas corrélé à une évolution vers un stade avancé du cancer pulmonaire (243).

2.6.1 Mécanismes de régulation présents au sein de l'exon 14

L'exon 14 de MET code principalement pour le domaine juxtamembranaire du récepteur, région clé de la régulation de son activité. Ce segment contient notamment la sérine 985 (S985), dont la phosphorylation par la PKC diminue l'activité kinase de MET, selon un mécanisme encore mal élucidé. La déphosphorylation de S985 par la phosphatase PP2A permet un retour à l'état basal du récepteur (244). Une diminution de la phosphorylation de S985, corrélée à une augmentation de la réponse mitogène induite par HGF, a été observée au cours de la régénération hépatique post-lésionnelle (245).

L'exon 14 de MET contient également le résidu Y1003, qui est phosphorylé après activation du récepteur, permettant le recrutement de l'ubiquitine ligase CBL. Cette interaction déclenche l'ubiquitination de MET, conduisant à son internalisation suivie de sa dégradation par les protéases lysosomales (246). Bien que CBL puisse également se fixer de manière indirecte à l'extrémité carboxy-terminale de MET via GRB2, la mutation du résidu Y1003 suffit à altérer cette interaction, compromettant la dégradation du récepteur (247). La perte du site régulateur Y1003 entraîne également une activation accrue des métalloprotéinases matricielles (MMP), par persistance de la signalisation RAS–ERK, favorisant la motilité et l'invasion cellulaire (248) (Figure 14).

Enfin, le domaine juxtamembranaire de MET contient également un site de clivage par les caspases, générant le fragment intracellulaire p40MET (177). Ce clivage dissocie le domaine catalytique de la portion extracellulaire du récepteur. Le fragment p40MET possède une activité pro-apoptotique, amplifiant les mécanismes de mort cellulaire via l'activation des flux calciques et la perméabilisation de la membrane mitochondriale (178,180).

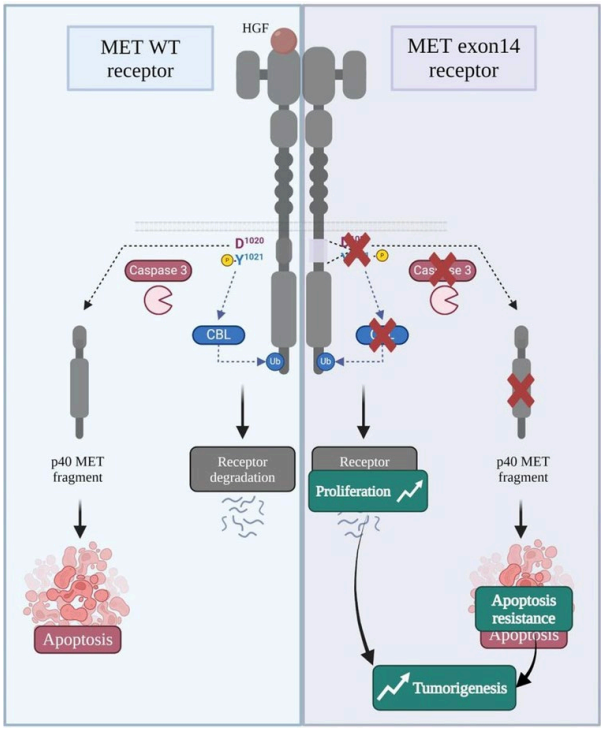


Figure 14 : Régulation de MET en condition physiologique et en présence du saut de l'exon 14

Physiologiquement, le récepteur MET est phosphorylé sur la Y1021, ce qui permet le recrutement de la ligase CBL. Le récepteur est alors ubiquitiné ce qui favorise internalisation puis sa dégradation. En l'absence de stimulation par son ligand et dans des conditions de stress, MET peut être clivé par les caspases, générant p40 MET capable d'amplifier la mort cellulaire par perméabilisation de la mitochondrie. La délétion du domaine juxtamembranaire, codé par l'exon 14, empêche ces mécanismes de régulation négative. Ce qui confère à MET une stabilisation, une augmentation de son activité tyrosine kinase et une résistance à l'apoptose. Par conséquent, ces altérations favorisent une prolifération accrue, une augmentation de la motilité, de la migration et de l'invasion cellulaires, ainsi qu'une meilleure survie cellulaire.
Issu de Guérin & Tulasne, 2024

Ainsi, contrairement à la plupart des mutations activatrices décrites pour les récepteurs à activité tyrosine kinase, qui induisent généralement une activation constitutive indépendante du ligand, le saut de l'exon 14 de MET ne supprime pas la dépendance du récepteur à la stimulation par son ligand, HGF. Cette observation suggère l'existence de configurations d'activation distinctes du récepteur, modulées par la disponibilité locale du ligand au sein du microenvironnement tumoral. En mobilisant les approches de biologie des systèmes et l'analyse des réseaux de régulation génique, nous avons construit un réseau transcriptionnel spécifique du cancer du poumon, fondé sur l'intégration de données transcriptomiques issues de lignées cellulaires pulmonaires. Ce réseau a ensuite servi de base à une modélisation dynamique de la signalisation induite par METex14del, en conditions basales ou après stimulation par HGF, dans le cadre de notre collaboration avec l'équipe du Dr Tulasne (249).

3. Biologie des systèmes

L'oncologie moderne est confrontée à la complexité intrinsèque des systèmes biologiques impliqués dans le développement et la progression des tumeurs. Face à la diversité des altérations génétiques, épigénétiques, transcriptionnelles et métaboliques observées dans les cancers, la biologie des systèmes s'impose comme une approche intégrative incontournable. En combinant modélisation mathématique, biologie computationnelle et données expérimentales multi-omiques, elle vise à élucider les interactions complexes entre les composants d'un système biologique. En oncologie, elle permet de décrypter les réseaux de signalisation, de régulation génique et de communication cellulaire qui sous-tendent les processus tumoraux, d'identifier des mécanismes moléculaires centraux et de révéler des vulnérabilités thérapeutiques potentielles. En offrant une vision holistique du cancer, la biologie des systèmes ouvre la voie à une médecine de précision plus rationnelle, mécaniste et stratifiée.

3.1 Modéliser les phénotypes tumoraux

3.1.1 Appréhender le destin cellulaire

La régulation du phénotype cellulaire est un processus complexe, opéré à différentes échelles spatio-temporelles. Spatiales, car différentes zones d'un système peuvent présenter une spécialisation fonctionnelle, comme cela a été démontré dans le lobule hépatique (250). Temporelles, car les processus de différenciation sont évolutifs et impliquent le passage par plusieurs états transitoires avant d'atteindre un état différencié stable. La plasticité cellulaire, définie comme la capacité des cellules à adopter différentes identités le long d'un spectre phénotypique, constitue un phénomène central, reconnu pour son rôle clé dans la résistance thérapeutique et les métastases, deux causes majeures de mortalité liée au cancer. Au cours de l'évolution tumorale, les cellules subissent une forte pression sélective exercée par leur microenvironnement. Cette pression favorise la diversification des compartiments malins et non malins (endothéliaux, stromaux, immunitaires) du microenvironnement tumoral, conduisant à une hétérogénéité intratumorale accrue, propice à la progression tumorale et à la résistance thérapeutique (251). Dans le but d'étudier le destin cellulaire, Conrad Waddington a introduit dès 1957 le concept de paysage épigénétique, conçu comme un intermédiaire entre génotype et phénotype (252). Il convient de souligner que le terme « épigénétique », tel que défini par Waddington, renvoie à tout mécanisme échappant au contrôle génétique classique. Il diffère ainsi de l'acception moderne, qui désigne plus spécifiquement les processus de régulation tels que la méthylation de l'ADN ou les modifications de la chromatine. Par conséquent, dans ce manuscrit, le terme « épigénétique » se réfère au sens original proposé par Waddington.

Le paysage épigénétique est représenté comme une surface inclinée formée de vallées séparées par des crêtes, définissant des trajectoires stables qu'un système en développement peut emprunter pour atteindre un état transitoire ou final. Ces trajectoires développementales, nommées « chréodes » par Waddington, reflètent des options dynamiques façonnées par l'état initial du système, considéré comme un équilibre métastable (Figure

15). Pour Waddington, les interactions complexes et souvent invisibles entre les gènes et leurs produits sculptaient la surface du paysage épigénétique, à l'image de la tectonique des plaques façonnant le relief terrestre(252) de la même manière que la tectonique des plaques souterraine sculpte la surface de la Terre. Dans cette perspective, les activités individuelles des gènes importent moins que leurs effets collectifs dans la dynamique du développement.

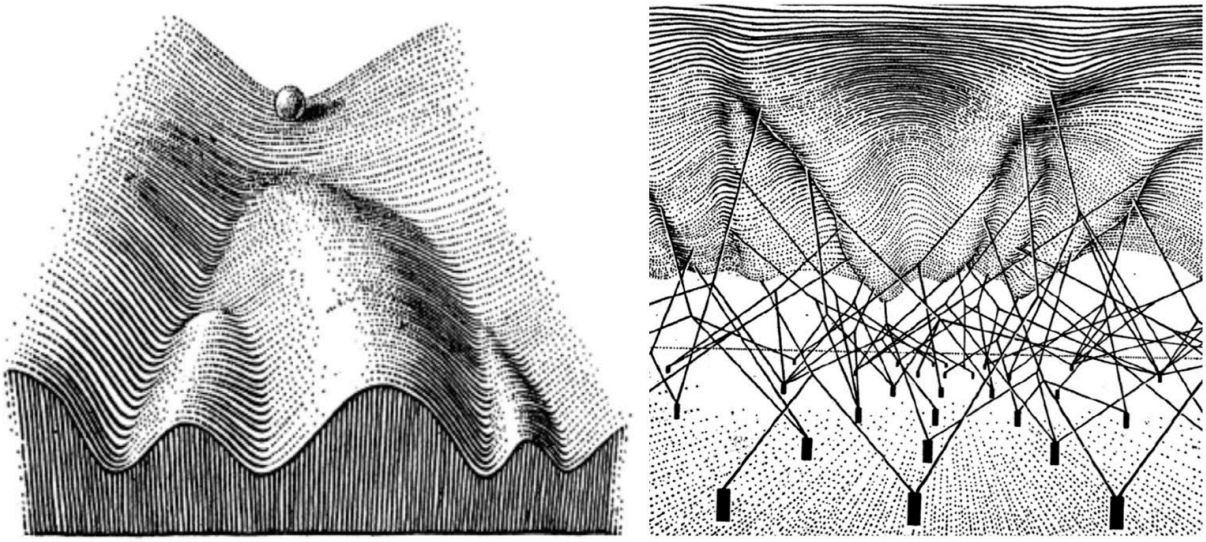


Figure 15: Le paysage épigénétique.

Gauche : la sphère au sommet est dans un équilibre instable compte tenu de la pente du paysage, et elle peut représenter une cellule qui fait face à des pressions internes et externes dont les possibilités sont représentées par les vallées ramifiées définissant des voies (chréodes) sur la surface du paysage. Droite : les chevilles fixées dans une surface souterraine cachée représentent les gènes individuels tandis que les cordons représentent les conséquences de leur expression qui exercent une traction vers le bas, modelant ainsi la surface du paysage
Illustration de Waddington, 1957

Appliqué à la cellule, ce modèle permet de comprendre qu'en fonction de ses caractéristiques intrinsèques — épigénétiques (au sens moderne du terme), génomiques, métaboliques et transcriptionnelles — et extrinsèques — facteurs physiques ou influences du microenvironnement immunitaire, stromal ou cellulaire immédiat —, le phénotype cellulaire tend à se stabiliser dans des états spécifiques, tout en conservant la capacité d'évoluer vers d'autres états cellulaires stables.

3.1.2 Les réseaux de régulations des états cellulaires stables

La tumeur, considérée comme un système complexe multicellulaire sous l'influence de facteurs intrinsèques et extrinsèques, peut donc être étudiée selon ce paradigme. Les stimuli extracellulaires auxquels la cellule est exposée activent des voies de signalisation intracellulaires, qui à leur tour induisent des programmes transcriptionnels spécifiques via l'action de facteurs de transcription. Au cours de la transformation d'une cellule normale en cellule cancéreuse, des mutations ponctuelles perturbent les voies de signalisation en place, entraînant une dérégulation des programmes transcriptionnels associés. Ainsi, plutôt que d'être une maladie des gènes, le cancer peut être envisagé comme une maladie de la régulation génique, et donc de la cellule elle-même.

Il devient alors essentiel de considérer les dynamiques collectives de l'expression génique au sein d'un réseau régulateur contrôlant le comportement cellulaire (253).

Ce réarrangement du réseau entraîne l'émergence de phénotypes anormaux ou malins dans les cellules cancéreuses. Lorsqu'elles sont évaluées dans le contexte d'un paysage épigénétique, les cellules cancéreuses et les cellules normales présentent des motifs d'attracteurs complètement différents (Figure 16). Les réseaux biologiques complexes, riches en boucles de rétroaction interconnectées, présentent souvent une multistabilité et une redondance fonctionnelle dans leur architecture, assurant le maintien de la cellule au sein de puits d'attracteurs stables (254,255).

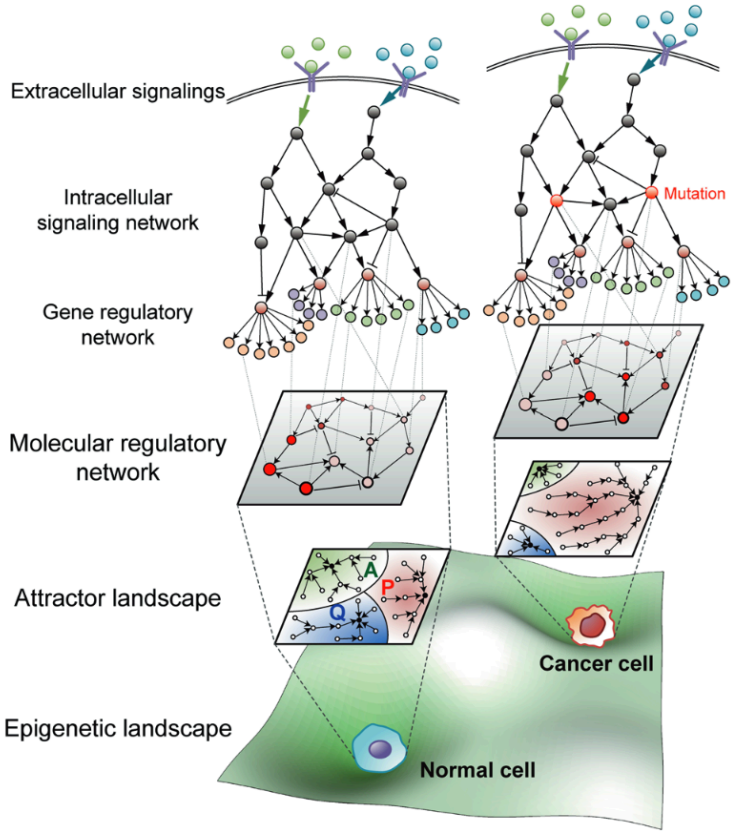


Figure 16 : Décomposition du paysage épigénétique par réseau de régulation génique.

Les propriétés phénotypiques des cellules peuvent être considérées comme des paysages d'attracteurs des réseaux de régulation moléculaire, incluant les voies de signalisation et les réseaux de régulation génique. Les cellules cancéreuses et les cellules normales présentent des paysages d'attracteurs différents en raison du réarrangement du réseau médié par des mutations somatiques, comme l'augmentation de la taille du bassin d'attraction lié à la prolifération. Par conséquent, les cellules normales et cancéreuses sont représentées comme des vallées distinctes dans un paysage épigénétique. A apoptose, P prolifération, Q quiescence.
Issu de Shin & Cho, 2023

La modélisation des réseaux de régulation génique constitue un moyen pertinent d'étudier les états tumoraux à travers une représentation en faible dimension, facilitant l'interprétation et l'analyse des processus cellulaires sous-jacents. Les GRN sont des modèles computationnels interprétables de la régulation de l'expression génique, représentés sous forme de réseaux ou de graphes. Un graphe est composé de nœuds — ici

représentant des gènes ou des facteurs de transcription — reliés entre eux par des arêtes définissant leurs relations régulatrices. Il est ainsi possible d'injecter des connaissances a priori lors de la construction du réseau, en s'appuyant sur des bases de données référençant les interactions protéine-protéine (256–258) ou les sites de fixation des facteurs de transcription (259–261).

L'intégration de données issues de plusieurs niveaux d'analyse de la tumeur — telles que l'expression génique (RNA-seq), la détection de variants génétiques (génotypage), l'accessibilité de la chromatine (ATAC-seq), la fixation des protéines à l'ADN (ChIP-seq), l'organisation tridimensionnelle du génome (Hi-C ou Capture-C), ou encore des perturbations fonctionnelles ciblées (CRISPR/Cas9) — permet d'élargir la couverture des réseaux de régulation génique et d'en affiner la structure en y incorporant des connaissances a priori ainsi que des observations expérimentales spécifiques.

Une première étape consiste à construire un réseau de coexpression génique en s'appuyant sur des mesures de corrélation d'expression entre les gènes, obtenues à partir d'une matrice d'expression génique (Figure 17, étape 1). L'intégration de bases de données d'interactions géniques permet d'orienter les relations au sein du réseau — représentées par des flèches — et d'éliminer les connexions biologiquement non plausibles (Figure 17, étape 2). L'apport de données omiques permet de spécialiser le GRN obtenu en le raffinant par une sélection conditionnelle des interactions. Par exemple, seules les relations comportant un motif de fixation connu peuvent être conservées (Figure 17, étape 3), tandis que les interactions sont exclues si l'accessibilité de la chromatine empêche le facteur de transcription d'atteindre son gène cible (Figure 17, étape 4). Le GRN ainsi construit peut ensuite être appliqué à de nouveaux jeux de données, agissant comme un filtre ou un crible, pour en extraire directement des états cellulaires spécifiques.

En biologie des systèmes, la construction de réseaux de régulation repose sur une étape clé appelée inférence. Ce processus vise à déduire des relations causales, fonctionnelles ou statistiques entre entités biologiques, à partir de jeux de données expérimentales souvent complexes, massifs et bruités. Ce terme, emprunté à la logique et aux statistiques, implique la formulation d'hypothèses robustes à partir d'observations partielles et incomplètes. Dans le contexte des GRN, l'inférence vise à reconstruire la structure du réseau sous-jacent — c'est-à-dire les interactions entre facteurs de transcription, gènes cibles et autres régulateurs — à partir de données d'expression génique, souvent dérivées de profils transcriptomiques à grande échelle. Cette étape permet de générer des modèles interprétables, jouant ainsi un rôle central dans la formalisation et l'exploration des hypothèses biologiques.

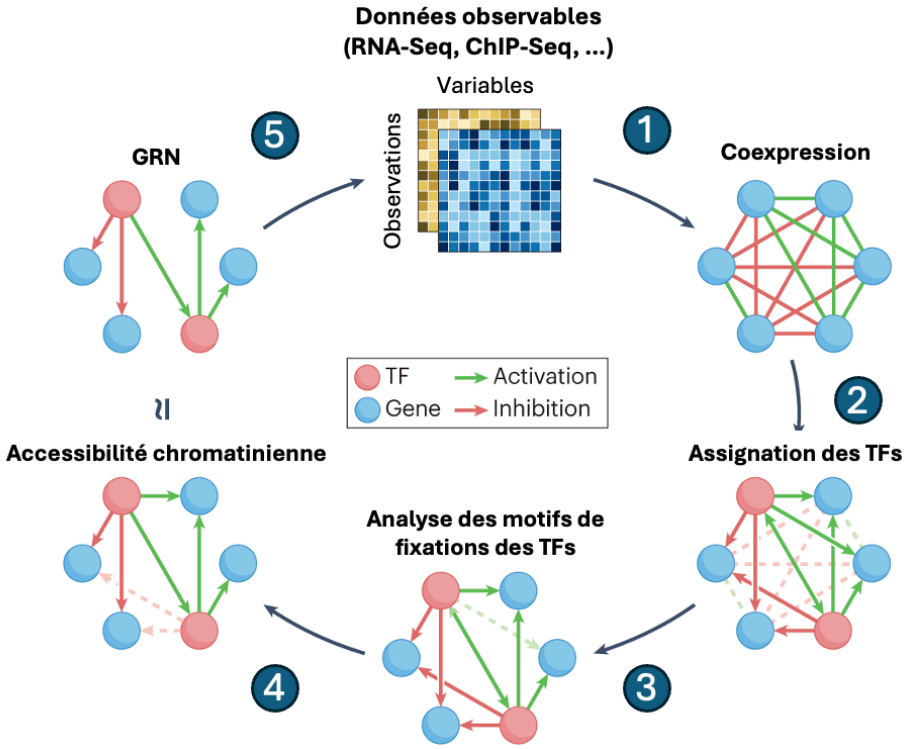


Figure 17 : Construction d'un GRN

Les GRNs peuvent être inférés à partir de données omiques mesurées et, grâce à la modélisation d'informations supplémentaires telles que les prédictions de liaison des facteurs de transcription ou l'accessibilité de la chromatine, peuvent être affinés pour mieux ressembler au GRN sous-jacent véritable. Les nœuds du GRN sont les facteurs de transcription et leurs gènes régulés, et les arêtes entre les nœuds indiquent le mode de régulation (activation ou inhibition).
Adapté de Badia-i-Mompel et al., 2023

Au cours des deux dernières décennies, les réseaux de régulation génique (GRN) sont devenus des outils computationnels incontournables pour l'analyse des données moléculaires à haut débit, notamment transcriptomiques et protéomiques (262). Les GRN fournissent un cadre conceptuel à la fois pratique et interprétable pour modéliser la complexité des interactions de régulation génique à l'échelle cellulaire. Cette approche permet d'identifier des régulateurs clés dont la perturbation pourrait avoir un effet thérapeutique systémique, plus marqué que celle d'un composant isolé d'une voie linéaire. Toutefois, en raison de la redondance fonctionnelle et des interactions croisées entre voies de signalisation, l'inhibition d'un seul composant peut être contournée par l'activation de voies alternatives, un mécanisme bien documenté de résistance aux thérapies ciblées (263). L'étude de la topologie et de la dynamique des GRN constitue une stratégie essentielle pour comprendre comment l'identité tumorale est établie et maintenue, avec des implications majeures pour l'appréhension de l'histoire naturelle du cancer.

De vastes projets de séquençage à haut débit menés sur de larges cohortes de tumeurs humaines et de lignées cellulaires ont permis la création de bases de données de référence. À titre d'exemple, le Genomic Data Commons (GDC), maintenu par le NIH, regroupe 86 projets d'étude de néoplasmes malins, totalisant 44 736 patients, et propose une grande diversité de données cliniques et moléculaires, incluant la protéomique, la génomique, la transcriptomique et l'analyse du méthylome (264). Le portail DepMap, issu du projet original CCLE, vise à représenter la diversité génétique et moléculaire des cancers à travers un vaste panel de modèles

cellulaires, incluant des essais de dépendance génique (CRISPR), des tests pharmacologiques et l'ensemble de leurs données omiques associée (265).

La section suivante présentera et comparera différentes approches d'inférence des réseaux de régulation génique.

3.2 Méthodes d'inférence

3.2.1 Modèles reposant sur la théorie de l'information

Les réseaux fondés sur la théorie de l'information figurent parmi les approches les plus fréquemment utilisées en raison de leur simplicité computationnelle. Également appelés réseaux de coexpression, ils reposent sur l'établissement de relations entre gènes à partir d'un seuil appliqué à une mesure de corrélation d'expression. Les principales mesures employées pour quantifier ces dépendances incluent les coefficients de corrélation de Pearson, Spearman et Kendall. Plus le seuil est strict, plus la coexpression requise pour établir un lien entre deux gènes devra être élevée, conduisant à un réseau de régulation plus clairsemé. D'autres mesures, telles que la distance euclidienne ou l'information mutuelle, sont également utilisées pour inférer les GRN. La distance euclidienne quantifie la similarité globale des profils d'expression des gènes, en mesurant leur éloignement dans un espace multidimensionnel, à la manière d'une distance entre deux points. L'information mutuelle, quant à elle, évalue la dépendance non linéaire entre deux gènes, c'est-à-dire dans quelle mesure la connaissance de l'expression de l'un renseigne sur celle de l'autre, même en l'absence de relation linéaire.

Ces modèles sont adaptés pour capturer différents aspects du fonctionnement cellulaire, y compris des processus complexes intra- et intercellulaires. L'un des principaux avantages des modèles fondés sur la théorie de l'information est leur capacité à inférer de vastes réseaux de régulation génique à partir de données transcriptomiques de faible taille, grâce à une complexité computationnelle réduite. Parmi les méthodes reposant sur ce type d'inférence, on peut citer REVEAL (266), RELEVANCE (267), ARACNe (Algorithm for the Reverse Engineering of Accurate Cellular Network) (268) ainsi que ses variantes (269), CLR (270), MRNET (271), ou encore FyNE (272). De manière analogue, les modèles graphiques gaussiens (GGM) représentent les dépendances conditionnelles entre gènes en supposant que leurs niveaux d'expression suivent une distribution normale, permettant d'identifier les interactions directes en tenant compte des effets des autres gènes. Les modèles fondés sur l'information mutuelle estiment les interactions régulatrices à partir des dépendances statistiques entre paires de gènes. Deux algorithmes emblématiques de cette approche sont ARACNe (273) et CLR (274). Globalement, ces modèles se distinguent par leur simplicité, leur faible coût computationnel et leur capacité à fonctionner avec un nombre limité d'échantillons.

Ces approches sont particulièrement utiles pour identifier les gènes co-régulés ou potentiellement impliqués dans des régulations croisées. Toutefois, elles présentent une limite majeure : elles génèrent des réseaux statiques, dépourvus de dynamique temporelle. Elles ne prennent pas en compte les transitions d'état cellulaire, la hiérarchie des régulateurs, ni la dynamique spatio-temporelle des régulations : tous les gènes sont considérés comme équivalents en termes d'importance régulatrice. En d'autres termes, ces réseaux identifient des corrélations globales, sans indiquer si ces relations sont contextuellement actives, ni permettre de

hiérarchiser les régulateurs. Ils manquent de contextualisation et risquent de négliger des régulations conditionnelles essentielles.

3.2.2 Réseaux booléens

Les réseaux booléens permettent de capturer le comportement dynamique des GRN et sont simples à implémenter. Ils sont capables de décrire des phénomènes biologiques tels que les oscillations, les événements multi-stationnaires, les corrélations à longue portée, les comportements en commutation, la stabilité et l'hystérésis (retard de l'effet sur la cause) (275). Au sein des réseaux de régulation booléens les gènes sont représentés par des variables et leur niveau d'expression est discrétisé en valeurs booléennes binaires selon un seuil défini : 0 pour les valeurs faibles traduisant des gènes silencieux ou s'exprimant à bas bruit ou 1 pour les gènes avec de fortes valeurs correspondant à des gènes activés. Les relations entre les variables booléennes alors créées sont exprimées sous forme de fonctions logiques avec les opérateurs logiques AND, OR et NOT. Contrairement aux approches statistiques classiques, les réseaux booléens permettent de simuler l'évolution temporelle de l'état des gènes, d'établir une hiérarchie entre les gènes en disant au modèle quels sont les gènes considérés comme facteurs de transcription par exemple. Étant donné l'étape de discrétisation, les réseaux booléens sont plus résistants au bruit et robustes aux fluctuations expérimentales. Cette étape de discrétisation est aussi leur point faible. Selon la méthode discrétisation adoptée, avec cette simplification, des détails importants du comportement du système peuvent être perdus. La méthode de discrétisation influencera fortement les résultats. Cette méthode d'inférence fait face à un risque d'explosion combinatoire, l'espace de toutes les règles logiques possibles devient très grand dès qu'on a beaucoup de gènes, rendant l'inférence difficile sans contraintes supplémentaires. Néanmoins, les réseaux booléens sont faciles à interpréter et offrent une approche dynamique simple pour modéliser GRN.

3.2.3 Modèles par équations différentielles

Les approches fondées sur les équations différentielles ordinaires (ODE) conservent les niveaux d'expression génique sous forme continue, sans recourir à une discrétisation, ce qui permet de représenter plus précisément les dynamiques sous-jacentes de la régulation génique (276). Les équations différentielles permettent de modéliser les variations continues de l'expression d'un gène en fonction de celle d'autres gènes et, potentiellement, de facteurs environnementaux, offrant ainsi une modélisation quantitative plus fidèle au comportement réel du système biologique (277). La dynamique de l'expression d'un gène peut être modélisée par une équation différentielle de la forme suivante :

$$\frac{dx_i}{dt} = f_i(x_1, x_2, \dots, x_N, p, u)$$

où x est l'expression du gène considéré noté i à un instant t, N est le nombre de gènes analysés, u représente les perturbations externes au système, p aux paramètres propres au système, f est la fonction qui décrit le taux de changement des variables d'état en fonction de p.

L'obtention d'une solution unique nécessite l'imposition de contraintes traduisant des connaissances a priori, telles que des simplifications, des approximations ou des régulations déjà établies. Dans les modèles ODE,

les variables continues sont soumises à des contraintes, notamment l'interdiction de valeurs négatives, reflétant l'impossibilité biologique d'une expression négative pour certaines entités telles que les ARNm ou les protéines (278). Un autre inconvénient des ODE réside dans leur tendance à modéliser principalement des systèmes linéaires, alors que les processus de régulation sont fréquemment gouvernés par des dynamiques non linéaires complexes. Ces modèles sont également difficiles à appliquer directement à de grands réseaux (au-delà de 30 à 50 gènes), en raison de limitations liées à la disponibilité des données expérimentales et à la complexité computationnelle. En conséquence, leur usage est généralement réservé à des sous-systèmes bien caractérisés, ou bien combiné à des approches hybrides, non linéaires ou approximatives – telles que les NeuralODEs ou les méthodes d'inférence intégrées – afin d'élargir leur applicabilité à des réseaux plus complexes (279,280).

3.2.4 Réseaux bayésiens

Les approches bayésiennes d'inférence des GRN reposent sur l'intégration de modèles probabilistes avec la théorie des graphes. Les réseaux bayésiens représentent les relations probabilistes entre gènes à l'aide de graphes orientés acycliques (DAG), où la probabilité d'expression d'un gène dépend conditionnellement de celle de ses gènes parents (281). Ces méthodes sont particulièrement appréciées pour leur capacité à intégrer des connaissances a priori, à traiter des données discrètes ou continues, à modéliser l'incertitude expérimentale, et à inférer la structure du réseau même en présence de variables partiellement observées Toutefois, l'inférence bayésienne est associée à une complexité computationnelle élevée, ce qui restreint son application à des réseaux de petite à moyenne taille (283). De plus, la structure acyclique des DAGs empêche la modélisation explicite des boucles de rétrocontrôle, bien qu'elles soient courantes dans la régulation biologique et essentielles au contrôle des processus cellulaires (284). Cette contrainte peut être partiellement levée par l'utilisation de réseaux bayésiens dynamiques, capables de capturer des relations cycliques à travers le temps et de modéliser des données transcriptomiques longitudinales (285). Ces approches dynamiques permettent également d'améliorer la sensibilité et la spécificité dans la détection des relations régulatrices, en particulier lorsqu'elles sont couplées à des méthodes de validation croisée et à des données issues de groupes phénotypiques distincts.

3.2.5 Méthodes de deep learning

Au-delà des approches classiques d'apprentissage automatique appliquées aux réseaux, l'apprentissage profond (Deep Learning) ouvre de nouvelles perspectives d'analyse, notamment par l'utilisation de réseaux de neurones artificiels. Les réseaux de neurones ont été initialement conçus pour simuler le fonctionnement des neurones biologiques : l'information y transite via des couches successives de nœuds (ou "neurones") interconnectés et pondérés (286). Le modèle de réseau neuronal le plus simple est composé de trois couches : une couche d'entrée recevant les données brutes, une couche cachée appliquant une transformation non linéaire vers une représentation latente, et une couche de sortie produisant une prédiction (classification ou régression), en fonction de la tâche à accomplir (287). Un réseau neuronal profond se distingue par la présence de multiples couches cachées (Figure 18), la "profondeur" désignant précisément leur nombre. Ces couches sont connectées de manière séquentielle, chaque couche apprenant une représentation de plus haut niveau à partir de la sortie transformée de la couche précédente. Plus un réseau contient de couches cachées, plus il est théoriquement capable de modéliser des relations complexes et non linéaires dans les données, ce qui en fait un outil puissant

pour des tâches d'apprentissage supervisé ou non supervisé. L'un des principaux inconvénients du Deep Learning réside dans son besoin en données massives pour l'entraînement, ce qui représente une contrainte importante dans de nombreuses études biologiques. Cette exigence s'explique par la nécessité d'ajuster un grand nombre de paramètres répartis sur les multiples couches du réseau. Par ailleurs, la complexité de l'architecture et du processus d'apprentissage compromet l'interprétabilité des prédictions produites par ces modèles. En effet, malgré la maîtrise de certains paramètres comme les données d'entrée ou la structure du réseau, le fonctionnement interne demeure largement opaque. Cette opacité nuit à l'interprétabilité et peut limiter l'utilisation de ces modèles pour l'exploration mécanistique des processus biologiques sous-jacents (288).

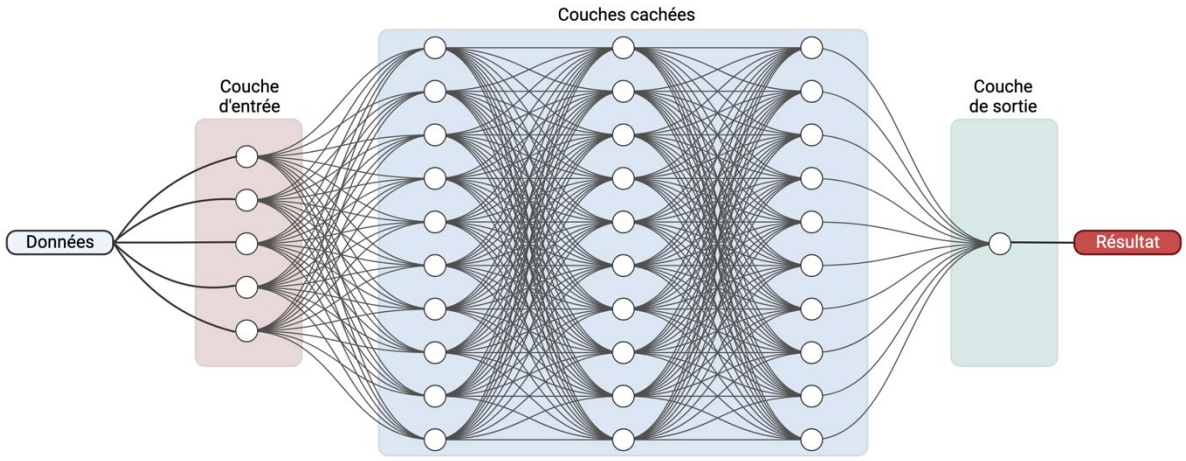


Figure 18 : Exemple d'une structure d'un modèle de Deep learning

Le modèle est constitué d'une couche d'entrée composée de cinq neurones, trois couches cachées de dix neurones chacune et une couche de sortie d'un seul neurone.

Les récentes avancées en matière de réseaux de neurones profonds transforment en profondeur les domaines de la biologie, de la bioinformatique et de l'analyse des réseaux biologiques (289–293). DeepWalk constitue l'un des premiers modèles d'apprentissage profond appliqué aux graphes ; il vise à apprendre des représentations vectorielles latentes des nœuds d'un graphe en adaptant des méthodes issues du traitement automatique du langage naturel. Il s'appuie sur le cadre word2vec (294), développé initialement pour le traitement du langage naturel, afin d'intégrer chaque nœud du graphe dans un espace vectoriel. Le processus repose sur la génération de marches aléatoires à partir de chaque nœud, suivie de l'optimisation de la fonction objectif Skip-gram. L'objectif de l'algorithme Skip-gram est d'apprendre une représentation vectorielle pour chaque nœud en maximisant la probabilité de prédire ses voisins dans la marche aléatoire, selon un principe analogue à celui de word2vec, qui optimise la prédiction des mots contextuels à partir d'un mot cible. Ainsi, DeepWalk transpose les principes du traitement du langage naturel à la modélisation de graphes, permettant une représentation des nœuds qui reflète efficacement leur contexte topologique (295). Ce type d'algorithme et ses dérivés ont été appliqués à la prédiction de fonctions protéiques (296–298) ou à l'inférence des propriétés physico-chimiques de molécules à partir de leur structure (299).

Dans une approche analogue à celle des réseaux de neurones, Scarselli et al. ont proposé en 2009 une méthode permettant de calculer une moyenne pondérée des informations issues des voisins de chaque nœud

d'un graphe (300). Depuis, plusieurs modèles de réseaux de neurones appliqués aux graphes ont été développés, tels que les autoencodeurs graphiques (GAE) (301), les réseaux neuronaux convolutifs sur graphes (GCN) (302) et les réseaux neuronaux récurrents adaptés aux graphes (GNN) (303). Les GAE sont constitués d'un encodeur chargé de capter à la fois la structure topologique et les attributs des nœuds, et d'un décodeur visant à reconstruire le graphe à partir des représentations latentes apprises. Récemment, les GAE ont démontré leur efficacité dans divers domaines, notamment l'identification cellulaire à partir de données scRNA-Seq (304,305), la prédiction d'interactions protéine–protéine et médicament–protéine (306), l'établissement d'associations entre ARN circulants et une centaine de pathologies (307), ainsi que dans l'inférence de réseaux de régulation génique(308). Ces résultats suggèrent que les GAE possèdent une capacité d'adaptation à différents contextes biologiques, tout en fournissant des performances satisfaisantes. Néanmoins, les GAE présentent certaines limitations, la principale étant leur manque d'interprétabilité, inhérent à leur nature de modèle à boîte noire. Les GCN reposent sur l'application de filtres de convolution au niveau des nœuds d'un graphe, consistant par exemple à calculer la moyenne des valeurs de chaque nœud et de ses voisins, avant traitement par un réseau de neurones. L'utilisation des GCN est particulièrement répandue en pharmacologie, notamment pour la prédiction des propriétés des composés (309–312), de la réponse aux traitements (313–316), pour le repositionnement thérapeutique (317–320), ou encore la prédiction d'interactions biomoléculaires complexes (321–325). L'architecture des réseaux de neurones récurrents (RNN) s'inspire de l'idée que la cognition humaine repose sur la mémoire des événements passés. Les RNN sont ainsi appelés car la sortie actuelle dépend des sorties précédentes, ce qui les rend adaptés à l'apprentissage sur des graphes dynamiques. Ils peuvent être combinés à d'autres modèles pour capturer des dépendances spatio-temporelles dans les réseaux. Ces modèles ont notamment été appliqués à l'analyse métagénomique (326) et à l'identification de protéines essentielles (327).

La figure 19 reprend les algorithmes principaux d'apprentissage profonds utilisés.

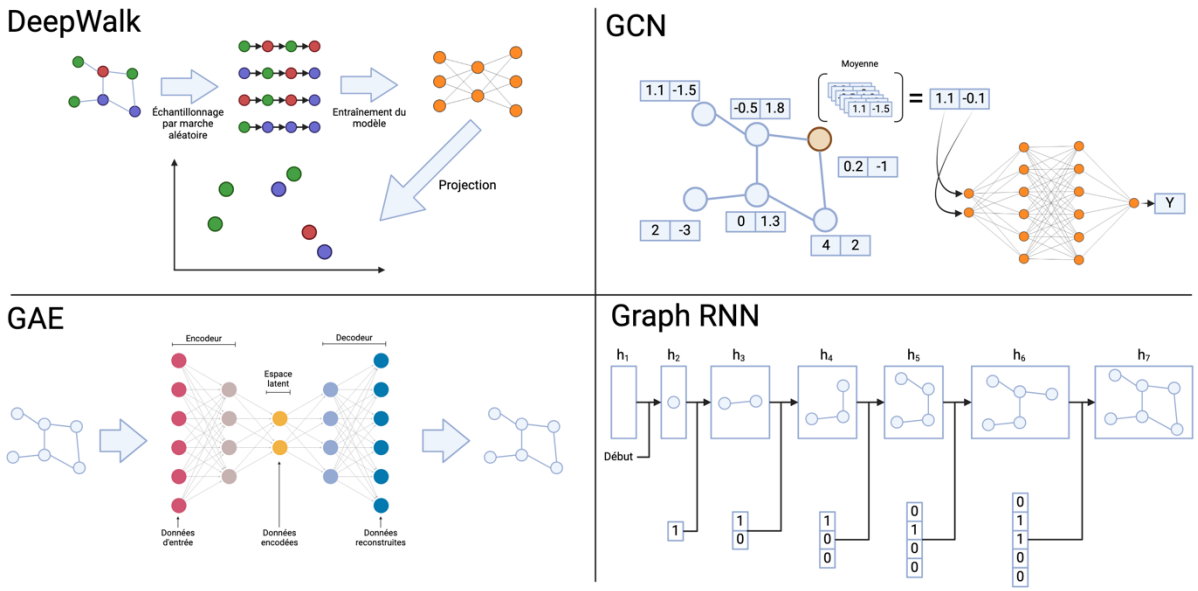


Figure 19 : Principaux algorithmes d'apprentissage profonds appliqués aux graphes

3.2.6 Méthodes intégratives

Les approches intégratives combinent plusieurs types de données — transcriptomiques, épigénétiques, protéomiques — afin de surmonter les limites inhérentes aux méthodes unimodales, souvent entravées par le bruit et un manque d'information contextuelle. Elles permettent ainsi de capturer la régulation conjointe exercée par divers facteurs biologiques, tels que les variations du nombre de copies, la méthylation de l'ADN ou les interactions protéine–protéine (328). Par exemple, l'algorithme iRafNet, fondé sur des forêts aléatoires intégratives, fusionne différentes sources de données biologiques et a montré une performance supérieure aux approches standards pour la prédiction des cibles transcriptionnelles (329). D'autres travaux ont recours à la factorisation matricielle pour intégrer de manière cohérente les données transcriptomiques et post-transcriptionnelles dans un cadre inférentiel unifié, améliorant ainsi la précision des interactions déduites (330). Ces approches sont particulièrement prometteuses pour l'étude des pathologies, en particulier des cancers, car elles permettent d'inférer des réseaux régulateurs à la fois pertinents contextuellement et biologiquement interprétables (331). L'algorithme PANDA constitue un exemple marquant d'approche intégrative reposant sur un mécanisme de message-passing pour l'inférence de réseaux de régulation génique (332). Contrairement aux méthodes classiques fondées sur une modalité unique (expression, ChIP-seq ou PPI), PANDA intègre simultanément les interactions protéine–protéine, les données transcriptomiques et les motifs de liaison transcriptionnels pour construire des réseaux spécifiques à une condition biologique. Le modèle évalue dynamiquement la cohérence entre trois réseaux — régulation, co-régulation et coopération protéique — en itérant sur les responsabilités (flux d'information du facteur de transcription vers le gène) et les disponibilités (du gène vers le facteur), jusqu'à convergence. Comparé aux méthodes classiques telles que CLR, ReMoDiscovery ou SEREND, PANDA offre de meilleures performances prédictives, une robustesse accrue face au bruit, et une capacité à identifier des modules régulateurs spécifiques à des conditions biologiques, tout en restant interprétable.

Dans l'article « GRaNIE and GRaNPA: inference and evaluation of enhancer-mediated gene regulatory networks », Kamal et al. introduisent un pipeline répondant à deux lacunes majeures de la génomique : (i) la reconstruction d'eGRN (réseaux de régulation génique médiés par les enhancers) tenant explicitement compte de la couche enhancer et (ii) l'évaluation de leur pertinence biologique (333). GRaNIE exploite la covariance inter-individus entre les données d'accessibilité de la chromatine (ATAC-seq) et d'expression des gènes (RNA-seq) pour associer, via un filtrage FDR, les motifs de facteurs de transcription à des régions ouvertes, puis à des gènes cibles, générant ainsi des réseaux tripartites TF–enhancer–gène spécifiques d'un contexte donné.

GRaNPA évalue la pertinence biologique de tout GRN centré sur un facteur de transcription, via un test de performance prédictive : un modèle de forêt aléatoire prédit les $\log_2$FC d'un jeu de données différentielle indépendant, et compare le R^2 obtenu à deux contrôles permutés — l'un conservant le degré du réseau, l'autre injectant du bruit — pour calculer un z-score. Il identifie ensuite les TF capables de prédire la réponse transcriptionnelle spécifique d'un type cellulaire donné à des perturbations. Appliquées à des macrophages infectés, à divers sous-types de leucémie myéloïde aiguë et à des cellules T stimulées, les méthodes surpassent ARACNe, SCENIC et DoRothEA, tout en révélant le rôle inattendu du facteur PURA dans la polarisation inflammatoire. L'atout majeur du couple GRaNIE–GRaNPA réside dans l'intégration explicite de la couche

enhancer : en corrélant accessibilité chromatinienne et transcription au niveau inter-individuel, le cadre d'analyse révèle des interactions régulatrices inaccessibles aux approches basées uniquement sur l'ARN. De plus, GRaNPA fournit une métrique quantitative indépendante de tout « gold standard » potentiellement biaisé ; les réseaux sont évalués sur leur capacité prédictive réelle, permettant des comparaisons objectives entre méthodes. En revanche, la dépendance à des données ATAC-seq rigoureusement appariées constitue une contrainte expérimentale importante, et la nature purement corrélative du modèle ne permet pas d'établir la causalité des interactions inférées. En outre, le classifieur Random Forest utilisé par défaut agit comme une « boîte noire » : il ne fournit ni information sur le sens (activation/répression) des interactions, ni prise en compte explicite des covariables, limitant ainsi l'interprétabilité biologique.

3.3 Description de l'inférence du réseaux de régulation génique

Les mécanismes de régulation cellulaires sont multiples et interviennent à différents niveaux, rendant la modélisation de la régulation génique particulièrement complexe lorsqu'elle est envisagée dans sa globalité (334,335). Les méthodes d'inférence des GRN visent généralement à prédire les interactions de régulation existant entre les gènes. Ici, un niveau supplémentaire est introduit en déterminant les relations de régulation entre un facteur de transcription (ou régulateur) et ses gènes cibles. Dans l'objectif de modéliser un réseau de corégulation spécifique au cancer du poumon, l'algorithme h-LICORN (Hybrid-Learning Co-operative Regulation Networks) a été utilisé.

h-LICORN appartient à la famille des méthodes booléennes d'inférence de réseaux de régulation, tout en intégrant une approche hybride visant à renforcer la robustesse des prédictions. (336,337). Il s'appuie sur un algorithme précurseur, LICORN, qui identifie les meilleurs programmes de régulation liant un gène cible à des ensembles de coactivateurs et de co-inhibiteurs agissant de manière coopérative. La recherche peut être restreinte à une liste de régulateurs connus, si celle-ci est fournie. La première étape consiste à projeter les profils d'expression continue sur trois états discrets (-1, 0, +1), transformant ainsi les matrices RNA-seq en une table d'événements booléens. LICORN explore alors l'espace des combinaisons de facteurs de transcription à l'aide d'un algorithme de fouille d'itemsets : il recherche les ensembles de TF dont l'association avec le gène cible dépasse un seuil de support prédéfini dans la cohorte d'échantillons. Chaque combinaison candidate est considérée comme un programme de régulation ; sa cohérence est évaluée en comparant, pour chaque patient, l'état du gène prédit par la logique à l'état réellement observé. Le programme offrant le meilleur compromis entre exactitude prédictive et parcimonie est retenu. En sortie, LICORN produit un réseau biparti dont les arêtes relient des ensembles de TF aux gènes qu'ils régulent, fournissant une lecture explicite des circuits d'activation et de répression (336).

L'algorithme h-LICORN (337) conserve le concept de programme de régulation introduit par LICORN, tout en y ajoutant deux éléments clés : une sélection quantitative et une stabilisation par bootstrap. Il reprend les premières étapes de LICORN, mais là où ce dernier s'arrêtait, h-LICORN injecte ensuite les niveaux d'expression continue dans un modèle de régression linéaire : pour chaque programme candidat, les coefficients

β reliant les TF au niveau d'expression du gène cible sont estimés, et seuls les programmes minimisant l'erreur quadratique sont retenus, ce qui permet d'éliminer la majorité des faux positifs induits par les corrélations discrètes. Cette régression est répétée sur des rééchantillonnages bootstrap des échantillons ; une interaction régulateur–gène n'est conservée que si sa fréquence d'apparition dépasse un seuil défini (par exemple, 60 à 70 % des tirages), fournissant ainsi un score de support interprétable comme un niveau de confiance Au total, h-LICORN génère un réseau plus parcimonieux, pondéré et statistiquement robuste, tout en conservant l'un des atouts majeurs de LICORN : la lisibilité explicite des modules coopératifs. Le pseudo-code de l'algorithme est présenté ci-après :

Entrée requise : Deux matrices d'expression numérique réelle :

- `OR` pour les gènes régulateurs
- `OG` pour les gènes cibles

Sortie attendue : `EG`, ensemble des réseaux de régulation génétique (GRN) pour chaque gène cible de l'ensemble $\mathcal{G}$

Algorithme :

1. Discrétiser les matrices d'expression :
   `MR ← Discrétiser(OR) ; MG ← Discrétiser(OG)`
2. Initialiser l'ensemble des réseaux extraits :
   `EG ← ∅`
3. Pour chaque gène cible `g` dans l'ensemble $\mathcal{G}$ faire :
4. Pour chaque itération i allant de 1 à Ns faire :
5. Sélectionner aléatoirement un sous-échantillon Si depuis notre ensemble S
6. Extraire un réseau via LICORN :
   `Ei ← LICORN_GRN_Extraction(MRi, MGi)`
7. Filtrer les régulateurs via régression :
   `EGi ← RegressionSelect(ORi, OGi, Ei)`
8. Ajouter ce sous-réseau à l'ensemble temporaire :
   `EGNs ← EGNs ∪ EGi`
9. Fin pour
10. Ajouter l'ensemble `EGNs` à `EG` :
    `EG ← EG ∪ EGNs`
11. Fin pour
12. Classer les réseaux obtenus :
    `Rank(EG, OR, OG)`

L'évaluation des capacités de prédiction du modèle repose sur plusieurs métriques : la sensibilité (ou rappel) correspond au taux de vrais positifs, c'est-à-dire la capacité du modèle à identifier correctement les cas positifs parmi l'ensemble des cas réellement positifs. La spécificité correspond à la proportion de vrais négatifs correctement identifiés parmi tous les cas réellement négatifs, ce qui reflète la capacité du modèle à éviter les faux positifsLa précision, quant à elle, mesure la proportion de cas effectivement positifs parmi ceux prédits comme tels, ce qui reflète la fiabilité des prédictions positives. Le rappel (identique à la sensibilité) met l'accent sur la capacité du modèle à limiter les faux négatifs. Enfin, le score F1 correspond à la moyenne harmonique entre la précision et le rappel, selon la formule suivante : $F1 = 2 * \frac{precision * recall}{precision + recall}$, synthétise ces deux dimensions en un seul indicateur en offrant une métrique équilibrée (Figure 20).

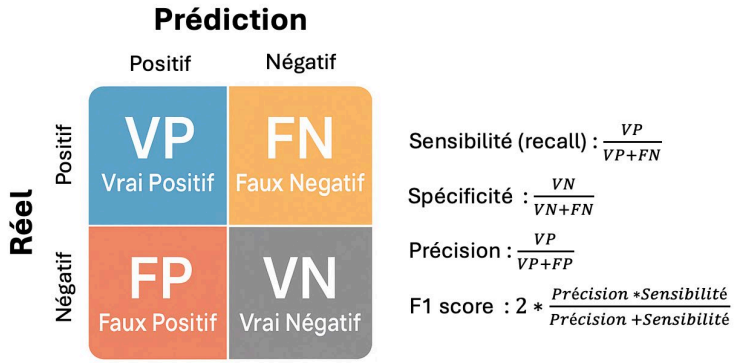


Figure 20 : Mesures des performances du modèle SVM

Le modèle SVM a été évalué selon plusieurs métriques, Sensibilité (recall) : taux de vrais positifs ; Spécificité : taux de vrais négatifs ; Précision : fiabilité des prédictions positives ; Score F1 : moyenne harmonique précision-recall.. VP : Vrai positif ; VN : Vrai négatif; FP : Faux positif ; FN : Faux négatif.

Les performances de h-LICORN ont été évaluées à l'aide du jeu de données du cinquième challenge DREAM5 (338), qui consiste en la reconstruction des réseaux de régulation de trois organismes différents. L'ensemble des interactions régulatrices y est considéré comme connu, constituant ainsi une référence permettant d'évaluer la performance des modèles prédictifs. Les performances du modèle sont mesurées à l'aide de la courbe ROC (Receiver Operating Characteristic), qui évalue les performances d'un classifieur binaire en fonction du taux de vrais positifs (TPR) et du taux de faux positifs (FPR), ce qui permet d'estimer la capacité à distinguer les interactions vraies des fausses. La courbe ROC est une représentation graphique de la relation existante entre la sensibilité et la spécificité d'un test, calculée pour toutes les valeurs seuils possibles. En complément, la courbe précision/rappel (Precision-Recall curve) est utilisée pour évaluer la pertinence des interactions prédites (précision) et la proportion des interactions réelles correctement identifiées (rappel.

Pour évaluer ses performances prédictives, h-LICORN a été comparé à GENIE3, l'algorithme vainqueur du challenge DREAM5 basé sur les forêts aléatoires (339), ainsi qu'à ARACNE, une méthode reposant sur l'information mutuelle réputée pour sa robustesse dans ce type de problématique (340,341). h-LICORN affiche des performances comparables à GENIE3 et surpasse ARACNE (Figure 21), tout en présentant l'avantage d'une grande simplicité d'implémentation, facilitant ainsi son déploiement.

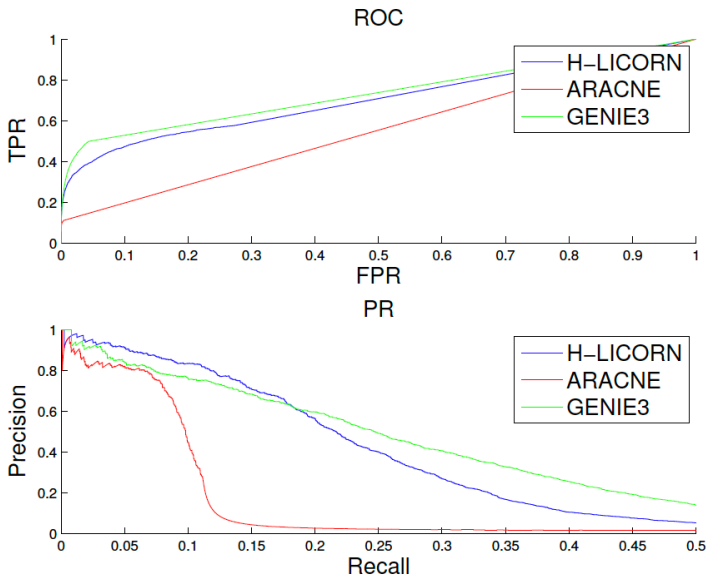


Figure 21 : Performances du modèle h-LICORN

Courbes de performance selon ses prédictions sur le jeu de données DREAM5 (338) par rapport à ses deux modèles concurrents ARACNE (268) et GENI3 (339).
D'après Chebil et al. (337)

La capacité des interactions prédites par h-LICORN à refléter des interactions protéine-protéine entre régulateurs a été comparée à celles inférées par les algorithmes ARACNE, GENIE3, CLR (342) et WGCNA (343). La base de données STRING répertorie les interactions protéine-protéine annotées selon leur niveau de confiance (prédiction ou validation expérimentale) et leur type (voisinage génomique, gène de fusion, co-occurrence), couvrant plus de 12 000 organismes (258). Dans ce cadre, la méthode h-LICORN surpasse les performances des autres algorithmes, comme le montre la Figure 22

.

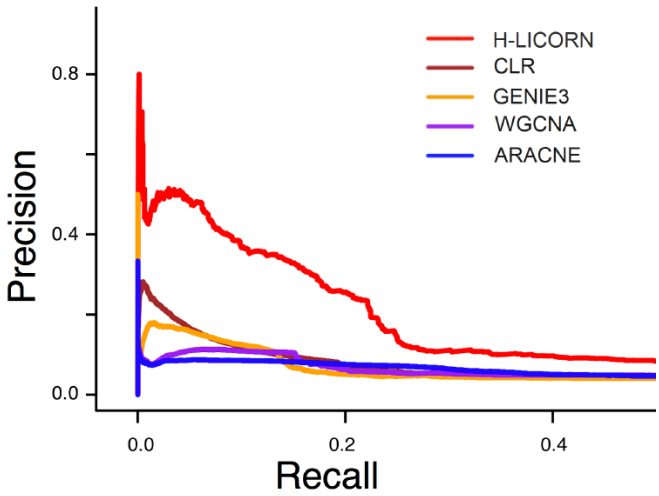


Figure 22 : Performances de h-LICORN face à d'autres algorithmes

Courbe précision/rappel des algorithmes h-LICORN, CLR, GENIE3, WGCNA et ARACNE pour la prédiction des interactions protéine-protéine entres corégulateurs. La base de données STRING a été prise comme référence.

Nous avons utilisé le package R CoRegNet, disponible sur la plateforme Bioconductor, qui implémente l'algorithme h-LICORN (344). Le réseau de corégulation est construit à partir de données d'expression issues de lignées cellulaires cancéreuses, spécifiques du type tumoral étudié, afin de capturer les mécanismes de régulation propres aux cellules tumorales, sans interférence stromale. Les données transcriptomiques des modèles cellulaires tumoraux ont été récupérées via le portail DepMap (345).

CoRegNet implémente une chaîne complète d'inférence de réseaux de corégulation à partir des programmes de régulation identifiés par h-LICORN. Le résultat brut est encapsulé dans un objet coregnet, contenant la matrice d'adjacence, les modules coopératifs et les paramètres d'inférence. Ensuite, la fonction addEvidences permet d'enrichir chaque interaction avec des données externes (ChIP-seq, motifs PWM, interactions protéine-protéine, signal ATAC-seq), lesquelles sont re-pondérées et intégrées dans un score de confiance global. La fonction refine applique ensuite un test d'enrichissement (test de Fisher corrigé par FDR), sélectionnant soit le programme de régulation le plus informatif pour chaque gène, soit l'ensemble des arêtes dont le score dépasse un seuil optimisé, réduisant ainsi les faux positifs.

Une spécificité majeure de notre approche réside dans l'agrégation du signal transcriptomique en un score d'activité régulatrice pour chaque nœud du réseau inféré. Cette approche, précurseur, permet de caractériser le phénotype tumoral d'un échantillon à travers l'activité de ses régulateurs clés, rendant possible l'analyse des tumeurs sous l'angle des coopérations entre processus cellulaires, tout en assurant une interprétabilité biologique robuste. L'influence, ou activité régulatrice, est quantifiée via un test t de Welch, implémenté dans CoRegNet :

$$\text{Influence}\,(\mathbf{r}) = \frac{\overline{\mathbf{E}(\mathbf{A_r})} - \overline{\mathbf{E}(\mathbf{R_r})}}{\sqrt{\frac{\boldsymbol{\mu}^2_{\mathbf{A_r}}}{\mathbf{A_r}}} + \sqrt{\frac{\boldsymbol{\mu}^2_{\mathbf{R_r}}}{\mathbf{R_r}}}}$$

où : r est le régulateur étudié, A_r et R_r l'ensemble des gènes activés et réprimés par ce régulateur, $\overline{E(A_r)}$ et $\overline{E(R_r)}$ sont l'espérance moyenne de l'expression des gènes activés et réprimés et $\mu^2_{A_r}$ et $\mu^2_{A_r}$ leurs variances. Ce test évalue la différence de distribution entre les niveaux d'expression des gènes cibles activés et réprimés par le régulateur. Un régulateur est considéré comme actif ou influent dans une condition donnée lorsque l'on observe une différence significative entre les niveaux d'expression de ses cibles activées et réprimées (Figure 23). En quantifiant ainsi l'activité régulatrice, CoRegNet améliore l'interprétation biologique au niveau échantillon et réduit significativement la complexité des données. Plutôt que de générer un réseau de régulation massivement dimensionné décrivant des milliers de gènes, CoRegNet extrait un sous-réseau compact de quelques centaines de corégulateurs, reflétant les caractéristiques phénotypiques spécifiques à l'échantillon.

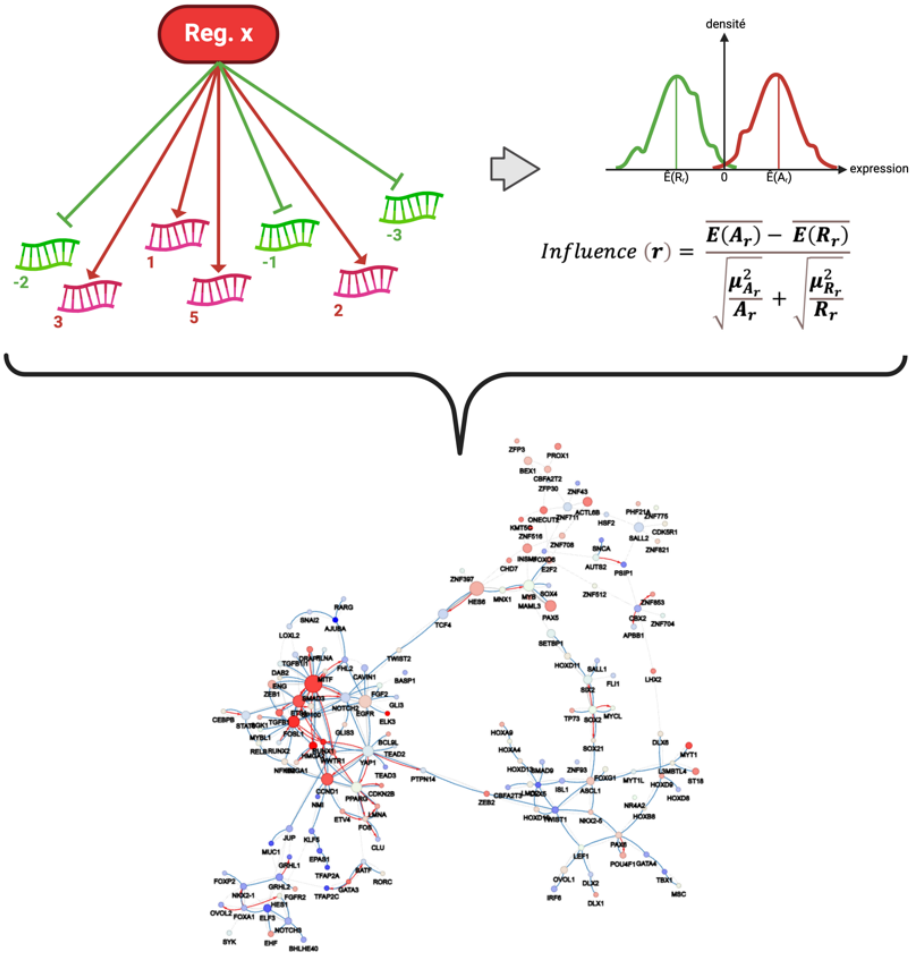


Figure 23 : Calcul de la valeur d'influence d'un régulateur et visualisation

Pour un régulateur on effectue un test t de Welch qui compare les distributions d'expression des gènes cibles activés et réprimés. La statistique de décision constitue la mesure d'influence du régulateur.

La démarche complète d'inférence du réseau de corégulation est présentée dans la figure 24.

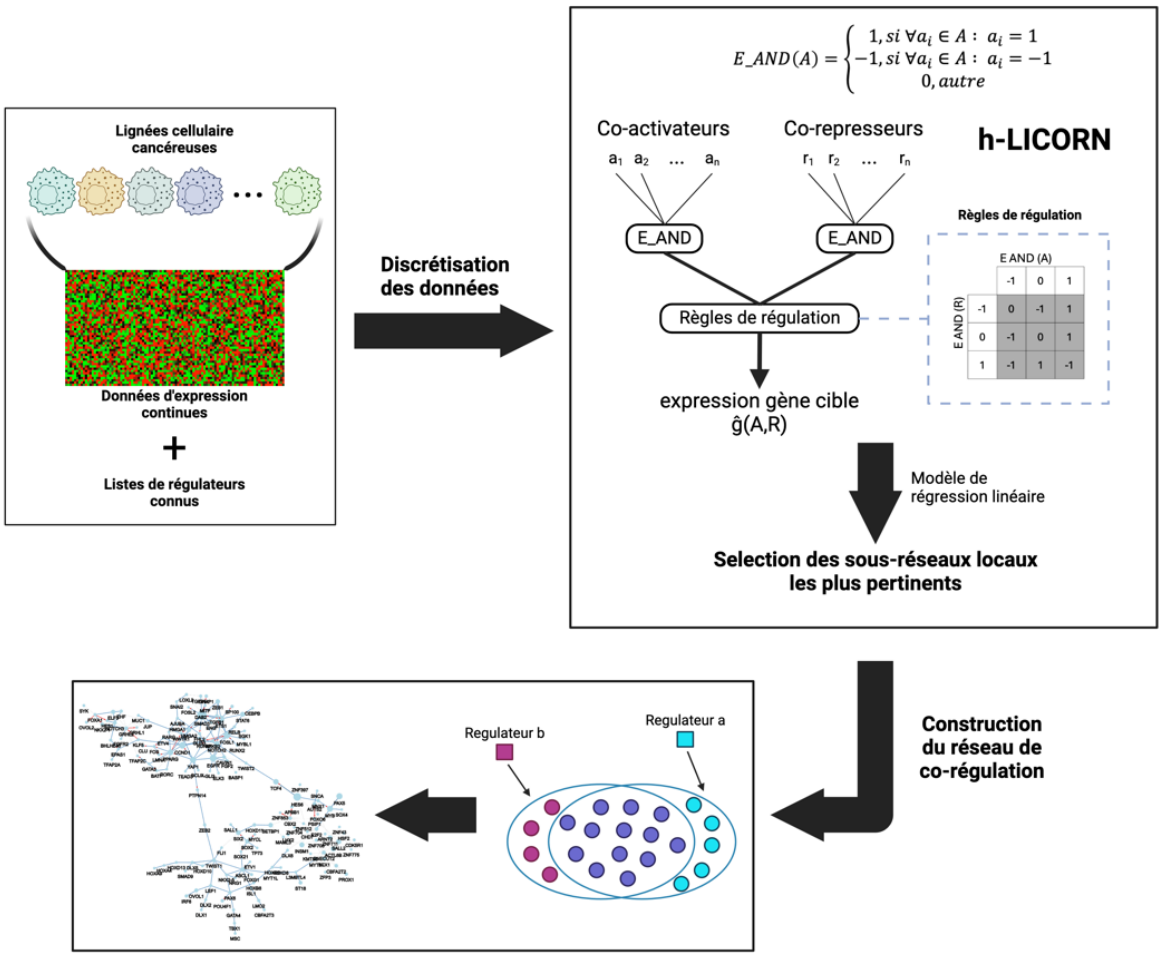


Figure 24 : Inférence d'un réseau de régulation avec le package R CoRegNet

À partir de données d'expression génique continues issues de lignées cellulaires cancéreuses et d'une liste de régulateurs connus, les profils transcriptomiques sont discrétisés en états d'expression. Le modèle h-LICORN identifie les règles de régulation en intégrant l'effet combiné de coactivateurs et de corépresseurs via des opérateurs logiques (E_AND), et évalue l'expression du gène cible à l'aide d'un modèle de régression linéaire. Les sous-réseaux les plus pertinents sont sélectionnés, puis intégrés via l'algorithme CoRegNet pour construire un réseau de corégulation permettant de visualiser les interactions transcriptionnelles entre régulateurs et gènes cibles.

Objectifs de thèse

Durant cette thèse nous avons cherché à comprendre ce qui caractérise la spécificité du phénotype tumoral à travers l'exploitation des réseaux de régulation géniques. En corolaire de nos études, nous avons eu comme volonté de rendre nos modèles de réseaux de régulation génique accessible à la communauté scientifique.

Le premier objectif de la thèse a été de développer un portail interactif sous la forme d'une application web publiquement disponible constituée d'un premier module sur lequel l'utilisateur peut interagir avec notre de réseau de régulation et identifier ses régulateurs clés, déterminer des modules de régulation spécifiques aux sous-groupes de tumeurs. Via un deuxième module, il peut accéder à une carte de tumeurs classées selon la qualité de leur réseau, explorer les clusters au travers des données d'activité régulatrices des nœuds du réseau et des informations cliniques et biologique. Un troisième module rend possible projection de données externes, issues d'une base de données externe ou directement issues de l'utilisateur pour tester d'éventuelles hypothèses scientifiques

Le second objectif est un lien direct avec le premier, puisque nous avons eu ensuite pour objectif de déterminer des sous-classes de cancers du poumon à partir de métacohorte de données de patients en nous appuyant sur les méthodologies mises à disposition sur le site web. En effet, les cancers pulmonaires sont classés sur des bases immunohistochimique et moléculaire permettant notamment de proposer la stratégie thérapeutique la plus adaptée. Des tentatives de classification plus fines à partir du profilage de l'expression génique des tumeurs pulmonaires existent, néanmoins l'absence de standardisation rend difficile la comparaison directe ou la fusion de ces classifications et toutes ne s'alignent pas parfaitement. Ainsi, mon deuxième objectif a été de cartographier une métacohorte de patients atteints d'un cancer pulmonaire avec une approche par réseau de régulation génique. Dans le but de déterminer des sous-groupes caractérisés cliniquement, histologiquement et moléculairement par le biais des régulateurs de notre GRN, de leurs gènes cibles et des réseaux de signalisations activés en aval. Puis, pour chaque sous-groupes, nous avons cherché à identifier des régulateurs clés, potentiellement responsables du phénotype tumoral

L'identification de marqueurs d'addiction oncogéniques chez les patients atteints d'un NSCLC permet une prise en charge par des thérapies ciblées. Le récepteur MET est l'un de ces marqueurs oncogéniques et dans environ 3% des cas de NSCLC des mutations engendrent un saut de l'exon 14 résultant en la perte de son domaine régulateur juxtamembranaire. Néanmoins, les mécanismes d'addiction oncogénique associés à METex14Del vis-à-vis de son ligand sont encore mal compris, l'activation de METex14Del nécessiterait tout de même la liaison à son ligand. Lors d'un précédent travail, l'équipe du Dr Tulasne a démontré que le pouvoir transformant est strictement dépendant de la présence de HGF, aussi bien in vitro qu'in vivo. Ainsi nous avons collaboré avec l'équipe du Dr Tulasne pour identifier et proposer les régulateurs clés du réseau de corégulation du cancer du poumon, ainsi que leurs gènes cibles, impliqués dans la réponse de METex14Del à l'HGF dans un modèle de

lignées cellulaires 16HBE. La pertinence des régulateurs identifiés a été ensuite évalué par les conséquences biologique et fonctionnelle de la suppression de leur l'expression.

Résultats

1. Décloisonner la recherche et diffuser les modèles au moyen d'une application interrogeable

Au cours de cette thèse, nous avons développé un portail web interactif afin de maximiser le potentiel scientifique de notre travail et d'en faciliter la diffusion au sein de la communauté scientifique. L'application a été développée à l'aide du framework R Golem, qui fournit un environnement structuré favorisant la maintenabilité du code et garantissant la robustesse des applications produites (346). L'application, encapsulée sous forme de package R, est ensuite conteneurisée à l'aide de Docker, un système de virtualisation légère par conteneur (347). La conteneurisation de l'application R Shiny avec Golem assure une standardisation de l'environnement d'exécution, une reproductibilité accrue, une gestion simplifiée des mises à jour, et facilite son déploiement. L'application est hébergée sur Google Cloud Platform, une solution offrant une mise à l'échelle automatique des ressources en fonction de la charge, tout en garantissant la conformité au RGPD.

L'interface développée se structure en plusieurs modules. Le premier, CoRegNet, permet d'explorer le réseau de corégulation du cancer du poumon et d'y projeter les valeurs moyennes d'influence des régulateurs. Le second module, CoRegMap, établit une correspondance entre le réseau de corégulation et la projection UMAP de la cohorte. La projection 2D permet d'explorer les sous-groupes tumoraux, les types histologiques, les profils moléculaires, les valeurs d'influence d'un régulateur donné, ainsi que les données cliniques, avec la possibilité de superposer les lignées cellulaires. Le dernier module, CoRegQuery, étend notre modèle en autorisant l'intégration de données externes – locales ou issues de GEO (Gene Expression Omnibus) – directement projetées sur la carte UMAP. L'utilisateur peut ainsi explorer la distribution des valeurs d'influence sur les nœuds du réseau de régulation inféré (Figure 25).

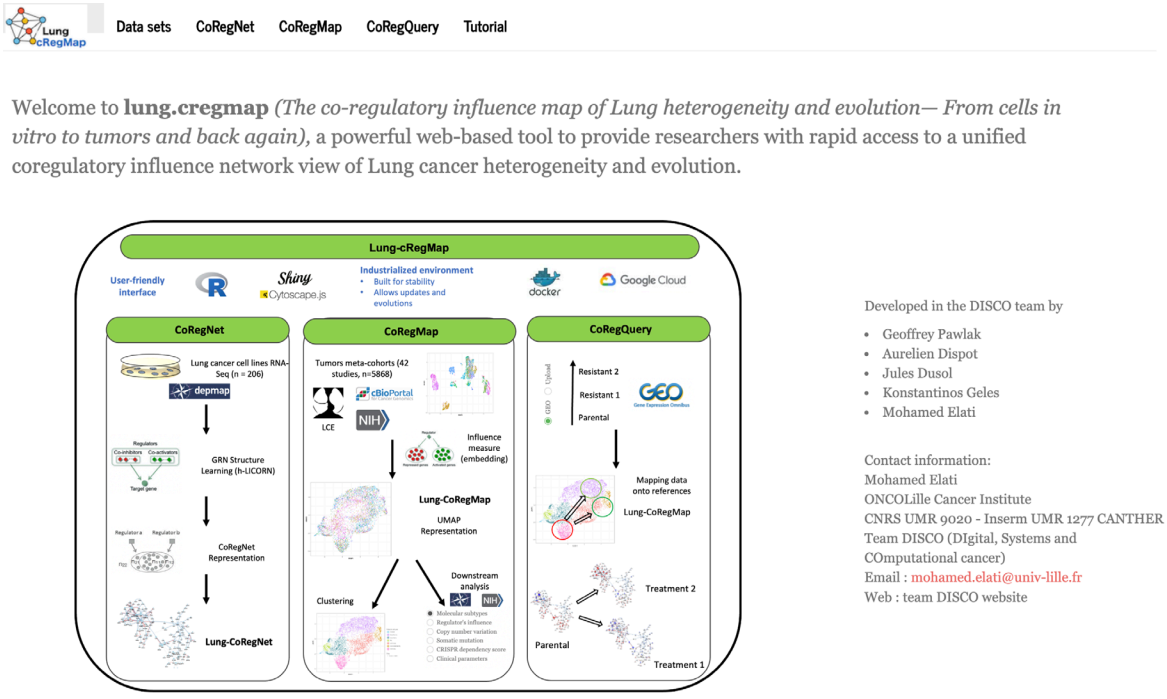


Figure 25 : Vue générale du fonctionnement de l'application cRegMap

L'application lung.cregmap est constituée de 3 modules. CoRegNet, siège de l'interaction avec notre GRN. CoRegMap, où l'exploration de notre cohorte de 5487 patients au travers des données cliniques, moléculaire et d'influence est possible. CoRegQuery, un module d'importation et de projection de données externe sur notre modèle.

1.1 Interagir avec le GRN : CoRegnet

1.1.1 Tout d'abord, construire le GRN

À partir de données transcriptomiques de lignées cellulaires et d'une liste de facteurs de transcription humains connus, CoRegNet (344), comme décrit précédemment, utilise l'algorithme h-LICORN (337) pour inférer les interactions de régulation entre régulateurs et gènes cibles. Les données de transcriptomiques de lignées cellulaires sont obtenues à partir du projet Cancer Cell Line Encyclopedia (CCLE) téléchargées à partir du portail Cancer Dependency Map (DepMap Broad (2024) ; https://depmap.org/portal.). Au total, 206 lignées cellulaires ont été utilisées, incluant 135 lignées NSCLC, 50 lignées SCLC et 21 d'autres types histologiques.

L'inférence du GRN repose sur des connaissances a priori, notamment une liste de 2 375 régulateurs, comprenant 1 639 facteurs de transcription humains avec une spécificité de liaison à l'ADN validée expérimentalement, et 752 cofacteurs transcriptionnels issus des bases Human Transcription Factor Catalog et TcoF-DB v2. Les sites de liaison des facteurs de transcription ont été récupérés via MotifDB (R/Bioconductor) (348) ; et les séquences promotrices identifiées à l'aide de PWMEnrich (R/Bioconductor) (349). Les preuves de régulation TF–gène ont été agrégées à partir de ChEA2 (350), du package tftargets (https://github.com/slowkow/tftargets), et de plusieurs bases complémentaires (TRED, ITFP, ENCODE, BEDOPS, TRRUST), aboutissant à un total de 2 256 674 interactions TF-cible Les interactions protéine-protéine entre régulateurs (coopération TF–TF), soit 1 257 053 interactions, ont été extraites via le package iRefR (R/Bioconductor) (351), à partir des bases BioGRID, HIPPIE, STRING et HPRD.

Le réseau inféré contient 502 régulateurs connectés à 4 143 gènes cibles, pour un total de 18 053 interactions régulatrices.

1.1.2 Représenter et interagir avec le GRN

Les visualisations du réseau de corégulation ont été réalisées à l'aide du package R visNetwork (352). Cette solution interactive permet de zoomer dans la représentation du réseau et de déplacer les nœuds pour une exploration plus intuitive. L'interface comprend un panneau de sélection dans la partie supérieure, permettant de choisir les données d'influence régulatrice et la stratification de la cohorte de patients ou des lignées cellulaires pulmonaires. L'utilisateur peut sélectionner soit la classification histologique standard, soit la classification interne cRegMap. Il peut ensuite explorer un sous-type tumoral spécifique. Le panneau inférieur affiche la projection interactive du réseau de corégulation, correspondant au contexte sélectionné (Figure 26).

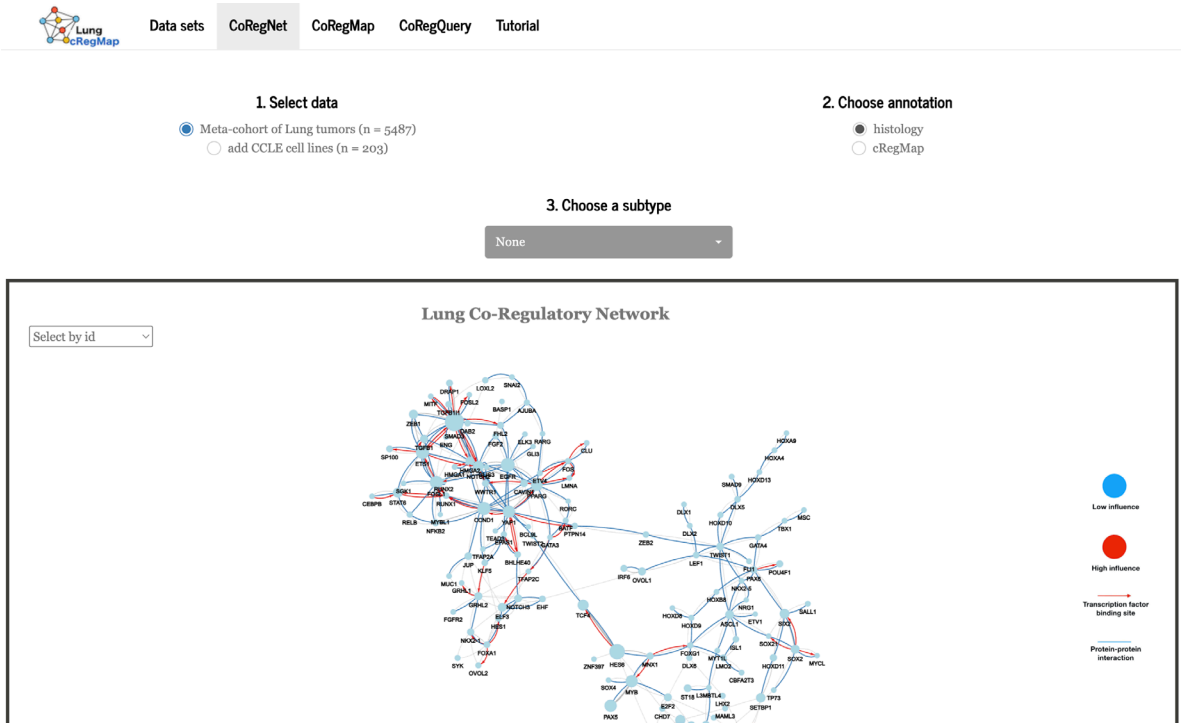


Figure 26 : Onglet CoRegNet

Le panneau supérieur permet à l'utilisateur de choisir ses options de visualisation (Origine des données, type de classification et sous-groupe). Le panneau inférieur permet de visualiser le réseau, d'interagir avec lui et d'afficher la légende.

Dans la visualisation (Figure 27), un régulateur est coloré en rouge lorsqu'il présente une influence positive, et en bleu lorsqu'il exerce une influence négative. L'échelle des couleurs est normalisée sur l'ensemble des conditions étudiées. La taille des nœuds est proportionnelle à la fréquence d'apparition de chaque régulateur dans les réseaux locaux de corégulation, définis par des triplets corégulateurs–gène cible Les arêtes reliant les régulateurs sont colorées selon les preuves de régulation disponibles : gris si l'interaction est prédite par l'algorithme h-LICORN, bleu en cas d'interaction protéine-protéine documentée, et rouge directionnel (de A vers B) si le régulateur A interagit avec un site de fixation du régulateur B.

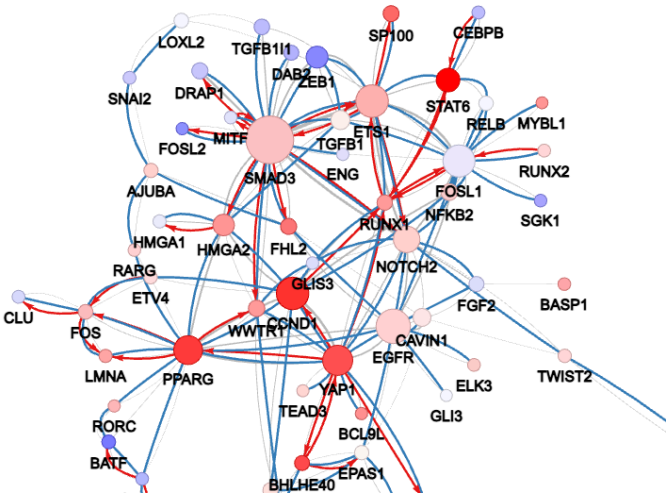


Figure 27 : Visualisation du GRN sur le module CoRegNet

L'influence moyenne des nœuds est représentée selon leur couleur, rouge et bleu si positivement ou négativement influence. Un lien gris entre régulateur signifie que l'algorithme hLICORN a détecté un lien entre les deux régulateur, un lien bleu représente une interaction protéine-protéine entre les régulateurs et une flèche rouge signifie l'existence d'un site de fixation de facteur de transcription. Par exemple les liens entre CEBPB et STAT6 en haut à droit de la figure regroupent les trois types de liens.

1.2 Explorer notre cohorte : CoRegMap

1.2.1 Réduire les dimension vers une représentation 2D

Pour explorer la structure intrinsèque des données d'influence et identifier d'éventuels clusters de patients, nous avons appliqué une analyse de projection non linéaire via UMAP. Cette approche, non linéaire, est particulièrement adaptée à la visualisation de données de grande dimension en raison de sa sensibilité élevée et de sa capacité à préserver à la fois les relations locales et globales. Elle permet une représentation bidimensionnelle compacte tout en conservant la structure topologique initiale des échantillons.

L'algorithme UMAP partage des similarités avec t-SNE (353). Il construit un graphe pondéré en haute dimension, où le poids des arêtes encode la proximité relative entre les points. Une étape d'optimisation ultérieure cherche une projection en dimension réduite qui préserve autant que possible la structure du graphe de proximité initial (Figure 28).

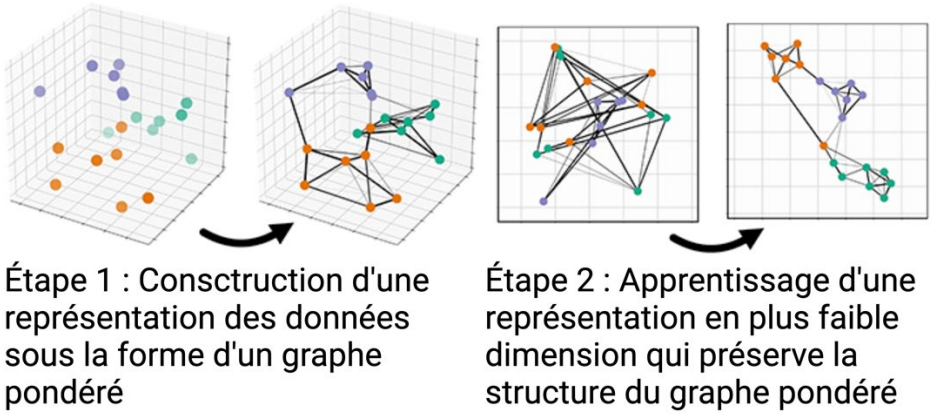


Figure 28 : Étapes principales de l'algorithme UMAP

Le processus se déroule en deux étapes principales : (1) construction d'un graphe pondéré à partir des données initiales, représentant les similarités locales entre les points ; (2) apprentissage d'une représentation en plus faible dimension qui préserve au mieux la structure du graphe initial. Cette approche est typique d'algorithmes tels que UMAP ou t-SNE, utilisés pour la visualisation de données de haute dimension tout en conservant les relations locales.

À partir du résultat obtenu après réduction de dimension via l'algorithme UMAP, il est possible d'appliquer une stratégie de clustering non supervisée visant à identifier des groupes cohérents d'échantillons. Cette approche s'appuie sur la représentation des données sous forme de graphe afin d'y détecter des communautés de nœuds partageant des similarités locales. La modularité est le concept central de ces méthodes de clustering : elle évalue la qualité d'un partitionnement des nœuds d'un graphe en communautés, en comparant la densité d'arêtes intra-communautés à celle attendue dans un graphe aléatoire équivalent. Cette métrique a été introduite par Newman et Girvan en 2004 (354), et s'exprime selon la formule suivante :

$$Q = \frac{1}{2n}\sum_{i,j}\left(A_{i,j} - \frac{k_i k_j}{2m}\right)\delta(c_i, c_j)$$

où c_i est la communauté assignée au nœud i, $A_{i,j}$ signifie s'il existe un lien ou non entre les nœuds i et i , $A_{i,j} = 1$ si oui sinon $A_{i,j} = 0$, k_i est le degré du nœud i, c'est-à-dire le nombre de liens reliés à ce nœud et m est le nombre total de liens dans le graphe. Plus la valeur de la modularité est haute plus les communautés au sein du graphe sont distinctes. L'objectif des algorithmes de clustering sur graphes est de maximiser cette fonction de modularité.

Une méthode largement utilisée pour optimiser la modularité dans les graphes est l'heuristique de déplacement local (local moving heuristic). Cette approche repose sur le déplacement itératif de nœuds individuels d'une communauté vers une autre, dans le but d'augmenter la modularité à chaque étape. Les nœuds sont explorés dans un ordre aléatoire, et pour chacun, l'algorithme évalue si son transfert vers une autre communauté — y compris une communauté vide — permet un gain de modularité. Si tel est le cas, le nœud est affecté à la communauté optimisant le gain. Ce processus se poursuit jusqu'à ce qu'aucune amélioration ne soit possible par des déplacements unitaires.

Cette heuristique constitue la brique centrale de l'algorithme de Louvain (Figure 29), une méthode gloutonne particulièrement efficace pour la détection de communautés dans les grands réseaux (355).

L'algorithme opère en deux phases répétées : dans un premier temps, il optimise localement la modularité en évaluant les gains possibles liés au transfert de chaque nœud vers la communauté d'un de ses voisins. Si le gain est positif, le changement est conservé ; sinon, il est annulé. Ensuite, le graphe est compressé en un super-graphe où chaque communauté devient un nœud unique, et un nouvel ensemble d'arêtes est calculé. Ces deux étapes sont réitérées jusqu'à ce qu'aucun gain supplémentaire de modularité ne soit obtenu ou qu'un optimum local soit atteint.

L'algorithme de Louvain présente une complexité en temps quasi-linéaire, de l'ordre de O(n + m), où n est le nombre de nœuds et m le nombre d'arêtes (356) ce qui en fait un outil robuste pour l'analyse de graphes biologiques de grande taille.

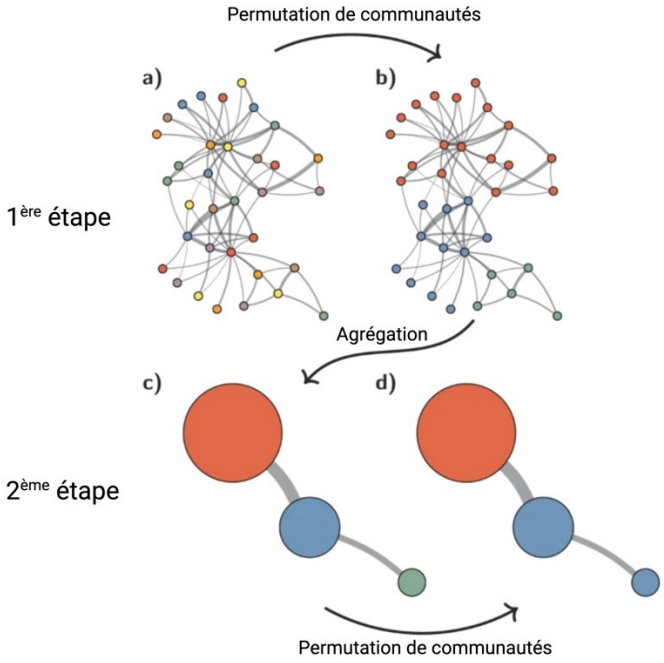


Figure 29 : Principe de l'algorithme de Louvain

(a) L'algorithme déplace les nœuds individuels d'une communauté à une autre afin de trouver une modularité optimale. (b) Sur la base de cette partition, un réseau agrégé est créé. (c) L'algorithme déplace ensuite les nœuds individuels dans le réseau agrégé. (d) Ces étapes sont répétées jusqu'à ce que la qualité de la partition ne puisse plus être améliorée.
D'après (357)

La méthode utilisée dans ces travaux applique l'algorithme de Louvain sur un graphe des k plus proches voisins partagés (shared nearest neighbor graph, SNN), construit à partir de la matrice projetée par UMAP, en utilisant l'indice de Jaccard comme mesure de similarité entre les paires de nœuds (Figure 30).

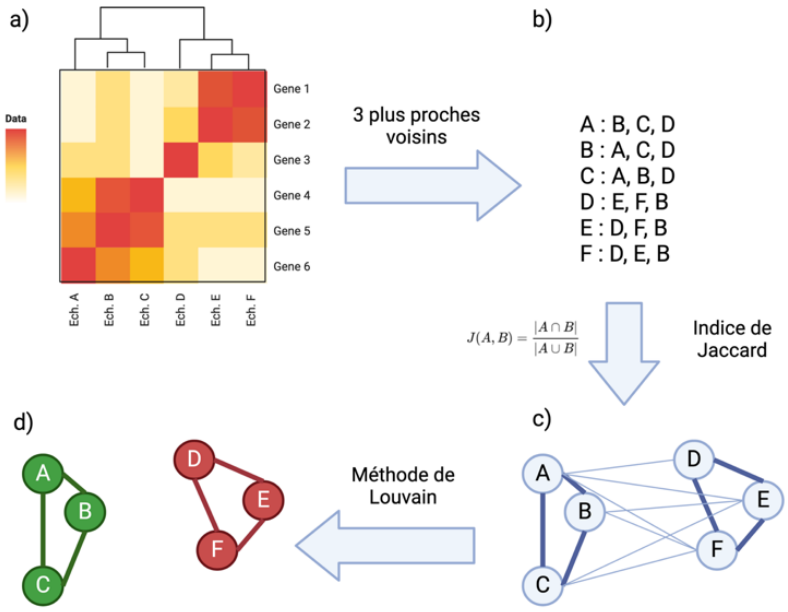


Figure 30 : Illustration de la méthode de Louvain pour l'identification de clusters

a) Matrice de distance illustrant la similarité entre six gènes (Genes 1 à 6). Les couleurs indiquent la proximité (rouge foncé : forte proximité, jaune clair : faible proximité). b) Pour chaque gène, sélection des trois voisins les plus proches selon la matrice.

c) Construction du graphe pondéré basé sur l'indice de Jaccard calculé à partir des listes de voisins communs. d) Détermination des clusters (ou communautés) à l'aide de l'algorithme de Louvain, permettant de regrouper les gènes présentant des profils de similarité communs.

1.2.2 Explorer la cohorte de patients par le modèle UMAP

L'onglet CoRegMap est dédié à l'exploration de la métacohorte de patients. Un bandeau supérieur permet à l'utilisateur de choisir s'il souhaite projeter uniquement les données des patients ou y superposer les lignées cellulaires. Selon le type d'information sélectionné, plusieurs modalités d'affichage sont proposées : une classification (pour les sous-types moléculaires), un régulateur (pour projeter les scores d'influence), un gène (pour visualiser les altérations génomiques), ou encore des variables cliniques telles que l'âge au diagnostic, le statut tabagique, la survie ou le sexe du patient. La zone inférieure est consacrée à la visualisation ; les graphiques y sont générés à l'aide du package ggplot2 (358). Un bouton interactif permet le téléchargement des figures obtenues (Figure 31).

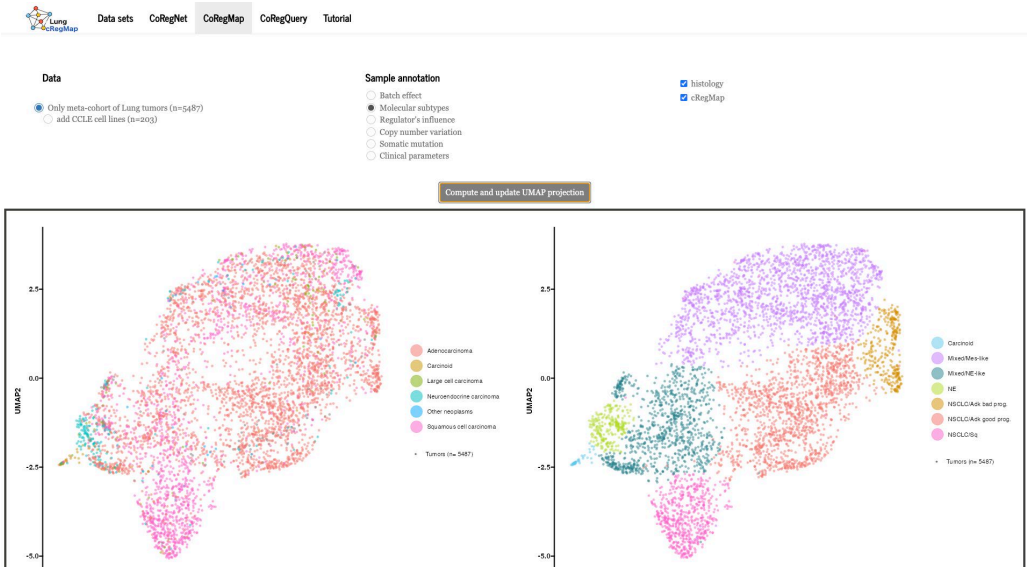


Figure 31 : Onglet CoRegMap de lung.cregmap

Le panneau supérieur regroupe les options de visualisations possibles. Le panneau inférieur est dédié à la visualisation de la projection. Ici sont sélectionnées uniquement les données des patients et les deux classifications (histologique et cRegMap) sont représentées.

C'est dans cet onglet que l'utilisateur peut aisément constater la suppression de l'effet de batch grâce à l'utilisation des scores d'influence. Lorsque l'on applique une réduction de dimension par UMAP sur les données transcriptomiques brutes, la représentation obtenue est très granulaire : chaque jeu de données forme un cluster distinct, empêchant toute comparaison directe entre patients. En revanche, le calcul des scores d'influence des régulateurs du GRN, suivi d'une réduction de dimension par UMAP, permet d'harmoniser les jeux de données et de les projeter sur un même plan d'analyse (Figure 32).

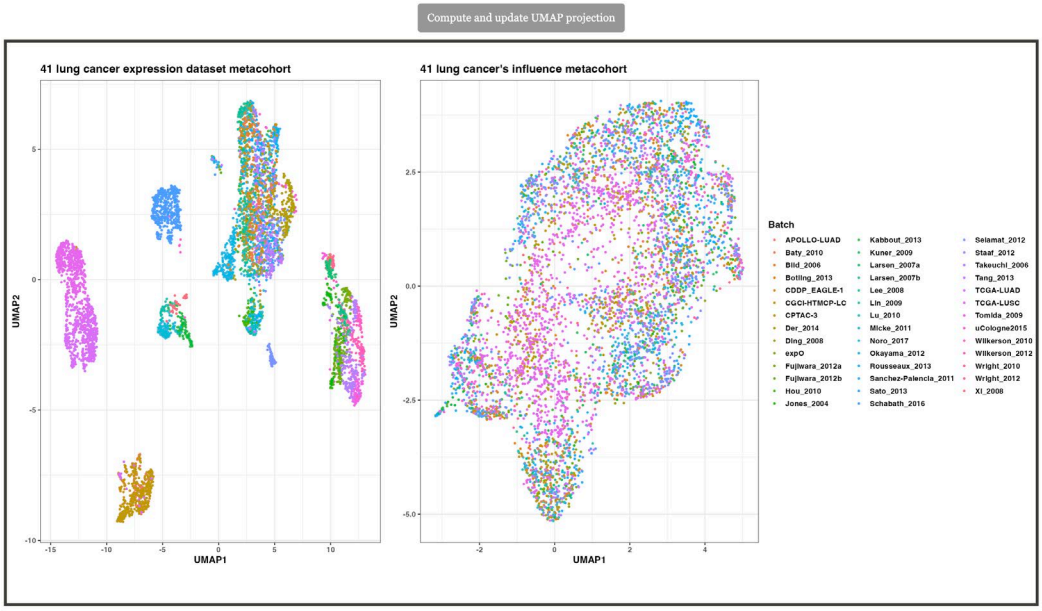


Figure 32 : Visualisation de la suppression de l'effet de batch au sein de l'application cRegMap

La superposition des lignées cellulaires sur le modèle permet de projeter les informations de dépendance génique associées à chaque lignée. Ces données sont issues du pipeline CRISPR du projet DepMap, qui produit des criblages à grande échelle de haute qualité, dans le but d'identifier les dépendances génétiques dans un large éventail de lignées cellulaires cancéreuses humaines.

Un score ajusté de dépendance génique (score CERES) est estimé en intégrant deux sources majeures de biais : (i) les altérations du nombre de copies, qui, dans les régions amplifiées, induisent une accumulation de cassures double-brin via Cas9, provoquant une réponse cellulaire non spécifique et générant des faux positifs ; (ii) la variabilité d'efficacité des sgRNA, responsable d'une hétérogénéité du pouvoir de knock-out selon les guides utilisés. L'algorithme CERES corrige ces biais en intégrant un score d'activité spécifique aux guides.(359) Ces scores permettent d'identifier les gènes dont l'inactivation inhibe la prolifération ou induit la mort cellulaire, révélant ainsi les vulnérabilités spécifiques à chaque type tumoral. La figure 33 illustre la projection du score de dépendance à FOSL1, révélant un gradient de dépendance à ce gène à travers l'espace UMAP.

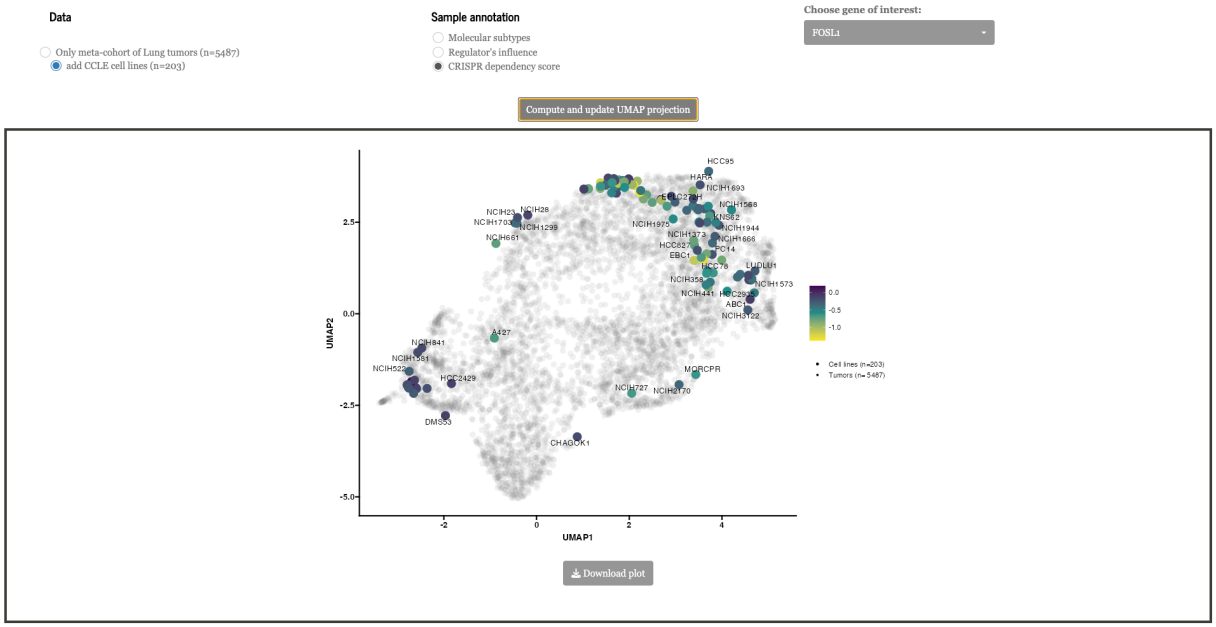


Figure 33 : Projection de la dépendance à FOSL1 des lignées cellulaires cancéreuses pulmonaires

La projection de dépendances géniques à FOSL1 des lignées cellulaires utlisées pour l'inférence du GRN permet d'évaluer l'importance d'un éventuel régulateur clés d'un sous-groupe dans la prolifération et la survie des lignées cellulaires.

1.3 Projeter de nouvelles données : CoRegQuery

1.3.1 Développer les méthodes pour ajouter de nouvelles données

En complément des données issues de la métacohorte et des lignées cellulaires, l'application permet d'appliquer l'ensemble des analyses précédemment décrites à des jeux de données externes. Celles-ci peuvent être soit importées localement, soit directement récupérées via une connexion au dépôt public GEO (Gene Expression Omnibus), hébergé par le National Center for Biotechnology Information (NCBI). Ce dépôt international archive et diffuse librement des données de puces à ADN, de séquençage de nouvelle génération et d'autres types de données de génomique fonctionnelle à haut débit soumises par la communauté scientifique. En 2024, GEO comptabilisait environ 200 000 études et 6,5 millions d'échantillons (360).

Les données sont récupérées grâce à une adaptation du code du projet GEOexplorer (361). Les noms de gènes sont ensuite convertis au format HGNC, afin d'assurer leur correspondance avec les identifiants utilisés dans notre modèle. L'étape suivante consiste à calculer l'influence des régulateurs, puis à prédire les coordonnées UMAP des nouveaux échantillons à l'aide des fonctions FindTransferAnchors et MapQuery du package R Seurat (362), en utilisant notre métacohorte comme jeu de référence.

Au-delà de la prédiction des coordonnées UMAP des nouvelles données, nous estimons leur appartenance aux classes définies dans notre classification. Pour cela, nous appliquons un modèle de Support Vector Machine (SVM) afin de prédire les clusters à partir des profils d'influence régulatrice.

Les SVM, développés dans les années 1990 par Vladimir Vapnik et ses collaborateurs, ont été introduits dans leur publication fondatrice de 1995 intitulée "Support Vector Method for Function Approximation, Regression Estimation, and Signal Processing" (363). Ces modèles sont largement utilisés pour les tâches de

classification supervisée, où ils cherchent à identifier l'hyperplan de séparation optimal maximisant la marge entre les points critiques des classes opposées, appelés vecteurs de support (Figure 34). La représentation géométrique de cet hyperplan dépend de la dimensionnalité des données : il s'agit d'une ligne en deux dimensions, d'un plan en trois dimensions, ou d'un hyperplan dans les espaces de dimension supérieure. Lorsqu'il existe plusieurs solutions de séparation linéaire, l'algorithme sélectionne celle qui maximise la distance entre les vecteurs de support. Cette propriété garantit une capacité de généralisation élevée sur des données non vues et une robustesse accrue des prédictions.

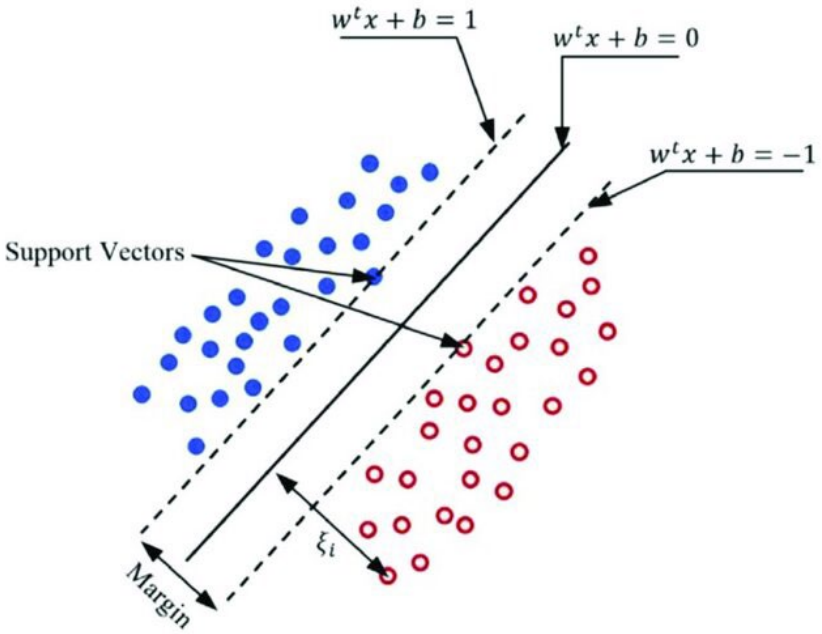


Figure 34 : Principe de fonctionnement d'un modèle SVM

L'hyperplan de décision (trait plein) est défini par l'équation $w^t x+b=0$, tandis que les hyperplans marginaux $w^t x+b = +1$ et $w^t x+b = -1$ (traits pointillés) délimitent la marge maximale. Les points situés sur ces hyperplans marginaux correspondent aux vecteurs de support, lesquels déterminent de manière critique la position optimale de l'hyperplan. Les variables d'écart ξi (flèches) illustrent la tolérance aux violations de marge, permettant un compromis entre largeur de la marge et erreurs de classification.
Issu de (364)

Dans notre cas, le modèle SVM a été choisi en raison de sa meilleure capacité de prédiction des classes cRegMap, en l'entraînant sur les données de notre cohorte selon une méthode de validation croisée à 10 blocs. La validation croisée à k blocs (ou k-fold cross-validation) consiste à répartir aléatoirement l'ensemble des données en k sous-ensembles de taille à peu près égale. À chaque itération, on retient l'un de ces blocs comme jeu de test, et on entraîne le modèle sur les k-1 blocs restants, puis on répète l'opération jusqu'à ce que chaque bloc ait servi de test exactement une fois. On calcule ensuite la performance moyenne sur les k évaluations pour obtenir une estimation plus fiable de la capacité de généralisation du modèle. Ce procédé permet à la fois de maximiser l'utilisation des données disponibles et de limiter le risque de biais lié à un découpage unique, tout en offrant un bon compromis entre coût de calcul et variance de l'estimation.

Le SVM applique une séparation linéaire des données, mais souvent les données ne sont pas linéairement séparables. Le kernel trick dans les SVM permet de gérer des données non linéairement séparables en projetant implicitement ces données dans un espace de dimension supérieure. Plutôt que d'effectuer explicitement une transformation complexe des données vers un espace de très grande dimension (ce qui serait long et difficile à gérer), on calcule directement les similarités entre chaque paire d'échantillons dans l'espace de dimension élevée, à l'aide d'une simple fonction mathématique appelée noyau (Figure 35). Parmi les noyaux couramment employés, on distingue le noyau linéaire, adapté aux données initialement linéaires ; le noyau polynomial, capable de capturer des relations non linéaires complexes grâce à une transformation polynomiale ; le noyau radial, très performant pour modéliser des données distribuées radialement dans l'espace par mesure de la distance euclidienne entre deux positions ; et le noyau sigmoïde, inspiré des fonctions d'activation neuronales et utile dans certains contextes précis. Le choix judicieux de ces noyaux permet ainsi aux SVM d'offrir une grande flexibilité dans la classification de données biologiques complexes.

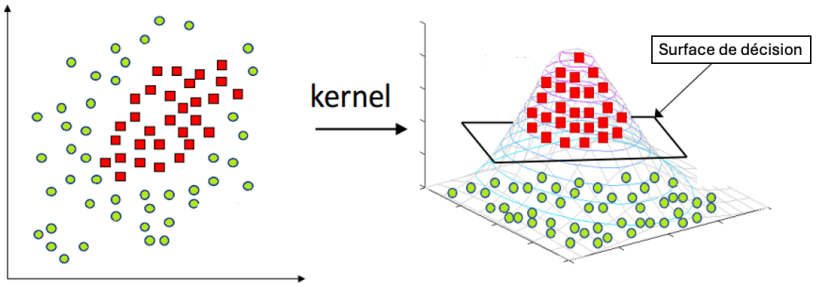


Figure 35 : Illustration du kernel trick utilisé dans les SVM

À gauche : dans l'espace d'origine, les deux classes (rouges et vertes) ne sont pas linéairement séparables. À droite : en appliquant une fonction noyau (ici un noyau radial), les données sont projetées implicitement dans un espace de plus grande dimension, dans lequel les deux classes deviennent séparables par un hyperplan linéaire. Cette transformation permet de définir une surface de décision non linéaire dans l'espace initial, sans avoir à calculer explicitement les coordonnées dans l'espace projeté.

Le modèle SVM a été testé face à d'autres modèles de classification supervisée : les algorithmes de forêts aléatoires (Random Forest) et XGBoost. Random Forest construit plusieurs arbres en parallèle sur des sous-échantillons aléatoires pour réduire la variance. XGBoost, basé sur le boosting, construit les arbres de manière séquentielle en corrigeant les erreurs précédentes, avec régularisation et optimisation par descente de gradient. Le modèle SVM à noyau radial surpasse les autres modèles en termes de sensibilité, spécificité, précision et score F1 (Figure 36).

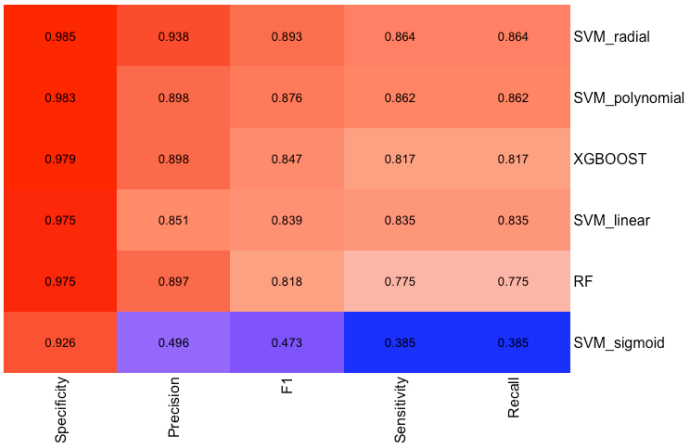


Figure 36 : Évaluation comparative des performances de plusieurs modèles de classification

Heatmap des métriques de performance (spécificité, précision, F1-score, sensibilité, et rappel) pour six modèles : SVM avec noyaux radial, polynomial, linéaire et sigmoïde, XGBoost, et Random Forest (RF). Les valeurs affichées dans chaque cellule correspondent aux scores obtenus pour chaque métrique. On observe que le modèle SVM_radial présentent les meilleures performances globales, tandis que SVM_sigmoid montre des résultats nettement inférieurs, notamment en termes de sensibilité et de F1-score.

1.3.2 Tester de nouvelles données

L'onglet CoReQuery (Figure 37) donne d'abord le choix à l'utilisateur entre envoyer des données locales (d'une taille maximale de 30 Mo) à l'application, ou avoir recours au service GEO pour récupérer des jeux de données externes. En cas de données locales, une boîte d'information peut être affichée pour indiquer le format à respecter afin que l'application puisse accomplir les différentes étapes de prétraitement. Il est aussi possible d'envoyer une classification personnelle des échantillons envoyés pour visualiser leur projection sur notre représentation 2D.



Figure 37: Onglet CoRegQuery de chargement de données locales

L'utilisateur peut choisir un fichier de son ordinateur où les noms d'échantillons sont en colonnes, les noms de gènes dans la première colonne, le séparateur est une virgule et le caractère décimal est un point. En seconde possibilité il peut envoyer une classification personnelle des échantillons sous la forme d'un tableau avec deux colonnes, la première correspondant au noms des échantillons et la seconde à la classe qui leur est associée. L'ensemble des informations de formatage sont disponibles en cliquant sur les boutons « How to format my data ».

La seconde possibilité consiste à utiliser l'option « GEO » de l'onglet CoRegQuery, qui permet de récupérer automatiquement les données d'expression d'une étude publiée à partir de son identifiant GEO. L'application télécharge alors le jeu de données d'expression correspondant, calcule l'influence des régulateurs

à partir de celui-ci, prédit les classes cRegMap, puis estime les coordonnées UMAP en prenant comme référence le modèle UMAP de la métacohorte (Figure 38).

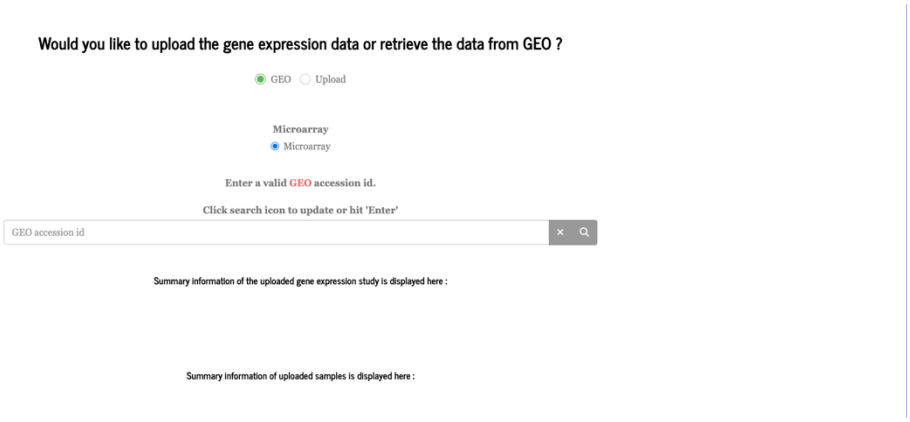
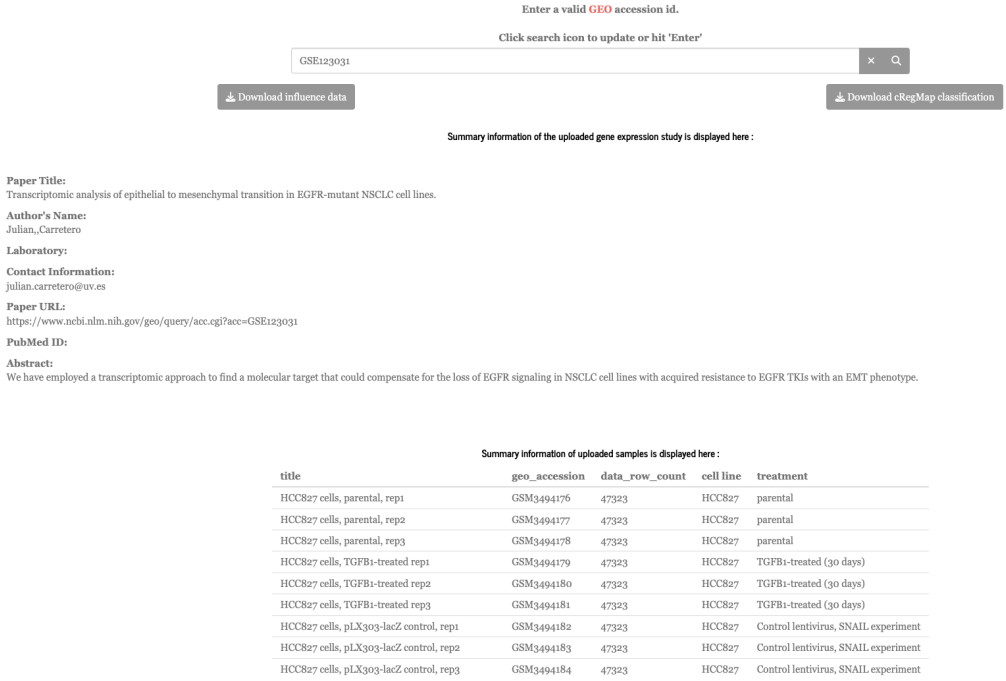


Figure 38 : Importation de données externes dans l'onglet CoRegQuery via l'option « GEO »

Une fois le jeu de données téléchargé et prétraité par l'application, un résumé de l'étude est affiché en bas de la page de l'onglet (Figure 39)



title	geo_accession	data_row_count	cell line	treatment
HCC827 cells, parental, rep1	GSM3494176	47323	HCC827	parental
HCC827 cells, parental, rep2	GSM3494177	47323	HCC827	parental
HCC827 cells, parental, rep3	GSM3494178	47323	HCC827	parental
HCC827 cells, TGFB1-treated rep1	GSM3494179	47323	HCC827	TGFB1-treated (30 days)
HCC827 cells, TGFB1-treated rep2	GSM3494180	47323	HCC827	TGFB1-treated (30 days)
HCC827 cells, TGFB1-treated rep3	GSM3494181	47323	HCC827	TGFB1-treated (30 days)
HCC827 cells, pLX303-lacZ control, rep1	GSM3494182	47323	HCC827	Control lentivirus, SNAIL experiment
HCC827 cells, pLX303-lacZ control, rep2	GSM3494183	47323	HCC827	Control lentivirus, SNAIL experiment
HCC827 cells, pLX303-lacZ control, rep3	GSM3494184	47323	HCC827	Control lentivirus, SNAIL experiment

Figure 39 : Résumé de l'étude téléchargée après récupération des données du GEO

Une fois les données récupérées, l'utilisateur peut retourner dans l'onglet CoRegNet, où il peut sélectionner l'identifiant GSE comme source des données et visualiser la projection de l'influence des classes prédites sur le GRN. Dans l'onglet CoRegMap, il peut superposer ces nouvelles données sur la représentation 2D et visualiser la colocalisation des nouveaux échantillons avec nos sous-groupes (Figure 40).



Figure 40 : Onglet CoRegMap où un GSE a été projeté

L'onglet CoRegQuery permet de projeter des résultats d'expérimentations scientifiques sur la UMAP de notre cohorte. Les nouvelles données sont indiquées par leur nom d'échantillons.

Cette contribution, qu'est l'ajout du module CoRegQuery, a été valorisée dans une étude sur le glioblastome, à laquelle j'ai participé en réalisant les analyses bio-informatiques. Un portail web interactif a également été spécialement développé afin d'offrir à la communauté scientifique l'opportunité d'explorer notre modèle.

2. Gene regulatory network analysis identifies prognostic subgroups of lung adenocarcinoma driven by distinct master regulators

Dès leur découverte, les cancers du poumon ont été classés selon leur histologie, puis selon l'expression de marqueurs spécifiques par immunohistochimie, et enfin, cette classification a été affinée par la caractérisation moléculaire des cancers. Le cancer du poumon, en particulier les NSCLC, a largement bénéficié de l'intégration de la médecine de précision en pratique clinique, grâce à l'introduction systématique des tests moléculaires permettant d'orienter les patients vers des thérapies ciblées. Ces thérapies ont significativement contribué à la réduction initiale de la mortalité, se traduisant par une amélioration du taux de survie à cinq ans, désormais proche de 27 %, contre 19 % en 2020, même s'il reste parmi les taux de survie associés aux cancers les plus faibles (365,366). Ce faible taux de survie s'explique notamment par l'inefficacité des traitements de première intention, liée à une compréhension encore insuffisante de l'hétérogénéité tumorale au niveau moléculaire (367,368). L'hétérogénéité des tumeurs est principalement déterminée par leur profil génétique et épigénétique. Ainsi, l'identification précise des sous-types moléculaires à partir de données moléculaires apparaît comme une étape essentielle pour mieux comprendre la physiopathologie du cancer du poumon, dans l'espoir d'optimiser l'utilisation des stratégies thérapeutiques existantes et d'améliorer la prise en charge des patients (369,370).

Les récentes avancées des technologies de séquençage ont fait basculer les analyses génomiques et transcriptomiques à grande échelle dans l'ère du big data. Dans ce contexte, l'étude exhaustive du transcriptome, par des approches bio-informatiques, a permis d'identifier des sous-types moléculaires distincts du cancer du poumon. Dès les années 2000, l'étude du transcriptome de tumeurs pulmonaires a identifié des sous-groupes d'adénocarcinomes différenciés par leur survie et le caractère invasif des tumeurs (371). À ce jour, la classification la plus largement acceptée reste celle proposée en 2006 par Hayes et al., qui décrit trois sous-types transcriptionnels consensuels : bronchioïde, squamoïde et magnoïde (372). Ces sous-types ont été confirmés plus tard par une étude multiomique d'envergure du NSCLC menée par le consortium TCGA, toutefois renommés « proximal-proliferative », « proximal-inflammatory » et « terminal respiratory unit » (373) et servent encore de référence dans des études récentes (374,375). Inamura et al. ont identifié deux sous-types de carcinomes épidermoïdes, caractérisés par une expression différentielle de gènes impliqués dans le processus de prolifération cellulaire, associé à une différence nette de survie (376). Dans une étude plus approfondie publiée en 2010, Wilkerson et al. (377) ont proposé une classification du carcinome épidermoïde en quatre sous-types : primitif, sécrétoire, basal et classique. Ces sous-groupes présentaient des associations significatives avec la survie, le grade tumoral, ainsi qu'avec des processus biologiques distincts, et ont été validés par une caractérisation moléculaire plus fine menée par le consortium TCGA en 2012 (378).

Dans cette étude, nous avons appliqué notre approche en réseau de corégulation pour agréger de multiples jeux de données en une métacohorte unique et y identifier, de façon non supervisée, des sous-types de cancers du poumon. J'ai orchestré la récupération puis l'alignement des données cliniques et

transcriptomiques, exploité le GRN du poumon pour inférer, à l'échelle patient, l'influence des régulateurs, déterminé les clusters via une approche non-supervisée, puis construit un modèle de classification supervisée reliant ces profils d'influence aux classes. J'ai projeté les influences sur le réseau afin de mettre en évidence les modules régulateurs actifs spécifiques à ces classes. Les sous-groupes obtenus présentent une forte cohérence avec les précédentes classifications histologiques, mutationnelles et moléculaires ; deux sous-groupes riches en adénocarcinomes se distinguent par une différence marquée de médiane de survie, et nous identifions un sous-groupe de NSCLC dominé par des signatures imputables au stroma et au microenvironnement tumoral. J'ai caractérisé chaque cluster cliniquement et fonctionnellement via les données de survie des patients, par analyses par ontologies et par déconvolution cellulaire. J'ai identifié, cluster par cluster, des régulateurs potentiellement actionnables en mobilisant les données de dépendance génique issues d'essais de viabilité consécutif à du silencing de gènes par CRISPR-Cas9 issues du projet DepMap, afin de relier directement les régulateurs influents à des vulnérabilités thérapeutiques testables.

Pour rendre cette approche exploitable par la communauté et transposable à de nouveaux jeux de données, j'ai conçu et déployé le portail interactif cRegMap dédié au poumon. Ce portail permet d'explorer le modèle de GRN, d'inspecter les nœuds régulateurs majeurs, d'interroger les groupes définis à partir d'une métacohorte de plus de 4 000 patients, d'identifier les nœuds spécifiques à chaque groupe, et de projeter des données cliniques et biologiques sur une carte 2D des patients. Il autorise également la projection de nouvelles données expérimentales, qu'elles proviennent du dépôt public Gene Expression Omnibus (GEO) ou d'uploads utilisateurs, et leur assignation automatique à une classe via un classifieur supervisé (SVM) entraîné sur les profils d'influence et les classes. L'application est développée en R dans un cadre respectant les bonnes pratiques de génie logiciel pour assurer maintenabilité et fiabilité, conteneurisée via Docker puis déployée sur Google Cloud Platform, ce qui garantit scalabilité et stabilité opérationnelle pour l'exploration et la réanalyse reproductible des nouvelles classes et au-delà.

Gene regulatory network analysis identifies prognostic subgroups of lung adenocarcinoma driven by distinct master regulators

Geoffrey Pawlak1, David Tulasne2, Mohamed Elati1@

1 Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, UMR9020 – UMR1277 - Canther - Cancer Heterogeneity, Plasticity and Resistance to Therapies, Target team, F-59000 Lille, France
2 Univ. Lille, CNRS, Inserm, CHU Lille, UMR9020 – UMR1277 - Canther - Cancer Heterogeneity, Plasticity and Resistance to Therapies, Plasticity team, F-59000 Lille, France

@ To whom correspondence should be addressed.
@ Corresponding author: Prof. Mohamed Elati
Address: ONCOLille, Boulevard du Professeur Jules Leclercq
59000 LILLE, France
E-Mail: mohamed.elati@univ-lille.fr

Introduction

Lung cancer remains a paramount global health challenge, accounting for a substantial portion of cancer morbidity and mortality worldwide. With nearly 2.5 million new cases and over 1.8 million deaths reported in 2022, lung cancer was responsible for approximately 12.4% of all cancer diagnoses and 18.7% of cancer-related deaths, making it the leading cause of cancer mortality globally (1,2). In the United States alone, an estimated 238,340 new cases and 127,070 deaths were projected for 2023, highlighting the lung cancer's prime importance in global burden of cancer. The primary etiological factor for lung cancer is cigarette smoking, which contributes to approximately 80% of cases. However, other risk factors such as radon exposure, asbestos, air pollution, and genetic predisposition also play crucial roles. The disease is primarily categorized into non-small cell lung cancer (NSCLC), which represents 80-85% of cases, and small cell lung cancer (SCLC), comprising the remaining 15%. Despite advances in early detection and therapeutic strategies, the prognosis for lung cancer remains dismal, with a 5-year survival rate of only 10-20% (3). A better understanding of the disease as a heterogeneous entity with relevant tools is necessary to raise the bar of disease control and long-term survival.

The advent of high-throughput sequencing technologies has revolutionized our understanding of the cellular and molecular landscapes of cancers. Several studies have identified distinct molecular subgroups based on gene expression profiling, highlighting their differences in survival, mutational landscape, immunological and stromal signature. Initially, the Cancer Genome Atlas Research Network (TCGA) characterized the molecular pathways involved in lung cancer and classified lung squamous cell carcinoma (LUSC) into four subtypes: classical, basal, secretory, and primitive (4). Lung adenocarcinoma (LUAD) was similarly stratified into terminal respiratory unit, proximal inflammatory, and proximal proliferative subtypes (5). Recently, new single and multiomics approaches

have been proposed to identify molecular subtypes of lung cancer to better understand the molecular mechanisms at work and determine new therapeutic targets (6–9). Although these efforts have significantly advanced our understanding of cancer biology, translating these findings into clinically applicable high-throughput approaches remains a major challenge. First, many of these studies relies on multiomics approach, the heterogeneity of publicly available multi-omics datasets, both in terms of data type and cohort composition, poses a significant barrier to the comprehensive analysis of lung tumors across all histological subtypes. This disparity limits the ability to capture the full spectrum of tumor heterogeneity necessary for robust and generalizable conclusions. Second, most previous studies have concentrated on individual lung cancer subtypes, such as LUAD, LUSC, large cell lung cancer (LCLC), or small cell lung cancer (SCLC), thereby neglecting the possibility of intratumoral coexistence of distinct histological subtypes (10,11). Moreover, such an approach overlooks dynamic processes like NSCLC-to-SCLC transformation, which are increasingly recognized as key mechanisms of therapeutic resistance (12,13).

Advances in transcriptomic sequencing technologies have generated an unprecedented wealth of gene expression profiling data. In parallel, cancer is increasingly understood not as the result of alterations in a single set of genes, but rather as a disease of the genome and of disrupted molecular interaction networks and regulatory control. Consequently, the reconstruction of gene regulatory networks (GRN) in both healthy and diseased tissues has become essential for deciphering cancer phenotypes and for the development of more effective therapeutic strategies (14). To date, the computational approaches that have been most successful in this field are those enabling the reverse-engineering of context-specific gene regulatory networks, such as ARACNe (15) and hLICORN (16) based on bulk transcriptomic data, and SCENIC (17) using single-cell transcriptomics. A second major advancement in the exploration of regulatory networks, exemplified by CoRegNet (18) and VIPER (19), was the shift towards evaluating regulator activity, inferred from the expression profiles of their target genes, rather than solely assessing regulator expression, thereby enabling the identification of master regulators. Bansal et al. applied this methodology in adenocarcinoma to predict the therapeutic response and synergy of action of treatment with tyrosine kinase inhibitors (TKI) targeting EGFR and MET (20). Kostyrko et al. applied the ARACNe and VIPER algorithms to investigate the role of the epigenetic regulator UHRF1 in KRAS-driven lung cancer (21). Their study identified UHRF1 as a critical mediator of tumor cell growth. Functional experiments further demonstrated that UHRF1 knockout selectively impaired proliferation and induced apoptosis in KRAS-mutant lung cancer models. These findings highlight UHRF1 as a KRAS-specific vulnerability and a promising therapeutic target in lung cancer. The integration of clinical, molecular, and histological data through network-based approaches offers a powerful tool for uncovering the regulatory mechanisms that define tumor phenotypes. Studies have shown that network-based stratification can effectively cluster patients into distinct prognostic groups, correlating with survival outcomes and clinical characteristics (22–24).

In this study, we employed a coregulatory network analysis using the CoRegNet algorithm to integrate data from 42 studies, assembling a cohort of 5,487 lung cancer patients spanning both NSCLC and SCLC subtypes. This

integrative approach enabled the classification of lung cancers into seven distinct groups based on transcriptomic profiles, with strong correlations to survival outcomes, clinical features, histological classifications, and molecular characteristics. Then, we identified unique key regulators for each group by associating them with the underlying biological processes and validate them through experimental data from cell viability assays employing CRISPR-Cas9 genome editing

Results

To produce a highly specific lung regulatory network and to prevent non-specific and tumor-microenvironment related signals we have used transcriptomic data of 206 lung cancer cell lines from CCLE project and 2,375 regulatory elements from curated list transcription factor (TF) and co-factor (co-TFs) (See Methods) to generate a lung gene regulatory network (Lung-GRN). The resulting GRN (Figure 1) is consisted of 502 regulators (Supplementary Table S3) with 4143 target genes and 18053 regulatory interactions with a significant enrichment in validated TF-gene interactions and transcription factor-binding sites (p-value <1e-130). From the Lung-GRN and pairs of regulators based on their shared target genes (Supplementary Table S4) we built a co-regulatory network (Lung-CoRegNet; Figure 1) with the CoRegNet R package which incorporate 2203 cooperative protein-protein and transcription-factor binding site enriched interaction (p-value < 1e-130).

From coregulatory network to metacohort subtyping

To aggregate the 42 lung cancer cohorts into one unique lung cancer metacohort we computed regulatory influence of each lung coregulatory network's regulators in a sample-independent way (See Methods) avoiding batch effect in result (Supplementary Figure S1) and reaching dimensionality reduction (~20 000 genes to 211 regulators). With its capability to preserve global structure and capture global and local connectivity, Uniform Manifold Approximation and Projection (UMAP) has been applied to produce a two-dimension embedded space visualization from influence data henceforth noted Lung-CoRegMap.

A density-based clustering algorithm (See Methods) was applied on Lung-CoRegMap to identify relevant cluster below-mentioned cRegMap subgroups (Figure 2A). Determined clusters show strong adequation with main histologies (adenocarcinomas, squamous cell carcinomas and neuroendocrine), but with some differences: Large cell carcinomas and neuroendocrine are disseminated between two cRegMap subtypes and due to widespread distribution of adenocarcinomas and squamous cell carcinomas some cRegMap subgroups are made up of mixed histological entities (Figure 2B). More detailed clustering enables the proposal of a refined classification of lung cancer, with its histological, mutational, clinical, and molecular features extensively characterized in the subsequent sections of this study.

Two cRegMap subtypes are characterized by pure adenocarcinoma histology with 91.2% and 85.8% of adenocarcinomas (Supplementary table S5) and represents 26.04% and 5.69% of total metacohort samples. They are clearly differentiated by their very different survival data. Indeed, the first group stands out for its good prognosis (96 months, 95%CI: 76 - 112 months), noted in the following as NSCLC/Adk good prog. and the second

for the worst overall survival rate over NSCLC groups (34 months, 95%CI: 28 – 46 months), noted as NSCLC/Adk bad prog (Figure 3). Mutational analysis shows a high proportion of EGFR and KRAS mutated samples in NSCLC/Adk good prog. (27.1% each), two most-frequently proto-oncogene in adenocarcinoma (Fig 4) (25,26). In addition, when we project LUAD TCGA subtypes, 45% of the terminal respiratory unit (TRU) class fall into NSCLC/Adk good prog. cluster (Figure 5B). On the other hand, NSCLC/Adk bad prog. display a more aggressive phenotype, with KEAP1, STK11 and SMARCA4 mutations (32.4%, 17.6% and 14.7% individually) associated with poor prognosis outcomes (27,28). Moreover, NSCLC/Adk bad prog. differs from NSCLC/Adk group on its absence of the targetable molecular alteration EGFR (3%), and absence of mutation of the migration and invasion associated gene LRP1B (9%) (29). As reported by several studies (30,31), our evidence confirm the poor prognosis NSCLC/bad prog. group with a higher proliferation signature (Figure 5A) in NSCLC/Adk bad prognosis than NSCLC/Adk group (Hallmark ontology, adjusted p-value < 1.10-3). Ontology analysis performed on differentially expressed genes between NSCLC/Adk, NSCLC/Adk bad prog. and other cRegMap groups revealed upregulation KRAS signaling pathway (Figure 5C; Hallmark ontology, p-value < 0.01).

Some of the squamous cell lung tumors defines a highly homogeneous class with 90% squamous cell carcinoma (95%CI: 87.3-92.2]; Figure 2B) henceforth named as NSCLC/Sq. NSCLC/Sq group has the highest proportion of smokers (89.6%, 95%CI [85.5-92.7]) (Supplementary figure 3; Supplementary table S6) and median overall survival of 61.9 months [95%CI : 52-86] (Figure 3), these values correlate favorably lung squamous cell carcinoma epidemiological data (32). Mutational landscape display a majority of CSMD3 mutated samples (53.9%), a tumor suppressor gene (33) , KMT2D (31.3%) an epigenetic modulator which loss is involved in lung tumorigenesis (34) (Figure 4), KEAP1 and NFE2L2 which are frequently mutated in squamous cell carcinomas and involved in metastasis, cancer progression and radio/chemo-resistance and associated with shorter time to treatment failure (TTF) and overall survival (OS). The high proliferative phenotype of NSCLC/Sq is supported by the highest proliferation score among NSCLC groups and demonstrates high purity. In correlation with Figure 5A, NSCLC/Sq has the highest proportion of desert MFP subtype, a subtype characterized by its highest proportion of malignant cells percentage (Figure 5A). Forty-two percent of classical LUSC TCGA subtype fall into NSCLC/Sq, this is in line with the group's mutational characteristics (Figure 5B). NSCLC/Sq’s differentially expressed genes affect extracellular matrix organization, inflammatory response, EMT and KRAS signaling pathway (Figure 5C).

Another proportion of squamous cell tumors flows to a cRegMap subgroups made of 62.2% adenocarcinomas (95%CI: 60.0-64.4) and 31.4% squamous cell carcinomas (95%CI: 29.3-33.5) (Figure 2B). This group has a median overall survival like NSCLC/Sq subgroup: 60.8 months with a 55-64 months 95% confidence interval (Figure 3). Moreover, its stroma score and tumor purity make this group highly infiltrated by stroma with the lowest scores across all cRegMap subgroups. The low purity phenotype is correlated with a strong epithelial-mesenchyme transition (EMT) signature confirmed by the top matrix remodeling, CAF and EMT signature score. These signaling pathways have been shown to be activated by tumor-extracellular matrix (ECM) interactions (35). Mixed/Mes-like subgroup has the lowest desert MFP subtype and the highest fibrotic and immune-enriched – fibrotic MFP

subgroup, two subgroups characterized by increased vascularization and EMT process and high percentage of cancer-associated fibroblasts (CAF) (Figure 5A). Furthermore, functional enrichment analysis unveils a strong EMT-related pathways enrichment with EMT, ECM organization and cell adhesion activated pathways (Figure 5C) compared to all other groups and cell cycle. Consistently with its mesenchymal-like phenotype this cRegMap subgroup will be noted as Mixed/Mes-like

Remaining squamous cell lung tumors flows to a cRegMap subgroup which comprise also residuals adenocarcinomas tumors: 57.7% adenocarcinomas (95%CI: 54.4-60.9), 34.9% (95%CI: 31.8-38.0) and 4.3% neuroendocrine tumors (95%CI: 3.2-5.9) (Figure 2B). Its median overall survival is close to Mixed/Mes-like subgroup (60.8 months [95%CI: 54.7-68]) (Figure 3). As NSCLC/Adk bad prog. it shows bad prognostic-related mutations with KEAP1+ (17%) and STK11+ (16%) mutations (Figure 4). Hence, this group is distinguished from Mixed/Mes-like group by its strong proliferation signature (Figure 5A), it includes 70% of proximal-proliferative LUAD TCGA subtypes and 70% of primitive LUSC TCGA subtypes (Figure 5B) and over-representation of KRAS signaling pathway and neuron-related signatures (Figure 5C) giving its neuroendocrine signature, hereinafter referred to as Mixed/NE-like subgroup.

One cRegMap subgroup is characterized by most carcinoid tumors (64.4%; 96%CI: 49.4-77.4) amid neuroendocrine tumors (16.7%; 96%CI: 8.0-30.8) and adenocarcinomas (10.4%; 96%CI : 9.3-23.4), it is therefore named Carcinoid. Accordingly, to lung cancer literature, carcinoid subtype demonstrate the best overall survival rate with a median way above 14 years (Figure 3) and comprise young non-smokers patients (median age: 51 years, IC95% [36-64.5], non-smokers percentage: 66.7 %, IC95% [38.7-80.7], Supplementary Table S5). Furthermore, this subgroup exhibits the highest purity score, the lowest stroma score of all cRegMap subgroups (Figure 5A), significant neural-related signature and no KRAS signaling pathway activation (Figure 5C).

The last cRegMap subgroup is mainly composed of neuroendocrine tumors (41.4%; 95%CI: 35.2-47.9) after adenocarcinomas (26.6; 95%CI: 21.3-32.7), squamous cell carcinoma (21.7%; 95%CI: 16.8-27.5) and large cell carcinomas (7.0%; 95%CI: 4.2-11.1) (Supplementary Table S5). This groups have the worst survival rate (32 months, IC95% [27-39]) (Figure 3), harbor the neuroendocrine-marker RB1 mutation (50%; Figure 4) and the strongest proliferative signature and a high proportion of desert MFP subtype (72%) (Figure 5A). This groups harbor also a strong neuroendocrine-related signature (Figure 5C). Taken together, these results reinforce the strong neuroendocrine phenotype of the NE group.

Identify subtype-specific key regulators

Further analyses were conducted to identified cRegMap subgroups key regulators, for each cRegMap subgroup we compared mean regulator influence versus mean group null model through 25000 permutation tests (See Methods).

Key regulators seem to be distributed in modules, mainly tree module can be extracted with an adenocarcinoma module (Figure 6A) specific to NSCLC/Adk good prog. and NSCLC/Adk bad prog., an ECM-activity module (Figure 6B) linked to Mixed/Mes-like subgroup and a neuroendocrine-like module (Figure 6C) including NSCLC/Sq, Mixed/NE-like, NE and Carcinoid subgroups.

Adenocarcinomas' key regulators are associated with proliferation, tumor growth, matrix remodeling signature and tumor progression and cell cycle, with FOXA1, MUC1, HES1, NOTCH3 and NKX2-1. Key master regulators of NSCLC/Adk bad prog. group are primarily associated with proliferation and tumor growth such as very well-known genes NOTCH3, CCND1, IRF6 and MUC1. Considering Mixed/Mes-like's key regulators, the strong EMT process is driven by ETS1, SMAD3, TGFB1I1, TFGB1 through TGF□pathway together with EGFR, ZEB1, implicated in EMT-driven resistance, LOXL2, involved in cell surface matrix remodeling (36), YAP1, also participating in proliferation, drug resistance and invasion (37), and its paralog WWTR1 (38).
NSCL/Sq key regulators are involved in tumor progression and proliferation like CDK5R1, IRF6 or HES6 explaining its high proliferation score among NSCLC groups. Neuroendocrine groups (Mixed/NE-like, NE and Carcinoid) share a group of key regulators: INSM1, a very well know SCLC marker, required for neuroendocrine differentiation, MYB which has an influence on cell fate (39), PAX5 involved in cell viability in SCLC (40) and HES6, correlated with neuroendocrine phenotype (39).
Mixed/NE-like group differs from other NE groups with key regulators implicated in cell migration, invasion and proliferation such as MYT1L which also drive neuronal differentiation (41), FOXA2 (42), FLI1 and FGFR2 (43,44). Besides, tumor growth and differentiation were identified the neuroendocrine marker ASCL1 which drive pulmonary neuroendocrine differentiation and stimulates tumor progression.
NE group exhibits proliferation enhancing, metastasis and tumor growth regulators with HOXC10 and SYK (45). We found also DLX2 as NE master regulator involved in glycolytic switch, tumor invasion, apoptosis resistance and overexpressed in NE samples with neuroendocrine markers, neuronal developmental markers, and neurotransmitters (46,47).

To biologically validate our assumptions on group-specific key master regulators, some of the key regulators were validated through a gene dependency analysis: cRegMap subgroup were assign to lung cancer cell lines with a SVM supervised classifier. Thus, for each cRegMap cell lines' subgroup and for each key regulators we compared key regulator dependency score vs other subgroups' dependency score, a key regulator was considered as essential was its dependency score was significantly lower than other groups' dependency score. NSCLC/Adk good prog. subgroups show a common dependency on KLF5 regulator with NSCLC/Adk bad prog. whereas only NSCLC/Adk good prog. has a strong dependency on KLF5 regulator. Analysis confirm Mixed/Mes-like subgroup EMT phenotype with validation of WWTR1, FOSL1 and YAP1 dependencies. NSCLC/Sq had only MYB regulator confirmed by CRISPR/Cas9 experiments. Mixed/NE-like subgroups rely mainly on CHD7 MYB and CBFA2T2 regulators and finally, NE subgroup, three key regulators were experimentally confirmed, SOX4, MYB and INSM1 which is a SCLC marker (Figure 6)

Discussion

In this study, we built a lung cancer specific coregulatory network and determined lung cancer subgroups from lung tumors transcriptomics data. We characterized subgroups according to clinical and molecular profile, tumor purity, microenvironment, expression profile and key regulators. The groups determined proved to be highly histologically distinct and consistent with the current classification of lung cancers and their survival. Their key regulators reveal a heterogeneity of signaling pathways involved. Adenocarcinomas were split between NSCLC/Adk good prog., NSCLC/Adk bad prog. and Mixed/Mes-like groups. NSCLC/Adk good prog. malignancy seems to be based on cell proliferation, migration and invasion supported by two key regulator KLF5 and FOXA1. KL5 has an ambivalent, it has been proved to be associated FOXA1 in squamous cell carcinoma of the lung with distant metastasis and an unfavourable survival rate (48) and with favorable prognosis (49,50) in cancers other than lung cancer. In lung adenocarcinomas, FOXA1 promotes NSCLC progression (51). KLF5 plays a pivotal role in regulating cell proliferation, differentiation, and programmed cell death and can be used as a prognostic indicator for both OS and disease-free survival in LUAD (52) it has been also associated with apoptosis inhibition (53).

Proliferation, migration, invasion and cell growth as well as tumorigenesis are the main processes in which the key regulators identified in the NSCLC/Adk bad prog group are involved, along with ELF3, NOTCH3, KLF5, CCDN1, MUC1 and IRF6. Previously described as a tumor suppressor gene, ELF3 is required for tumor growth, Enfield et al. have demonstrated that ELF3 is associated with poor survival and proliferation but needs other molecular aberration to transform cells (54), supporting the idea of key regulator cooperation trough a coregulation network. Together with NOTCH3, and consistently with the high KRAS mutation rate inside NSCLC/Adk bad prog. group, they are found in KRAS-mediated lung adenocarcinoma to be responsible of tumor initiation and maintenance by fixation of ELF3 on NOTCH3 promoter after its phosphorylation by PKC (55). MUC1 and CCND1 are key regulators strongly related to proliferation, MUC1 has been found overexpressed preferentially in NSCLC adenocarcinomas and associated with signaling in tumor cell proliferation and angiogenesis (56). CCND1 is an oncogene and a well-documented regulator of the G1/S transition, associated with poor prognostic indicated in lung cancer (57). Furthermore, CCND1 is implicated in the proliferation of LUAD cells through the long noncoding RNA PTPRG-AS1, which enhances its expression, thereby promoting cell cycle progression and proliferation66. Those evidence comforting our results about this highly KRAS-mutated aggressive group. Moreover, IRF6 is involved in cell proliferation since its inhibition by miR-320, an endogenous inhibitor of IRF6, leads to the inhibition of proliferation and migration of lung cancer cells (58). Moreover, its dysregulation has also been noted in other squamous cell carcinomas, such as vulvar squamous cell carcinoma, where hypermethylation of the IRF6 promoter correlates with cancer progression (59). Epithelial-mesenchymal transition plays a crucial role in lung cancer progression, particularly in the context of tumor stroma interactions and have a strong signature inside Mixed/Mes-like group. ZEB1 is a pivotal transcription factor that plays a significant role in the EMT and progression of lung cancer by regulating various lncRNAs like ZEBA-AS1 (60). Additionally, ZEB1 promotes EMT by downregulating epithelial markers and upregulating mesenchymal markers, facilitating invasive capabilities in lung cancer cells (61). Besides, the transcriptional co-activator with PDZ-binding motif (TAZ or WWTR1) is a key effector of the Hippo signaling pathway, which plays a significant role in organ growth, tumorigenesis, and EMT.

Furthermore, YAP is also frequently activated in lung cancer and can drive with TAZ cancer cell survival, proliferation, and metastasis. ZEB1, another transcription factor interacts directly with YAP1, switching its function to a transcriptional co-activator, which is associated with aggressive cancer behavior and poor prognosis. This interaction links the EMT-inducing properties of ZEB1 with the oncogenic effects of YAP1, promoting metastasis and therapy resistance (62). These studies reinforce the relevance of our findings, making WWTR1 and ZEB1 experimentally validated key regulators within the Mixed-Mes-like group. NE rich groups (Mixed/NE-like, NE and Carcinoid) exhibit specific NE key regulators such as Insulinoma-associated protein 1 (INSM1) a highly sensitive and specific marker for neuroendocrine differentiation (63) and is crucial for neuroendocrine differentiation in lung cancer. It regulates the expression of other neuroendocrine markers such as BRN2, chromogranin A, and synaptophysin (64). Factor 4 (SOX4) belong to NE group's key regulators validated by CRISPR-Cas9 data. SOX4 is significantly upregulated in small cell lung cancer (SCLC) and large cell neuroendocrine carcinomas (LCNEC), correlating with poor prognosis and increased tumor aggressiveness. SOX4 is involved in the regulation of genes associated with neuronal development and is highly expressed in SCLC, suggesting a functional role in the neuroendocrine characteristics of these tumors (65). Consistently with literature, Carcinoid group as the best overall survival rate among all cRegMap groups which is reflected by its master regulators

In conclusion, we identified non-small cell lung cancer (NSCLC) and small-cell lung cancer (SCLC) specific clusters such as NSCLC/Adk, NSCLC/Sq, NSCLC/Adk Advanced, SCLC and Carcinoid with highly homogeneous subtypes and Mixed/Mes-like and Mixed/SCLC as heterogeneous but still specific subgroups, highlighting the potential of our classification to produce innovative hypothesis on lung-cancer subtype mechanisms. The cRegMap clusters determined with CoRegNet biological network approach managed to be extremely coherent with 2021 WHO lung histological classification, which allow the possibility to detect vulnerability points / therapeutic insights and refine previously defined histological groups. Further studies are needed to overcome current limitations such as developing specific lung cancer cell lines models and validate suggested cRegMap group's key regulators. All results are publicly available at www.lung.cregmap.com.

Materials and methods

Lung-cRegMap development

Public transcriptome data collection. Transcriptomic and clinical data were collected from Lung Cancer Explorer Project (LCE) (66), Genomic Data Commons (GDC) portal (67) and cBioPortal (68), resulting in 42 datasets for 5868 samples, 5487 tumorous and 381 healthy samples. (Supplementary data S1). Additional transcriptomic data from cell line were downloaded through the Cancer Dependency Map (DepMap) project (69) resulting on 206 cell lines. All gene identifiers were converted into HUGO Gene Nomenclature Committee (HGNC) symbols (70).

Public data collection of regulatory elements and interactions. The human transcription factor catalogue (71) and the TcoF-DB v2 database (72) yielded a list of 2,375 regulators corresponding to 1,639 transcription factors (TF) with experimentally validated DNA binding specificity and 752 transcription cofactors (TcoF) with experimental validation information. The human transcription factor binding site (TFBS) was modelled using

MotifDB R/Bioconductor package (73) and promoter sequences were searched using the PWMEnrich R/Bioconductor package (74). ChIP-seq data from the ChEA2 database (75) and the tftargets R package (76) and various databases (e.g. TRED, ITFP, ENCODE, BEDOPS, TRRUST) provided all regulatory evidence from TF to target gene, yielding a list of 2,256,674 TF-target interactions. The collaborative evidence between TFs (TF-TF) list containing 1,257,053 protein-protein interactions (PPIs) from different PPI databases (e.g., BioGrid, HIPPIE, STRING, and HPRD) was obtained using the iRefR R package (77).

Gene regulatory network inference. Gene regulatory networks were inferred with the CoRegNet Bioconductor package (18). Starting from transcriptomic data and a list of human regulators, CoRegNet uses the hLICORN algorithm (16) to capture regulatory interactions between regulators and target genes in four steps: (1) Transcriptomic data is discretized into values of -1, 0 and 1 according to per-gene distribution. Genes present in the transcriptomic dataset were split into regulators and target genes, and only those with minimum support for non-zero values after discretization were retained. (2) A frequent itemset algorithm (a priori) identifies potential sets of co-activators and co-inhibitors from the list of regulators. (3) For each target gene, a list of candidate co-activator and co-inhibitor sets (GRN) was selected according to a regulatory program (association rule) metric. (4) Each GRN candidate for each gene was scored based on the regression between the expression of the GRN set regulators and the expression of the gene concerned. For each gene, we retained the top ten GRN candidates according to R2 score. CoRegNet can then be used to select the best GRN for each gene based on the interaction evidence data. Evidence of regulation and cooperation was incorporated into each candidate GRN with an integrative selection algorithm, yielding an R2 score for each of the integrated datasets. The GRN with the highest merged score was selected. CoRegNet builds a coregulatory network, GBM-CoRegNet, from the GRNs obtained, by setting a cooperativity relationship between a pair of regulators, TFi and TFj, if they have a minimum of five target genes in common and the relationship is significant ($p \leq 0.01$) according to the Jaccard similarity co-efficient formula: : $\frac{(TF_i \cap TF_j)}{(TF_i \cup TF_j)}$.

Network-based regulatory influence signal quantification. The regulatory network structure provides a set of genes that are activated and repressed under reference conditions. Based on this structure, we can capture the influence of each regulator, a latent signal of the regulator activity in each sample based on its observed effect on downstream entities. For each regulator, Welch's t-test was performed to compare the distribution of activated (Ar) and repressed (Ir) genes. The influence of a regulator r is calculated as follows:

"

$$\text{Influence}(r) = \frac{\overline{E(A^r)} - \overline{E(I^r)}}{\sqrt{\frac{\mu^2_{A^r}}{A^r} + \frac{\mu^2_{I^r}}{I^r}}}$$

where $E(A^r)$ and $E(I^r)$ are the expression of the activated and repressed genes in the samples, respectively. . $\overline{E(A^r)} \wedge \overline{E(I^r)}$ are their respective means, and μAr and μIr are their standard deviations. A regulator is active

only when it activates Ar and represses Ir in accordance with the expectations of the regulatory network model, giving a positive t value in the Welch t-test. Higher $Influence(r)$ values indicate higher levels of regulator r activity in the sample. In this work, we calculated the $Influence$ of a regulator only if it has activated or repressed at least five genes. Due to the strict thresholds used in network construction, some local gene networks may be filtered out, resulting in some regulators having activated or repressed target gene sets with fewer than five members. In such cases, we fill the smaller set (fewer than 5 target genes) with the target genes for which the regulator concerned has the highest R2 regression score. This makes it possible to capture the $Influence$ values for a maximum number of regulators. In addition, for some regulators, expression values may be missing for some of the target genes in some tumor samples for which we wish to assess the influence. In this case, the missing expression values are estimated with the LatNet method (78), using expression levels in cell lines as a reference dataset. It may not be possible to estimate the expression of some target genes for a particular sample, and this may preclude calculation of the influence of a particular regulator. In this (rare) case, the median of the obtained influence values for the regulator in the other samples is used to replace the missing value.

Dimension reduction. Seurat R package (79) was used to embedded influence data into a 2D representation. Minimum distance and number of neighbors were set to 0.1 and 25 respectively.

Clustering. FindClusters and FindSubCluster functions from Seurat R packag were applied to detected cluster upon tumorous samples. Resolution, prune and k. neighbors were set to 0.1, 0.05 and 50 respectively.

The Lung-cRegMap webserver. Lung.cRegMap web application is written in the R language with the golem R framework and RShiny package (80,81). To avoid update issues and ensure stability to our app the application was containerized with Docker and deploy on Google Cloud Platform with Cloud Run service and publicly available at www.lung.cregmap.com. Lung.cregmap includes three main tabs: “CoRegNet”, “CoRegMap” and “CoRegQuery”. In the “CoRegNet” tab the user can interact with an intuitively comprehensible representation of Lung.CoRegNet network throught visNetwork R package (82) and explore the coregulatory network and to identify core active coregulators in a single sample (e.g., cell line) or a particular phenotype (e.g., tumor subtype) for further analysis. The “CoRegMap” tab allow user to annotate 2D UMAP lung metacohort representation with various projection including histological and cRegMap classifications, specific regualtor influence, somatic mutations, clinical data and a CRISPR/Cas9 screen-based DepMap gene essentiality score for the survival/proliferation of cell lines. Finally, the CoRegQuery grants user the possibility to upload their expression datasets or to retrieve expression datasets from the Gene Expression Omnibus (GEO) with ease. As input, expression data can be uploaded in CSV format or as a GEO series ID, also known as a GEO accession code, from which GBM-cRegMap extracts the gene expression dataset and experimental information. In only a few seconds, Lung-cRegMap calculates the influence of regulators on the new data, displaying the results in the reference GBM-CoRegNet and GBM-CoRegMap, using the machine-learning procedure of mapping and annotating the query dataset. Resulting projection can be visualized in the CoRegMap tab.

Lung-cRegMap subclasses characterization.

Clinical variable. Statistical tests on clinical variables were performed with R and we used Kruskal-Wallis non-parametric test and pairwise comparison Dunn's test for quantitative variable and homogeneity of proportion between groups Chi-squared test for qualitative variables, chi-squared post-hoc investigations were performed with z-test pairwise proportion comparison for two-levels chi-squared test of independence and accordingly to Sharpe study (83) with the ±1.96 standardized residual value criteria for large contingency table. Multiple tests were adjusted following Bonferroni correction.

Survival analysis. Survival analysis was perform with survival R package (84), progression-free survival and overall survival differences were assessed with Cox regression model with control of proportional hazards assumption. Kaplan-Meier plots are generated with ggsurvfit R package (85).

Gene differential expression and functional enrichment analysis. Differentially expressed genes were identified with linear models for microarray data (limma) R package (86) and p-Value were adjusted by Benjamini, Hochberg and Yekutieli method. Gene ontology (GO) analysis was done with clusterProfiler R package (87) on Kyoto Encyclopedia of Genes and Genomes (KEGG), Molecular Signatures Database (MSigDB), and REACTOME pathway databases. Results were filtered with a p-value < 0.01 and gene-ratio > 5%.

Genomic alterations. Somatic mutations were manually downloaded for cBioPortal dataset and collected with GenomicDataCommons R package (88) for GDC datasets. Somatic Somatic mutations were publicly available for uCologne_2015, CDDP_EAGLE-1, CPTAC-3, HCMI-CMDC, TCGA-LUAD and TCGA-LUSC for a total of 1477 samples. All somatic mutation files were aligned with the hg38 sequence and read, preprocessed and analysed with maftools R package (89). We performed an over-representation analysis of the mutations characterizing each subgroup by testing the observed number of mutations for each gene and each emerging cluster with Z-tests against their respective null models. Null models have been built via bootstrap resampling of the cRegMap clustering labels and provide the statistically expected abundance of each variable in each cluster given the distribution of the variable itself over the whole dataset and the numerosity of samples in each cluster. In particular, the cRegMap labels have been resampled 2000 times, finally computing the average proportion of each mutation for each cluster and its standard deviation. The Z-score is built as the difference between observed and relative abundance in units of standard deviation.

Tumor purity, stroma score and molecular functional portrait. Tumor purity and stroma scores were computed with immunedeconv R package with ESTIMATE method (90) on all samples' expression data. Mean score pairwise comparison was performed with t-test adjusted with Bonferroni method. Molecular functional portrait (MFP) analysis was conducted following Bagaev et al. procedure (91) to assess tumor microenvironment subtypes. Dunn test was employed for pairwise comparison adjusted by Bonferroni procedure.

cRegMap classes classifier. An SVM model with radial kernel was trained on metacohort influence data following a 10-fold cross-validation procedure with a mean F1-score equal to 90.58% with e1071 R package (92).

Identification and validation of subclass-specific regulators. To identify cluster-specific over-influent regulator we applied the same methodology as mutational analysis. The null model was computed with 500 permutations and mean regulator's influence was compared to null distribution with Z-score measurement. P-values were adjusted with FDR procedure.

To validate DIRs we have selected DIRs with negative difference between cluster vs rest dependency before computing one-sided Wilcoxon-Mann-Withney non-parametric tests between the mean CERES dependency score of the cell lines assigned to the subclass and the mean dependency score of the rest. CERES scores were obtained with the depmap R package (69). All tests were adjusted following Bonferroni correction procedure.

Code availability

The Lung-cRegMap web application was deployed with Google Cloud Run and is publicly available from https://lung.cregmap.com. The underlying code is available in GitHub/i3bionet/lung.cregmap and can be accessed via this link http://github.com/i3bionet/lung-cregmap under the Apache 2 Licence.

References

1.Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-63.

2.Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: current status and future trends. Nat Rev Clin Oncol. sept 2023;20(9):624-39.

3.Allemani C, Matsuda T, Di Carlo V, Harewood R, Matz M, Nikšić M, et al. Global surveillance of trends in cancer survival 2000-14 (CONCORD-3): analysis of individual records for 37 513 025 patients diagnosed with one of 18 cancers from 322 population-based registries in 71 countries. Lancet Lond Engl. 17 mars 2018;391(10125):1023-75.

4.Cancer Genome Atlas Research Network. Comprehensive genomic characterization of squamous cell lung cancers. Nature. 27 sept 2012;489(7417):519-25.

5.Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 31 juill 2014;511(7511):543-50.

6.Lehtiö J, Arslan T, Siavelis I, Pan Y, Socciarelli F, Berkovska O, et al. Proteogenomics of non-small cell lung cancer reveals molecular subtypes associated with specific therapeutic targets and immune-evasion mechanisms. Nat Cancer. nov 2021;2(11):1224-42.

7.Gillette MA, Satpathy S, Cao S, Dhanasekaran SM, Vasaikar SV, Krug K, et al. Proteogenomic Characterization Reveals Therapeutic Vulnerabilities in Lung Adenocarcinoma. Cell. 9 juill 2020;182(1):200-225.e35.

8.Song KJ, Choi S, Kim K, Hwang HS, Chang E, Park JS, et al. Proteogenomic analysis reveals non-small cell lung cancer subtypes predicting chromosome instability, and tumor microenvironment. Nat Commun. 23 nov 2024;15(1):10164.

9.Chen Z, Mei K, Tan F, Zhou Y, Du H, Wang M, et al. Integrative multi-omics analysis for identifying novel therapeutic targets and predicting immunotherapy efficacy in lung adenocarcinoma. Cancer Drug Resist. 14 janv 2025;8(0):N/A-N/A.

10.Deng P, Hu C, Zhou L, Li Y, Huang L. Clinical characteristics and prognostic significance of 92 cases of patients with primary mixed-histology lung cancer. Mol Clin Oncol. 1 sept 2013;1(5):863-8.

11.Ruffini E, Rena O, Oliaro A, Filosso PL, Bongiovanni M, Arslanian A, et al. Lung tumors with mixed histologic pattern. Clinico-pathologic characteristics and prognostic significance. Eur J Cardiothorac Surg. 1 nov 2002;22(5):701-7.

12.Zeng J, Ding X, Ding J, Wang X. Histological transformation into SCLC: An important resistance mechanism of NSCLC upon immunotherapy. Front Immunol. 2023;14:1275957.

13.Liang X, Lin A, Wang Q, Zhang J, Luo P. Cell plasticity in patients with NSCLC: The controversial origins of transformed SCLC. Biomed Pharmacother. 1 mai 2022;149:112909.

14.Sonawane AR, Platig J, Fagny M, Chen CY, Paulson JN, Lopes-Ramos CM, et al. Understanding Tissue-Specific Gene Regulation. Cell Rep. 24 oct 2017;21(4):1077-88.

15.Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Dalla Favera R, et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. BMC Bioinformatics. 20 mars 2006;7 Suppl 1(Suppl 1):S7.

16.Chebil I, Nicolle R, Santini G, Rouveirol C, Elati M. Hybrid method inference for the construction of cooperative regulatory network in human. IEEE Trans Nanobioscience. juin 2014;13(2):97-103.

17.Aibar S, González-Blas CB, Moerman T, Huynh-Thu VA, Imrichova H, Hulselmans G, et al. SCENIC: single-cell regulatory network inference and clustering. Nat Methods. nov 2017;14(11):1083-6.

18.Nicolle R, Radvanyi F, Elati M. CoRegNet: reconstruction and integrated analysis of co-regulatory networks. Bioinforma Oxf Engl. 15 sept 2015;31(18):3066-8.

19.Alvarez MJ, Shen Y, Giorgi FM, Lachmann A, Ding BB, Ye BH, et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. Nat Genet. août 2016;48(8):838-47.

20.Bansal M, He J, Peyton M, Kustagi M, Iyer A, Comb M, et al. Elucidating synergistic dependencies in lung adenocarcinoma by proteome-wide signaling-network analysis. Chellappan S, éditeur. PLOS ONE. 7 janv 2019;14(1):e0208646.

21.Kostyrko K, Román M, Lee AG, Simpson DR, Dinh PT, Leung SG, et al. UHRF1 is a mediator of KRAS driven oncogenesis in lung adenocarcinoma. Nat Commun. 5 juill 2023;14(1):3966.

22.Chudasama D, Bo V, Hall M, Anikin V, Jeyaneethi J, Gregory J, et al. Identification of cancer biomarkers of prognostic value using specific gene regulatory networks (GRN): a novel role of RAD51AP1 for ovarian and lung cancers. Carcinogenesis. 8 mars 2018;39(3):407-17.

23.He Z, Zhang J, Yuan X, Liu Z, Liu B, Tuo S, et al. Network based stratification of major cancers by integrating somatic mutation and gene expression data. PLOS ONE. 16 mai 2017;12(5):e0177662.

24.Li B, Nabavi S. A multimodal graph neural network framework for cancer molecular subtype classification. BMC Bioinformatics. 15 janv 2024;25(1):27.
25.Castellanos E, Feld E, Horn L. Driven by Mutations: The Predictive Value of Mutation Subtype in EGFR-Mutated Non-Small Cell Lung Cancer. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. avr 2017;12(4):612-23.
26.Goulding RE, Chenoweth M, Carter GC, Boye ME, Sheffield KM, John WJ, et al. KRAS mutation as a prognostic factor and predictive factor in advanced/metastatic non-small cell lung cancer: A systematic literature review and meta-analysis. Cancer Treat Res Commun. 2020;24:100200.
27.Dempke WCM, Reck M. KEAP1/NRF2 (NFE2L2) mutations in NSCLC - Fuel for a superresistant phenotype? Lung Cancer Amst Neth. sept 2021;159:10-7.
28.Pang LL, Zhou HQ, Zhang YX, Zhuang WT, Pang F, Chen LJ, et al. SWI/SNF family mutations in advanced NSCLC: genetic characteristics and immune checkpoint inhibitors' therapeutic implication. ESMO Open. 3 juin 2024;9(6):103472.
29.Gonias SL, Campana WM. LDL Receptor–Related Protein-1: A Regulator of Inflammation in Atherosclerosis, Cancer, and Injury to the Nervous System. Am J Pathol. 1 janv 2014;184(1):18-27.
30.Ekman S, Wynes MW, Hirsch FR. The mTOR Pathway in Lung Cancer and Implications for Therapy and Biomarker Analysis. J Thorac Oncol. 1 juin 2012;7(6):947-53.
31.Papadimitrakopoulou V. Development of PI3K/AKT/mTOR Pathway Inhibitors and Their Application in Personalized Therapy for Non–Small-Cell Lung Cancer. J Thorac Oncol. 1 août 2012;7(8):1315-26.
32. Reuss JE, Zaemes J, Gandhi N, Walker P, Patel SP, Xiu J, et al. Comprehensive molecular profiling of squamous non-small cell lung cancer reveals high incidence of actionable genomic alterations among patients with no history of smoking. Lung Cancer. 19 janv 2025 ; 200 : 108101.
33.Zhang P, Kang B, Xie G, Li S, Gu Y, Shen Y, et al. Genomic sequencing and editing revealed the GRM8 signaling pathway as potential therapeutic targets of squamous cell lung cancer. Cancer Lett. 1 févr 2019;442:53-67.
34.Pan Y, Han H, Hu H, Wang H, Song Y, Hao Y, et al. KMT2D deficiency drives lung squamous cell carcinoma and hypersensitivity to RTK-RAS inhibition. Cancer Cell. 9 janv 2023;41(1):88-105.e8.
35.Winkler J, Abisoye-Ogunniyan A, Metcalf KJ, Werb Z. Concepts of extracellular matrix remodelling in tumour progression and metastasis. Nat Commun. 9 oct 2020;11(1):5120.
36.Peng DH, Ungewiss C, Tong P, Byers LA, Wang J, Canales JR, et al. ZEB1 induces LOXL2-mediated collagen stabilization and deposition in the extracellular matrix to drive lung cancer invasion and metastasis. Oncogene. 6 avr 2017;36(14):1925-38.
37.Li L, Liu D, Chen T, Wei C, Qiao Y, Liu W, et al. Hypoxia-enhanced YAP1-EIF4A3 interaction drives circ_0007386 circularization by competing with CRIM1 pre-mRNA linear splicing and promotes non-small cell lung cancer progression. J Exp Clin Cancer Res. 20 juill 2024;43(1):200.
38.Lo Sardo F, Pulito C, Sacconi A, Korita E, Sudol M, Strano S, et al. YAP/TAZ and EZH2 synergize to impair tumor suppressor activity of TGFBR2 in non-small cell lung cancer. Cancer Lett. 1 mars 2021;500:51-63.
39.Zhang W, Girard L, Zhang YA, Haruki T, Papari-Zareei M, Stastny V, et al. Small cell lung cancer tumors and preclinical models display heterogeneity of neuroendocrine phenotypes. Transl Lung Cancer Res. févr 2018;7(1):32-49.

40.Kanteti R, Nallasura V, Loganathan S, Tretiakova M, Kroll T, Krishnaswamy S, et al. PAX5 is expressed in small-cell lung cancer and positively regulates c-Met transcription. Lab Invest. mars 2009;89(3):301-14.
41.Yu Z, Liu Y, Li Y, Zhang J, Peng J, Gong J, et al. miRNA-338-3p inhibits glioma cell proliferation and progression by targeting MYT1L. Brain Res Bull. 1 févr 2022;179:1-12.
42.Jin S, He J, Zhou Y, Wu D, Li J, Gao W. LncRNA FTX activates FOXA2 expression to inhibit non–small-cell lung cancer proliferation and metastasis. J Cell Mol Med. 2020;24(8):4839-49.
43.Miao B, Bauer AS, Hufnagel K, Wu Y, Trajkovic-Arsic M, Pirona AC, et al. The transcription factor FLI1 promotes cancer progression by affecting cell cycle regulation. Int J Cancer. juill 2020;147(1):189-201.
44.Li L, Zhang S, Wei L, Wang Z, Ma W, Liu F, et al. FGF2 and FGFR2 in patients with idiopathic pulmonary fibrosis and lung cancer. Oncol Lett. 1 août 2018;16(2):2490-4.
45.Fang J, Wang J, Yu L, Xu W. Role of HOXC10 in Cancer. Front Oncol [Internet]. 25 mai 2021 [cité 11 août 2024];11. Disponible sur: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2021.684021/full
46.Lee SY, Jeon HM, Ju MK, Jeong EK, Kim CH, Park HG, et al. Dlx-2 and glutaminase upregulate epithelial-mesenchymal transition and glycolytic switch. Oncotarget. 11 janv 2016;7(7):7925-39.
47.Pedersen N, Mortensen S, Sørensen SB, Pedersen MW, Rieneck K, Bovin LF, et al. Transcriptional gene expression profiling of small cell lung cancer cells. Cancer Res. 15 avr 2003;63(8):1943-53.
48.Deutsch L, Wrage M, Koops S, Glatzel M, Uzunoglu FG, Kutup A, et al. Opposite roles of FOXA1 and NKX2-1 in lung cancer progression. Genes Chromosomes Cancer. 2012;51(6):618-29.
49.Albergaria A, Paredes J, Sousa B, Milanezi F, Carneiro V, Bastos J, et al. Expression of FOXA1 and GATA-3 in breast cancer: the prognostic significance in hormone receptor-negative tumours. Breast Cancer Res. 23 juin 2009;11(3):R40.
50.Abe Y, Ijichi N, Ikeda K, Kayano H, Horie-Inoue K, Takeda S, et al. Forkhead box transcription factor, forkhead box A1, shows negative association with lymph node status in endometrial cancer, and represses cell proliferation and migration of endometrial cancer cells. Cancer Sci. avr 2012;103(4):806-12.
51.Chen G, Wei RS, Ma J, Li XH, Feng L, Yu JR. FOXA1 prolongs S phase and promotes cancer progression in non-small cell lung cancer through upregulation of CDC5L and activation of the ERK1/2 and JAK2 pathways. Kaohsiung J Med Sci. nov 2023;39(11):1077-86.
52.Fang Q, Xu M, Yao W, Wu R, Han R, Kawakita S, et al. The role of KLF5 in gut microbiota and lung adenocarcinoma: unveiling programmed cell death pathways and prognostic biomarkers. Discov Oncol. 5 sept 2024;15(1):408.
53.Li X, Liu X, Xu Y, Liu J, Xie M, Ni W, et al. KLF5 promotes hypoxia-induced survival and inhibits apoptosis in non-small cell lung cancer cells via HIF-1α. Int J Oncol. 1 oct 2014;45(4):1507-14.
54.Enfield KSS, Marshall EA, Anderson C, Ng KW, Rahmati S, Xu Z, et al. Epithelial tumor suppressor ELF3 is a lineage-specific amplified oncogene in lung adenocarcinoma. Nat Commun. 28 nov 2019;10(1):5438.
55.Ali SA, Justilien V, Jamieson L, Murray NR, Fields AP. Protein Kinase Cι Drives a NOTCH3-dependent Stem-like Phenotype in Mutant KRAS Lung Adenocarcinoma. Cancer Cell. 14 mars 2016;29(3):367-78.

56.Lan Y, Ni W, Tai G. Expression of MUC1 in different tumours and its clinical significance (Review). Mol Clin Oncol. 1 déc 2022;17(6):1-10.
57.Yao Y, Luo J, Sun Q, Xu T, Sun S, Chen M, et al. HOXC13 promotes proliferation of lung adenocarcinoma via modulation of CCND1 and CCNE1. Am J Cancer Res. 1 sept 2017;7(9):1820-34.
58.Liu Y, Shao G, Yang Z, Lin X, Liu X, Qian B, et al. Interferon regulatory factor 6 correlates with the progression of non-small cell lung cancer and can be regulated by miR-320. J Pharm Pharmacol. 27 mars 2021;73(5):682-91.
59.Botti E, Spallone G, Moretti F, Marinari B, Pinetti V, Galanti S, et al. Developmental factor IRF6 exhibits tumor suppressor activity in squamous cell carcinomas. Proc Natl Acad Sci. 16 août 2011;108(33):13710-5.
60.Tan N, Tang J, Chen G, Jiang W, Liu Z. ZEB1-AS1 mediates bone metastasis through targeting miR-320b/BMPR1A axis in lung cancer. Clin Respir J. 2024;18(5):e13770.
61.Perez-Oquendo M, Manshouri R, Tian Y, Fradette JJ, Rodriguez BL, Kundu ST, et al. ZEB1 Is Regulated by K811 Acetylation to Promote Stability, NuRD Complex Interactions, EMT, and NSCLC Metastasis. Mol Cancer Res MCR. 1 août 2023;21(8):779-94.
62.Lehmann W, Mossmann D, Kleemann J, Mock K, Meisinger C, Brummer T, et al. ZEB1 turns into a transcriptional activator by interacting with YAP1 in aggressive cancer types. Nat Commun [Internet]. 2016 [cité 15 août 2024];7. Disponible sur: https://consensus.app/papers/zeb1-turns-activator-interacting-yap1-cancer-types-lehmann/4b789a068a875a84816007871e9a0303/
63.Staaf J, Tran L, Söderlund L, Nodin B, Jirström K, Vidarsdottir H, et al. Diagnostic Value of Insulinoma-Associated Protein 1 (INSM1) and Comparison With Established Neuroendocrine Markers in Pulmonary Cancers. Arch Pathol Lab Med. 1 sept 2020;144(9):1075-85.
64.Chen C, Breslin M, Lan M. Sonic hedgehog signaling pathway promotes INSM1 transcription factor in neuroendocrine lung cancer. Cell Signal. 2018;46:83-91.
65.Castillo SD, Matheu A, Mariani N, Carretero J, Lopez-Rios F, Lovell-Badge R, et al. Novel transcriptional targets of the SRY-HMG box transcription factor SOX4 link its expression to the development of small cell lung cancer. Cancer Res. 1 janv 2012;72(1):176-86.
66.Cai L, Lin S, Girard L, Zhou Y, Yang L, Ci B, et al. LCE: an open web portal to explore gene expression and clinical associations in lung cancer. Oncogene. avr 2019;38(14):2551-64.
67.Heath AP, Ferretti V, Agrawal S, An M, Angelakos JC, Arya R, et al. The NCI Genomic Data Commons. Nat Genet. mars 2021;53(3):257-62.
68.de Bruijn I, Kundra R, Mastrogiacomo B, Tran TN, Sikina L, Mazor T, et al. Analysis and Visualization of Longitudinal Genomic and Clinical Data from the AACR Project GENIE Biopharma Collaborative in cBioPortal. Cancer Res. 1 déc 2023;83(23):3861-7.
69.Gatto L, Killian T. depmap: Cancer Dependency Map Data Package. 2022.
70.Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L, et al. Genenames.org: the HGNC resources in 2023. Nucleic Acids Res. 6 janv 2023;51(D1):D1003-9.
71.Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 8 févr 2018;172(4):650-65.

72.Schmeier S, Alam T, Essack M, Bajic VB. TcoF-DB v2: update of the database of human and mouse transcription co-factors and transcription factor interactions. Nucleic Acids Res. 4 janv 2017;45(D1):D145-50.
73.Shannon P, Richards M. MotifDb: An Annotated Collection of Protein-DNA Binding Sequence Motifs. 2022.
74.Stojnic R, Diez D. PWMEnrich: PWM enrichment analysis. 2022.
75.Kou Y, Chen EY, Clark NR, Duan Q, Tan CM, Ma'ayan A. ChEA2: Gene-Set Libraries from ChIP-X Experiments to Decode the Transcription Regulome. In: Cuzzocrea A, Kittl C, Simos DE, Weippl E, Xu L, éditeurs. Availability, Reliability, and Security in Information Systems and HCI. Berlin, Heidelberg: Springer; 2013. p. 416-30.
76.Slowikowski. tftargets: Human Transcription Factor Targets. 2016.
77.Mora A, Donaldson IM. iRefR: an R package to manipulate the iRefIndex consolidated protein interaction database. BMC Bioinformatics. 24 nov 2011;12:455.
78.Dhifli W, Puig J, Dispot A, Elati M. Latent network-based representations for large-scale gene expression data analysis. BMC Bioinformatics. 4 févr 2019;19(Suppl 13):466.
79.Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. févr 2024;42(2):293-304.
80.Fay, Colin, Vincent Guyader, Sébastien Rochette, and Cervan Girard. Golem: A Framework for Robust Shiny Applications. 2021.
81.C. Sievert. Interactive Web-Based Data Visualization with R, plotly, and shiny. Chapman and Hall/CRC Florida. 2020.
82.Almende B.V. and Contributors, Thieurmel B. visNetwork: Network Visualization using « vis.js » Library. 2022.
83.Sharpe D. Chi-Square Test is Statistically Significant: Now What? Pract Assess Res Eval [Internet]. 2 janv 2015 [cité 12 juill 2024];20(1). Disponible sur: https://openpublishing.library.umass.edu/pare/article/id/1485/
84.Therneau T. A Package for Survival Analysis in R. 2023.
85.Sjoberg D, Baillie M, Fruechtenicht C, Haesendonckx S, Treis T. ggsurvfit: Flexible Time-to-Event Figures [Internet]. Disponible sur: https://www.danieldsjoberg.com/ggsurvfit/
86.Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 20 avr 2015;43(7):e47.
87.Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innov Camb Mass. 28 août 2021;2(3):100141.
88.Morgan M, Davis S. GenomicDataCommons: NIH / NCI Genomic Data Commons Access. [Internet]. Disponible sur: http://github.com/Bioconductor/GenomicDataCommons,
89.Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. Genome Res. nov 2018;28(11):1747-56.
90.Yoshihara K, Shahmoradgoli M, Martínez E, Vegesna R, Kim H, Torres-Garcia W, et al. Inferring tumour purity and stromal and immune cell admixture from expression data. Nat Commun. 11 oct 2013;4(1):2612.
91.Bagaev A, Kotlov N, Nomie K, Svekolkin V, Gafurov A, Isaeva O, et al. Conserved pan-cancer microenvironment subtypes predict response to immunotherapy. Cancer Cell. 14 juin 2021;39(6):845-865.e7.
92.Meyer D, Dimitriadou E, Hornik K, Weingessel A, Leisch F. e1071: Misc Functions of the Department of Statistics, Probability Theory Group. 2022.

Figures legend

Figure 1

Graphical representation (top) of the co-operativity network inferred from the transcriptome of 206 lung cancer-derived cell lines (Lung-CoRegNet). Nodes represent transcription factors and cofactors (TFs/co-TFs). Node size is proportional to the number of targets of the TF/co-TF. Coregulatory interactions between nodes are indicated: protein-protein interactions with published evidence are shown in blue, transcriptional regulation interactions with published evidence are shown in red, and interactions defined solely by the h-LICORN algorithm are shown in gray.

Figure 2

(A) left : UMAP visualization of the metacohort of 5 487 tumors, illustrating the difference between the use of transcriptomic (top) and influence (bottom) data from different datasets. Influence data produce a batch effect-free metacohort from the original RNA expression values; right : UMAP visualization of the Lung-cRegMap metacohort, using sample annotation colors derived from histological classification (top) and cRegMap classification (bottom); (B) Alluvial plot displaying the Lung-cRegMap metacohort samples with histological classifications of lung.

Figure 3

Overall survival probability stratified by cRegMap subtypes according to Kaplan-Meier method. Survival analysis was performed with a Cox regression model on the available clinical data from 4319 patients.;

Figure 4

Oncoplot of main lung cancer mutations stratified by cRegMap subtypes. Mutational data were acquired through maftools R package for 1438 patients. Only significative proportions are displayed and colors are representative of p-values.

Figure 5

(A): Metacohort heatmap for cRegMap subgroups, histology, ESTIMATE analysis (tumor purity, immune and stroma scores), MFP analysis (EMT, CAF, matrix remodeling, proliferation rate signatures and MFP classes) and top 5 differentially influent regulators for each cRegMap clusters. (B): Heatmap of TCGA samples with projection of LUAD and LUSC subtypes, ESTIMATE and MFP analysis results and mutations projection.
(C): Ontologies enrichment analysis of cRegMap classes performed on up and down DIRs' target genes.

Figure 6

Impact on cell viability (CERES dependency score) for each of the predicted subclasses. For each regulator, two values are shown vertically: the colored dot is the mean CERES score for the cell lines assigned to the subclass,

and the gray dot is the mean CERES score for cell lines not assigned to the subclass. A gene may have a higher CERES score (indicating that its knockout would reduce the survival of the corresponding cell line) or a lower score (indicating that its deletion might increase the survival of the cell line

Figure 7

Summary of the main characteristics of the Lung-cRegMap classes. From top to bottom, the following information is displayed: Class abbreviation and name; Main histology; Overall survival; Tumour purity; Stroma influence; Smokers proportion; Matrix influence; EMT signature; Angiogenesis; Proliferation signature; Overall signature; Characteristic mutations; Key regulators; Representative cell lines.

Figure 1

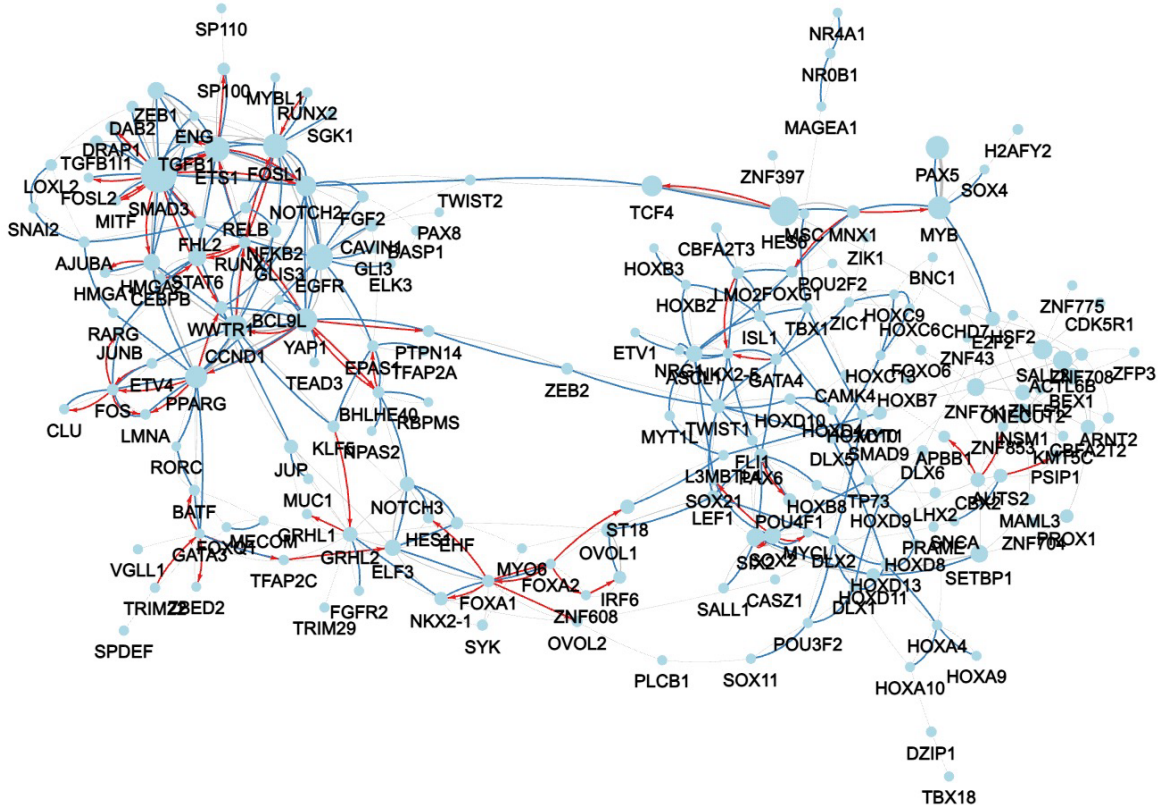
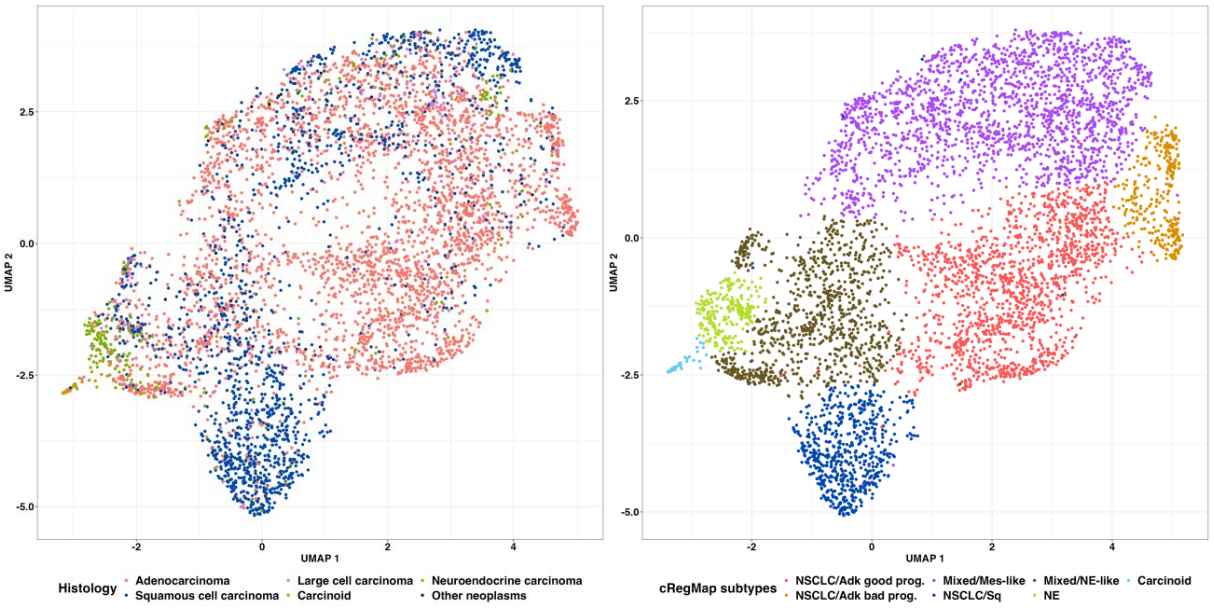

SP110
SP100
MYBL1
RUNX2
ZEB1
DAB2
ENG
SGK1
DRAP1
TGFB1I1
TGFB1
ETS1
FOSL1
LOXL2
FOSL2
SMAD3
NOTCH2
TWIST2
SNAI2
MITF
RELB
FGF2
PAX8
FHL2
NFKB2
CAVIN1
BASP1
AJUBA
RUNX1
GLI3
GLIS3
STAT6
EGFR
ELK3
HMGA1
CEBPB
BCL9L
RARG
WWTR1
YAP1
PTPN14
JUNB
CCND1
EPAS1
TFAP2A
ETV4
TEAD3
PPARG
FOS
BHLHE40
RBPMS
CLU
LMNA
KLF6
RPAS2
RORC
JUP
MUC1
NOTCH3
BATF
GRHL1
MECOM
FOXQ1
HES1
EHF
GATA3
GRHL2
ELF3
VGLL1
TFAP2C
MYO6
FOXA2
TRIM22
BED2
FGFR2
FOXA1
IRF6
TRIM29
NKX2-1
ZNF608
SPDEF
SYK
OVOL2
NR4A1
NR0B1
MAGEA1
H2AFY2
ZNF397
PAX5
SOX4
TCF4
MSC
MNX1
MYB
HES6
CBFA2T3
ZIK1
HOXB3
BNC1
POU2F2
HOXB2
LMO2
FOXG1
ZNF775
CDK5R1
HOXC9
ZIC1
TBX1
ISL1
CHD7
HSF2
E2F2
ETV1
ZNF43
NRG1
ASCL1
NKX2-5
GATA4
FOXO6
HOXC13
ZFP3
ZEB2
CAMK4
HOXB7
ACTL6B
REX1
MYT1L
TWIST1
HOXD10
INSM1
ARNT2
SMAD9
APBB1
DLX5
DLX6
KMT5C
PSIP1
FLI1
L3MBTL1
PAX6
SOX21
HOXB8
TP73
CBX2
AUTS2
LEF1
POU4F1
HOXD9
LHX2
MAML3
ST18
OVOL1
SOX2
MYCL
PRAME
GNG4
ZNF704
PROX1
DLX2
HOXD8
SETBP1
HOXD13
CASZ1
SALL1
HOXD11
DLX1
POU3F2
HOXA4
PLCB1
SOX11
HOXA10
HOXA9
DZIP1
TBX18

Figure 2

A


UMAP 2
UMAP 1
2.5
0.0
-2.5
-5.0
-2
0
2
4
Histology
Adenocarcinoma
Squamous cell carcinoma
Large cell carcinoma
Carcinoid
Neuroendocrine carcinoma
Other neoplasms
cRegMap subtypes
NSCLC/Adk good prog.
NSCLC/Adk bad prog.
Mixed/Mes-like
NSCLC/Sq
Mixed/NE-like
NE
Carcinoid

B

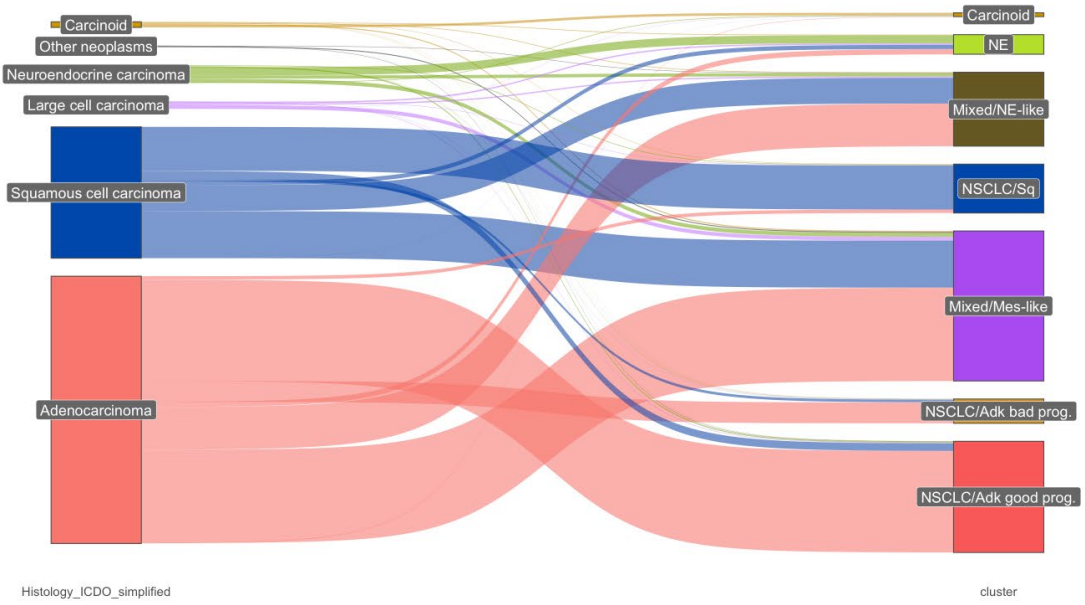

Carcinoid
Other neoplasms
Neuroendocrine carcinoma
Large cell carcinoma
Squamous cell carcinoma
Adenocarcinoma
Carcinoid
NE
Mixed/NE-like
NSCLC/Sq
Mixed/Mes-like
NSCLC/Adk bad prog.
NSCLC/Adk good prog.
Histology_ICDO_simplified
cluster

Figure 3

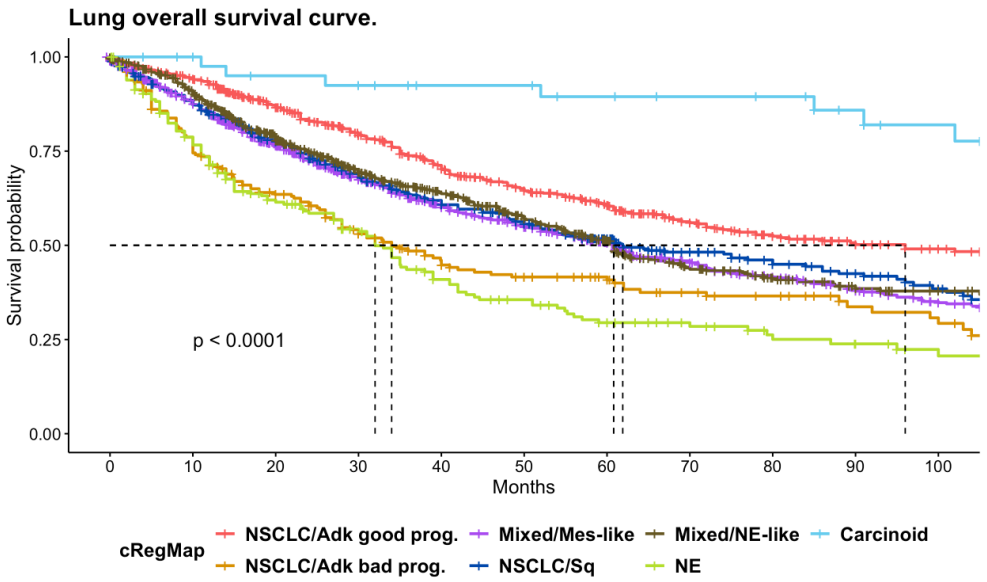

Lung overall survival curve.
Survival probability
1.00
0.75
0.50
0.25
0.00
p < 0.0001
0
10
20
30
40
50
60
70
80
90
100
Months
cRegMap
NSCLC/Adk good prog.
NSCLC/Adk bad prog.
Mixed/Mes-like
NSCLC/Sq
Mixed/NE-like
NE
Carcinoid

Figure 4

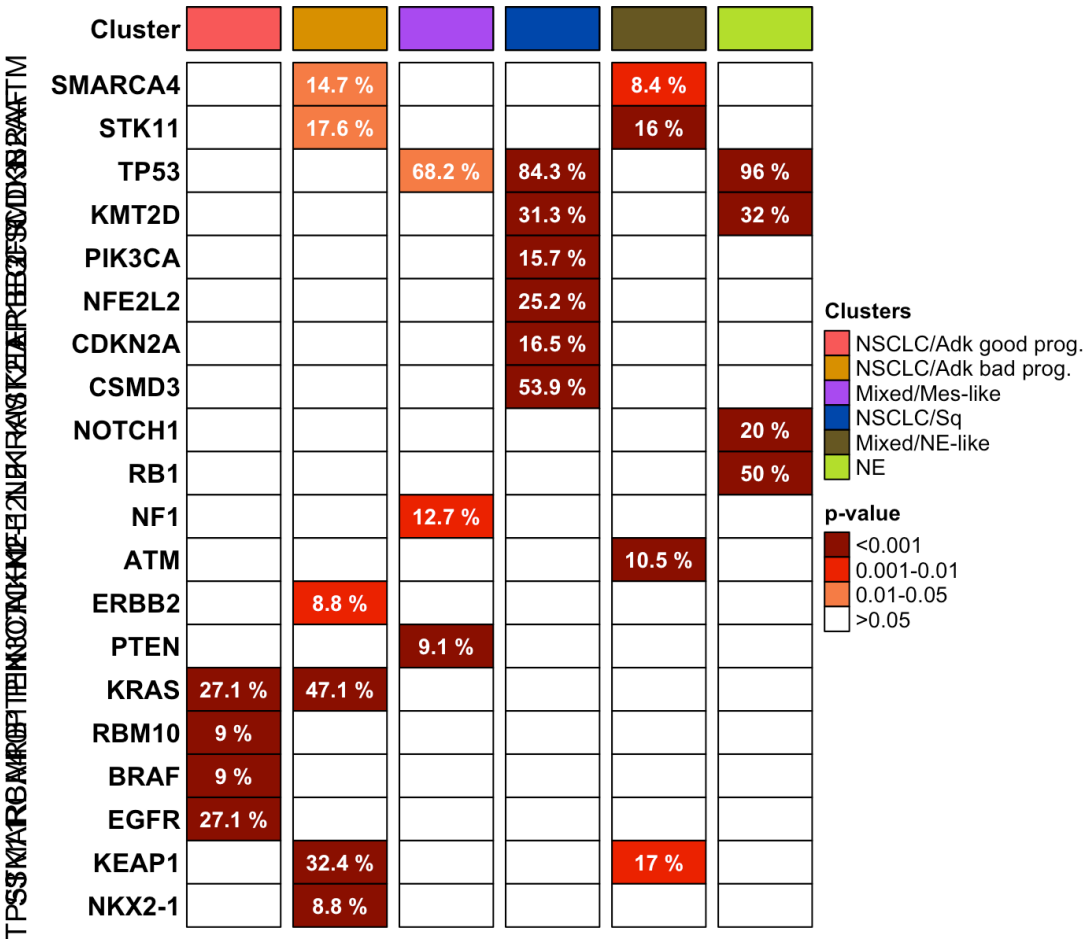

Cluster
SMARCA4
STK11
TP53
KMT2D
PIK3CA
NFE2L2
CDKN2A
CSMD3
NOTCH1
RB1
NF1
ATM
ERBB2
PTEN
KRAS
RBM10
BRAF
EGFR
KEAP1
NKX2-1
14.7 %
17.6 %
68.2 %
84.3 %
31.3 %
15.7 %
25.2 %
16.5 %
53.9 %
12.7 %
8.8 %
9.1 %
27.1 %
47.1 %
9 %
9 %
27.1 %
32.4 %
8.8 %
8.4 %
16 %
10.5 %
17 %
96 %
32 %
20 %
50 %
Clusters
NSCLC/Adk good prog.
NSCLC/Adk bad prog.
Mixed/Mes-like
NSCLC/Sq
Mixed/NE-like
NE
p-value
<0.001
0.001-0.01
0.01-0.05
>0.05

Figure 5

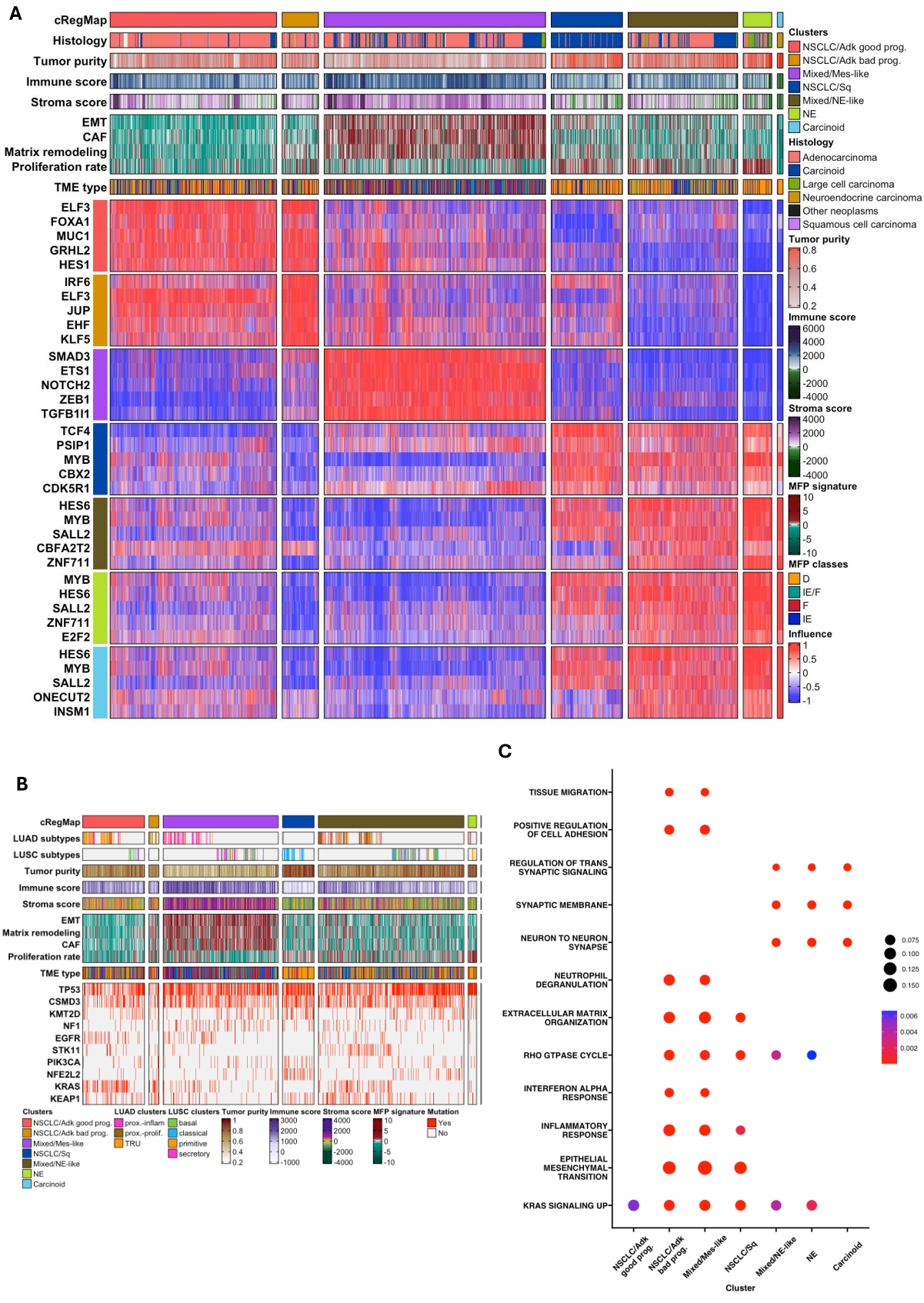

A
cRegMap
Histology
Tumor purity
Immune score
Stroma score
EMT
CAF
Matrix remodeling
Proliferation rate
TME type
ELF3
FOXA1
MUC1
GRHL2
HES1
IRF6
ELF3
JUP
EHF
KLF5
SMAD3
ETS1
NOTCH2
ZEB1
TGFB1I1
TCF4
PSIP1
MYB
CBX2
CDK5R1
HES6
MYB
SALL2
CBFA2T2
ZNF711
MYB
HES6
SALL2
ZNF711
E2F2
HES6
MYB
SALL2
ONECUT2
INSM1
Clusters
NSCLC/Adk good prog.
NSCLC/Adk bad prog.
Mixed/Mes-like
NSCLC/Sq
Mixed/NE-like
NE
Carcinoid
Histology
Adenocarcinoma
Carcinoid
Large cell carcinoma
Neuroendocrine carcinoma
Other neoplasms
Squamous cell carcinoma
Tumor purity
0.8
0.6
0.4
0.2
Immune score
6000
4000
2000
0
-2000
-4000
Stroma score
4000
2000
0
-2000
-4000
MFP signature
10
5
0
-5
-10
MFP classes
D
IE/F
F
IE
Influence
1
0.5
0
-0.5
-1
B
cRegMap
LUAD subtypes
LUSC subtypes
Tumor purity
Immune score
Stroma score
EMT
Matrix remodeling
CAF
Proliferation rate
TME type
TP53
CSMD3
KMT2D
NF1
EGFR
STK11
PIK3CA
NFE2L2
KRAS
KEAP1
Clusters
NSCLC/Adk good prog.
NSCLC/Adk bad prog.
Mixed/Mes-like
NSCLC/Sq
Mixed/NE-like
NE
Carcinoid
LUAD clusters
prox.-inflam
prox.-prolif.
TRU
LUSC clusters
basal
classical
primitive
secretory
Tumor purity
1
0.8
0.6
0.4
0.2
Immune score
3000
2000
1000
0
-1000
Stroma score
4000
2000
0
-2000
-4000
MFP signature
10
5
0
-5
-10
Mutation
Yes
No
C
TISSUE MIGRATION
POSITIVE REGULATION OF CELL ADHESION
REGULATION OF TRANS SYNAPTIC SIGNALING
SYNAPTIC MEMBRANE
NEURON TO NEURON SYNAPSE
NEUTROPHIL DEGRANULATION
EXTRACELLULAR MATRIX ORGANIZATION
RHO GTPASE CYCLE
INTERFERON ALPHA RESPONSE
INFLAMMATORY RESPONSE
EPITHELIAL MESENCHYMAL TRANSITION
KRAS SIGNALING UP
0.075
0.100
0.125
0.150
0.006
0.004
0.002
NSCLC/Adk good prog.
NSCLC/Adk bad prog.
Mixed/Mes-like
NSCLC/Sq
Mixed/NE-like
NE
Carcinoid
Cluster

Figure 6

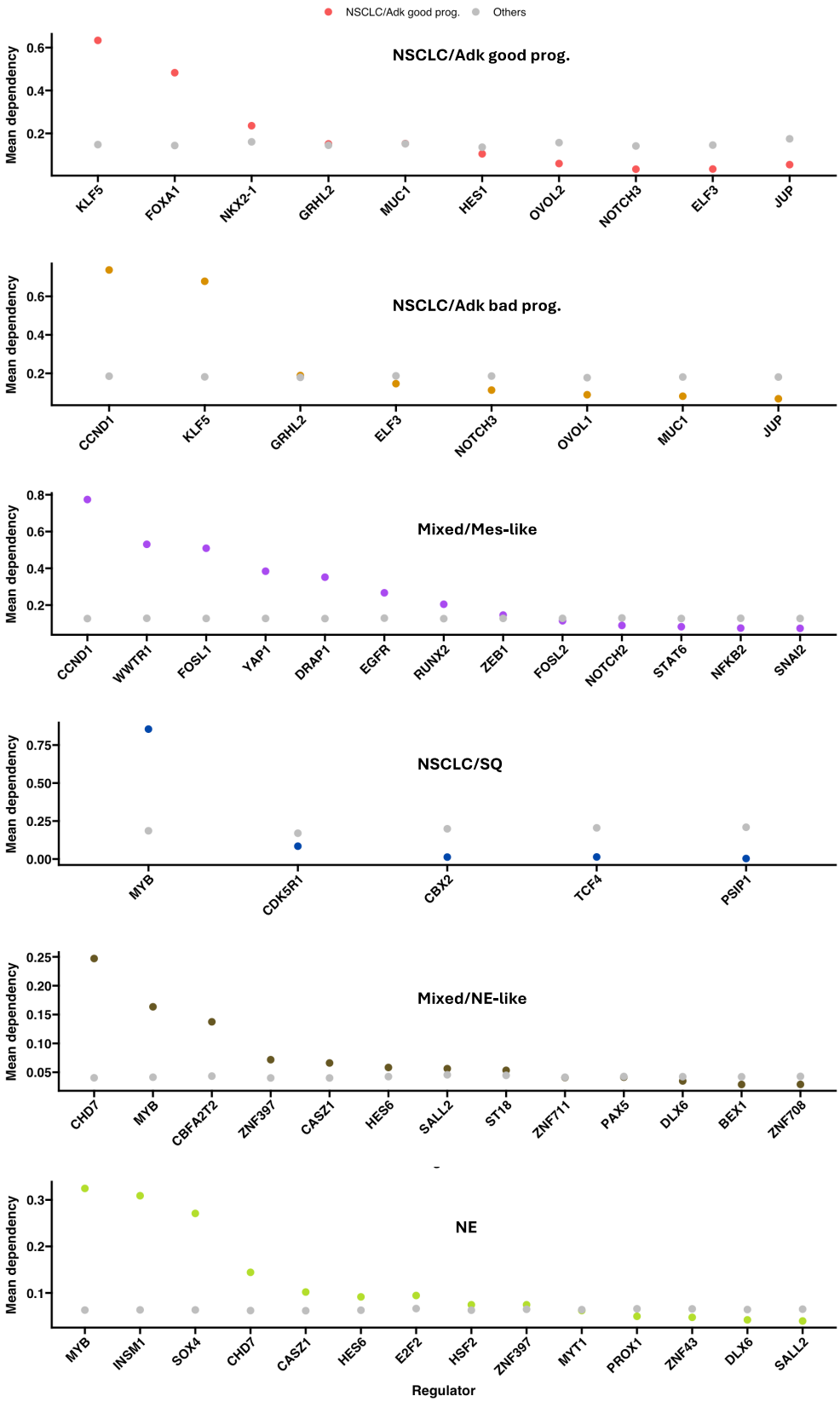

NSCLC/Adk good prog.
Others
NSCLC/Adk good prog.
Mean dependency
0.6
0.4
0.2
KLF5
FOXA1
NKX2-1
GRHL2
MUC1
HES1
OVOL2
NOTCH3
ELF3
JUP
NSCLC/Adk bad prog.
Mean dependency
0.6
0.4
0.2
CCND1
KLF5
GRHL2
ELF3
NOTCH3
OVOL1
MUC1
JUP
Mixed/Mes-like
Mean dependency
0.8
0.6
0.4
0.2
CCND1
WWTR1
FOSL1
YAP1
DRAP1
EGFR
RUNX2
ZEB1
FOSL2
NOTCH2
STAT6
NFKB2
SNAI2
NSCLC/SQ
Mean dependency
0.75
0.50
0.25
0.00
MYB
CDK5R1
CBX2
TCF4
PSIP1
Mixed/NE-like
Mean dependency
0.25
0.20
0.15
0.10
0.05
CHD7
MYB
CBFA2T2
ZNF397
CASZ1
HES6
SALL2
ST18
ZNF711
PAX5
DLX6
BEX1
ZNF708
NE
Mean dependency
0.3
0.2
0.1
MYB
INSM1
SOX4
CHD7
CASZ1
HES6
E2F2
HSF2
ZNF397
MYT1
PROX1
ZNF43
DLX6
SALL2
Regulator

Figure 6

	NSCLC/Adk	NSCLC/Adk bad prog.	Mixed/Mes-like	NSCLC/Sq	Mixed/NE-like	NE	Carcinoid
Histology	Adenocarcninoma	Adenocarcninoma	Adenocarcninoma / Squamous cell carcinoma	Squamous cell carcinoma	Adenocarcninoma / Squamous cell carcinoma	Neuroendocrine	Carcinoid
Survival (month [95%CI])	96 [76-112]	34 [28-46]	68.8 [55-64]	61.9 [52-86]	60.8 [54.7-68]	32 [27-39]	> 179
Tumor purity	+	+	- -	++	++	++	+++
Stroma score	+	+	+++	- -	- -	- -	- - -
Smokers proportion (% [95%CI])	64 [60.9-67]	74.7 [67.7-80.7]	80 [77.6-82.1]	89.6 [85.5-92.7]	84.5 [81.6-87.1]	80.9 [72.9-87]	33.3 [13-61.3]
Matrix signature	- - -	++	+++	+	+	- - -	- - -
EMT signature	- - -	-	+++	+	+	+	+
Angiogenesis signature	-	++	+++	- -	-	- -	- -
Proliferation signature	- -	+	-	+++	++	+++	- - -
Signature	EMT, TNFα, IFNγ	EMT, TNFα, IFNγ, KRAS	EMT, TNFα, IFNγ, KRAS, Cell cycle	Cell cycle, proliferation	Cell cycle, proliferation	Cell cycle, proliferation	Cell cycle
Mutations	KRAS+ (27%) EGFR+ (27%)	LRP1B- (9%) KRAS+ (27%) EGFR+ (27%) CTNND2- (2%) KEAP1+ (39%) STK11+ (18%) SMARCA4+ (15%)	/	CSMD3+ (54%) KMT2D+ (31%) KEAP1+ (18%) NFE2L2+ (25%)	KEAP1+ (17%) STK11+ (11%)	RB1+ (50%)	/
Key regulators	E2F2 ONECUT2 MYB	ELF3 NOTCH3 KLF5 CCND1 IRF6 EHF MUC1 OVOL1	WWTR1 ZEB1 SMAD3 TGFB1I1 TGFB1 EGFR LOXL2 YAP1	TCF4 DLX6 CBX2 CDK5R1 IRF6 HES6	HOXD9 MYT1L FOXA2 FLI1 FGFR2 FOXG1 AEBP1 ASCL1	HOXD13 SOX4 SOX11 HOXC10 SYK DLX2	ZNF439 ZNF677 RORC IRAK2 SOX21
Cell lines	NCIH3122 ABC1	NCIH1944 MORCPR NCIH1437	NCIH2286 NCIH838 SW1573 NCIH1915 EPLC272H NCIH650 NCIH1793 CALU6 LU99 HCC2450 NCIH2030 NCIH1703 NCIH460 NCIH1693 NCIH2882	LK2 HARA HCC95	NCIH1581 NCIH522 NCIH841 HCC2429 A427	NCIH1048 NCIH211 CORL311 LU135 SHP77 SCLC22H NCIH446 NCIH526 NCIH510 NCIH82 CORL279	/

3. Transcriptional-program-based deciphering of the MET exon 14 skipping regulation network

Le pronostic des patients atteints de cancer du poumon reste généralement mauvais, avec un taux de survie à cinq ans d'environ 20 % (379,380). Ce faible taux de survie s'explique en partie par un diagnostic à un stade avancé, car de nombreux patients sont diagnostiqués lorsque la maladie est déjà très évoluée (381). Bien que les avancées récentes, notamment en matière de thérapies ciblées et d'immunothérapie, aient amélioré les résultats pour certains patients, la majorité continue de faire face à des défis liés à la résistance aux traitements et aux récidives (382,383). Par exemple, la présence de mutations génétiques spécifiques, telles que celles des gènes EGFR et KRAS, peut fortement influencer l'efficacité des traitements et le pronostic des patients (384,385).

Le gène MET, qui code pour le récepteur du facteur de croissance des hépatocytes (MET), joue un rôle clé dans la pathogenèse du cancer du poumon, en particulier du cancer du poumon non à petites cellules. L'implication de MET dans le cancer du poumon est multifactorielle et englobe divers mécanismes, notamment les mutations, les amplifications et les interactions avec d'autres voies de signalisation. Ces altérations de MET contribuent à la croissance tumorale, à la prolifération cellulaire et à la survie des cellules cancéreuses, tout en facilitant des processus tels que l'invasion, la migration et la formation de métastases. L'interaction entre MET et d'autres récepteurs à activité tyrosine kinase, en particulier le récepteur du facteur de croissance épidermique, constitue un aspect essentiel de son implication dans le cancer du poumon. Les recherches montrent que MET peut être activé par la signalisation de l'EGFR, ce qui augmente le potentiel invasif des cellules de NSCLC (149,386). Ce dialogue croisé est médié par plusieurs voies intracellulaires, notamment la voie MAPK, qui favorise davantage la progression tumorale et les métastases (149). Cette activation de MET en réponse à la signalisation de l'EGFR joue un rôle crucial dans la résistance aux thérapies ciblant l'EGFR. Cela met en évidence l'importance de MET en tant que cible thérapeutique dans le NSCLC, car son inhibition pourrait potentiellement surmonter la résistance observée chez certains patients traités avec des inhibiteurs de l'EGFR (202,387,388). Le ciblage simultané de ces voies de signalisation pourrait offrir de nouvelles opportunités thérapeutiques pour améliorer l'efficacité des traitements dans les cas résistants.

L'une des altérations les plus critiques associées à MET dans le cancer du poumon est la mutation d'épissage de l'exon 14. Cette mutation entraîne la production d'une protéine MET tronquée, dépourvue d'un domaine régulateur essentiel, ce qui conduit à une activation constitutive du récepteur. Des études ont montré que les altérations de l'exon 14 de MET surviennent dans environ 3 % des adénocarcinomes pulmonaires et sont associées à des stades avancés de la maladie (389,390). La présence de cette mutation est liée à une classe moléculaire unique de NSCLC, caractérisée par un comportement agressif et un pronostic défavorable (248,390,391).

Nous avons collaboré avec l'équipe du Dr Tulasne, qui a développé une expertise dans l'étude du récepteur MET (200,392,393), afin, d'une part, d'évaluer la capacité du GRN du cancer du poumon à rendre

compte des perturbations biologiques induites durant la phase expérimentale. Et, d'autre part, de disséquer la réponse biologique induite par le récepteur METex14Del exposé à son ligand, le HGF.

J'ai exploité le GRN du poumon comme colonne vertébrale de l'étude, en opérant le passage rigoureux des profils d'expression vers des scores d'influence à l'échelle des échantillons, puis en projetant ces influences sur le réseau afin de visualiser l'activation des modules de regulateurs. Cette stratégie m'a permis d'identifier les régulateurs clés impliqués dans la cascade de réponse du récepteur METex14del suite à la liaison à son ligand et d'intégrer des données de transcriptomique issues d'essais d'inhibition pharmacologique pour tester in silico la cohérence des signatures observées. En combinant ces projections d'influence avec les expériences biologiques menées par l'équipe du Dr Tulasne, nous avons validé la robustesse du modèle réseau et obtenu une vision intégrée reliant la signalisation en amont, l'activité régulatrice en aval et les réponses cellulaires mesurables.

Title: Comprehensive map of the regulatory network triggered by MET exon 14 skipping reveals important involvement of the RAS-ERK signaling pathway

Running Title: Regulatory network of MET exon 14 skipping

Authors: Marie-José Truong[1], Geoffrey Pawlak[2], Jean-Pascal Meneboo[3], Shéhérazade Sebda[3], Marie Fernandes[1], Martin Figeac[3], Mohamed Elati[2] and David Tulasne[1*]

Affiliations:

[1] Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, UMR9020 – UMR1277 - Canther - Cancer Heterogeneity, Plasticity and Resistance to Therapies, Target team, Lille, France

[2] Univ. Lille, CNRS, Inserm, CHU Lille, UMR9020 – UMR1277 - Canther - Cancer Heterogeneity, Plasticity and Resistance to Therapies, Disco team, Lille, France

[3] Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, US 41 - UAR 2014 - PLBS, Lille, France

[*]Corresponding author. Email: david.tulasne@inserm.fr

Abstract

The MET exon 14 skipping mutation (named METex14Del) described in lung cancer leads to prolonged activation of signaling pathways and aberrant cell responses, but the link between HGF signaling and cell responses remains unclear. A putative lung cancer regulatory network of influential transcription factors was constructed from the transcriptomes of lung cancer cell lines. Transcriptomic data from METex14Del-expressing cells, stimulated or not by HGF, were mapped onto this lung cancer reference network and revealed activation of a major regulatory node composed mainly by the highly influential transcription factors ETS1, FOSL1 and SMAD3. HGF activation of METex14Del receptor induced the expression and phosphorylation of these three master regulators and the expression of their predicted target genes involved in migration and invasion. All these molecular and biological effects were inhibited by trametinib, a MEK inhibitor, which was potentiated by combination with capmatinib, a MET inhibitor. New mapping with transcriptomic data from trametinib-treated METex14Del cells validated the key role of the RAS-ERK pathway signaling in the activation of ETS1, FOSL1 and SMAD3 regulators and the induction of their target genes in HGF-activated METex14Del receptor. Thus, we report an original and powerful strategy to uncover key regulators, including transcription factors that have not been widely described in METex14Del signaling, such as SMAD3. These factors are activated by specific signaling pathways and could provide a novel therapeutic strategy involving a combination of receptor and signaling inhibitors.

INTRODUCTION

Receptor tyrosine kinases (RTKs) provide an interface between the extracellular and intracellular environments. Once activated by an extracellular growth factor, RTKs can activate intracellular signaling pathways that relay information within the cell, triggering various cell responses such as proliferation, differentiation, migration/invasion, and changes in the cell death/survival balance [1]. Signaling pathways transmit information to various cell compartments, including the nucleus, where regulation of gene expression is one of the most important mechanisms influencing cell responses [2]. The transcriptional program, which can be deciphered by transcriptomic analyses, provides important information about biological responses, as gene expression profiles can be linked to key biological processes. However, how signaling pathways are integrated at the transcriptional level remains elusive.

In some cancers, RTK activation by genomic alterations can lead to oncogene addiction, in which cancer cell growth and/or survival is dependent on a single oncogene [3]. In this situation, targeted therapies against RTKs, mainly tyrosine kinase inhibitors (TKIs), are effective and have profoundly improved patient care [4]. In addicted cancer cells, overactive RTKs induce aberrant signaling leading to a deregulated transcriptional program. Thus, understanding oncogene addiction requires a deep understanding of signaling pathway integration at the transcriptional level. In addition, the efficacy of targeted therapies against RTKs is limited by primary or acquired resistance, often supported by activation of alternative signaling pathways that can bypass the targeted inhibition [5]. In this situation, understanding signaling pathway integration at the transcriptional level is also an important issue.

MET, a member of the RTK family, is an emerging target in cancer. In lung cancer, the *MET* gene has been found to be mutated in approximately 3% of patients, with point mutations, deletions or insertions affecting the splice sites of exon 14. Such mutations result in in-frame skipping of this exon (METex14Del) [6,7] and thus deletion of the juxtamembrane regulatory domain of the receptor [8,9].

Experiments using ectopic expression of METex14Del or its reconstitution by genome editing in normal lung epithelial cells and cancer cell lines have shown that activated METex14Del induces enhanced and sustained activation of downstream signaling pathways, including the RAS-ERK and PI3K-AKT pathways [10–13]. In normal epithelial cells, METex14Del activation depends on stimulation by HGF, its high-affinity ligand, and the mutant receptor can promote experimental tumor growth only in humanized mice expressing human HGF [14]. Thus, in contrast to most RTK-activating mutations described to date, which result in ligand-independent RTK activation, MET exon 14 skipping does not eliminate MET

receptor dependence on ligand stimulation. This suggests the existence of different activation states depending on ligand availability.

The juxtamembrane domain of MET contains several negative regulatory sites: (i) phosphorylated serine 985, involved in downregulation of MET tyrosine kinase activity [15,16], (ii) phosphorylated tyrosine 1003, involved in recruitment of the E3 ubiquitin ligase CBL and thus important for MET degradation [17–19], (iii) aspartic acid 1002, a caspase cleavage site involved in regulation of the death/survival balance [20–23]. Downregulation of MET signaling appears to be mediated primarily by the CBL binding site [10].

Activated METex14Del can also induce a much more complex transcriptional program than its WT counterpart. Gene ontology analyses are consistent with the migration and invasion responses induced by METex14Del [14]. However, this approach based on mRNA levels does not take into account the complexity of gene regulation, which involves the coordinated action of multiple transcription factors and co-factors (TF/co-TFs) and the influence of signaling pathways on their regulation.

The recently developed software solution CoRegNet [24], which can be visualized with the interactive tool Cytoscape Widget [25], is based on expression correlations between known cooperative TFs/co-TFs and target genes extracted from transcriptomic data and on described interactions between these regulators. CoRegNet has already proven its ability in complex organisms to identify 12 microglia-specific transcriptional regulators [26], 10 key regulators driving the transition in non-alcoholic liver disease [27], and five master regulators specific to rheumatoid arthritis synovial fibroblasts [28].

Here we constructed a lung cancer-specific regulatory network based on the aggregation of extensive transcriptomic data from lung cancer cell lines. This network was then used to dynamically model the network of METex14Del, stimulated or not by HGF. A “dialogue” between biological experiments and modeling confirmed the robustness of the proposed network and allowed an integrative view of METex14Del signaling from its downstream signaling pathways to gene expression, which in turn regulates cell responses.

RESULTS

Transcriptome-based modeling of the putative regulatory network of HGF-stimulated METex14Del

The Cancer Cell Line Encyclopedia provided transcriptomic data for 206 lung cancer cell lines, including 135 NSCLC cell lines, 50 SCLC cell lines and 21 other cell lines. Of the 206 cell lines, 105 were derived from primary tumors and 101 from metastatic tumors. Using these gene

expression data and the CoRegNet package **[24]**, we constructed a lung cancer regulatory network and then enriched it with TF binding sites, ChIP-seq data and protein interactions found in the CoRegNet-embedded databases. The lung cancer-specific regulatory network consists of 502 key regulators that potentially regulate 4 143 target genes which are connected by 18 053 regulator-target gene regulatory interactions. The key regulators are connected by 2 014 described protein-protein interactions ($p < 1e-130$) and 61 272 transcription factor binding sites on their target genes ($p < 1e-300$). CoRegNet can identify the most influential regulators based on their influence score (Supplementary Table 1) and co-regulatory pairs in our network that share common target genes (Supplementary Table 2).

To identify which of the TFs have a strong influence induced by HGF binding to the MET receptor, normalized expression levels of differentially expressed genes (DEGs) from transcriptomic datasets of 16HBE cells expressing MET WT (Supplementary Fig. S1A, B) or 16HBE METex14Del (Fig. 1A and Supplementary Fig. S1C, D) in the presence and absence of HGF **[11]**, were mapped onto this lung cancer regulatory network. HGF-stimulated 16HBE cells expressing METex14Del reveal a large node of influential regulators formed by six positively interacting regulators: ETS1, FOSL1, SMAD3, HMGA2, CCND1 and RUNX1 with a high number of target genes (Fig. 1A and Supplementary Fig. S1D). ETS1, FOSL1 and SMAD3 interact with each other and binding sites were found in ETS1 and FOSL1 for SMAD3 and in SMAD3 for ETS1. The representation by the mean difference in influence (Fig. 1B) and the high value of the mean group influence (Supplementary Table 1) confirm the activation state of these six key regulators. Looking at the mRNA levels for these six regulators in the transcriptomic data, only FOSL1 and HMGA2 appeared to be significantly up-regulated under HGF stimulation (Fig. 1C), whereas by quantitative RT-PCR all six regulators showed significantly increased mRNA expression in the presence of HGF, with a stronger up-regulation for FOSL1 and HMGA2 (Fig. 1D).

Evidence for a regulatory relationship between ETS1, FOSL1 and SMAD3 and their predicted target genes

We then focused on the three most influential TFs: ETS1, FOSL1 and SMAD3. The transcriptional regulatory network of 16HBE METex14Del constructed with the microarray datasets allowed visualizing the putative influence of ETS1, FOSL1 and SMAD3 on their respective or common target genes (Fig. 2A). The 50 most regulated targets are shown in a heatmap (Fig. 2B), and the complete list of regulated target genes by specific or combined TFs was compiled in Supplementary Table 2. The predicted differential expression by HGF

activation was confirmed at the mRNA expression level for ten HGF up-regulated (VIM, NOG, SERPINE2, SERPINA1, SH2D5, PTX3, CHGB, LETM2, DCBLD2, ABCA1) and two HGF down-regulated (IRS1, SLC15A3) target genes (Fig. 3A) in 16HBE METex14Del cell line. Similar results were obtained for ten up-regulated (ADMA19, KRTAP2-3, G0S2, NGEF, COL13A1, THBD, TSPAN5, ARHGAP22, CAVIN3, RAC2) and two down-regulated (CAV1, CLDN11) additional target genes (Supplementary Fig. S2).

To assess the regulatory relationships between ETS1, FOSL1 and SMAD3 and their targets, these TFs were silenced by RNA interference, separately or in combination, and the effect on expression of 12 original target genes was evaluated. The efficacy of each siRNA pair and the pooled one was confirmed at the mRNA level (Fig. 3B) and at the protein level (Fig. 3C), under stimulation or not by HGF in 16HBE METex14Del cell line. Among the HGF upregulated target genes, VIM, NOG, SERPINE2, SERPINA1, PTX3, CHGB, ABCA1 and LETM2 showed reduced expression in the presence of the three silencers, alone or in combination (Fig. 3D). The effects of ETS1 silencing on PTX3 expression and FOSL1 silencing on LETM2 expression were particularly pronounced. Reduced expression of SH2D5 was observed with either ETS1 or FOSL1 silencing alone and with the three silencers in combination. Reduced expression of DCBLD2 was observed with either the FOSL1 or SMAD3 silencer used alone and with the three silencers combined. Of the two downregulated genes, IRS1 and SLC15A3 showed restored expression in the presence of the three silencers. SLC15A3 expression was strongly increased in the presence of the FOSL1 silencer alone (Fig. 3D).

Silencing of ETS1, FOSL1 and SMAD3 slightly affects wound healing and scattering

Knockdown of ETS1, FOSL1 or SMAD3 alone in 16HBE METex14Del cell line did not affect wound healing, but combined knockdown slightly reduced both basal and HGF-stimulated responses (Fig. 4A). Similarly, in the presence of Matrigel, combined knockdown of all three TFs reduced HGF-induced wound invasion (Fig. 4B). HGF did not increase cell proliferation, but ETS1 knockdown enhanced this process both in the presence and absence of HGF (Fig. 4C). In addition, ETS1 knockdown appeared to increase HGF-induced scattering, whereas FOSL1 and SMAD3 knockdown decreased it (Fig. 4D). These results show that combined knockdown of the three TFs slightly reduces HGF-induced migration, invasion and scattering of 16HBE METex14Del cell line.

The RAS-ERK signaling pathway controls ETS1, FOSL1 and SMAD3 phosphorylation and cellular wound healing

Since ETS1, FOSL1 and SMAD3 silencing did not fully inhibit known biological functions induced by HGF, we investigated upstream regulators to understand how these 3 TFs might act together in the regulatory node. The heatmap of the top 50 differentially expressed target genes of 16HBE MET WT vs. 16HBE METex14Del upon HGF stimulation (Fig. 5A) and the complete gene ontology (GO) enrichment analysis revealed a significant overrepresentation of annotations such as epithelial-mesenchymal transition (EMT), KRAS signaling up or positive regulation of locomotion (Fig. 5B and Supplementary Table 3). Given our previous study showing HGF/MET signaling induces activation of ETS1 through ERK-dependent phosphorylation of its threonine 38 (Thr^{38}) [35,36], we examined the induction of ETS1, FOSL1 and SMAD3 phosphorylation upon stimulation by HGF (Fig. 5C, D, E). As expected, ETS1 was found to be phosphorylated at Thr^{38} in 16HBE METex14delMETex14Del cells, an effect detected up to 8 h after stimulation. No phosphorylation of this TF was detected in their WT counterparts (Fig. 5C). Similar sustained ETS1 phosphorylation was observed in ZORG cells (Fig. 5C) and H596 cells (Supplementary Fig. S3A) derived from NSCLC patients harboring the METex14Del variant. A slight increase in ETS1 expression was observed in METex14Del cells 8 hours after HGF stimulation (Fig. 5C). A strong induction of FOSL1 expression and phosphorylation at Ser^{265} was found in our three cell models peaking at 8 h post-simulation and persisting until 24 h (Fig. 5D and Supplementary Fig. S3B). SMAD3 was phosphorylated at Ser^{208} in our three cell models with a maximum phosphorylation at 30 min post-stimulation (Fig. 5E and Supplementary Fig. S3C). The expression of ETS1, FOSL1, SMAD3 mRNA was also assessed in a similar HGF time course stimulation. In 16HBE METex14Del, ZORG and H596 cells, HGF stimulation induced weak ETS1 mRNA expression, which was only detected at 8 h post-stimulation. FOSL1 mRNA expression increased from 30 min post-stimulation, reaching a maximum expression at 8 h and a decrease at 24 h in all cell lines. Expression was stronger in 16HBE METex14Del cells compared to its WT counterpart. SMAD3 mRNA expression remained unchanged during the course of HGF stimulation in any of the cell lines (Supplementary Fig. S3D, E, F, G). Taken together, the mRNA expression of ETS1, FOSL1 and SMAD3 is consistent with their protein expression.

To confirm involvement of the RAS-ERK signaling, trametinib was used to specifically inhibit MEK. The addition of trametinib completely reversed the HGF-induced phosphorylation of ETS1 at Thr^{38}, FOSL1 at Ser^{265} and SMAD3 at Ser^{208} in 16HBE METex14Del and ZORG cells (Fig. 6A, B, C). U0126, another MEK inhibitor, was also effective in the 2 cell models (Supplementary Fig. S4A, B, C). The effect of trametinib on the biological responses of 16HBE METex14Del cells was also investigated. Trametinib was found to inhibit HGF-induced wound

healing both in the absence (Fig. 6D) and presence of Matrigel (Fig. 6E), although less effectively than the MET TKI capmatinib used as a control. Co-treatment with both capmatinib and trametinib resulted in more potent inhibition (Fig. 6D, E, F). Similar results were observed in cell scattering assays (Fig. 6F). Cell proliferation was only slightly affected by the co-treatment (Fig. 6G). Wound healing experiments in presence or absence of Matrigel have been also performed in both ZORG and H596 lung cancer cells and confirmed the inhibitory effect of the capmatinib and trametinib (Supplementary Fig. S5A, B, C, D).

To determine whether another oncogene driver activates the regulatory node composed of ETS1, FOSL1 and SMAD3, four cell lines harboring an activating mutation in the EGFR gene were treated with Osimertinib (a third generation EGFR TKI) or trametinib. As expected, the four cell lines displayed constitutive EGFR and ERK phosphorylation, which was inhibited by osimertinib. Trametinib inhibited ERK phosphorylation but not EGFR (Supplementary Fig. S6A, B). ETS1 was found to be constitutively phosphorylated in PC9, HCC0827 and H1975 cells and this was inhibited by both osimertinib and trametinib. ETS1 expression was not detected in H3255 cells (Supplementary Fig. S6A). FOSL1 expression and phosphorylation were inhibited by both inhibitors in PC9 and HCC0827 cells. In H1975 cells, the inhibitors did not reduce FOSL1 expression or phosphorylation and in H3255 cells, FOSL1 expression and phosphorylation were weak and remained unchanged by treatment (Supplementary Fig. S6B). SMAD3 was found to be phosphorylated in H1975 and H3255 cells and was inhibited by both osimertinib and trametinib (Supplementary Fig. S6B). Overall, only one or two transcription factors were found to be phosphorylated and inhibited by both inhibitors in each different cell line.

Involvement of the RAS-ERK signaling pathway in the positive regulation of the METex14Del regulatory network

A new RNA-seq transcriptomic analysis was performed on 16HBE METex14Del cells stimulated or not with HGF and treated or not with capmatinib (CAPM) or trametinib (TRAM). The heatmap of CAPM treatment shows a restoration of the transcriptional program corresponding to the unstimulated state (Fig. 7A) and the heatmap of TRAM treatment shows a drastic change in all differentially expressed genes under both basal and HGF-stimulated conditions (Fig. 7A). To go further, these RNAseq datasets were mapped onto the lung cancer-specific reference network (Fig. 7B and Supplementary Fig. S7) and showed that HGF treatment alone resulted in an activated network profile similar to Fig. 1A, with activation of the same six positively influential TFs of the major regulatory node: ETS1, FOSL1, SMAD3,

HMGA2, CCND1 and RUNX1 (Fig. 7B). As expected, CAPM completely inhibited their positive influence and stronger effect was found with TRAM treatment (Supplementary Table 4). Furthermore, when the expression of the previously analyzed regulated target genes of ETS1, FOSL1 and SMAD3 were plotted in a box-and-whisker plot (Fig. 7C), VIM, NOG, SERPINE2, SERPINA1, SH2D5, CHGB, LETM2, DCBLD2, IRS1 and SLC15A3 showed the expected up- or down-regulation under HGF stimulation and were reversed by these two inhibitor treatments. The exception was ABCA1, which showed only weak activation (Fig. 7C).

DISCUSSION

Disease phenotypes, including those related to disease progression and response to therapy, are maintained by small groups of TFs and co-TFs [37,38]. Therefore, it is important to focus on the identification of these regulators. There are several bioinformatic methods for inferring gene regulatory networks from high-throughput data to enable the discovery of disease-related genes and/or pathways. To date, successful approaches for reverse-engineered construction of context-specific networks have been ARACNe, LICORN or GENIE3 with enrichment of interaction evidence such as protein-protein interactions and/or transcriptional regulation [39]. Our regulatory network concept is based on the representation of numerous TFs and co-regulators with an indication of cooperation/interaction [40,41] and also the level influence of regulators on target genes [41].

Our reverse engineering approach to infer the first lung cancer-specific gene regulatory network from the lung cancer cell dataset without any a priori knowledge allowed us to distinguish two different states of cell activation: HGF-stimulated or not. Notably, we and others have demonstrated that METex14Del activation requires HGF stimulation in various cell lines expressing METex14Del, via genome editing [42–44], by ectopic expression [45,46] or cells derived from patients [44,47]. The dependence of tumor formation *in vivo* on HGF has been demonstrated using human 16HBE METex14Del cells, which induce tumor growth only when xenografted in HGF-humanized mice. Interestingly, mice infected with a lentivirus expressing mouse METex15del (equivalent to human METex14Del) also develop lung tumors. However, the contribution of HGF to tumorigenesis was not elucidated in this experiment, as mouse HGF could also contribute to tumor growth[8]. By contrast, cells displaying both MET exon 14 skipping and MET gene amplification, such as Hs746T, receptor activation is ligand-independent, likely due to the high expression of MET. Nevertheless, the ligand dependence of METex14Del remains debated, regardless of its overexpression, since receptor phosphorylation

and downstream signaling activation were observed without ligand stimulation in some models in which METex14Del was reconstituted by genome editing [48].

Our transcriptomic data of HGF-stimulated METex14Del, interpretated by the regulatory network, revealed three interconnected transcription factors: FOSL1, ETS1 and SMAD3, the last two only highlighted by the network representation but not by transcriptomic analysis. Furthermore, the network showed that the relationship between ETS1, SMAD3 and FOSL1 was characterized by protein-protein interactions between them and by the fact that the transcription factors ETS1 and SMAD3 have binding sites on each other's genes and FOSL1 has a binding site for SMAD3. Several studies have already documented interactions between these differentially influential regulators (DIRs): co-regulation of gene expression by ETS1 and SMAD3 [49–51] and the existence of active transcriptional complexes containing them [52]; an association between FOSL1 and SMAD3 leading to co-regulation of target genes has also been demonstrated [53], as has a potential cooperation between ETS1, SMAD3 and AP1 to regulate the angiopoietin-like 4 gene [54]. Therefore, although the interplay between these transcription factors had already been documented, our study shows that they actively cooperate in the context of an integrated biological response (HGF stimulation of the METex14Del receptor) to regulate a set of target gene. This demonstrates that they belong to the regulatory node.

Transcription factors can be activated by several mechanisms, including expression, phosphorylation state, localization, and interaction with co-activators. In our model, we found that mRNA and protein expression of FOSL1 is strongly increased under HGF stimulation, but ETS1 and SMAD3 expression is only weakly increased. This suggests a different activation mechanism for the latter two, independent of protein expression, as confirmed here by the phosphorylation/activation state for the 3 TFs under HGF stimulation. Note that ETS1 and FOSL1 phosphorylations are known to be ERK-dependent [55–58] and can be induced by HGF stimulation in cells expressing the WT form of MET [59,60]. SMAD3 phosphorylation can also be ERK-dependent, on serine residues within the linker domain [61] and plays different roles depending on the cell context [62]. As expected, direct blockade of the MET receptor with capmatinib did not allow HGF/MET to induce these stimulations and blocking only the RAS-ERK signaling pathway with trametinib prevented the activation of some specific HGF-activated DIRs and the expression of related target genes involved in cell motility and invasion. Lung cancer regulatory networks from trametinib-treated cells confirmed the strong involvement of the RAS-ERK signaling pathway and its positive influence on ETS1, FOSL1 and SMAD3. However, METex14Del is able to induce sustained activation of other signaling

pathways including PI3K/AKT, which should also be investigated. The network also highlights the potential involvement of a variety of other TFs, such as YAP1 of the HIPPO pathway, which is known to be involved in tumorigenesis [63]. Epithelial-to-mesenchymal transition (EMT) is an important step in cell migration and invasion, specifically through reduction of intercellular adhesion, loss of apical-basal polarity and gain of motility [64] notably upon binding of TGFβ, HGF and EGF to their respective receptors. ETS1, FOSL1 and especially SMAD3 are known to induce EMT [65–67]. In our study, several EMT target genes tested (VIM, SERPINE2, PTX3 and ITGA2) were found to be regulated by HGF and their expression was reduced by knockdown of the combination of the 3 TFs. Functionally, significant inhibition of METex14Del-induced cell motility and invasion was achieved by knocked down of ETS1, FOSL1 and SMAD3, suggesting that these three regulators work together to regulate migration and invasion. However, additional transcription factors may be involved including HMGA2, ZEB1 and TWIST2, known to be involved in lung cancer [64,68].

Taken together, our transcriptomics-based regulatory network and the functional experiments performed to support our interpretations highlight a number of key transcription factors that act together to regulate the biological responses induced by METex14Del. In particular, inhibition of a single TF of the node did not perturb the biological responses, and only simultaneous knockdown of three of them resulted in significant but not complete, inhibition. Furthermore, inhibition or restoration of target gene expression required knockdown of two or three of them. Thus, the regulation of genes by multiple TFs forming an interconnected node could ensure the robustness of the network, with the knock-out of one regulator being insufficient to disrupt the biological outcome. To go further, we interpreted "influence" as post-transcriptional activation, i.e. phosphorylation, of the transcription factors, rather than simply increased expression. This led us to identify the RAS-ERK pathway as an important signaling pathway that can positively regulate both the entire major regulatory node and associated biological responses. Interestingly, activation of ETS1, FOSL1 or SMAD3 was also observed in several cell lines harboring mutation of EGFR. However, activation of these three transcription factors was not found together in the same cell type. This suggests that the complexity and specificity of the regulatory network activated by an oncogene driver depends on the expression or availability of key regulatory factors. Therefore, importantly, unlike the TFs, this pathway is readily targeted by pharmacological approaches, opening the way to novel therapeutic strategies, particularly in the context of incomplete response to MET TKIs in patients harboring METex14Del mutations.

MATERIALS AND METHODS

Cell Lines

The parental 16HBE cell line (MET WT) and derived cells (METex14Del) [11] were maintained in GIBCO MEM. The ZORG cell line was derived from a pleural effusion of a MET-TKI resistant patient and were maintained in RPMI1640 [69]. The H596, PC9 (EGFR del E746_A750), HCC0827 (EGFR del E746_A750), H1975 (EGFR T790M L858R) and H3255 (EGFR L858R) cell lines were obtained from the American Type Culture Collection (Manassas, VA) and were maintained in RPMI1640. All media were supplemented with 10% FBS (Sigma-Aldrich, Merck KGaA, Darmstadt). All cells were routinely tested for mycoplasma using MycoAlertTM (Lonza, Basel, Switzerland). Cells were cultured at 37°C under a humidified controlled atmosphere of 5% CO2 in air. For MET activation experiments, cells were starved in serum-free medium for 16 h after overnight adherence. Cells were then pretreated with inhibitors for 3 h before activation by HGF for 30 min, 8 h or 24 h.

RNA interference

The ETS1 (VHS40612 and VHS40614), FOSL1 (HSS188462 and s15585), and SMAD3 (VHS41114 and VHS41111) siRNAs and the Stealth siRNA negative control Lo GC (12935200) and Silencer® Select Negative Control #1 siRNA (#4390843) were purchased from Thermo Fischer Scientific. Transfections were performed using Lipofectamine™ RNAiMAX (Thermo Fisher Scientific) with 50nM siRNA per well according to manufacturer's protocol. Cells were used for further experiments 48 h after transfection.

Immunoblotting

The Nucleospin RNA/Protein Kit (Macherey-Nagel, Düren, DE) was used to extract RNA and proteins from siRNA validation experiments. All other proteins were extracted with RIPA buffer supplemented with protease and phosphatase inhibitor cocktail (1% aprotinin, 1 mM PMSF, 1 µM leupeptine, 1 mM Na3VO4, 20 mM β-glycerophosphate), sonicated for 20 s, and centrifuged at 1500 g for 15 min. Protein samples was quantified using BCA protein assay kit (Thermo Fischer Scientific) and equal amounts of protein were resolved on NuPAGE Bis-Tris polyacrylamide gel (Invitrogen, Thermo Fischer Scientific) and transferred to Immobilon-P polyvinylidene fluoride membranes (Merk Millipore, Darmstadt, DE). After blocking with casein buffer (0.2% casein in PBS containing 0.1% Tween 20), the membranes were sequentially incubated with the indicated primary antibodies and HRP-conjugated secondary

antibodies for 1 h before detection (Supplementary Table 5). The chemiluminescence signal was detected using SuperSignal West Dura or Femto (ThermoFisher Scientific) and captured using LAS 3000 (Fujifilm, Ratingen, DE). For reprobing, primary antibodies were removed from membranes using Antibody stripping buffer (Gene Bio-Application, Israel) according to manufacturer's instructions. Representative results of *n*=3 independent experiments are shown. The full length uncropped original western blots used in the manuscript are provide as part of their original submission (Supplemental Material file).

RT-qPCR analysis

After total RNA extraction using the Nucleospin RNA/Protein Kit (Macherey-Nagel) and quantification using NanoDrop™ 2000 (Thermo Scientific), cDNA was reverse transcribed using the High Capacity cDNA Reverse Transcriptase Kit (Invitrogen). Quantitative PCR was performed using Fast SYBR Green Master Mix on QuantStudio3 (Applied Biosystems, Thermo Fisher Scientific). All conditions were performed in triplicate of n=2, 3 or 4 independent experiments (see figure comments). Cycle threshold (ΔΔCt) values were calculated by normalization to b2m and the gene expression levels were compared using the 2-ΔΔCt method. Primers sequenced are listed in Supplementary Table 6.

Scratch wound healing assay

After 24 h of attachment, cells (transfected or not) were either starved with 0.1% FBS medium for 4 h or treated with mitomycin C (Sigma-Aldrich) for 2 h (ZORG) before scratch wounds were made using a 96-well WoundMaker (Essen BioScience, Sartorius, Göttingen, Germany) according to manufacturer's instructions. For the invasion assay, 50 µl of Matrigel matrix was added to each well and allowed to solidify for 30 min in a 37°C CO2 incubator. Cells were then treated with or without inhibitor in the presence or absence of HGF in 0.1% FBS medium or in 10% FBS medium (ZORG) (*n*=3 to 4 independent experiments; see figure comments). Images of the wounds were captured automatically in the CO2 incubator using IncuCyte Zoom or SX5 software (Essen BioScience) every 2 h. Data were analyzed for wound confluence and calculated using the IncuCyte software package.

Proliferation and scattering cell imaging

After 24 h of attachment, cells (transfected or not) were starved for 4 h in 0.1% FBS medium before treatment or not with inhibitor in the presence or absence of HGF for 16 h. Cell growth and viability were quantified using AlamarBlue Cell Viability Reagent (Invitrogen) and fluorescence was measured using a Multiskan RC Microplate Reader (ThermoLabsystems)

411 with 560/590 (ex/em) wavelength filter settings. Each result was expressed as the average
412 optical density (6 replicates of n=3 to 4 independent experiments; see figure comments). For
413 scattering, images were acquired automatically in the CO2 incubator using the IncuCyte system
414 every 2 h.
415
416 **Inference of gene regulatory networks**
417 The Cancer Cell Line Encyclopedia project [70] provides transcriptomic data for lung cancer
418 cell lines. The Human Transcription Factor Catalog [71] and the TcoF-DB v2 database [72]
419 provided a list of 2 375 regulators corresponding to 1 639 TFs with experimentally validated
420 DNA binding specificity and 752 co-TFs. Human transcription factor binding sites were
421 modeled using MotifDB R/Bioconductor [73] and promoter sequences were searched using the
422 PWMEnrich R/Bioconductor package [74]. ChIP-seq data from the ChEA2 database [75] and
423 the tftargets R package and various databases (e.g. TRED, ITFP, ENCODE, BEDOPS,
424 TRRUST) provided all regulatory evidence from TF to target gene, resulting in a list of 2 256
425 674 TF-target interactions. The list of TF-TF cooperation containing 1 257 053 protein-protein
426 interactions from different databases (e.g. BioGrid, HIPPIE, STRING, and HPRD) was
427 obtained using the R package iRefR [76]. Gene regulatory networks were inferred using the
428 bioconductor package CoRegNet [24]. Starting with transcriptomic data and a list of human
429 regulators, CoRegNet uses the hLICORN algorithm [77] to capture regulatory interactions
430 between regulators and target genes.
431
432 **Quantification of regulatory influence signals**
433 The regulatory network structure provides a set of genes that are activated or repressed under
434 reference conditions. Based on this structure, we can capture the influence of each regulator, a
435 latent signal of regulator activity in each sample based on its observed effect on downstream
436 entities. For each regulator, Welch's t-test was performed to compare the distribution of
437 activated *(A^r)* and repressed *(I^r)* genes. The influence of a regulator r is calculated as
438 Influence(r) = (¯(E(A^r))-¯(E(I^r)))/√((μ_(A^r)^2)/A^r +(μ_(I^r)^2)/I^r) where *$E(A^r)$* and
439 *$E(I^r)$* are the expressions of the activated and repressed genes in the samples, respectively,
440 ¯(E(A^r))∧¯(E(I^r)) are their respective means, and μA^r and μI^r are their standard deviations.
441 A regulator is active only if it activates A^r and represses I^r according to the expectations of the
442 regulatory network model, resulting in a positive *t*-value in the Welch *t*-test. The *influence* of a
443 regulator is only present if it activates or represses at least five genes. In the case that expression
444 values of target genes might be missing in some of the tumor samples, their expression values

were estimated using the LatNet method [78], with expression levels in cell lines as a reference dataset.

Identification of group-specific regulators and target genes

Gene expression analysis was performed on the normalized gene expression dataset and DIRs were identified using the linear models for microarray data (LIMMA) R package [79], and *p*-values were adjusted using the Benjamini-Hochberg method. Regulators with an adj. p-value < 0.05 were considered DIRs.

Functional enrichment analysis of DEGs

GO enrichment analysis was performed on Hallmarks, Molecular Function and Reactome using the R package msigdbr [80], with overrepresentation analysis methods using the R packages clusterProfiler [81] and enrichR [82].

Library construction and sequencing

3'RNA-Seq libraries were prepared using the QuantSeq 3' mRNA-Seq Library Prep Kit FWD (Lexogen, Greenland, US) with 200ng of total RNA according to manufacturer's instructions. Libraries are purified and loaded onto a High Sensitivity DNA Chip controlled by an Agilent Bioanalyzer 2100 (Agilent, Santa Clara, CA). The concentration and size distribution of the libraries are checked. Each library is pooled equimolarly, and the final pool is sequenced on a Nova 6000 instrument (Illumina, San Diego, CA) with 100 cycles of chemistry.

Data analysis for transcriptome data

The *fastp* program was used to remove poor quality regions and poly(A) from the reads, and only reads with quality score threshold of 20 and reads above 25 pb were retained. Read alignments were performed using the *STAR* program with the human genome reference (GRCh38) and reference gene annotations (*Ensembl*). The UMI (Unique Molecular Index) allowed us to reduce errors and biases in the PCRs, which were processed using *fastp* and *umi tools*. Based on the read alignments and the UMI, we counted the number of molecules per gene using *FeatureCount* (from 3.73M to 11.9M molecules, average 6.82M). Other programs used for read quality control and workflow were Q*ualimap*, *fastp*, *FastQC* and *MultiQC*. Differential gene expression of RNA-seq was performed using the *R/Bioconductor* package *DESeq2*. The cut-off for differentially expressed genes was adj. p-value< 0.05 and absolute fold change >1.5.

Statistics

All results except Network inference and transcriptomic analysis are expressed as mean ± SEM. or SD for the indicated number of independent experiments. Data were analyzed using GraphPad Prism® 9 (San Diego, CA).

DATA AVAILABILITY

Transcriptomic data from 16HBE cells expressing MET WT or METex14Del, stimulated or not by HGF, are available in the Sequence Read Archive under accession number PRJNA764905 (GSE184514, raw data). The RNAseq data from 16HBE with or without treatment (CAPM/TRAM) were deposit in the Sequence Read Archive (SRA) under accession number PRJNA1123466.

REFERENCES

1. Lemmon MA, Schlessinger J. Cell Signaling by Receptor Tyrosine Kinases. Cell. 2010 Jun 25;141(7):1117–34.

2. Chakraborty J, Chakraborty S, Chakraborty S, Narayan MN. Entanglement of MAPK pathways with gene expression and its omnipresence in the etiology for cancer and neurodegenerative disorders. Biochim Biophys Acta Gene Regul Mech. 2023 Dec;1866(4):194988.

3. Weinstein IB. Cancer. Addiction to oncogenes--the Achilles heal of cancer. Science (New York, NY. 2002 Jul 5;297(5578):63–4.

4. Saraon P, Pathmanathan S, Snider J, Lyakisheva A, Wong V, Stagljar I. Receptor tyrosine kinases and cancer: oncogenic mechanisms and therapeutic approaches. Oncogene. 2021 Jun;40(24):4079–93.

5. Yang Y, Li S, Wang Y, Zhao Y, Li Q. Protein tyrosine kinase inhibitor resistance in malignant tumors: molecular mechanisms and future perspective. Sig Transduct Target Ther. 2022 Sep 17;7(1):1–36.

6. The cancer genome atlas research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 2014 Jul 31;511(7511):543–50.

7. Descarpentries C, Lepretre F, Escande F, Kherrouche Z, Figeac M, Sebda S, et al. Optimization of Routine Testing for MET Exon 14 Splice Site Mutations in NSCLC Patients. J Thorac Oncol. 2018 Dec;13(12):1873–83.

8. Frampton GM, Ali SM, Rosenzweig M, Chmielecki J, Lu X, Bauer TM, et al. Activation of MET via Diverse Exon 14 Splicing Alterations Occurs in Multiple Tumor Types and Confers Clinical Sensitivity to MET Inhibitors. Cancer discovery. 2015 Aug;5(8):850–9.

9. Cortot AB, Kherrouche Z, Descarpentries C, Wislez M, Baldacci S, Furlan A, et al. Exon 14 deleted MET receptor as a new biomarker and target in cancers. J Natl Cancer Inst. 2017 May;109(5).

10. Fernandes M, Paget S, Kherrouche Z, Truong MJ, Vinchent A, Meneboo JP, et al. Transforming properties of MET receptor exon 14 skipping can be recapitulated by loss of the CBL ubiquitin ligase binding site. FEBS Lett. 2023 Sep;597(18):2301–15.

11. Fernandes M, Hoggard B, Jamme P, Paget S, Truong MJ, Grégoire V, et al. MET exon 14 skipping mutation is a hepatocyte growth factor (HGF)-dependent oncogenic driver in vitro and in humanised HGF knock-in mice. Mol Oncol. 2023 Nov;17(11):2257–74.

12. Lu D, Nagelberg A, Chow JL, Chen YT, Michalchuk Q, Somwar R, et al. MET Exon 14 Splice-Site Mutations Preferentially Activate KRAS Signaling to Drive Tumourigenesis. Cancers (Basel). 2022 Mar 8;14(6):1378.

13. Liang Q, Hu Y, Yuan Q, Yu M, Wang H, Zhao B. MET exon 14 skipping mutation drives cancer progression and recurrence via activation of SMAD2 signalling. Br J Cancer. 2024 Feb;130(3):380–93.

14. Fernandes M, Hoggard B, Jamme P, Paget S, Truong MJ, Grégoire V, et al. MET exon 14 skipping mutation is a hepatocyte growth factor (HGF)-dependent oncogenic driver in vitro and in humanized HGF knock-in mice. Mol Oncol. 2023 Feb 17;

15. Gandino L, Longati P, Medico E, Prat M, Comoglio PM. Phosphorylation of serine 985 negatively regulates the hepatocyte growth factor receptor kinase. The Journal of biological chemistry. 1994;269(3):1815-20.

16. Nakayama M, Sakai K, Yamashita A, Nakamura T, Suzuki Y, Matsumoto K. Met/HGF receptor activation is regulated by juxtamembrane Ser985 phosphorylation in hepatocytes. Cytokine. 2013 Jun;62(3):446–52.

17. Peschard P, Ishiyama N, Lin T, Lipkowitz S, Park M. A conserved DpYR motif in the juxtamembrane domain of the Met receptor family forms an atypical c-Cbl/Cbl-b tyrosine kinase binding domain binding site required for suppression of oncogenic activation. The Journal of biological chemistry. 2004;279(28):29565–71.

18. Abella JV, Peschard P, Naujokas MA, Lin T, Saucier C, Urbe S, et al. Met/Hepatocyte growth factor receptor ubiquitination suppresses transformation and is required for Hrs phosphorylation. Mol Cell Biol. 2005 Nov;25(21):9632–45.

19. Peschard P, Fournier TM, Lamorte L, Naujokas MA, Band H, Langdon WY, et al. Mutation of the c-Cbl TKB domain binding site on the Met receptor tyrosine kinase converts it into a transforming protein. Molecular cell. 2001;8(5):995–1004.

20. Tulasne D, Deheuninck J, Lourenco FC, Lamballe F, Ji Z, Leroy C, et al. Proapoptotic function of the MET tyrosine kinase receptor through caspase cleavage. Mol Cell Biol. 2004;24(23):10328–39.

21. Foveau B, Leroy C, Ancot F, Deheuninck J, Ji Z, Fafeur V, et al. Amplification of apoptosis through sequential caspase cleavage of the MET tyrosine kinase receptor. Cell Death Differ. 2007 Apr;14(4):752–64.

22. Duplaquet L, Leroy C, Vinchent A, Paget S, Lefebvre J, Vanden Abeele F, et al. Control of cell death/survival balance by the MET dependence receptor. eLife [Internet]. 2020 Feb [cited 2020 Apr 6];9. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7039684/

556 23. Lefebvre J, Muharram G, Leroy C, Kherrouche Z, Montagne R, Ichim G, et al. Caspase-generated
557 fragment of the Met receptor favors apoptosis via the intrinsic pathway independently of its
558 tyrosine kinase activity. Cell Death Dis. 2013;4:e871.

559 24. Nicolle R, Radvanyi F, Elati M. CoRegNet: reconstruction and integrated analysis of co-
560 regulatory networks. Bioinformatics. 2015 Sep 15;31(18):3066–8.

561 25. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software
562 environment for integrated models of biomolecular interaction networks. Genome Res. 2003
563 Nov;13(11):2498–504.

564 26. Galatro TF, Holtman IR, Lerario AM, Vainchtein ID, Brouwer N, Sola PR, et al. Transcriptomic
565 analysis of purified human cortical microglia reveals age-associated changes. Nat Neurosci.
566 2017 Aug;20(8):1162–71.

567 27. Lou Y, Tian GY, Song Y, Liu YL, Chen YD, Shi JP, et al. Characterization of transcriptional modules
568 related to fibrosing-NAFLD progression. Sci Rep. 2017 Jul 6;7(1):4748.

569 28. Zerrouk N, Miagoux Q, Dispot A, Elati M, Niarakis A. Identification of putative master regulators
570 in rheumatoid arthritis synovial fibroblasts using gene expression data and network inference.
571 Sci Rep. 2020 Oct 1;10(1):16236.

572 29. Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database
573 in 2023: protein-protein association networks and functional enrichment analyses for any
574 sequenced genome of interest. Nucleic Acids Res. 2023 Jan 6;51(D1):D638–46.

575 30. Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness
576 and reliability of protein-protein interaction networks. Nucleic Acids Res. 2017 Jan
577 4;45(D1):D408–14.

578 31. Davis CA, Hitz BC, Sloan CA, Chan ET, Davidson JM, Gabdank I, et al. The Encyclopedia of DNA
579 elements (ENCODE): data portal update. Nucleic Acids Res. 2018 Jan 4;46(D1):D794–801.

580 32. Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. ChEA: transcription factor
581 regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics. 2010
582 Oct 1;26(19):2438–44.

583 33. Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness
584 and reliability of protein-protein interaction networks. Nucleic Acids Res. 2017 Jan
585 4;45(D1):D408–14.

586 34. Lachmann A, Xu H, Krishnan J, Berger SI, Mazloom AR, Ma'ayan A. ChEA: transcription factor
587 regulation inferred from integrating genome-wide ChIP-X experiments. Bioinformatics. 2010
588 Oct 1;26(19):2438–44.

589 35. Paumelle R, Tulashe D, Kherrouche Z, Plaza S, Leroy C, Reveneau S, et al. Hepatocyte growth
590 factor/scatter factor activates the ETS1 transcription factor by a RAS-RAF-MEK-ERK signaling
591 pathway. Oncogene. 2002 Apr;21(15):2309–19.

592 36. Yang BS, Hauser CA, Henkel G, Colman MS, Beveren CV, Stacey KJ, et al. Ras-mediated
593 phosphorylation of a conserved threonine residue enhances the transactivation activities of c-
594 Ets1 and c-Ets2. Molecular and Cellular Biology. 1996 Feb;16(2):538.

595 37. Bradner JE, Hnisz D, Young RA. Transcriptional Addiction in Cancer. Cell. 2017 Feb
596 9;168(4):629–43.

597 38. Lee TI, Young RA. Transcriptional regulation and its misregulation in disease. Cell. 2013 Mar
598 14;152(6):1237–51.

599 39. Linde J, Schulze S, Henkel SG, Guthke R. Data- and knowledge-based modeling of gene
600 regulatory networks: an update. EXCLI J. 2015;14:346–78.

601 40. Elati M, Neuvial P, Bolotin-Fukuhara M, Barillot E, Radvanyi F, Rouveirol C. LICORN: learning
602 cooperative regulation networks from gene expression data. Bioinformatics. 2007 Sep
603 15;23(18):2407–14.

604 41. Nicolle R, Radvanyi F, Elati M. CoRegNet: reconstruction and integrated analysis of co-
605 regulatory networks. Bioinformatics (Oxford, England). 2015 Sep 15;31(18):3066–8.

606 42. Frampton GM, Ali SM, Rosenzweig M, Chmielecki J, Lu X, Bauer TM, et al. Activation of MET via
607 Diverse Exon 14 Splicing Alterations Occurs in Multiple Tumor Types and Confers Clinical
608 Sensitivity to MET Inhibitors. Cancer discovery. 2015 Aug;5(8):850–9.

609 43. Wang F, Liu Y, Qiu W, Shum E, Feng M, Zhao D, et al. Functional Analysis of MET Exon 14
610 Skipping Alteration in Cancer Invasion and Metastatic Dissemination. Cancer Research. 2022
611 Apr 1;82(7):1365–79.

612 44. Fernandes M, Hoggard B, Jamme P, Paget S, Truong MJ, Grégoire V, et al. MET exon 14 skipping
613 mutation is a hepatocyte growth factor (HGF)-dependent oncogenic driver in vitro and in
614 humanised HGF knock-in mice. Mol Oncol. 2023 Nov;17(11):2257–74.

615 45. Fernandes M, Paget S, Kherrouche Z, Truong MJ, Vinchent A, Meneboo JP, et al. Transforming
616 properties of MET receptor exon 14 skipping can be recapitulated by loss of the CBL ubiquitin
617 ligase binding site. FEBS Letters [Internet]. 2023 Aug 8 [cited 2023 Aug 8];n/a(n/a). Available
618 from: https://onlinelibrary.wiley.com/doi/abs/10.1002/1873-3468.14702

619 46. Guérin C, Vinchent A, Fernandes M, Damour I, Laratte A, Tellier R, et al. MET variants with
620 activating N-lobe mutations identified in hereditary papillary renal cell carcinomas still require
621 ligand stimulation. Mol Oncol. 2025 Feb 20;

622 47. Cerqua M, Botti O, Arigoni M, Gioelli N, Serini G, Calogero R, et al. METΔ14 promotes a ligand-
623 dependent, AKT-driven invasive growth. Life Sci Alliance. 2022 Oct;5(10):e202201409.

624 48. Liang Q, Hu Y, Yuan Q, Yu M, Wang H, Zhao B. MET exon 14 skipping mutation drives cancer
625 progression and recurrence via activation of SMAD2 signalling. Br J Cancer. 2024
626 Feb;130(3):380–93.

627 49. Lindemann RK, Ballschmieter P, Nordheim A, Dittmer J. Transforming growth factor beta
628 regulates parathyroid hormone-related protein expression in MDA-MB-231 breast cancer cells
629 through a novel Smad/Ets synergism. J Biol Chem. 2001 Dec 7;276(49):46661–70.

630 50. Jinnin M, Ihn H, Asano Y, Yamane K, Trojanowska M, Tamaki K. Tenascin-C upregulation by
631 transforming growth factor-beta in human dermal fibroblasts involves Smad3, Sp1, and Ets1.
632 Oncogene. 2004 Mar 4;23(9):1656–67.

51. Van Beek JP, Kennedy L, Rockel JS, Bernier SM, Leask A. The induction of CCN2 by TGFbeta1 involves Ets-1. Arthritis Res Ther. 2006;8(2):R36.

52. Liu D, Wang K, Li K, Xu R, Chang X, Zhu Y, et al. Ets-1 deficiency alleviates nonalcoholic steatohepatitis via weakening TGF-β1 signaling-mediated hepatocyte apoptosis. Cell Death Dis. 2019 Jun 12;10(6):1–14.

53. Sundqvist A, Zieba A, Vasilaki E, Herrera Hidalgo C, Söderberg O, Koinuma D, et al. Specific interactions between Smad proteins and AP-1 components determine TGFβ-induced breast cancer cell invasion. Oncogene. 2013 Aug 1;32(31):3606–15.

54. Kaddatz K, Adhikary T, Finkernagel F, Meissner W, Müller-Brüsselbach S, Müller R. Transcriptional profiling identifies functional interactions of TGF β and PPAR β/δ signaling: synergistic induction of ANGPTL4 transcription. J Biol Chem. 2010 Sep 17;285(38):29469–79.

55. Yang BS, Hauser CA, Henkel G, Colman MS, Van Beveren C, Stacey KJ, et al. Ras-mediated phosphorylation of a conserved threonine residue enhances the transactivation activities of c-Ets1 and c-Ets2. Mol Cell Biol. 1996 Feb;16(2):538–47.

56. Rabault B, Roussel MF, Quang CT, Ghysdael J. Phosphorylation of Ets1 regulates the complementation of a CSF-1 receptor impaired in mitogenesis. Oncogene. 1996 Aug 15;13(4):877–81.

57. Gruda MC, Kovary K, Metz R, Bravo R. Regulation of Fra-1 and Fra-2 phosphorylation differs during the cell cycle of fibroblasts and phosphorylation in vitro by MAP kinase affects DNA binding activity. Oncogene. 1994 Sep;9(9):2537–47.

58. Basbous J, Chalbos D, Hipskind R, Jariel-Encontre I, Piechaczyk M. Ubiquitin-independent proteasomal degradation of Fra-1 is antagonized by Erk1/2 pathway-mediated phosphorylation of a unique C-terminal destabilizer. Mol Cell Biol. 2007 Jun;27(11):3936–50.

59. Paumelle R, Tulasne D, Kherrouche Z, Plaza S, Leroy C, Reveneau S, et al. Hepatocyte growth factor/scatter factor activates the ETS1 transcription factor by a RAS-RAF-MEK-ERK signaling pathway. Oncogene. 2002;21(15):2309-19.

60. Ramos-Nino ME, Blumen SR, Sabo-Attwood T, Pass H, Carbone M, Testa JR, et al. HGF mediates cell proliferation of human mesothelioma cells through a PI3K/MEK5/Fra-1 pathway. Am J Respir Cell Mol Biol. 2008 Feb;38(2):209–17.

61. Mori S, Matsuzaki K, Yoshida K, Furukawa F, Tahashi Y, Yamagata H, et al. TGF-β and HGF transmit the signals through JNK-dependent Smad2/3 phosphorylation at the linker regions. Oncogene. 2004 Sep;23(44):7416–29.

62. Ellenrieder V, Buck A, Gress TM. TGFβ-regulated transcriptional mechanisms in cancer. Int J Gastrointest Canc. 2002 May 1;31(1):61–9.

63. Franklin JM, Wu Z, Guan KL. Insights into recent findings and clinical application of YAP and TAZ in cancer. Nat Rev Cancer. 2023 Aug;23(8):512–25.

64. Debnath P, Huirem RS, Dutta P, Palchaudhuri S. Epithelial-mesenchymal transition and its transcription factors. Biosci Rep. 2022 Jan 28;42(1):BSR20211754.

65. Roberts AB, Tian F, Byfield SD, Stuelten C, Ooshima A, Saika S, et al. Smad3 is key to TGF-beta-mediated epithelial-to-mesenchymal transition, fibrosis, tumor suppression and metastasis. Cytokine Growth Factor Rev. 2006;17(1–2):19–27.

66. Dittmer J. The role of the transcription factor Ets1 in carcinoma. Semin Cancer Biol. 2015 Dec;35:20–38.

67. Sobolev VV, Khashukoeva AZ, Evina OE, Geppe NA, Chebysheva SN, Korsunskaya IM, et al. Role of the Transcription Factor FOSL1 in Organ Development and Tumorigenesis. Int J Mol Sci. 2022 Jan 28;23(3):1521.

68. Ma Q, Ye S, Liu H, Zhao Y, Mao Y, Zhang W. HMGA2 promotes cancer metastasis by regulating epithelial-mesenchymal transition. Front Oncol. 2024;14:1320887.

69. Jamme P, Fernandes M, Copin MC, Descarpentries C, Escande F, Morabito A, et al. Alterations in the PI3K Pathway Drive Resistance to MET Inhibitors in NSCLC Harboring MET Exon 14 Skipping Mutations. J Thorac Oncol. 2020 May;15(5):741–51.

70. Ghandi M, Huang FW, Jané-Valbuena J, Kryukov GV, Lo CC, McDonald ER, et al. Next-generation characterization of the Cancer Cell Line Encyclopedia. Nature. 2019 May;569(7757):503–8.

71. Lambert SA, Jolma A, Campitelli LF, Das PK, Yin Y, Albu M, et al. The Human Transcription Factors. Cell. 2018 Feb;172(4):650–65.

72. Schmeier S, Alam T, Essack M, Bajic VB. TcoF-DB v2: update of the database of human and mouse transcription co-factors and transcription factor interactions. Nucleic Acids Research. 2017 Jan 4;45(D1):D145–50.

73. Paul Shannon MR. MotifDb [Internet]. Bioconductor; 2017 [cited 2024 Oct 21]. Available from: https://bioconductor.org/packages/MotifDb

74. Robert Stojnic DD. PWMEnrich [Internet]. Bioconductor; 2017 [cited 2024 Oct 21]. Available from: https://bioconductor.org/packages/PWMEnrich

75. Kou Y, Chen EY, Clark NR, Duan Q, Tan CM, Ma'ayan A. ChEA2: Gene-Set Libraries from ChIP-X Experiments to Decode the Transcription Regulome. In: Cuzzocrea A, Kittl C, Simos DE, Weippl E, Xu L, editors. Availability, Reliability, and Security in Information Systems and HCI. Berlin, Heidelberg: Springer; 2013. p. 416–30.

76. Mora A, Donaldson IM. iRefR: an R package to manipulate the iRefIndex consolidated protein interaction database. BMC Bioinformatics. 2011 Nov 24;12(1):455.

77. Elati M, Neuvial P, Bolotin-Fukuhara M, Barillot E, Radvanyi F, Rouveirol C. LICORN: learning cooperative regulation networks from gene expression data. Bioinformatics. 2007 Sep 15;23(18):2407–14.

78. Dhifli W, Puig J, Dispot A, Elati M. Latent network-based representations for large-scale gene expression data analysis. BMC Bioinformatics. 2019 Feb 4;19(Suppl 13):466.

79. Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47.

80. Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database Hallmark Gene Set Collection. cels. 2015 Dec 23;1(6):417–25.

81. Wu T, Hu E, Xu S, Chen M, Guo P, Dai Z, et al. clusterProfiler 4.0: A universal enrichment tool for interpreting omics data. Innovation (Camb). 2021 Aug 28;2(3):100141.

82. Kuleshov MV, Jones MR, Rouillard AD, Fernandez NF, Duan Q, Wang Z, et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. Nucleic Acids Res. 2016 Jul 8;44(W1):W90-97.

ACKNOWLEDGEMENTS

The non-tumorigenic 16HBE14o-(RRID:CVCL_0112) (16HBE) cell line, developed by immortalization of primary human bronchial epithelial cells with SV40 large T antigen, was a kind gift from Pr. Dieter Gruenert. This work was supported by the Cancéropôle Nord-Ouest (Emergent 2022), INCa (NETMET-2022- 74012), and La Ligue Contre le Cancer (MET exoCancer-2021-74012).

AUTHOR CONTRIBUTIONS

MJT and DT directed the project and contributed to funding acquisition; conceived and designed the experiments and interpreted the main results; wrote the manuscript with the help of all authors. MJT, MFe, JPM, SS, DT, MFi contributed to the transcriptomic analysis (microarrays). MJT, JPM, SS, DT, MFi contributed to the transcriptomic analysis (RNAseq). GP and ME conceived and designed the co-regulatory networks. All authors provided critical feedback and helped shape the research, discussed the results, and commented on the manuscript.

COMPETITING INTEREST

The authors declare no competing interests.

FIGURE LEGENDS

Fig. 1 Regulatory network of 16HBE METex14Del in response to HGF. A Focus on the largest cluster of the co-regulatory network, showing the influence of TFs on METex14del cells in the presence and absence of HGF. Circles represent DIRs and the radius of the circle is proportional to the number of target genes regulated by the DIR. Co-regulatory interactions between DIRs are indicated: protein-protein interactions with published evidence (blue lines), transcriptional regulation interactions with published evidence (red arrows), and interactions defined only by the h-LICORN algorithm (gray lines). **B** DIRS were visualized in a volcano plot from interference network analysis: average difference in influence vs log10 transformed adj. P-value (regulators with an adj. p-value < 0.05 and an influence difference greater/smaller than 0.5 were considered up-/down-influenced (red/green). **C** DIRS were visualized in a volcano plot from transcriptomic analysis: $\log_2$ fold change vs p-value of differential mRNA-level expression (genes with an absolute adj. p-value < 0.05 and a fold change greater/smaller than 1.5 were considered upregulated/downregulated (red/green)). **D** mRNA-level expression of the six main influent DIRs present in the putative regulatory node, in the presence and absence of HGF (triplicates of *n*=3 independent experiments). Significance was determined by unpaired one-tailed *t*-test with Welch's correction and data are expressed as mean ± SD.

Fig. 2 Identification of differentially expressed target genes potentially regulated by ETS1, FOSL1, and SMAD3 on 16HBE METex14Del cell line. A Representations of the transcriptional regulatory network of ETS1, FOSL1 and SMAD3 with their putative influence on positively (red triangle) and negatively (green triangle) regulated DEGs. **B** Heatmaps showing putative ETS1, FOSL1 and SMAD3 target genes based on METex14Del vs HGF-activated METex14Del (adj p-value <0.05 and absolute fold change >1.5) (*n*=4 for each condition). Colors indicate high (red) and low (green) relative expression levels.

Fig. 3 Effects of ETS1, FOSL1 and SMAD3 silencing on the expression of their predicted target genes on 16HBE METex14Del cell line. A The mRNA-level expression of selected target genes up- or down-regulated in response to HGF was determined by RT-qPCR (triplicates of *n*=3 independent experiments). Significance was determined by unpaired one-tailed *t*-test with Welch's correction and data are expressed as mean ± S.D. **B, C** The silencing efficacy of each specific siRNA (siETS1, siFOSL1, siSMAD3), used individually or in combination (siPOOL) was assessed by (**B**) RT-qPCR for mRNA-level expression (triplicates of *n*=3 independent experiments) and (**C**) by Western blotting for protein-level expression (representative results). **D** mRNA-level expression of up- and down-regulated target genes induced by knockdown of ETS1, FOSL1 and SMAD3 expression (triplicates of *n*=2, 3 or 4 independent experiments). Significance was determined by one-way ANOVA test and data are expressed as mean ± SD in panels (**b**) and (**d**).

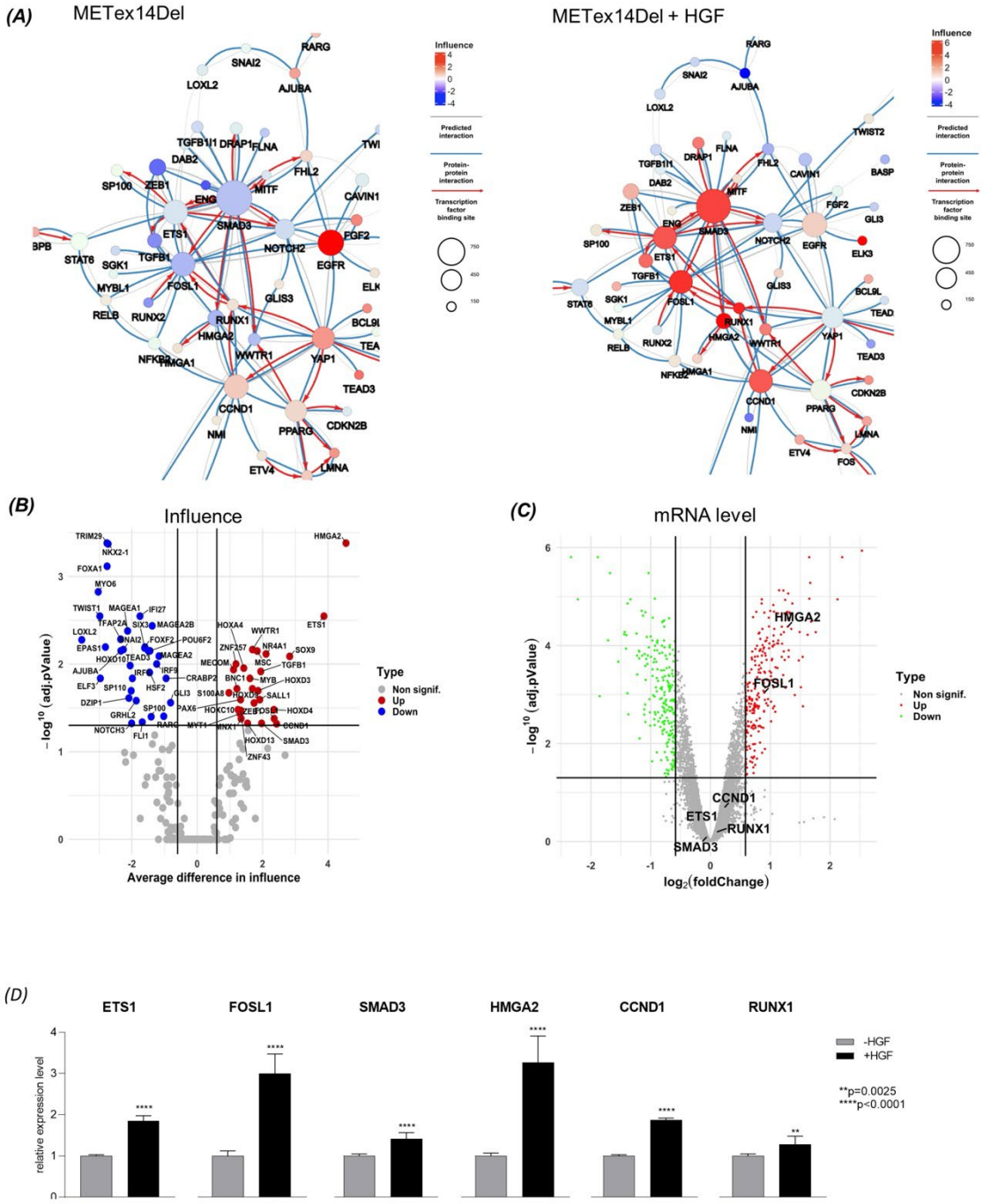
Fig. 4 Effect of ETS1, FOSL1, and SMAD3 silencing on HGF-induced migration, invasion, proliferation and scattering on 16HBE METex14Del cell line. Forty-eight h after transfection with siRNAs targeting ETS1, FOSL1 and SMAD3 alone or in combination, wound healing was monitored without (**A**) or with (**B**) Matrigel in the presence and absence of HGF (triplicates of *n*=4 independent experiments). In all graphs, only statistically significant differences between the negative control (siCTL) and the applied treatment (an siRNA used alone or pooled with the other siRNAs, siPOOL) in the presence or absence of HGF are indicated. Wound healing data are expressed as mean ± SEM and significance was determined by two-way ANOVA test. **C** The effect of siRNA on proliferation was determined by Alamar blue staining after 16 h of stimulation with HGF and relative growth was expressed as relative mean fluorescence (6 replicates of *n*=4 independent experiments). Proliferation data are expressed as mean ± SD and significance was determined by one-way ANOVA test. **D** Images of the effect of silencing on cell scattering were captured at different time points using the Incucyte system and representative images from baseline and 6 h after stimulation with HGF are shown.

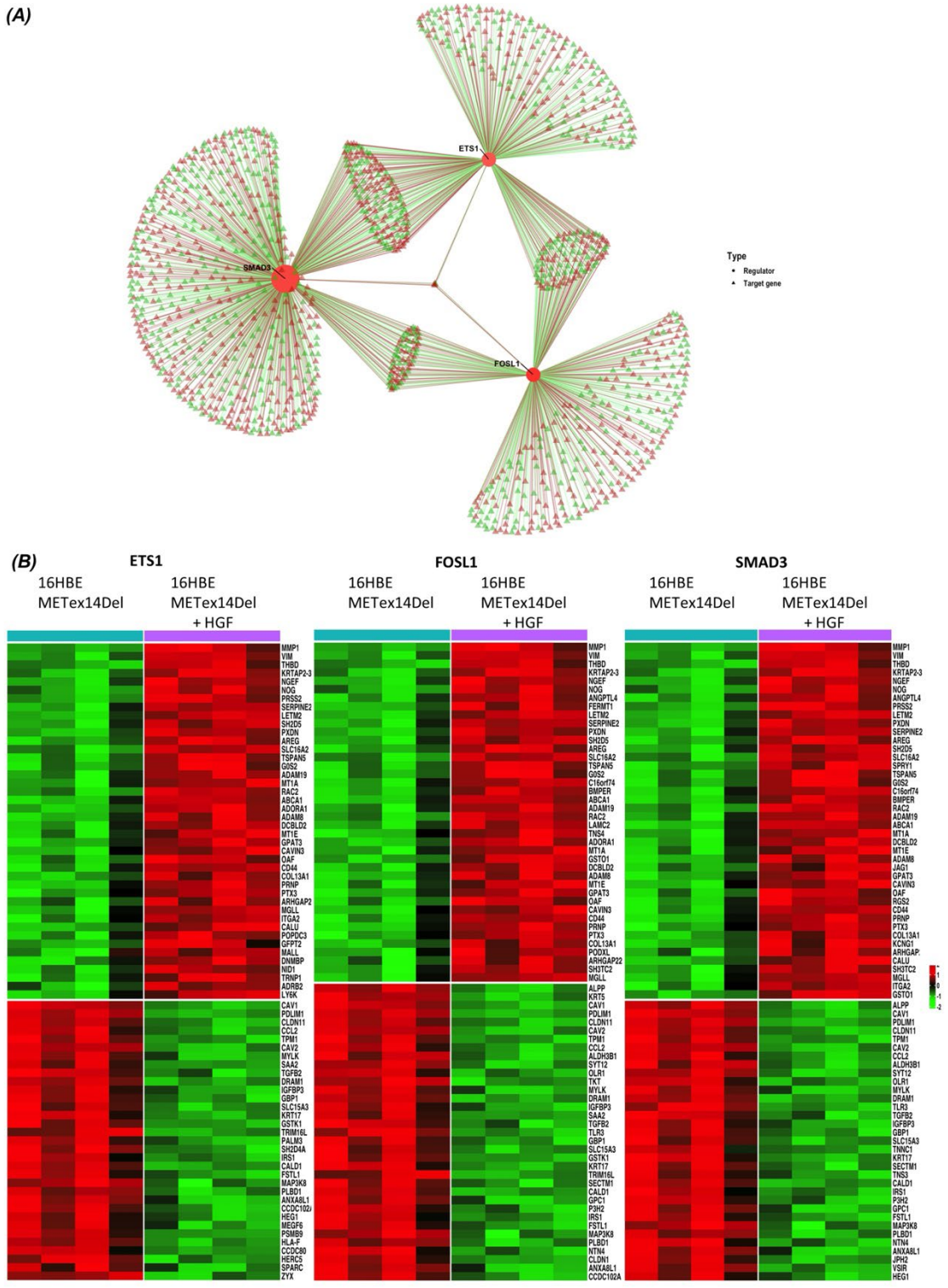
Fig. 5 Gene ontology (GO) annotation of target genes induced by HGF activation of METex14del. A Heatmap of the 50 most significantly differentially regulated target genes in response to HGF (adj. p-value < 0.05 and absolute fold change >1.5) between indicated conditions (*n*=4 per condition). Colors indicate high (red) and low (green) levels of relative expression. **B** Shown are gene ratios (%) of GO enrichment for the differentially expressed target genes according to Hallmarks, Biological Process, Molecular Functions and Reactome annotations. Overrepresentation analysis was performed using the hypergeometric test: ** p-value < 0.01, ***p-value < 0.001. **C, D, E** The 16HBE cells expressing MET WT or METex14Del and ZORG cells derived from lung cancer patients expressing METex14Del were stimulated with HGF in a time course experiment. Expression and activation of (**C**) ETS1 and its form phosphorylated at T^{38} (P-ETS1), (**D**) FOSL1 and its form phosphorylated at S^{265} (P-FOSL1), and (**E**) SMAD3 and its form phosphorylated at S^{208} (P-SMAD3) were analyzed by Western blotting (representative results).

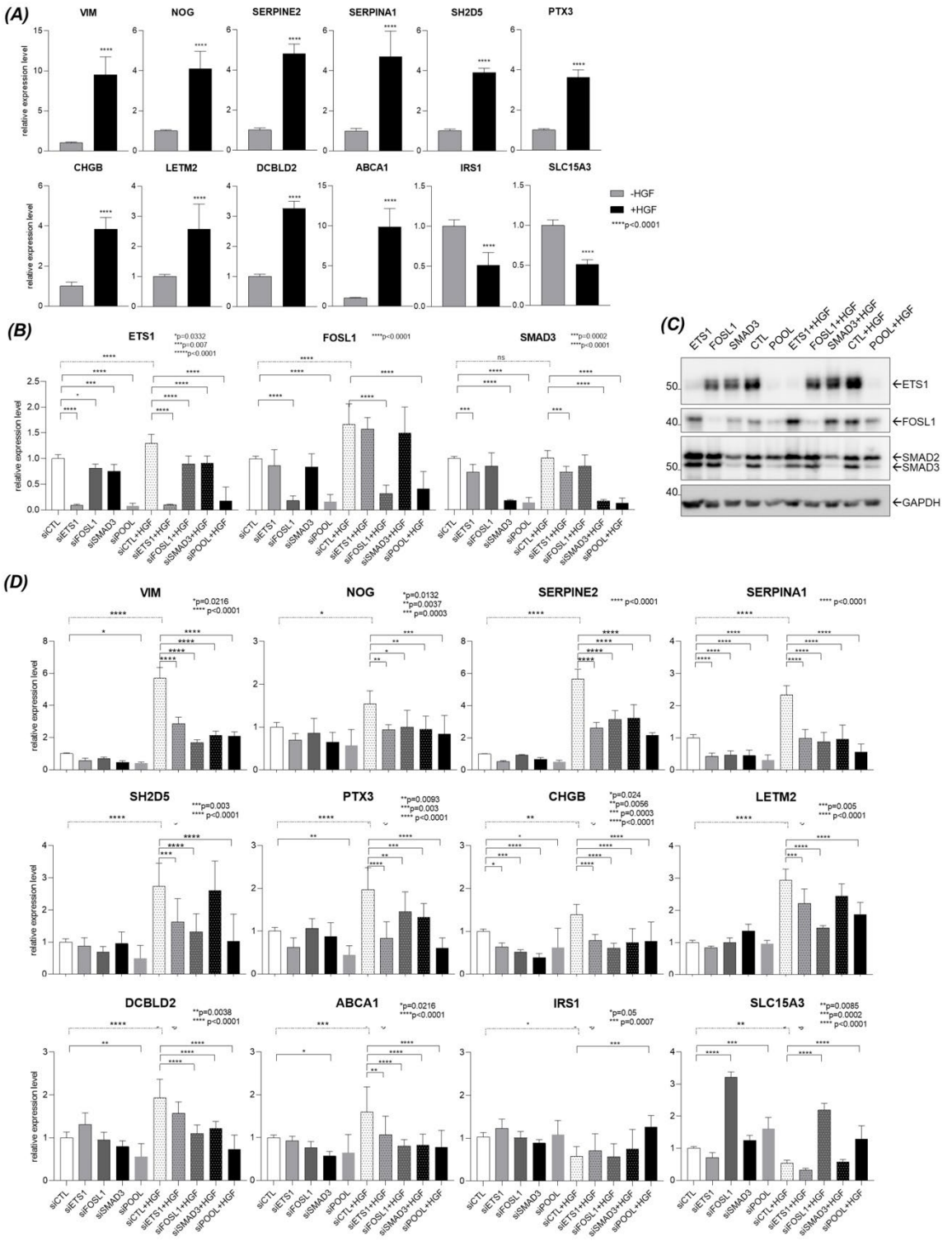
Fig. 6 Effect of MEK inhibitor on ETS1, FOSL1 and SMAD3 phosphorylation and cell responses induced by HGF. The effect of trametinib (TRAM, a MEK inhibitor) on the expression of (**A**) P-ETS1/ETS1, (**B**) P-FOSL1/FOSL1 and (**C**) P-SMAD3/SMAD3 in 16HBE METex14Del cells and ZORG cells stimulated or not with HGF was determined by Western blotting (representative results). The effects of trametinib and capmatinib (CAPM, a MET inhibitor), alone and in combination, on 16HBE METex14Del cell wound healing were evaluated without (**D**) and with (**E**) Matrigel (6 replicates of *n*=3 independent experiments); (**F**) scattering in 16HBE METex14Del cells stimulated or not with HGF (representative photo of baseline and 6 h after inhibition under HGF) and (**G**) proliferation (6 replicates of *n*=4 independent experiments). In all graphs, only statistically significant differences between the negative

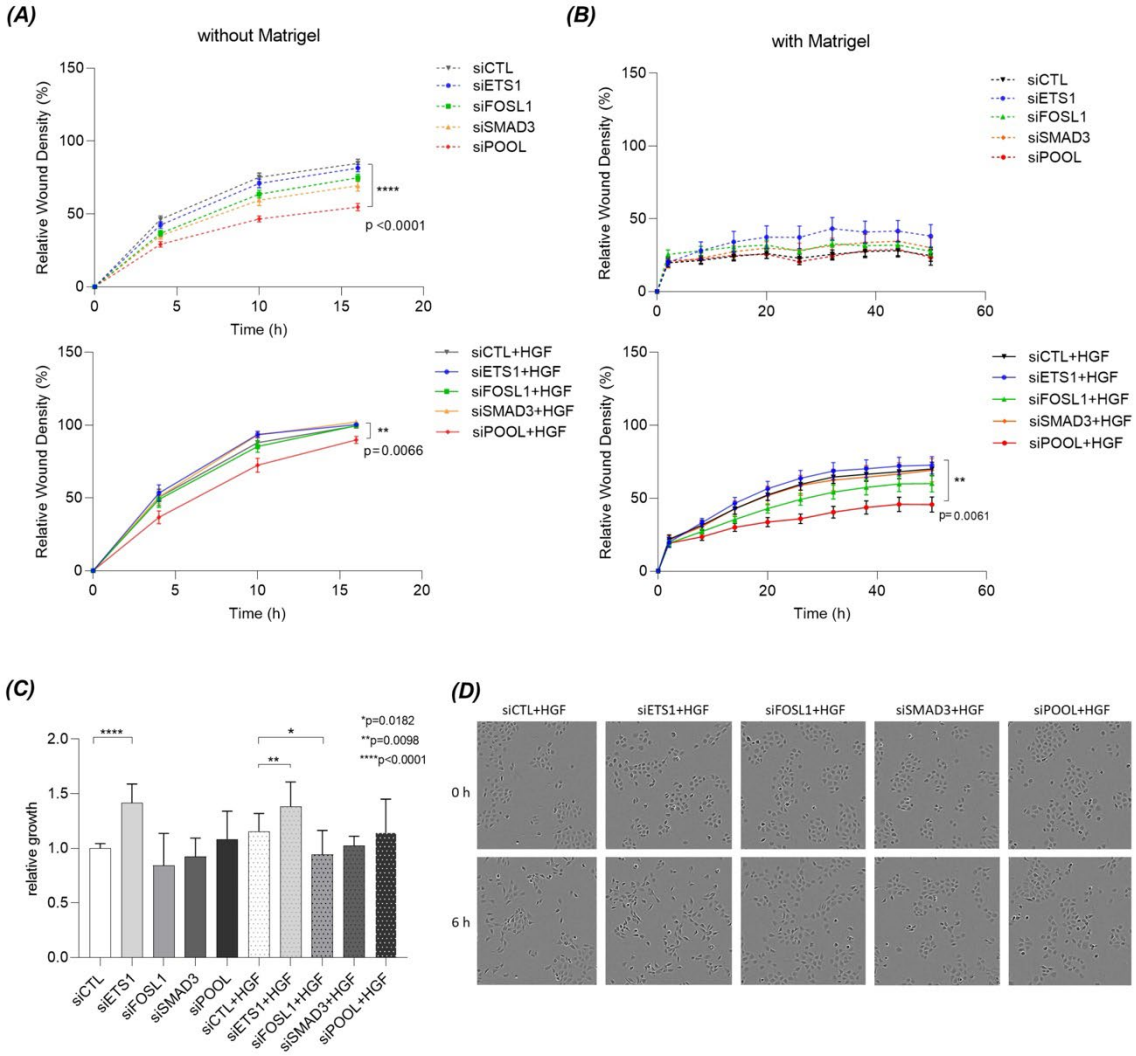
control without HGF (DMSO) and the different treatments (individual or combined inhibitors) in the presence or absence of HGF are indicated. Wound healing data in panels (**D, E**) are expressed as mean ± SEM and significance was determined by two-way ANOVA test. In panel (**G**), proliferation data are expressed as mean ± SD and significance was determined by one-way ANOVA test.

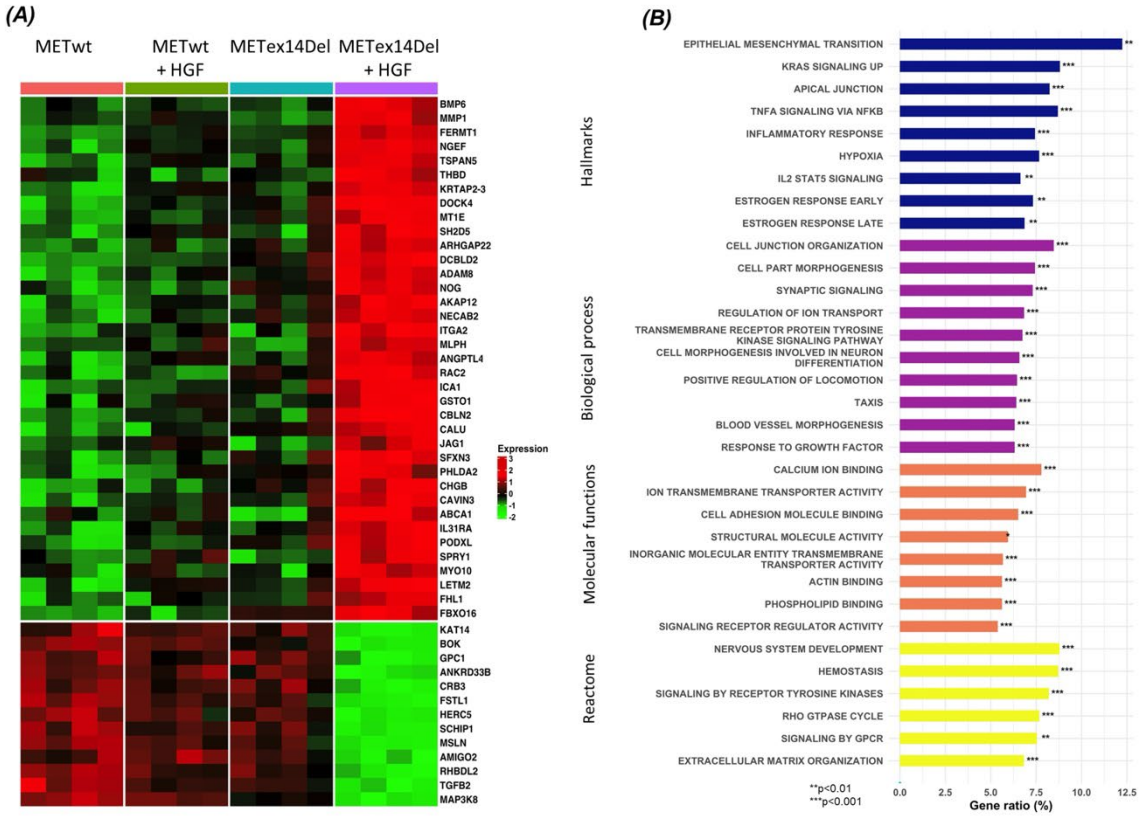
Fig. 7 Effect of MET and MEK inhibitors on the transcriptional program induced by HGF and on the predicted regulatory network. A Heatmaps of differentially expressed genes (adj. p-value < 0.05 and absolute fold change > 1.5) in 16HBE METex14Del cells treated with capmatinib (CAPM) or trametinib (TRAM) and stimulated or not with HGF are shown (*n*=4 per condition). **B** Zooms on the major regulatory node of the co-regulatory network of METex14Del cells in the presence and absence of inhibitors (CAPM, TRAM) upon stimulation with HGF. The radius of the circle is proportional to the number of target genes of the respective DIR. Co-regulatory interactions between transcription factors are indicated: protein-protein interactions with published evidence (blue line), transcriptional regulation interactions with published evidence (red arrow), and interactions defined only by the h-LICORN algorithm (gray line). **C** Expression of selected target genes regulated by ETS1, FOSL1, SMAD3 are shown in box-and-whisker plots. The box plot shows the 25th, 50th, and 75th percentiles, while the blue dots show the variant stabilization transformation (VST) of gene expression for each sample (*n*=4 per condition). The median is indicated by the line across the box.

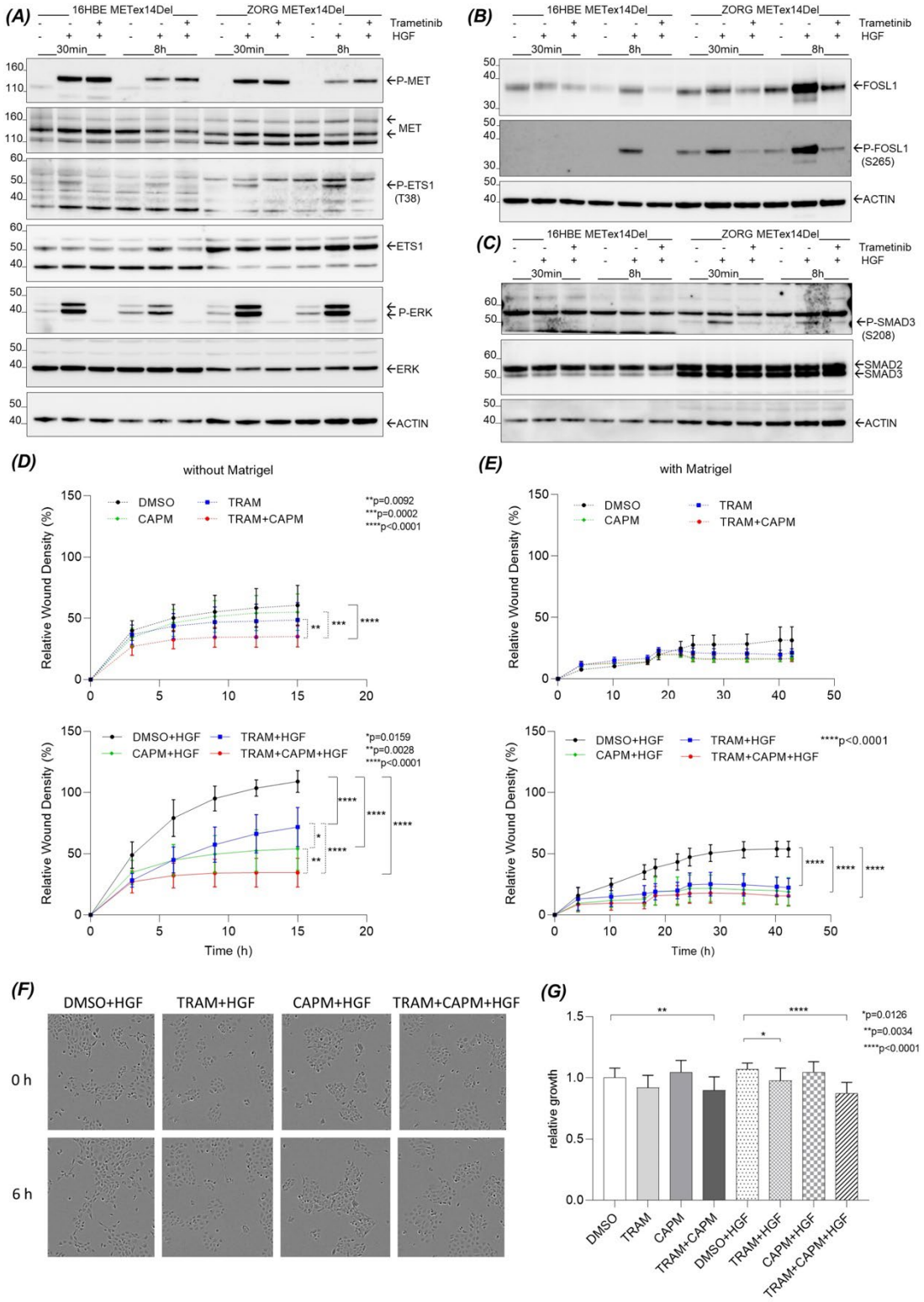

(A)
METex14Del
METex14Del + HGF
Influence
Predicted interaction
Protein-protein interaction
Transcription factor binding site
(B)
Influence
Average difference in influence
−log10 (adj.pValue)
Type
Non signif.
Up
Down
(C)
mRNA level
log2(foldChange)
HMGA2
FOSL1
CCND1
ETS1
RUNX1
SMAD3
(D)
ETS1
FOSL1
SMAD3
HMGA2
CCND1
RUNX1
relative expression level
-HGF
+HGF
**p=0.0025
****p<0.0001

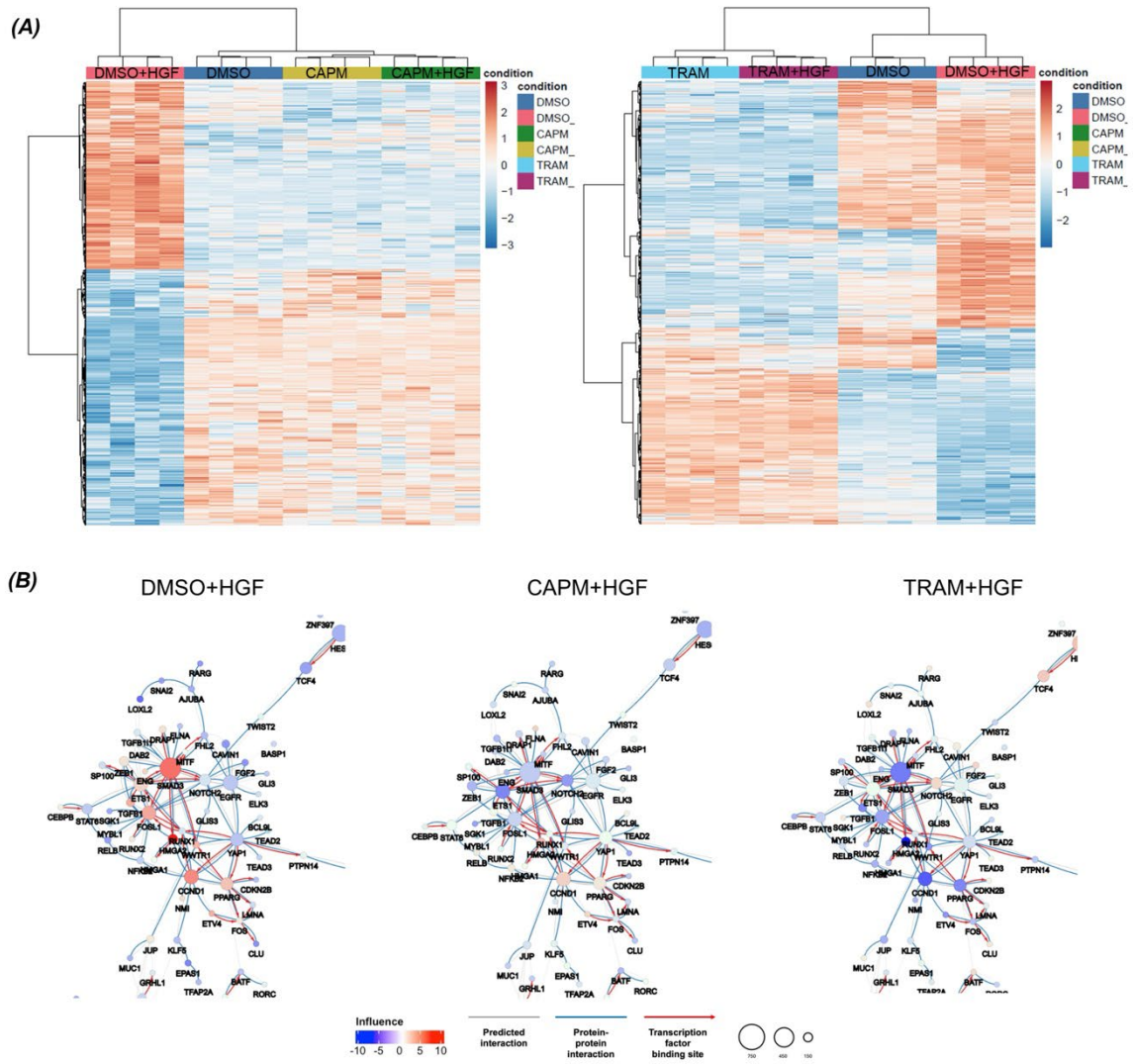

(A)
ETS1
SMAD3
FOSL1
Type
Regulator
Target gene
(B)
ETS1
FOSL1
SMAD3
16HBE
METex14Del
16HBE
METex14Del
+ HGF


(A)
VIM
NOG
SERPINE2
SERPINA1
SH2D5
PTX3
CHGB
LETM2
DCBLD2
ABCA1
IRS1
SLC15A3
relative expression level
-HGF
+HGF
****p<0.0001
(B)
ETS1
FOSL1
SMAD3
*p<0.0332
***p<0.007
*****p<0.0001
****p<0.0001
***p<0.0002
ns
siCTL
siETS1
siFOSL1
siSMAD3
siPOOL
siCTL+HGF
siETS1+HGF
siFOSL1+HGF
siSMAD3+HGF
siPOOL+HGF
(C)
ETS1
FOSL1
SMAD3
CTL
POOL
ETS1+HGF
FOSL1+HGF
SMAD3+HGF
CTL+HGF
POOL+HGF
50.
40.
←ETS1
←FOSL1
←SMAD2
←SMAD3
←GAPDH
(D)
VIM
*p=0.0216
**** p<0.0001
NOG
*p=0.0132
**p=0.0037
*** p=0.0003
SERPINE2
**** p<0.0001
SERPINA1
**** p<0.0001
SH2D5
***p=0.003
**** p<0.0001
PTX3
**p=0.0093
***p=0.003
**** p<0.0001
CHGB
*p=0.024
**p=0.0056
*** p=0.0003
****p<0.0001
LETM2
***p=0.005
**** p<0.0001
DCBLD2
**p=0.0038
**** p<0.0001
ABCA1
*p=0.0216
****p<0.0001
IRS1
*p=0.05
*** p=0.0007
SLC15A3
**p=0.0085
***p=0.0002
**** p<0.0001
relative expression level
siCTL
siETS1
siFOSL1
siSMAD3
siPOOL
siCTL+HGF
siETS1+HGF
siFOSL1+HGF
siSMAD3+HGF
siPOOL+HGF


(A)
without Matrigel
Relative Wound Density (%)
Time (h)
siCTL
siETS1
siFOSL1
siSMAD3
siPOOL
****
p <0.0001
siCTL+HGF
siETS1+HGF
siFOSL1+HGF
siSMAD3+HGF
siPOOL+HGF
**
p=0.0066
(B)
with Matrigel
Relative Wound Density (%)
Time (h)
siCTL
siETS1
siFOSL1
siSMAD3
siPOOL
siCTL+HGF
siETS1+HGF
siFOSL1+HGF
siSMAD3+HGF
siPOOL+HGF
**
p= 0.0061
(C)
relative growth
*p=0.0182
**p=0.0098
****p<0.0001
siCTL
siETS1
siFOSL1
siSMAD3
siPOOL
siCTL+HGF
siETS1+HGF
siFOSL1+HGF
siSMAD3+HGF
siPOOL+HGF
(D)
siCTL+HGF
siETS1+HGF
siFOSL1+HGF
siSMAD3+HGF
siPOOL+HGF
0 h
6 h


(A)
METwt
METwt + HGF
METex14Del
METex14Del + HGF
BMP6
MMP1
FERMT1
NGEF
TSPAN5
THBD
KRTAP2-3
DOCK4
MT1E
SH2D5
ARHGAP22
DCBLD2
ADAM8
NOG
AKAP12
NECAB2
ITGA2
MLPH
ANGPTL4
RAC2
ICA1
GSTO1
CBLN2
CALU
JAG1
SFXN3
PHLDA2
CHGB
CAVIN3
ABCA1
IL31RA
PODXL
SPRY1
MYO10
LETM2
FHL1
FBXO16
KAT14
BOK
GPC1
ANKRD33B
CRB3
FSTL1
HERC5
SCHIP1
MSLN
AMIGO2
RHBDL2
TGFB2
MAP3K8
Expression
(B)
Hallmarks
Biological process
Molecular functions
Reactome
EPITHELIAL MESENCHYMAL TRANSITION
KRAS SIGNALING UP
APICAL JUNCTION
TNFA SIGNALING VIA NFKB
INFLAMMATORY RESPONSE
HYPOXIA
IL2 STAT5 SIGNALING
ESTROGEN RESPONSE EARLY
ESTROGEN RESPONSE LATE
CELL JUNCTION ORGANIZATION
CELL PART MORPHOGENESIS
SYNAPTIC SIGNALING
REGULATION OF ION TRANSPORT
TRANSMEMBRANE RECEPTOR PROTEIN TYROSINE KINASE SIGNALING PATHWAY
CELL MORPHOGENESIS INVOLVED IN NEURON DIFFERENTIATION
POSITIVE REGULATION OF LOCOMOTION
TAXIS
BLOOD VESSEL MORPHOGENESIS
RESPONSE TO GROWTH FACTOR
CALCIUM ION BINDING
ION TRANSMEMBRANE TRANSPORTER ACTIVITY
CELL ADHESION MOLECULE BINDING
STRUCTURAL MOLECULE ACTIVITY
INORGANIC MOLECULAR ENTITY TRANSMEMBRANE TRANSPORTER ACTIVITY
ACTIN BINDING
PHOSPHOLIPID BINDING
SIGNALING RECEPTOR REGULATOR ACTIVITY
NERVOUS SYSTEM DEVELOPMENT
HEMOSTASIS
SIGNALING BY RECEPTOR TYROSINE KINASES
RHO GTPASE CYCLE
SIGNALING BY GPCR
EXTRACELLULAR MATRIX ORGANIZATION
**p<0.01
***p<0.001
Gene ratio (%)
(C)
16HBE MET WT
16HBE METex14Del
ZORG METex14Del
0 30' 8h 24h +HGF
P-MET
MET
P-ETS1 (T38)
ETS1
P-ERK
ERK
ACTIN
(D)
P-FOSL1 (S265)
FOSL1
(E)
P-SMAD3 (S208)
SMAD2
SMAD3


(A)
16HBE METex14Del
ZORG METex14Del
Trametinib
HGF
30min
8h
P-MET
MET
P-ETS1 (T38)
ETS1
P-ERK
ERK
ACTIN
(B)
FOSL1
P-FOSL1 (S265)
ACTIN
(C)
P-SMAD3 (S208)
SMAD2
SMAD3
ACTIN
(D)
without Matrigel
DMSO
TRAM
CAPM
TRAM+CAPM
**p=0.0092
***p=0.0002
****p<0.0001
Relative Wound Density (%)
DMSO+HGF
TRAM+HGF
CAPM+HGF
TRAM+CAPM+HGF
*p=0.0159
**p=0.0028
****p<0.0001
Time (h)
(E)
with Matrigel
****p<0.0001
(F)
DMSO+HGF
TRAM+HGF
CAPM+HGF
TRAM+CAPM+HGF
0 h
6 h
(G)
*p=0.0126
**p=0.0034
****p<0.0001
relative growth
DMSO
TRAM
CAPM
TRAM+CAPM
DMSO+HGF
TRAM+HGF
CAPM+HGF
TRAM+CAPM+HGF


(A)
DMSO+HGF
DMSO
CAPM
CAPM+HGF
condition
DMSO
DMSO_
CAPM
CAPM_
TRAM
TRAM_
TRAM
TRAM+HGF
DMSO
DMSO+HGF
(B)
DMSO+HGF
CAPM+HGF
TRAM+HGF
Influence
-10 -5 0 5 10
Predicted interaction
Protein-protein interaction
Transcription factor binding site
(C)
Read Counts (VST)
VIM
NOG
SERPINE2
SERPINA1
SH2D5
PTX3
CHGB
LETM2
DCBLD2
ABCA1
IRS1
SLC15A3
DMSO
DMSO+HGF
CAPM
CAPM+HGF
TRAM
TRAM+HGF

4. A coregulatory influence map of glioblastoma heterogeneity and plasticity – From cells in vitro to tumors and back again

Le glioblastome (GBM) demeure le cancer cérébral le plus fréquent et le plus létal, avec une survie médiane qui plafonne autour de 15–18 mois malgré le protocole de Stupp. Sa complexité vient d'une hétérogénéité transcriptionnelle et d'une plasticité marquées, depuis l'axe proneural–mésenchymateux jusqu'aux états « classical » et dérivés, que les approches réseau ont progressivement permis de décoder en dépassant la simple expression génique pour estimer l'activité des régulateurs maîtres au cœur des programmes tumoraux. C'est précisément sur ce terrain que s'inscrit l'article « A coregulatory influence map of glioblastoma heterogeneity and plasticity », qui propose GBM-cRegMap : une ressource intégrée combinant un réseau de corégulation spécifique aux cellules tumorales (GBM-CoRegNet) et une carte d'influence unifiée (GBM-CoRegMap) construite à partir d'une métacohorte de 1 612 tumeurs issues de 16 études, offrant un espace de référence sans effet de lot où les sous-types Verhaak et l'axe PN–MES se réorganisent de façon plus robuste, avec identification de sous-classes CL et de régulateurs candidats associés à la résistance thérapeutique et aux transitions d'état. En pratique, cette ressource crée une boucle « modèles cellulaires ↔ tumeurs » : elle aligne des données de transcriptomique hétérogènes sur un référentiel de régulation, améliore la prédiction inter-cohortes et fournit une interface web prête à l'emploi pour cartographier des données externes, explorer des régulateurs clés et rapprocher profils tumoraux et lignées in vitro.

Dans ce cadre, mes contributions structurent et amplifient la portée de l'étude. J'ai d'abord orchestré la récupération et l'harmonisation des données patient pour bâtir la métacohorte GBM, socle nécessaire à la carte d'influence et au repositionnement des états tumoraux au-delà des lots expérimentaux. Sur cette base, j'ai exploité le réseau de régulation génique inféré pour estimer, patient par patient, l'influence des régulateurs, basculant d'une lecture « gènes » à une lecture « régulateurs » mieux alignée avec la biologie. J'ai ensuite caractérisé les clusters résultants à l'aune des informations de survie et des régulateurs, afin de relier signatures de régulation, pronostic et hypothèses mécanistiques actionnables. Enfin, j'ai adapté le portail cRegMap aux données GBM : intégration de la métacohorte, des classifications de référence et des variables cliniques disponibles, afin d'offrir un outil de requêtage et de visualisation qui réduit drastiquement les frictions d'analyse pour la communauté et facilite la transposition vers des cas d'usage translationnels. Par cette combinaison d'ingénierie des

données, de modélisation réseau et de produit logiciel, l'introduction de GBM-cRegMap ne se limite pas à un « nouveau jeu de figures » ; elle installe un modèle interrogeable pour penser la plasticité du GBM à l'échelle des régulateurs et pour accélérer la génération d'hypothèses thérapeutiques, dans le prolongement des travaux qui ont établi la classification PN–MES, les sous-types CL et l'intérêt d'évaluer l'activité régulatrice plutôt que l'expression brute.

npj | precision oncology

Article

Published in partnership with The Hormel Institute, University of Minnesota

https://doi.org/10.1038/s41698-025-00890-0

A coregulatory influence map of glioblastoma heterogeneity and plasticity

Check for updates

Chloé Bernhard[1,6], Konstantinos Geles[2,6], Geoffrey Pawlak[2,6], Wajdi Dhifli[2], Aurélien Dispot[2], Jules Dusol[2], Maria Kondratova[2], Sophie Martin[1], Mélissa Messé[1], Damien Reita[1,3], David Tulasne[2], Isabelle Van Seuningen[2], Natacha Entz-Werle[1,4], Silvia Anna Ciafrè[5,7], Monique Dontenwill[1,7] & Mohamed Elati[2,7] ✉

We present GBM-cRegMap, an online resource providing a comprehensive coregulatory influence network perspective on glioblastoma (GBM) heterogeneity and plasticity. Using representation learning algorithms, we derived two components of this resource: GBM-CoRegNet, a highly specific coregulatory network of tumor cells, and GBM-CoRegMap, a unified network influence map based on 1612 tumors from 16 studies. As a widely applicable closed-loop system connecting cellular models and tumors, GBM-cRegMap will provide the GBM research community with an easy-to-use web tool (https://gbm.cregmap.com) that maps any existing or newly generated transcriptomic "query" data to a reference coregulatory network and a large-scale manifold of disease heterogeneity. Using GBM-cRegMap, we demonstrated the synergy between the two components by refining the molecular classification of GBM, identifying potential key regulators, and aligning the transcriptional profiles of tumors and in vitro models. Through the amalgamation of a vast dataset, we validated the proneural (PN)-mesenchymal (MES) axis and identified three subclasses of classical (CL) tumors: astrocyte-like (CL-A), epithelial basal-like (CL-B), and cilium-rich (CL-C). We revealed the CL-C subclass, an intermediate state demonstrating the plasticity of GBM cells along the PN-MES axis under chemotherapy. We identified key regulators, such as PAX8, and NKX2.5, potentially involved in temozolomide (TMZ) resistance. Notably, NKX2.5, more expressed in higher-grade gliomas, negatively impacts patient survival, and regulates genes involved in glucose metabolism.

Glioblastoma (GBM), the most common and aggressive tumor of the brain and central nervous system, represents a major challenge in neuro-oncology. According to the Central Brain Tumor Registry of the United States (CBTRUS) data repository[1], GBM accounts for 48.6% of malignant brain tumors and 57.7% of all gliomas, and is one of the most lethal and aggressive forms of cancer. Standard treatment involves surgery, chemotherapy and radiotherapy (Stupp protocol), but even with this treatment, most patients succumb within 15–18 months[1]. GBMs are intrinsically highly heterogeneous in terms of histological and molecular characteristics.

Previous studies attempting to classify GBM into molecular subtypes based on bulk gene expression profiling have met with various degrees of success. The Philips group first defined three major tumor subsets: proneural, mesenchymal, and proliferative[2]. Verhaak et al.[3] subsequently proposed a classification into four subtypes: proneural (PN), neural (NEU), classical (CL), and mesenchymal (MES). This classification was later revised, with exclusion of the neural subtype[4]. More recent single-cell studies have revealed additional layers of heterogeneity due to the plasticity of transcriptional programs within the same tumor and the dominating metabolic pathway. Neftel et al.[5] deconvoluted the phenotypic states of GBM cells into four major lineage-specific cellular identities: astrocyte-like (AC), mesenchymal-like (MES), neural progenitor cell-like (NPC) and oligodendrocyte progenitor cell-like (OPC). The Iavarone group[6,7] identified four

[1]UMR7021 CNRS, University of Strasbourg, Illkirch, France. [2]Univ. Lille, CNRS, Inserm, CHU Lille, UMR9020-U1277 – CANTHER – Cancer Heterogeneity Plasticity and Resistance to Therapies, Lille, F-59000, France. [3]Department of Cancer Molecular Genetics, Laboratory of Biochemistry and Molecular Biology, University Hospital of Strasbourg, 67200 Strasbourg, France. [4]Pediatric Onco-Hematology Unit, University Hospital of Strasbourg, 67098 Strasbourg, France. [5]Department of Biomedicine and Prevention, University of Rome "Tor Vergata", Rome, Italy. [6]These authors contributed equally: Chloé Bernhard, Konstantinos Geles, Geoffrey Pawlak.[7]These authors jointly supervised this work: Silvia Anna Ciafrè, Monique Dontenwill, Mohamed Elati. ✉e-mail: mohamed.elati@univ-lille.fr

THE HORMEL INSTITUTE
UNIVERSITY OF MINNESOTA

states, based on the most active pathways, along two axes: a metabolic axis including a mitochondrial (MTC) and a glycolytic/plurimetabolic (GPM) state, and a neurodevelopmental axis including a proliferative/progenitor (PPR) and a neuronal (NEU) state.

These observations suggest that one of the most promising ways to dissect GBM heterogeneity is to analyze the gene regulatory networks (GRNs) that give rise to different molecular subtypes and cell states. The computational approaches that have proved successful in this field to date are those that allow the reverse-engineered construction of context-specific GRNs (e.g., ARACNe[8] and LICORN[9,10] using a bulk-transcriptome, and SCENIC[11] using single-cell data). A second key breakthrough in regulatory network exploration made possible by CoRegNet[12,13] and VIPER[14] was the consideration of regulator activity, rather than just the production of regulators, based on evaluations of the expression of target genes with the aim of detecting master regulators. Several recent studies on GBM have inferred PN/MES subtype-specific regulatory networks, mostly from TCGA patients or single-cell data. Three important transcription factors (TFs) have been identified by SCENIC: *FOSL2, CEBPB*, and *EPAS1*, which predominate in mesenchymal cells[15]. Similarly, previous studies with ARACNe confirmed that CEBPB and STAT3, the master regulators, cooperate with *FOSL2* to mediate the PN to MES transition[16].

The application of -omics and systems biology approaches has, therefore, made a significant contribution to GBM research. However, the field now faces the challenge of dealing with data and knowledge distributed among a large number of studies and databases. Each study has entailed a marked increase in data size, complexity, and specificity. Consequently, fostering collaboration among biologists requires concerted efforts to centralize the data from these investigations into an intuitive system-level shared model. We have addressed this challenge by developing GBM-cRegMap, a robust web-based tool providing researchers with rapid access to a unified co-regulatory influence network view of GBM. GBM-cRegMap delves into representation learning, a pivotal aspect of machine learning in which algorithms extract significant patterns from raw data and transform them into more interpretable, accessible, and shared representations. The GBM-cRegMap tool incorporates the workflow illustrated in the Fig. 1, integrating data from 17 transcriptomic studies of tumors and cell lines (Supplementary Table 1). The GBM-cRegMap tool enables users to explore similarities and differences between subtypes, identify possible core regulators, detect rare subtypes, associate tumors with cell-line transcriptomic profiles, and define new targets associated with different tumor states and plasticity. GBM-cRegMap boasts an intuitive user interface (UI) that is easy to use, even for those with no expertise in bioinformatics. For all the networks and plots generated, users can run various annotations, add new data, and download the raw data required to generate plots for future analyses or publications. The wide range of computational network biology, machine learning, visualization, bioinformatics software and other resources required to achieve this are listed in Supplementary Data 1.

We demonstrated the utility of GBM-cRegMap by proposing an alternative description of GBM cancer heterogeneity based on regulatory network activities detectable across cohorts. Through the amalgamation of a vast dataset, we validated the proneural (PN)-mesenchymal (MES) axis and identified three subclasses of classical (CL) tumors: astrocyte-like (CL-A), epithelial basal-like (CL-B), and cilium-rich (CL-C). We also identified a neural normal-like subclass (NL). In addition, we identified new targets associated with different tumor states, plasticity, and resistance to therapy, showcasing the versatility and significance of GBM-cRegMap for advancing our understanding of GBM at the system level.

Results

Representation learning and evaluation of GBM-cRegMap reference components

We designed the GBM-cRegMap resource around two synergistic reference compounds: GBM-CoRegNet, a highly specific co-regulatory network of tumor cells, and GBM-CoRegMap, a unified network influence map based on 1612 tumors from 16 studies. To generate these reference compounds, we first used h-LICORN[9] to infer a gene regulatory network (GRN) from the transcriptomes of 42 GBM cancer cell lines from CCLE 2019Q1 with a curated list of regulators composed of TFs and co-factors (co-TFs) ($n = 2375$). We used transcriptomes from GBM tumor-derived cell lines to strengthen the signal specific to GBM cells and to prevent the introduction of possible bias due to the use of less homogeneous transcriptomic data from GBM tumors, which would include non-specific signals and signals associated with the tumor microenvironment. The resulting GRN was composed of 539 coregulators (Supplementary Data 2), 8269 target genes and 32360 regulatory interactions, with significant enrichment in validated TF-gene interactions and TF-binding sites ($P < 1e-100$). Based on the shared targets of each pair of TFs/co-TFs (Supplementary Data 3), the GBM-GRN was transformed into a co-regulatory network (GBM-CoRegNet) with the CoRegNet package[13]. GBM-CoRegNet (Fig. 2A) contains 2171 co-operative interactions enriched in protein-protein interactions (PPIs; $P < 1e-100$).

As for the source of GBM-CoRegMap, we compiled a GBM metacohort from 16 GBM tumor cohorts (Supplementary Table 1) and calculated sample-independent gene regulatory influences with the inferred GBM-GRN (see "Methods")[13]. We used minimalistic preprocessing pipelines for microarrays or RNA-sequencing (RNA-seq) data to achieve transparency and preprocessing parity (See Methods). The resulting metacohort influence data yielded both significant feature reduction and sample size augmentation (single GBM bulk expression cohort: <500 samples × ≃18,000 features vs. meta-cohort influence data: 1612 samples × 539 features). Based on its greater sensitivity, and its ability to capture both local and global relationships[17], we chose to use Uniform Manifold Approximation and Projection (UMAP)[18] to visualize the meta-cohort influence data in a two-dimensional embedded space, hereafter denoted "GBM-CoRegMap". We ensured that the resulting GBM-CoRegMap simulated tumor heterogeneity and did not result from potential confounding factors, by annotating the samples with Verhaak subtypes according to the corresponding published gene expression classifier. Using influence-based embedding, GBM-CoRegMap successfully reproduced a manifold that are not driven by batch effects, compared to the original expression space (Fig. 2B) and aligns with the known heterogeneity of GBM (Fig. 2C). Furthermore, relative to gene expression, this influence-based Verhaak subtype classifier yielded a mean 33% increase in cross-batch prediction performance, as evaluated by calculating the area under the curve (AUC) with the PAMR (prediction analysis for microarrays) classifier (Fig. 2D and Supplementary Table 2).

Reference-mapping and UI features of GBM-cRegMap

Gene expression profiling is routinely used by the GBM community to understand the disease. However, interpreting processed data to gain insights into biological mechanisms remains challenging, particularly for non-bioinformaticians. Many data visualization and network analysis tools require extensive data formatting, parameter tuning, significant running time, and are impractical with a limited number of samples. The concept of mapping newly generated biological "query" data to reliable references is a powerful idea that predates transcriptomic and network analysis. To advance this approach, we have created GBM-cRegMap (https://gbm.cregmap.com, http://github.com/i3bionet/gbm-cregmap), an open-source software package and web tool with GBM-CoRegNet and GBM-CoRegMap references and an interface for user-friendly analysis. The UI is composed of three functional components (see Fig. 1): "CoRegNet", "CoRegMap", and "CoRegQuery". The GBM-CoRegNet network is visualized via a Shiny applet in the R package visNetwork[19]. The CoRegNet network representations are intuitively understandable to biologists (Fig. 2A). They also facilitate the integration of several layers of information over nodes and edges. The CoRegNet tab allows users to investigate and evaluate the extent to which phenotypes of interest (e.g., tumor subtypes, wild-type vs. treated cell lines), each characterized by a set of active coregulators, colocalize within the reference GBM-CoRegNet network through an intuitive visualization. The second feature is GBM-CoRegMap, a unified influence map for 1612 tumors from 16 studies, enhanced by the inclusion of downstream analysis data, such as annotation of the map with molecular subtype, tumor purity,

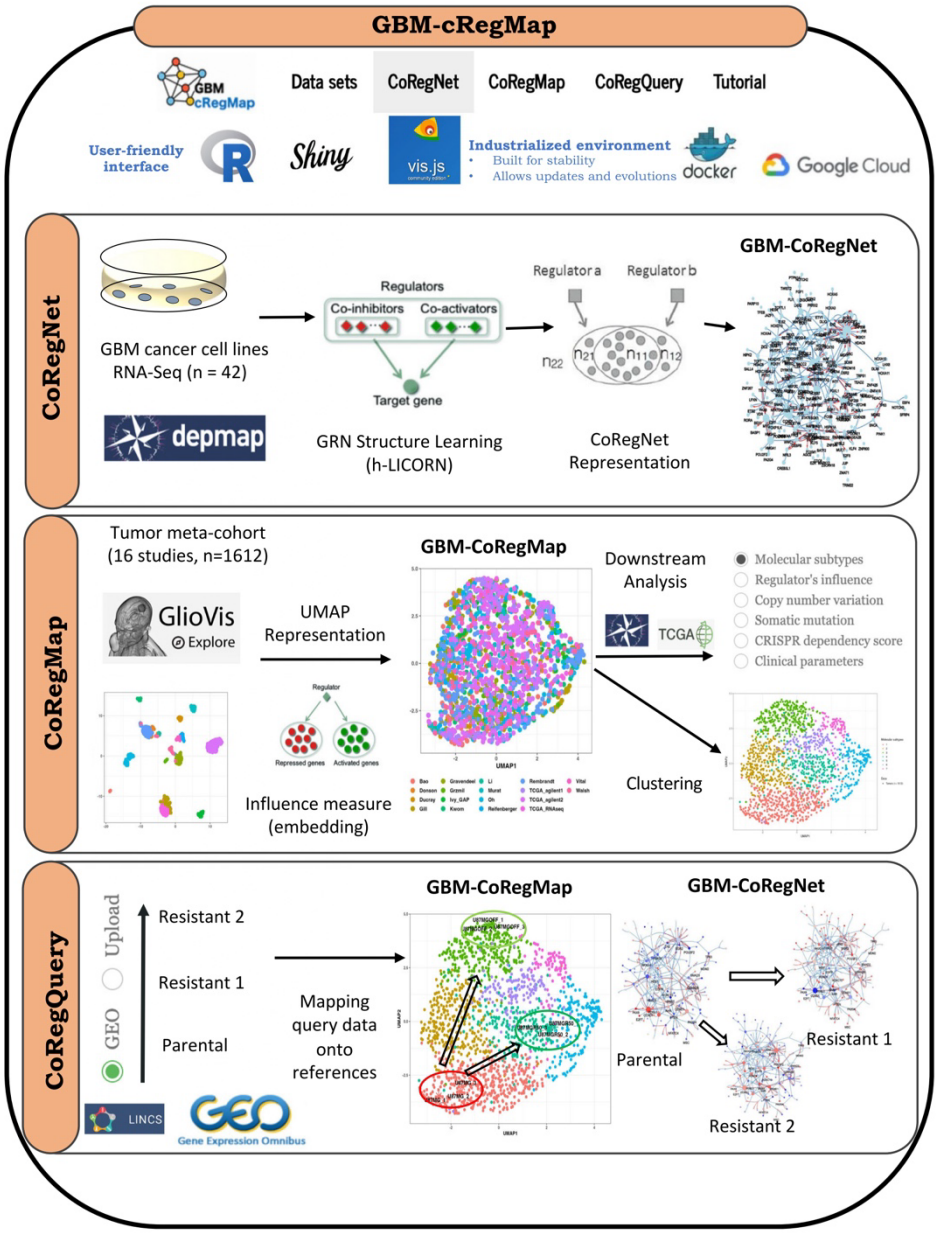
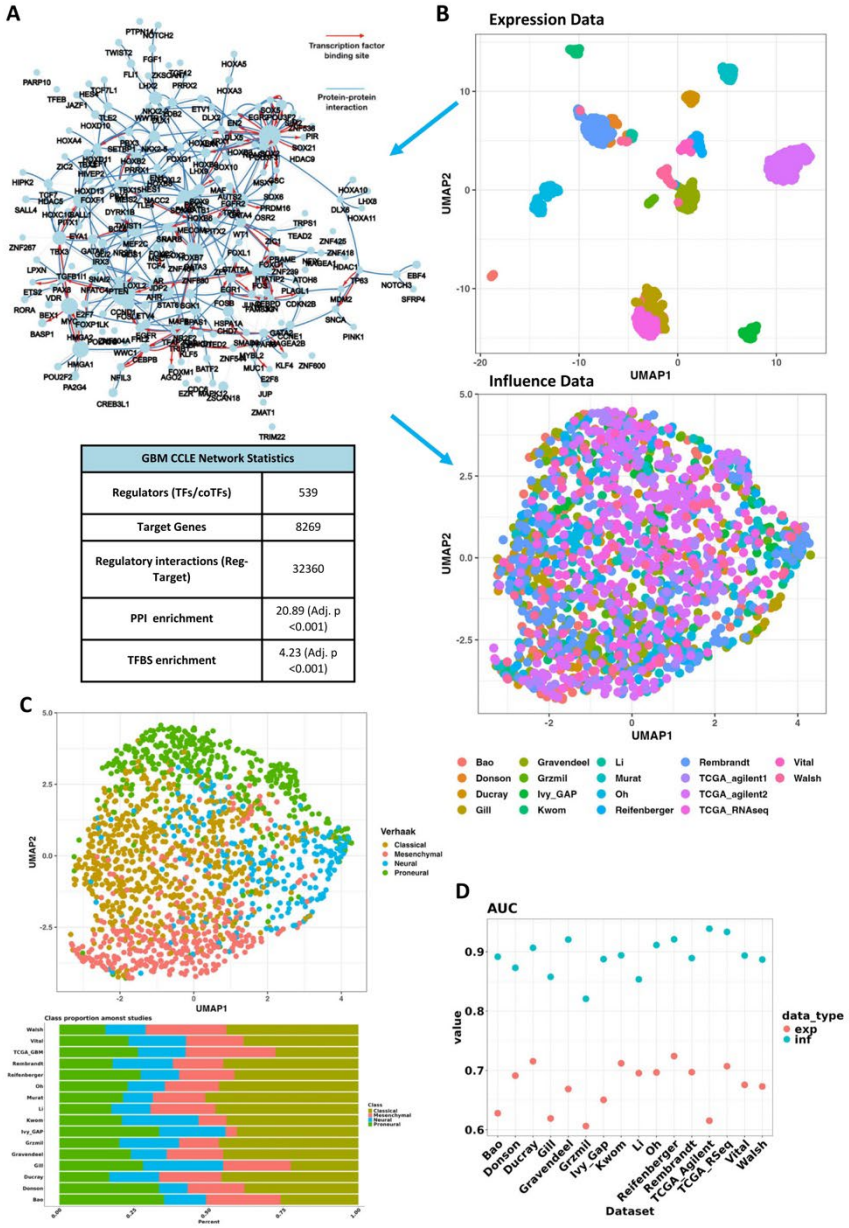


Fig. 1 | Overview of the GBM-cRegMap framework (from cells in vitro to tumors and back again) and its user interface (UI) features: CoRegNet, CoRegMap, and CoRegQuery. The GBM-cRegMap tool integrates three key components within its user interface: CoRegNet, CoRegMap, and CoRegQuery. These elements enable the exploration of glioblastoma (GBM) co-regulatory networks, the analysis of disease state landscapes, and the querying of user-provided data, respectively. The framework leverages a wide range of computational resources, including network biology algorithms, machine learning models, visualization tools, and bioinformatics software, as detailed in Supplementary Data 1. The 17 transcriptomic datasets used to generate the GBM-cRegMap reference compounds (GBM-CoRegNet and GBM-CoRegMap) are provided in Supplementary Table 1.



GBM CCLE Network Statistics	
Regulators (TFs/coTFs)	539
Target Genes	8269
Regulatory interactions (Reg-Target)	32360
PPI enrichment	20.89 (Adj. p <0.001)
TFBS enrichment	4.23 (Adj. p <0.001)

regulator influence, copy number variation, somatic mutations, clinical data, or gene essentiality scores. It can also be used for the alignment and colocalization of patient-derived cell models within the reference tumor heterogeneity map. Finally, query datasets can be mapped to GBM-cRegMap references with the "CoRegQuery" tab. GBM-cRegMap allows users to upload their expression datasets or to retrieve expression datasets from the Gene Expression Omnibus (GEO) with ease. As input, expression data can be uploaded in CSV format or as a GEO series ID, also known as a GEO accession code, from which GBM-cRegMap extracts the gene expression dataset and experimental information. GBM-cRegMap rapidly calculates regulator influences for new data, presenting the results in the GBM-CoRegNet compound and making it possible to investigate selected

Fig. 2 | Evaluation of GBM-cRegMap reference components. A Graphical representation (top) of the co-operativity network inferred from the transcriptome of 42 GBM cancer-derived cell lines (GBM-CoRegNet). Nodes represent TFs and co-TFs. Node size is proportional to the number of targets of the TF/co-TF. Co-regulatory interactions between nodes are indicated: protein-protein interactions with published evidence are shown in blue, transcriptional regulation interactions with published evidence are shown in red, and interactions defined solely by the h-LICORN algorithm are shown in gray. The table (bottom) summarize metrics from the GBM-CoRegNet. **B** Uniform Manifold Approximation and Projection (UMAP) visualization of the metacohort of 1612 tumors, illustrating the difference between the use of transcriptomic (top) and influence (bottom) data from different datasets. Influence data produce a batch effect-free metacohort from the original RNA expression values. **C** UMAP visualization of the GBM-cRegMap metacohort, using sample annotation colors derived from the Verhaak classification. **D** The Area Under the Curve (AUC) for the resulted classifier of Verhaak subtypes using gene expression or influence in cross-batch prediction. This cross-batch prediction involves training the classifier on one dataset and testing it on others. Used classifier: prediction analysis for microarrays (PAMR).

coregulators, interactions, and systems further for relevance to the study. Furthermore, a supervised machine learning procedure (see "Methods") increases the efficiency of the tool for predicting the localization of the query dataset samples within GBM-CoRegMap. This ensures consistent comparisons of tumor transcriptional patterns, facilitating potential explorations of tumor heterogeneity and the adaptability of the studied phenotypes.

A meta-cohort analysis of 1612 tumors refining the GBM molecular classification into seven classes of biological and clinical relevance

Consistent with other publications on GBM[4,6], the resulting metacohort influence data (1612 tumors × 539 regulators) demonstrate that the MES/PN subtypes were the most homogeneous, whereas the CL and NEU subtypes were the least homogeneous, as assessed by the lowest silhouette scores and UMAP visualization (Fig. 2C). This highlights the opportunity for a refinement of the Verhaak classification. The large sample size of GBM-CoRegMap provides greater statistical power for subtype identification and minimizes sampling bias[20].

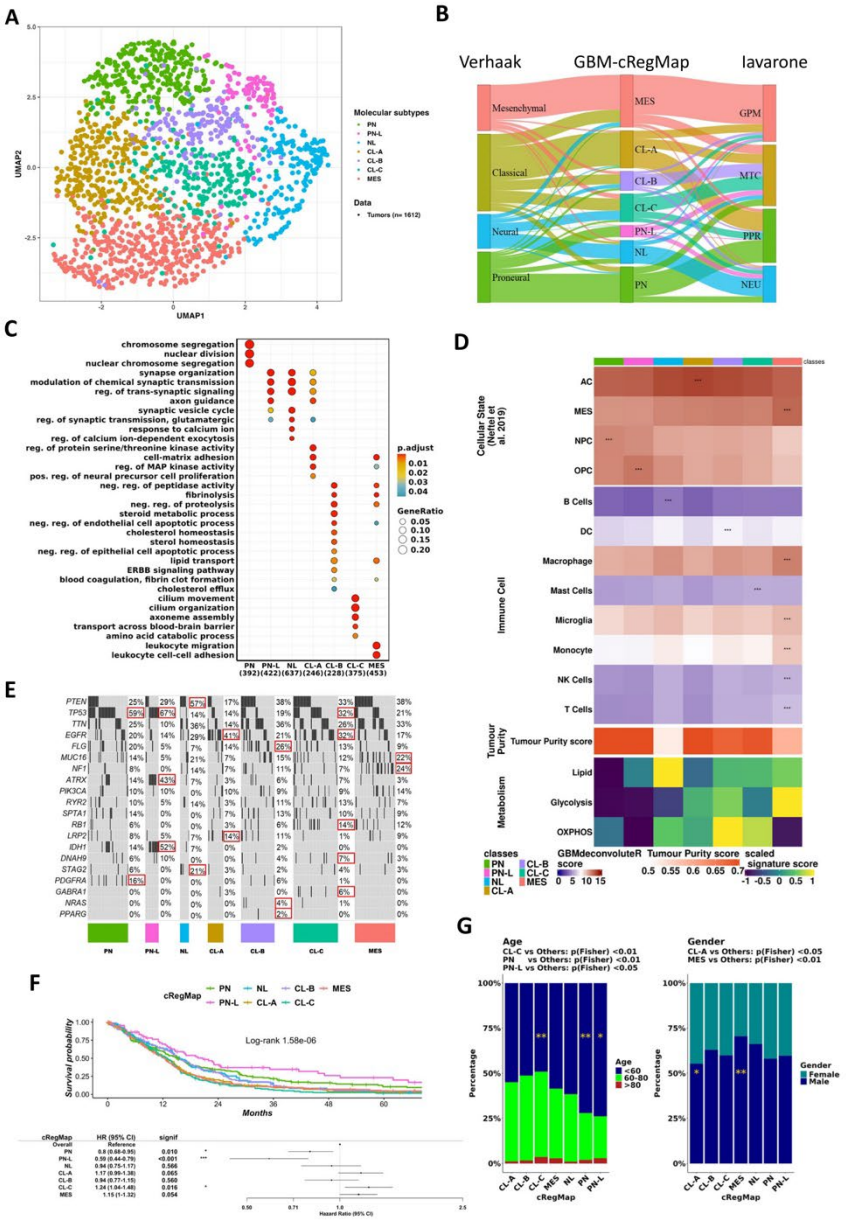
GBM-cRegMap-based Seurat clustering (see "Methods") resulted in seven distinct molecular classes (Fig. 3A). These classes were named on the basis of prior Verhaak subtyping (Fig. 3B). Sample size differed between classes, as follows: 17% proneural (PN), 6% proneural-low proliferative (PN-L), 11% neural normal-like (NL), 18% classical astrocyte-like (CL-A), 9% classical epithelial basal-like (CL-B), 14% classical cilium-rich (CL-C), and 25% mesenchymal (MES). The proportions of the seven GBM-cRegMap subtypes were almost identical in the different cohorts indicating the stability of the classification. We first characterized the GBM-cRegMap classes at the transcriptomic and genomic levels. At the transcriptomic level, we performed gene ontology enrichment analysis (Fig. 3C) and deconvolution analysis (Fig. 3D). At the genomic level, we analyzed mutation alterations associated with the GBM-cRegMap classes (Fig. 3E). Subsequently, we evaluated their clinical outcomes (Fig. 3F) and other relevant clinical data (Fig. 3G).

The PN, PN-L, and NL subclasses were characterized principally by attributes of developmental processes, displaying proneural, gliogenesis, and neural signatures. The PN and PN-L subclasses accounted for the majority of PN-Verhaak tumors, indicating a further subdivision of the proneural subtype into two distinct entities. These two subclasses were also associated with progenitor cell states (NPC- and OPC-like, respectively), as described by Neftel et al.[5]. The PN subclass was enriched in cell division and mitotic cell cycle pathways and displayed the highest levels of expression for the proliferation markers *MKI67* and *DLGAP5*. By contrast, PN-L tumors had low levels of proliferation. Both subclasses had a higher frequency of *TP53* mutations (>59%) than the other subtypes (0–32%). *PDGFRA* mutations were almost exclusive to the PN subclass (PN, 16%; PN-L, 0%), whereas *IDH1* mutations were more frequent in the PN-L subclass (PN-L, 52%; PN, 14%; NL, 7%; others, ~0%). Among IDH1-mutated tumors, the majority were distributed between the PN and PN-L classes at 35% and 55%, respectively. Although the 2021 WHO-CNS21 classification no longer categorizes IDH-mutant gliomas as GBM, we have chosen to retain IDH1-mutant gliomas in our analysis to ensure consistency across datasets and alignment with historical GBM molecular classifications. The NL subclass accounted for 42% and 49% of NEU_Verhaak and NEU_Iavarone subtype tumors, respectively. This subclass included a distinct cluster of GBM with a 'normal-like' transcriptomic profile enriched in neuron development-associated pathways and neuroglioma synaptic communication. We identified a common gene signature with the newly identified 'normal-like' IDH-WT subtype reported by Nguyen et al.[21], which included genes such as *SLC32A1* (vesicular gamma-aminobutyric acid transporter) and *SYT5* (synaptotagmin 5). NL tumors harbored *PTEN* mutations in 57% of the cases, and show a higher frequency (21%) of *STAG2* mutations compared to other subclasses. This latter finding is intriguing, in view of the role of STAG2 as a tumor suppressor linked to therapeutic response to PARP inhibitors[22]. Moreover, STAG2 has been recently shown to modulate sensitivity to pharmacologic EZH2 inhibition in GBM[23].

The CL-A, CL-B, and CL-C subclasses accounted for the majority of CL_Verhaak tumors. Unlike the CL-A and CL-B subclasses, the CL-C subclass was evenly distributed across the CL and NEU_Verhaak subtypes (Fig. 3B). Importantly, these three subclasses define previously unknown GBM subtypes based on tumor microenvironment and metabolic, mutational, and clinical information. The CL-A subclass was enriched in an astrocyte-like signature (Fig. 3D, Cellular State score[24]). It also displayed an upregulation of *MEOX2* homeobox TF and fatty-acid synthesis genes (*ELOVL2, ACSL3, PLA2G5*), and high levels of expression of the stem cell markers *NES* and *SOX9* (Supplementary Data 4). CL-B displayed enrichment in factors associated with stem cell survival (*SERPINB3*, *NANOG*, *SALL4*), the coagulation pathway, keratinization, and cell-cell adhesion—features commonly expressed in basal squamous epithelial cells[25–27]. Notably, pseudo-epithelial and epithelial morphologies, although uncommon in GBM, are recognized as subtypes in the 2021 WHO classification of CNS tumors[21] and are potentially associated with a poor prognosis[28].

Furthermore, CL-B subclass overlap with the MTC_Iavarone state (Fig. 3B), suggesting an association with a mitochondrial subtype, confirmed by the enrichment of CL-B subclass in cholesterol homeostasis and metabolism of steroids (Fig. 3C and Supplementary Data 5) and the oxidative phosphorylation (OXPHOS) pathway signature (Fig. 3D). This link is supported by shared metabolic intermediates and regulatory crosstalk between OXPHOS and cholesterol pathways in biological systems[29]. Notably, recent research underscores the critical role of cholesterol homeostasis in glioma stem cells[30]. Hence, targeting mitochondrial function and lysosomal cholesterol homeostasis could offer a promising therapeutic strategy for CL-B tumors, potentially leveraging specific mitochondrial inhibitors[7].

Unlike CL-A and CL-C tumors, CL-B tumors did not overexpress EGFR and had lower *EGFR* mutation rates (CL-A, 41%; CL-B, 21%; CL-C, 32%). By contrast, they displayed high rates of *FLG* mutation (CL-A, 14%; CL-B, 26%; CL-C, 13%) and were the only tumors in which infrequent *NRAS* (4%) and *PPARG* (2%) mutations are associated. Surprisingly, the CL-C (cilium-rich) subclass was found to be enriched in processes linked to cilia, including cilium assembly, movement, and organization (Fig. 2C), along with OXPHOS, lipid metabolism, and dopamine metabolism (Fig. 3D and Supplementary Data 5). The transcriptional landscape of CL-C tumors was further shaped by the activation of pivotal cell signaling pathways, mediated by *STAT3* and *RPS6KA1*, for example, and epigenetic regulators, such as *HDAC1, HDAC5, CBX4*, and *SUV39H2*. The CL-C subclass was also characterized by the upregulation of several prominent genes, including *DDIT4L*, *LGR6* (known to facilitate DNA repair and chemoresistance[31,32]), and *ETNPPL* (a negative regulator of glioma growth[33]). Unlike the other two



classical subclasses, CL-C tumors had a weaker proliferative signature ($P < 0.01$) and higher rates of *TP53* (CL-A, 14%; CL-B, 19%; CL-C, 32%) and *DNAH9* (CL-C, 7%) mutation (Fig. 3E). Recent studies have linked suppressed ciliogenesis in glioma stem cells to continuous self-renewal, whereas the restoration of ciliogenesis shifts glioma stem cells towards the differentiation state[34]. In addition, it has very recently been shown that primary cilia, through the regulation of the intracellular release of IL-6, contribute to GBM-driven immunosuppression[35]. Finally, the MES subclass was closely associated with the MES_Verhaak subtype and the GPM_Iavarone state (Fig. 3B). This association reflects the features linked to a mesenchymal identity reported, such as higher mutation rates and low levels of *NF1* expression[3]. The MES subclass is enriched in the epithelial-mesenchymal

Fig. 3 | Molecular and clinical characteristics of GBM-cRegMap classes. A UMAP visualization of the GBM-cRegMap metacohort, using sample annotation colors derived from the GBM-cRegMap classification. **B** Alluvial plot displaying the GBM-cRegMap metacohort samples with state-of-the-art classifications of GBM (Verhaak[3] and Iavarone[7] classifications on the left and right, respectively) and the corresponding classification proposed by GBM-cRegMap (in the middle). **C** Dot plot with significantly enriched Gene Ontology (GO) and Biological Processes (BP) for each GBM-cRegMap class derived from the clusterProfiler R package. GO/BPs with an adjusted P value < 0.05 were considered; P values were adjusted by the Benjamini–Hochberg method. **D** Heatmap displaying the deconvolution of cellular component scores with the GBMdeconvoluteR tool and metabolic gene signature scores with the singscore R package. Asterisks indicate the highest score per row. **E** Oncoplot showing the most important genomic alterations associated with cRegMap classes, based on available data from the TCGA-GBM dataset. Red squares indicate the highest rates of alteration per gene for each class. **F** Survival analysis and forest plot for cRegMap classes. Top: Kaplan–Meier overall survival curves stratified by cRegMap class. Survival analysis was performed using a Cox regression model and visualized with the Kaplan–Meier method via the R package *survival*. Bottom: Univariate Cox proportional hazards regression analysis comparing cRegMap groups with respect to overall survival in the metacohort. Results are displayed as a forest plot. **G** Gender and age differences in GBM-cRegMap subclasses.

transition (EMT) and immune-associated pathways (Fig. 3C). It also displayed enrichment in glycolysis/hypoxia-related functions and lipid metabolism but no enrichment in the OXPHOS signature (Fig. 3D and Supplementary Data 5).

For characterization of the cellular components of the tumor microenvironment in each GBM subclass, we inferred the fraction of cellular states and stromal/immune cells with GBMdeconvoluteR[24] and tumor cell purity by applying PUREE[36]. PN tumors were enriched in neural progenitor-like (NPC) cells, whereas PN-L tumors were associated with the presence of oligodendrocytes. CL-A tumors were enriched in astrocyte-like cells. The CL-B subtype had a strong activated dendritic cell gene signature, suggesting that this subtype could potentially be treated with dendritic cell vaccines. CL-C tumors were enriched in mast cells. MES tumors displayed infiltration with macrophages, monocytes, and other immune cells. The NL subclass was associated with the presence of B cells (Fig. 3D), consistent with previous findings[4], suggesting a distinctive immunological profile for NL gliomas. Tumor cell purity was lowest for NL, followed by MES and CL-B tumors, and tumor cell purity was highest for PN, CL-A, and CL-C tumors (Fig. 3D).

We evaluated the clinical impact of our refined GBM classification by analyzing outcomes across subgroups using hazard ratio analyses and log-rank tests. Our classification (Fig. 3F) was compared to Verhaak's reference classification (Supplementary Fig. 1). The global P value from the log-rank tests indicated significant overall survival differences among all subclasses (P = 1.58e-06), with incremental improvement compared to Verhaak's classification (P = 3.16e-05). Among the subgroups, the PN-L subclass exhibited the best overall survival (hazard ratio: 0.59 [0.44–0.79], P < 0.001), while the CL-C subclass had the poorest outcomes (hazard ratio: 1.24 [1.04–1.48], P = 0.01). Notably, Verhaak and Iavarone grouped the CL-B and CL-C subclasses together as CL and MTC, respectively, but our refined classification revealed a distinct survival difference between these two subclasses (CL-B hazard ratio: 1.24 [1.04–1.48]). To further validate our findings, we excluded IDH1-mutated tumors, confirming that the survival differences observed for the PN and PN-L subclasses are specific to GBM (Supplementary Fig. 2). We also confirmed the reported associations with sex and age (Fig. 3G and Supplementary Table 3), including the over-representation of MES subclass tumors among samples from male patients and the lower median age of patients for the PN and PN-L subclasses than for the other subclasses. Interestingly, the CL-A group displayed no sex-related differences with respect to the other subgroups.

Functional validation of subtypes-specific co-regulated subnetworks through DepMap database

We investigated the subtype-specific active co-regulatory networks by first inferring differentially Influencing regulators (DIRs) with the Seurat::findmarker function[37] and the default Wilcoxon rank-sum test (Fig. 4A and Supplementary Data 6). We then used publicly available data regarding the effect of CRISPR-Cas9-mediated knockout (Broad Institute, AVANA CRISPR-Cas9 dataset DepMap 2019Q4)[38] of the selected regulators in our defined GBM-derived cell lines for each subclass (Fig. 4B and Supplementary Data 7). As a starting step, we showed that we could employ CoRegQuery to assign a specific class with high confidence (class probability >0.75) and robustness (using various transcriptomic profiling techniques, including microarray and RNA-seq sourced from six different datasets) to 33 of the 42 GBM cancer cell lines (GBCCL), thus revealing a heterogeneous pattern of subclasses across the GBM cell-line panel (Supplementary Data 7). This analysis also showed that GBM cell lines can be subtyped in five out of seven classes, with the exception of the PN-L and the NL ones. Thus, we could not integrate the CRISPR-Cas9-mediated knockout data for these subtypes. In clinical samples, the PN co-regulatory subnetwork confirmed the highly proliferative state of these tumors, visible on enrichment in several regulators predominantly involved in cell division, cell cycle, and proliferation such *CDC6, FOXM1, FOSL1, BIRC5, E2F7, PA2G4, DEPDC1, ATAD2*, and *TFAP2C*. In parallel, CRISPR-Cas9 AVANA 19Q4 analysis of the PN cell lines revealed a strong dependence on *BRCA2* and *FOXM1*, and a moderate but significant dependence on *GLI2* and *PBX4* (Fig. 4C and Supplementary Data 6). The highest dependence of this subtype was for *PRMT5*, which appears as a strong regulator in this subtype and in PN-L and NL subtypes too. The PN-L class of tumors was characterized by high levels of activity of TFs showing high influence in the PN and the NL subtypes too, such as *SALL2, ZNF704, CAMK4*, and *CREB3L4*, but also by the uniquely high activity of *GBX2*, the Gastrulation Brain homeobox 2, recently described to act as a tumor suppressor in GBM cells, and inversely correlated to GBM grade[39]. This is in agreement with the most favorable clinical outcome shown by PN-L subtype of GBM. The NL co-regulated subnetwork comprised the *PRDM16, SP100, TNIP1, ZFP3, MEIS3, POU3F3, SNCA, SMAD3 and TRB23* regulators. The products of genes such as *PRDM16* and *MEIS3* are instrumental in cell-lineage development and fate decisions. *PRDM16*, in particular, is linked to stem-cell maintenance and cell differentiation[40]. *SP100* and *SNCA* are correlated with lower levels of malignancy in GBM cell lines[41,42], suggesting a potential role in mitigating malignancy in tumors of this class. Moreover, in the NL class, *SMAD3* has emerged as a regulator associated with favorable prognosis. Its inhibition, coupled with stimulation of the neuroglioma circuit, accelerates tumor progression[43]. Unsurprisingly, *SNAI2, PAX8, FOSL2, EPAS1, and RUNX2* were associated not only with the top-ranking influence scores for MES-specific regulators but also with the top dependency scores in MES GBM cell lines, highlighting their known roles in the mesenchymal state. Although *EGFR* alteration is a well-established marker of the CL subtype, our analysis revealed a notable regulatory role of EGFR in MES GBM, where it exhibited a significant influence and dependency score. This observation aligns with recent studies showing that EGFR activity contributes to mesenchymal-associated processes[44,45]. In addition, our co-regulated subnetwork for MES tumors confirmed the established role of the hypoxic tumor microenvironment, as suggested by the inclusion of *EPAS1* and *CEBPD*, both master TFs for hypoxia-regulated genes in GBM[46,47]. CRISPR-Cas9 AVANA 19Q4 analysis of the MES cell lines also revealed novel genes with a high influence and a strong dependence score, such as *ARHGEF5*, encoding for the Rho guanine nucleotide exchange factor 5, involved in the EMT process in other types of tumors[48], but never studied in GBM.

By contrast to the PN/MES axis, few TFs/co-TFs have been characterized so far as key regulators of CL_Verhaak tumors. Our analysis revealed that the CL-A co-regulated subnetwork included epigenetic modifiers, such as *DLX1* and *HDAC10*, as well as *NKX2.2* and *TLE2* (Fig. 4A

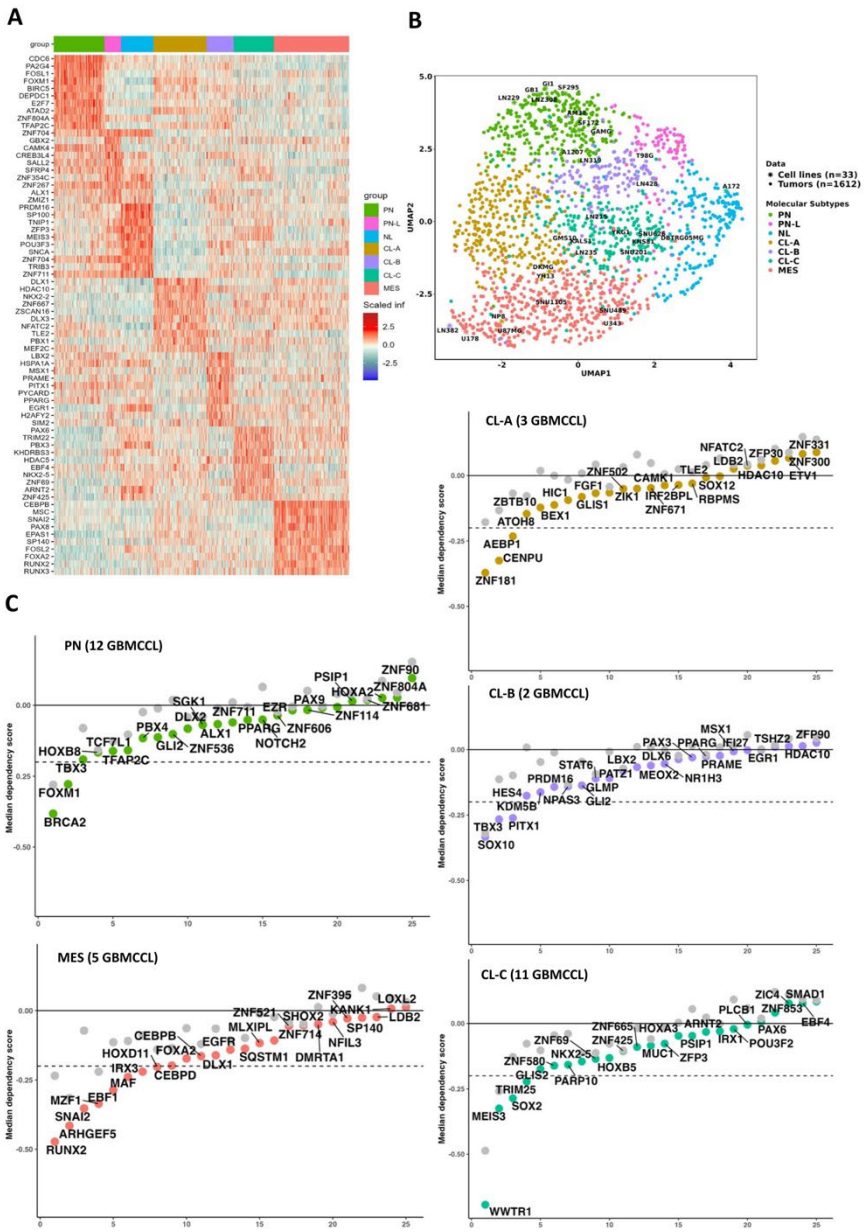


Fig. 4 | Identification of GBM subclass-matched cell-line repertoires and functional validation of specific differential subnetworks. A Heatmap showing the top 10 differentially Influencing regulators for each subclass. **B** Colocalization of the 42 GBM cancer cell lines on GBM-CoRegMap, as predicted by CoRegQuery, with labeling (color) according to the SVM-based classifier. **C** Impact on cell viability (CERES dependency score) for each of the predicted subclasses. For each regulator, two values are shown vertically: the colored dot is the mean CERES score for the cell lines assigned to the subclass, and the gray dot is the mean CERES score for cell lines not assigned to the subclass. A gene may have a negative CERES score (indicating that its knockout would reduce the survival of the corresponding cell line) or a positive score (indicating that its deletion might increase the survival of the cell line). The dotted horizontal line indicates the viability threshold used.

and Supplementary Data 6). However, the CRISPR-Cas9 analysis of CL-A cell lines uncovered a significant dependence on *ZNF181*, also showing a high influence score in CL-A samples, and of *CENPU*, encoding for the centromere-associated protein U, whose role has never been described in GBM so far (Fig. 4C and Supplementary Data 6). The CL-B co-regulated subnetwork included *PRAME, PPARG, SALL1*, and *STAT6*. CRISPR-Cas9 analysis confirmed that the main regulators of CL-B cell lines are *PITX1, HES4*, and *SOX10*. Notably, the CL-B tumor subclass displays high levels of activity for the TF *PPARG*, a master metabolic regulator implicated in tumor stromal-epithelial crosstalk and carcinogenesis[49]. Furthermore, the CL-B class relies heavily on the *PRAME* regulator (Fig. 4C), an oncoprotein linked to advanced tumor stages and poor prognosis, highlighting its clinical relevance[50]. Both *PPARG* and *PRAME* are among the TFs listed in the results of CRISPR-Cas9 screening in CL-B cell lines, further confirming the relevance of the data we obtained in the in vitro models (GBCCL), and its overlap with clinical sample results. The CL-C co-regulated subnetwork included *PAX6, HDAC5, POU3F2, SOX2, NKX2.5, EBF4, MEIS3*, and *GLIS2*. In CL-C cell lines, the CRISPR-Cas9 analysis (Fig. 4C) indicated *SOX2*, *GLIS2*, and *NKX2.5* as deeply involved in survival, together with *WWTR1* and *MEIS3*. While the involvement of the core stem cell TF *SOX2*[51,52] and of the well-known regulator of stemness and cell plasticity *WWTR1* are expected findings, our analysis uncovered the unprecedented importance of *NKX2.5*, *MEIS3*, and *GLIS2* in this subclass.

Note that the GBM-cRegMap tool allows the user to visualize active co-regulated subnetworks and to explore any gene viability score, while also providing the calculated influence scores. As a representative example of this application, Fig. 5 depicts the MES co-regulated subnetwork, where the *SNAI2* TF stands out for its influence, and, on the bottom, the influence and dependency score maps of *SNAI2* in the GBM cell lines.

Uncovering intermediate states in the PN-MES axis involved in chemotherapy resistance

Recent data have demonstrated the high degree of plasticity of GBM cells along a PN-MES axis[53]. Transitions between cellular states include intermediate states, especially during therapies[54]. We applied GBM-CoRegQuery to analyze transcriptomic data (GSE253458: $n = 9$ samples, Fig. 6A), focusing on the comparison between the untreated U87MG glioma cell line and its Temozolomide (TMZ)-resistant variants, U87MG-R50 (continuously exposed to 50 µM TMZ) and U87MG-OFF (representing the post-TMZ removal condition)[55]. Unlike the Verhaak classification, the three cell lines were classified into three clearly distinct GBM-cRegMap subclasses (Fig. 6B and Supplementary Table 4): U87MG to the MES subclass (as already described), U87MG-R50 to the CL-C subclass, and U87MG-OFF to the PN subclass. This categorization highlights the dynamic transcriptional change of GBM cells' behavior in response to chemotherapy and subsequent withdrawal of treatment, suggesting that CL-C subclass may represent an intermediate state in the PN-MES axis. The classification of U87MG-R50 within the CL-C subclass, characterized by a low proliferative signature, aligns with our prior in vitro studies, showing that U87MG-R50 exhibits a reduced proliferation rate compared to both U87MG and U87MG-OFF[55]. A crucial aspect of our analysis was the identification of regulators involved in the resistant states (Fig. 6C). In response to TMZ treatment and acquired resistance, the differentially activated subnetwork included MDM2 TF activation and corroborates our prior research on MDM2's involvement in TMZ resistance via the p53 pathway[55]. Focusing on U87MG-R50/CL-C subclass hallmarks, the influence of *PAX8* was reduced while that of *NKX2.5* grew in these cells, compared with non-treated U87MG (Fig. 6D). In the U87MG-OFF cells, characterized by TMZ resistance after treatment removal, we observed a tendency of both PAX8 and NKX2.5 influence to return to untreated U87MG levels (Fig. 6D).

Notably, the mRNA expression level of NKX2.5 correlated with its estimated influence but did not significantly differ across the three cell lines. In contrast, the mRNA expression level of PAX8 did not correlate with its estimated influence. Therefore, measuring the influence effect of regulators provides a new level of data interpretation. In line with this, we experimentally confirmed that PAX8 protein expression levels aligned with its predicted influence (Fig. 7A). Consequently, assessing the influence of master regulators provides a new level of data interpretation. Although very few explored, *PAX8* is detected in high-grade glioma and participates to glioma cell survival[56] and is related with EMT in other cancers[57]. *NKX2.5* is a TF essentially involved in cardiac development[58], and its expression/function has never been described in GBM so far. Interestingly, when we analyzed its mRNA expression levels in the CGGA (Chinese Glioma Genome Atlas) cohort[59] of primary ($n = 651$) and recurrent ($n = 332$) gliomas, we found that *NKX2.5* is significantly more expressed in grade IV gliomas compared to both grade III and grade II, and this difference holds true not only in primary but also in recurrent tumors (Fig. 7B). In the latter ones, a higher expression of *NKX2.5* negatively correlates with patients' survival, and particularly in recurrent GBM (Fig. 7C). This is particularly intriguing in view of the formation of TMZ-resistant, aggressive tumors after long-term treatments with this drug. Gene ontology (GO) enrichment analysis of the 282 *NKX2.5* putative target genes (175 activated and 107 repressed) revealed that *NKX2.5* negatively regulates genes implicated in glucose metabolism/glucose catabolic process to pyruvate (e.g., *HK2, GPI, GAPDH, LDHA*), (Fig. 7D and Supplementary Data 8). While endogenous validation of NKX2.5's role in U87MG-resistant cells is technically challenging due to its low expression, we instead verified the impact of NKX2.5 overexpression on the repression of predicted glycolytic genes in U87MG cells. A statistically significant decrease of all 4 genes mRNA was obtained accompanied by a decrease of the corresponding HK2 protein for example (Fig. 7E).

NKX2.5 was identified through GBM-cRegMap as a key player in metabolic reprogramming in GBM, warranting further study. Additional investigations are necessary to fully elucidate the molecular mechanisms underlying this regulation.

Furthermore, the metabolic profiles of our cell lines, as determined by extracellular flux analysis (Fig. 7F), were consistent with the metabolic signatures of their assigned subclasses. U87MG displayed a predominantly glycolytic profile, aligning with the glycolytic signature of the MES class. U87MG-R50 exhibited an OXPHOS-dominant profile, in line with the metabolic characteristics of the CL-C subclass. In contrast, U87MG-OFF was metabolically less active, reflecting the repressed metabolic signature associated with the PN subclass. Taking together, these results confirm the CL-C class represented by U87MG R50 as an important intermediate class in the response to therapies and shows that the associated transcriptomic program can be easily modeled in co-regulatory network, which is based on TF/coTFs influence and can therefore be used to identify putative master regulators of biological traits.

However, a limitation of our study is the absence of in vivo models, such as patient-derived xenografts (PDX), which would better capture the complexity of tumor-microenvironment interactions and provide further insight into the relevance of the CL-C class in the context of in vivo TMZ resistance.

Discussion

We have developed GBM-cRegMap, a powerful web-based tool, to provide researchers with rapid access to a unified co-regulatory influence network view of GBM. Many of the software tools used are based on techniques originating from artificial intelligence (AI), especially representation learning: h-LICORN[9], CoRegNet[13], LatNet[60], and Seurat[37] (Fig. 1 and Supplementary Data 1). We describe here the first-ever use of transcriptomes from GBM tumor-derived cell lines to produce a more reliable GBM-CoRegNet and to prevent the introduction of possible bias due to the use of less homogeneous transcriptomic data from GBM tumors presenting non-specific signals and signals associated with the tumor microenvironment. We then demonstrated the relevance of building a harmonized tumor heterogeneity map, GBM-CoRegMap, encompassing a meta-cohort analysis of 1612 tumors from 16 studies. We demonstrated that GBM-CoRegMap successfully removed batch effects and reproduced the known heterogeneity of GBM. More importantly, our comprehensive GBM-CoRegMap gave rise to a novel molecular classification of GBM into seven

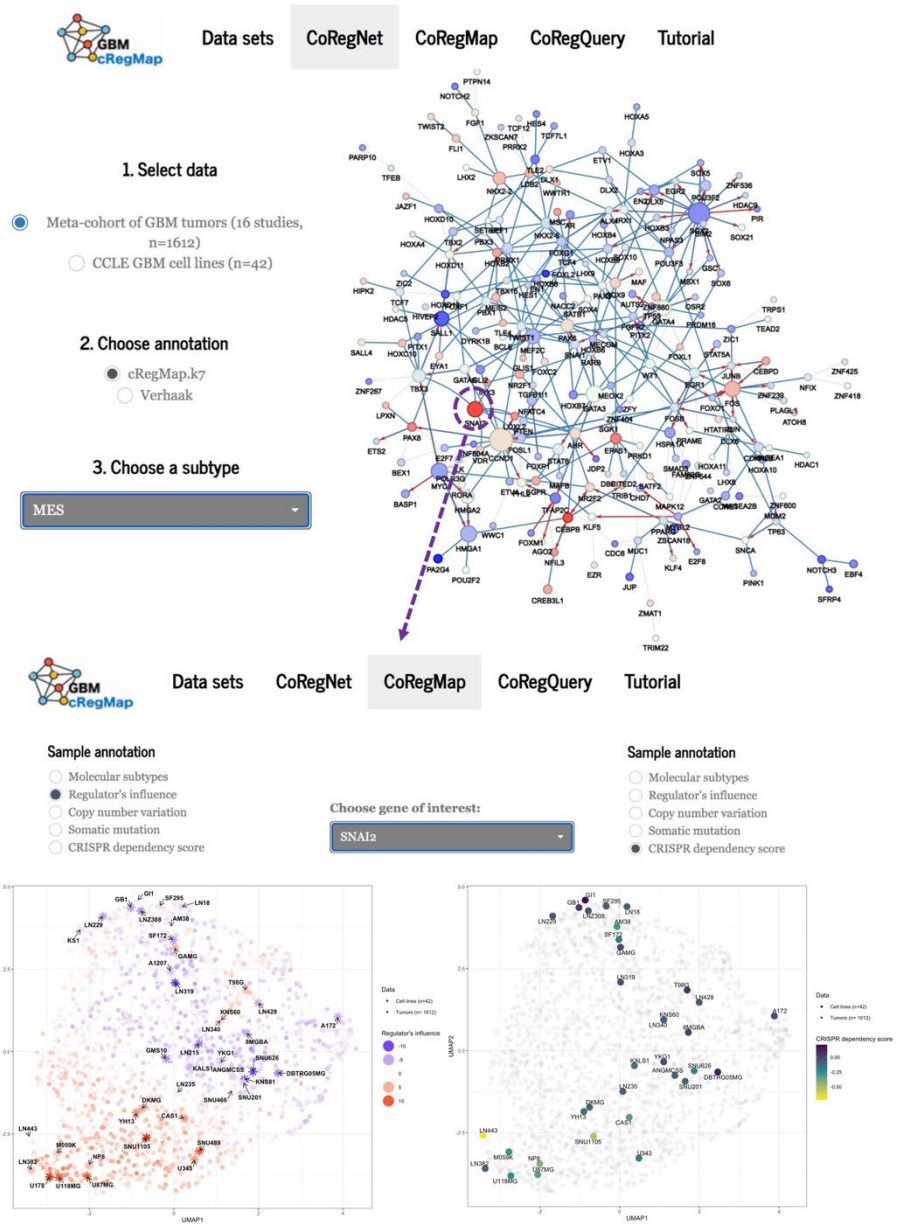


Fig. 5 | Example of exploration of MES co-regulated subnetwork using GBM-cRegMap synergistic compounds. The CoRegNet tab allows users to investigate the extent to which a phenotype of interest (e.g., MES subtypes), characterized by a set of active coregulators, colocalizes within the reference GBM-CoRegNet network through an intuitive visualization. Node color (red = high; blue = low) indicates the influence of the corresponding TF/coTF, with color intensity representing the strength of the influence. Node size and edge color follow the same patterns as indicated in Fig. 2A. The selected coregulators will then be analyzed sequentially, using the GBM-CoRegMap tab. The left plot displays the regulatory influence of the SNAI2 TF, while the right plot shows CRISPR dependency scores for SNAI2.

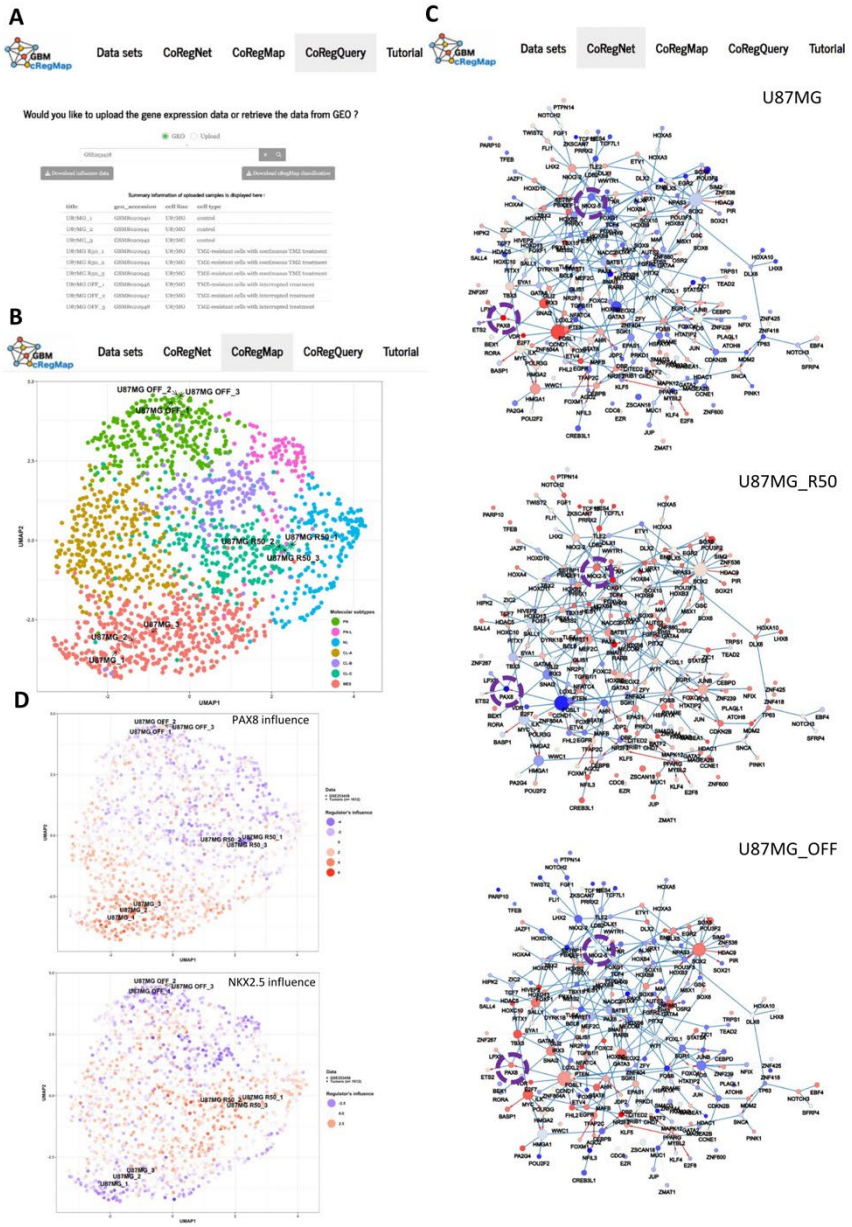
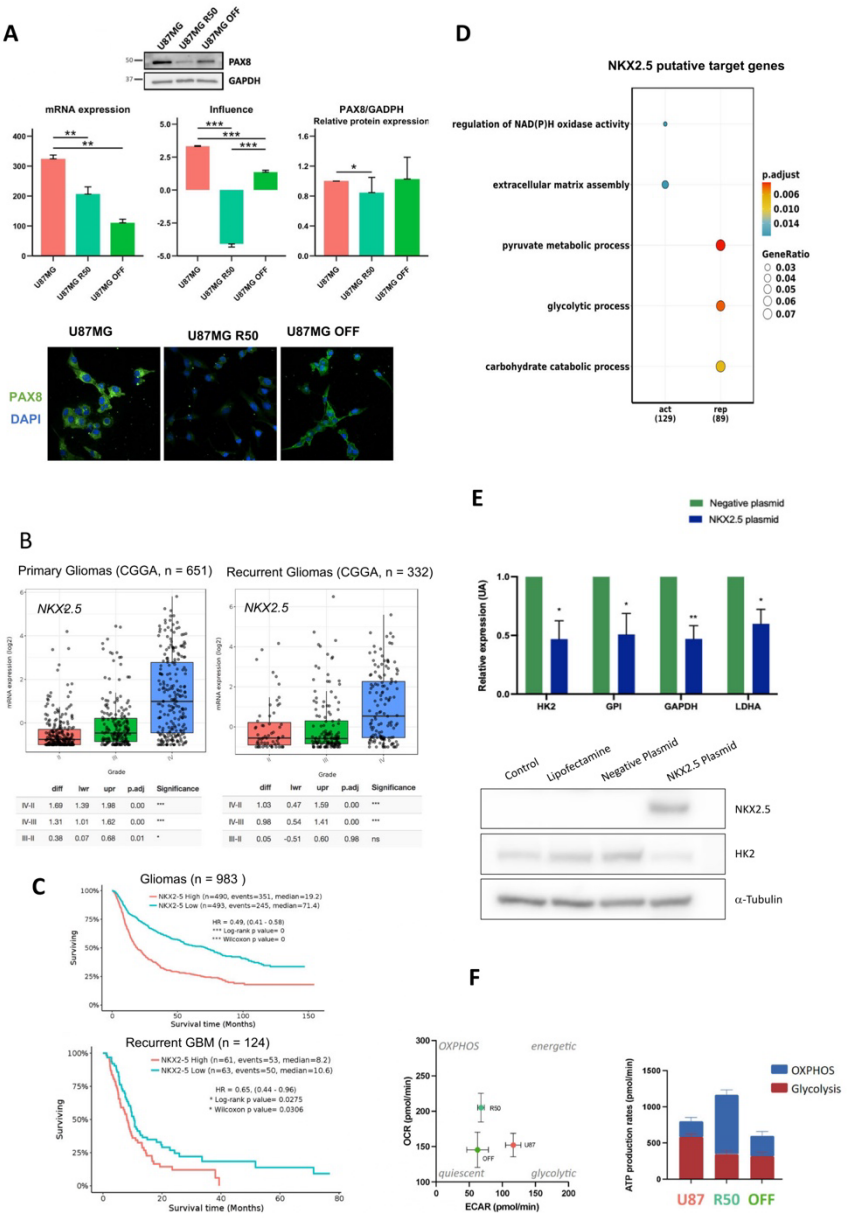


Fig. 6 | Example of reference-mapping and annotation of query dataset using CoRegQuery. A As input, users upload or simply provide a GEO series ID (GSE253458) of the query data using CoRegQuery tab. **B** The CoRegNet tab allows users to investigate the extent to which studied cells (three replicates each) for naive U87MG cells and TMZ-resistant cells, U87MG-R50 and U87MG-OFF, are characterized by a set of active coregulators. Node size and color follow the same patterns as indicated in Fig. 5. **C** The GBM-CoRegMap tab provides colocalization information for the nine cell lines, as predicted by CoRegQuery, with labeling (color) according to the SVM-based classifier. **D** The top plot displays the regula-tory influence of the PAX8 TF, while the bottom panel shows the regulatory influence of the NKX2.5 TF.



distinct classes. Each subtype is uniquely defined by the activity of a regulatory network with distinct biological and clinical associations, which are summarized in Fig. 8. We confirmed and expanded our knowledge of the PN and MES subtypes and distinguished three classical-like subtypes (CL-A: astrocyte-like, CL-B: epithelial basal-like, CL-C: cilium-rich), in addition to NL (neural normal-like) and weakly proliferating PN-L (proneural-Low) subtypes. The PN-L subclass includes IDH1-mutant gliomas, which are no longer classified as GBM under the 2021 WHO framework. However, these tumors were retained to ensure consistency with historical GBM classifications and dataset annotations. This approach aimed to capture the full spectrum of high-grade gliomas historically labeled as GBM, providing a comprehensive understanding of glioma heterogeneity. The discovery of different subtypes of classical GBM, such as CL-C, rich in ciliary features, and CL-B, with epithelial traits, has highlighted the roles of unique

Fig. 7 | Functional validation of predicted phenotypic plasticity upon chemotherapy. A Analysis of PAX8 mRNA expression, influence and relative protein expression in U87MG, U87MG-R50, and U87MG-OFF. Top panel: Representative Western blot (WB) detection of PAX8 protein and its predicted activity based on the regulatory network (colored dots under the WB), along with quantitative analysis ($n > 3$) of its relative protein level. Unprocessed scans of the WB are provided in Supplementary Fig. 3. Bottom panel: mRNA expression and influence barplots of PAX8 and detection of the PAX8 protein by immunofluorescence in the same samples analyzed by WB. **B** mRNA expression levels of NKX2.5 in primary and recurrent gliomas, as analyzed using the Chinese Glioma Genome Atlas (CGGA, http://www.cgga.org.cn/index.jsp) cohort. NKX2.5 expression is significantly higher in grade IV gliomas compared to grade III and grade II gliomas. **C** Survival analysis of recurrent GBM patients (grade IV) based on NKX2.5 expression levels. **D** Functional enrichment analysis of the repressed and activated genes of NKX2.5 using clusterProfiler. The plot visualizes the enriched Gene Ontology Biological Processes (GO-BP) terms, with the size of each dot representing the gene count and the color indicating the adjusted *P* value, reflecting the significance of the enrichment. **E** Top panel: RT-qPCR analysis showing the reduced expression of glycolytic genes (HK2, GPI, GAPDH, LDHA) in U87MG cells upon overexpression of NKX2.5. Bottom panel: Western Blot Analysis of NKX2.5 and HK2 protein expression in U87MG cells in four conditions: control (untreated cells), lipofectamine treated cells, negative control plasmid, and NKX2.5 plasmid transfected cells. Unprocessed scans of the WB are provided in Supplementary Fig. 4. **F** Metabolic signatures of assigned classes for each cell line (U87MG, U87MG-R50, and U87MG-OFF) and energy metabolism characterization using Seahorse XFp extracellular flux analysis. This includes ATP production rates and the contribution of glycolysis and OXPHOS, as well as an energy map, depicting both Oxygen Consumption Rates (OCR) and Extracellular Acidification Rates (ECAR). All Seahorse data are represented as means from three independent experiments ± SEM.

molecular and biological processes. The poor prognosis of the CL-C subtype, linked to its ciliary characteristics, highlights the potential role of primary cilia in GBM oncogenesis and resistance to treatment. This finding suggests that the targeting of cilium-associated pathways may be crucial, particularly for the treatment of challenging subtypes, such as CL-C. In this regard, the influence and dependence score of GLIS2 in the CL-C subtype is particularly interesting, as this protein is a transcriptional repressor, which localizes at the ciliary base, and works as a central regulator of stem and progenitor cell maintenance and differentiation[61]. Our findings provide evidence in support of the existence of a NL subtype, as identified by Nguyen et al.[21], which correlates with 42% of the NEU_Verhaak subtype tumors and 49% of the NEU_Iavarone state. The NL subtype is characterized by a "normal" transcriptomic profile enriched in neuronal development-associated pathways. There is increasing evidence to suggest that some GBMs hijack neuronal mechanisms to fuel their own growth[56]. This finding challenges the view put forward by Wang et al.[4], who suggested that the neural subtype defined by Verhaak might lack a tumor-specific signature, thereby calling into question the identification of neural tumors as a distinct tumor subtype. We also discovered alterations to the tumor suppressor gene *PTEN* in 57% of NL cases, together with a higher frequency of *STAG2* mutations in this subtype, highlighting the distinctive features of these tumors, and possibly indicating a therapeutic window specific for this subtype.

An additional achievement of our work, based on our analysis of the influence and dependence scores of TFs and coTFs, is the identification of key regulators for each subclass. These factors not only represent those which most strongly and widely regulate the gene expression typical of each subtype, but also those whose depletion mostly affects the survival of GBM cell lines belonging to a specific subtype. This, in turn, has the intriguing implication that they might be explored as potential subtype-specific targets. The key regulators of the PN subclass were *ZNF536, BRCA2, GLI2, PBX4,* and *PRMT5*. Notably, *ZNF536*, previously described as a repressor of neuron differentiation[62], has never been involved in glioma development. The very strong dependence score of the PN subtype on PRMT5 is in agreement with and integrates the findings of Sachamitr et al. [63] describing that PRMT5 is a key target for the inhibition of GBM proneural subtype. The high influence and dependence score assigned to PITX1 in the CL-B subtype may point to a particular tendency of this subtype to set up neuroglioma synaptic communications, where the contribution of *PITX1* was recently demonstrated[43]. Even in the already well-characterized MES subtype, we were able to confirm known regulators and a specific vulnerability, such as that on *RUNX2* expression[64], and we also uncovered a novel one, depending on *ARHGEF5*.

More interestingly, the CoRegQuery tab maps new data to a reliable tissue-specific co-regulatory network and a large-scale manifold of tumor heterogeneity references. Reference-based analysis shifts data interpretation from unsupervised to supervised, allowing accumulated information from prior experiments to aid in interpreting new data. The here-described use case of GBM cell phenotypic plasticity in response to TMZ treatment demonstrates the potential of the CoRegQuery algorithm to enhance the analysis and interpretation of small transcriptomic datasets. Our analysis reveals the CL-C subclass, an intermediate state demonstrating the plasticity of GBM cells along the PN-MES axis under chemotherapy. We identified key regulators, such as MDM2, PAX8, and NKX2.5, potentially involved in TMZ resistance. Notably, NKX2.5, more expressed in higher-grade gliomas, negatively impacts patient survival and represses genes involved in glucose metabolism. These findings suggest that the CL-C subclass is crucial in therapy response, providing insights into GBM treatment resistance and identifying potential targets for future research.

In conclusion, we provide a comprehensive new tool for use in studies of the co-regulatory networks driving the heterogeneity and plasticity of GBM. This offers great promise for the GBM clinical and research communities to better understand the complex landscape of GBM and opens up new possibilities for characterizing key regulators and suitable therapeutic targets according to the molecular heterogeneity and plasticity of GBM.

Methods

Public transcriptome data collection

Transcriptomic and clinical outcome data for GBM tumors were collected from GlioVis[65], including 16 datasets corresponding to 1,612 samples (Supplementary Table 1). Additional transcriptomic data for cell lines were downloaded from the Cancer Dependency Map (DepMap)[38] and Gene Expression Omnibus (GEO) databases, including seven datasets corresponding to 42 commercial GBM cell lines, and eight normal brain samples (Supplementary Data 7). For consistency, all gene identifiers were transformed into HUGO Gene Nomenclature Committee (HGNC) symbols. All the above datasets were also classified using the Verhaak classifier from the GlioVis web application. The RNA-seq datasets that had been downloaded from the GEO database were normalized with the edgeR[66] R package, with TMM normalization and log-transformed counts per million. Microarray data were normalized with the standard workflow of the limma[67] R package.

Public data collection for regulatory elements and interactions

A list of regulators corresponding to 2375 genes from two different sources (1639 TF) with experimentally validated DNA binding specificity[68] and 752 transcription cofactors (co-TFs) with experimental validation information[69] was used. The regulation evidence (TF-Target) list included 1,107,092 interactions retrieved with a wide range of bioinformatics tools and databases (listed in Supplementary Data 1) from (1) ChIP-seq data downloaded directly from the ChEA2 database and (2) human TF-binding site (TFBS) models, in the form of position weight matrices (PWM) recovered with MotifDB R/Bioconductor. The promoter sequences were scanned for these TFBS with the PWMEnrich R/Bioconductor package, and (3) completion of the TF-Target list with the tftargets R package and evidence from databases (e.g., TRED, ITFP, ENCODE, BEDOPS, TRRUST). The co-operative evidence (TF-TF) list contained 1,149,582 protein-protein interactions (PPI) from various PPI databases (e.g., BioGrid, HIPPIE, STRING, and HPRD) obtained with the iRefR R package.

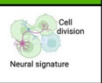
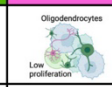
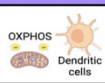
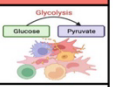
Class Name	Proneural (PN)	Proneural-low proliferative (PN-L)	Neural normal-like (NL)	Classical astrocyte-like (CL-A)	Classical epithelial basal-like (CL-B)	Classical ciliary-rich (CL-C)	Mesenchymal (MES)
							
% of GBM	17 %	6 %	11 %	18 %	9 %	14 %	25 %
Sex Ratio (F/M)	0,72	0,68	0,51	0,81	0,59	0,67	0,42
Median Age (years)	50	46.6	56.6	57	59.6	60.3	56.9
Hazard Ratio for Survival (HR, 95% CI)	0.8 (0.68-0.95)	0.59 (0.44-0.79)	0.94 (0.75-1.17)	1.17 (0.99-1.38)	0.94 (0.77-1.15)	1.24 (1.04-1.48)	1.15 (1-1.32)
Proliferation	High	Low	Low	High	Medium	Low	Medium
Lipid	Low	Medium	High	Medium	High	High	High
Glycolysis	Low	Low	Low	High	High	Medium	High
OXPHOS	Medium	Low	High	High	High	High	Low
Cellularity	NPC	OPC	B-cells	AC	Dendritic cells	Mast cells	Macrophages, Microglia, Monocytes
Mutation	TP53 (59%) PDGFRA (16%)	TP53 (67%) IDH1 (52%) ATRX (43%)	PTEN (57%) STAG2 (21%)	EGFR (41%) LRP2 (14%)	FLG (26%) NRAS(4%) PPARG(2%)	EGFR (32%) TP53 (32%) RB1 (14%) DNHA9 (7%)	NF1 (24%) MUC16 (22%) RB1 (12%)
Key Regulators (TFs/coTFs)	ATAD2 BIRC5 BRCA2 CDC6 E2F7 FOSL1 FOXM1 GLI2 PA2G4 PRMT5 TFAP2C TBX3	CAMK4 CREB3L4 GBX2 SALL2 ZNF704A	MEIS3 POU3F3 PRDM16 SMAD3 SNCA SP100 TNIP1 TRB23 ZFP3	CENPU DLX1 HDAC10 NKX2.2 TLE2 ZNF181	HES4 PITX1 PPARG PRAME SALL1 SOX10 STAT6	CBX4 GLIS2 HDAC5 MEIS3 NKX2.5 PAX6 POU3F2 RPS6KA1 SOX2 STAT3 TRIM25 WWTR1	ARHGEF5 CEBPB EGFR EPAS1 FOSL2 PAX8 RUNX2 SNAI2
Cell line Repertoire	A1207, AM38 GAMG, GB1 GI1, KS1 LN229, LNZ308			DKMG YH13	LN382 T98G	KALS1, KNS81 LN215, SNU201 SNU626, LN235 YKG1	SNU1105 SNU489 U178 U87MG

Fig. 8 | Summary of the main characteristics of the GBM-cRegMap classes. The figure displays, from top to bottom: class name and abbreviation; proportion of classes in the 1612-sample metacohort; clinical parameters (age, sex, and survival); and key biological characteristics, including proliferation and metabolic signatures (color-scaled), cellularity deconvolution, molecular alterations, active regulators, and cell line repertoire.

Gene regulatory network inference

Gene regulatory networks were inferred with the CoRegNet Bioconductor package[13]. Starting from transcriptomic data and a list of human regulators, CoRegNet uses the hLICORN algorithm[9] to capture regulatory interactions between regulators and target genes in four steps. First, transcriptomic data is discretized into values of −1, 0, and 1 according to per-gene distribution. Genes present in the transcriptomic dataset were split into regulators and target genes, and only those with minimum support for non-zero values after discretization were retained. Second, a frequent itemset algorithm identifies potential sets of co-activators and co-inhibitors from the list of regulators. Third, for each target gene, a list of candidate co-activator and co-inhibitor sets (GRN) was selected according to a regulatory program (association rule) metric. Fourth, each GRN candidate for each gene was scored based on the regression between the expression of the GRN set regulators and the expression of the gene concerned. For each gene, we retained the top ten GRN candidates according to R2 score. CoRegNet can then be used to select the best GRN for each gene based on the interaction evidence data. Evidence of regulation and cooperation was incorporated into each candidate GRN with an integrative selection algorithm, yielding an R2 score for each of the integrated datasets. The GRN with the highest merged score was selected. CoRegNet builds a co-regulatory network, GBM-CoRegNet, from the GRNs obtained, by setting a co-operativity relationship between a pair of regulators, TFi and TFj, if they have a minimum of five target genes in common and the relationship is significant ($P = <0.01$) according to the Jaccard similarity co-efficient formula:

$$\frac{TFi \cap TFj}{TFi \cup TFj} \quad (1)$$

Network-based regulatory influence signal quantification

The regulatory network structure provides a set of genes that are activated and repressed under reference conditions. Based on this structure, we can capture the influence of each regulator, a latent signal of the regulator

activity in each sample based on its observed effect on downstream entities. For each regulator, Welch's *t* test was performed to compare the distribution of activated (Ar) and repressed (Ir) genes. The influence of a regulator *r* is calculated as follows:

$$Influence(r) = \frac{\overline{E(A^r)} - \overline{E(I^r)}}{\sqrt{\frac{\mu^2_{A^r}}{A^r} + \frac{\mu^2_{I^r}}{I^r}}} \quad (2)$$

where $E(A^r)$ and $E(I^r)$ are the expression of the activated and repressed genes in the samples, respectively. $\overline{E(A^r)}$ and $\overline{E(I^r)}$ are their respective means, and μA^r and μI^r are their standard deviation. A regulator is active only when it activates A^r and represses I^r in accordance with the expectations of the regulatory network model, giving a positive *t* value in the Welch *t* test. Higher *Influence(r)* values indicate higher levels of regulator *r* activity in the sample. In this work, we calculated the *Influence* of a regulator only if it has activated or repressed at least five genes. Due to the strict thresholds used in network construction, some local gene networks may be filtered out, resulting in some regulators having activated or repressed target gene sets with fewer than five members. In such cases, we fill the smaller set (fewer than 5 target genes) with the target genes for which the regulator concerned has the highest R2 regression score. This makes it possible to capture the *Influence* values for a maximum number of regulators. In addition, for some regulators, expression values may be missing for some of the target genes in some tumor samples for which we wish to assess the influence. In this case, the missing expression values are estimated with the LatNet method[60], using expression levels in cell lines as a reference dataset. It may not be possible to estimate the expression of some target genes for a particular sample, and this may preclude the calculation of the influence of a particular regulator. In this (rare) case, the median of the obtained influence values for the regulator in the other samples is used to replace the missing value.

Seurat clustering

cRegMap subclass computation was performed with the Seurat R package[37], by a modularity optimization-based clustering approach. The smart local moving (SLM) graph clustering algorithm was applied to a shared nearest neighbor graph (SNN) plotted from the influence data. Before clustering with the Seurat package, the Seurat:: ScaleData function was used with default parameters to scale the influence data. Seurat::RunPCA results were called with 20 PCs, and all other parameters were set at default values. Clustering with the Seurat package was performed on all calculated principal components with the Seurat::FindNeighbors function with default parameters, followed by the Seurat::FindClusters function with a specified resolution of 0.8, resulting in seven clusters.

Reference-mapping and annotating query datasets

Query datasets are mapped and annotated in a multistep process. CoRegQuery first uses GBM-CoRegNet to compute the regulatory influences of the input dataset. A support vector machine (SVM) model is then used to classify the query samples. We used the e1071 R package[70], implementing an SVM with a radial kernel and a one-vs.-rest strategy for multiclass classification. The SVM model was trained on the influences and classifications derived from the tumor cohort. In the next step, CoRegQuery uses the Seurat::ProjectUMAP function to project the query dataset samples onto the UMAP (uniform manifold approximation and projection) coordinates of GBM-CoRegMap. The Seurat function identifies the nearest reference sample neighbors and their distances for each query sample and performs a UMAP projection by supplying the calculated neighbor set along with the UMAP model previously computed in GBM-CoRegMap.

GBM-cRegMap software implementation

GBM-cRegMap is written in the R language and uses the functions of several widely used packages available from Bioconductor and CRAN, based on the Golem framework (Supplementary Data 1). This framework simplifies the development and deployment of a stable and robust web application with R Shiny. A plotly graphics system was used to generate interactive visualizations, with interactions enabled by brushing or clicking on them in the Shiny framework. The GBM-cRegMap has been released on Docker with all the necessary packages to prevent update conflicts. The GBM-cRegMap web application was deployed with Google Cloud Run and is publicly available from https://gbm.cregmap.com. The source code is available from GitHub (http://github.com/i3bionet/gbm-cregmap) under the Apache 2 License.

Functional enrichment analysis of DEGs

For microarray gene expression analysis, the linear models for microarray data (LIMMA) R package[67] was used, and *P* values were adjusted by the Benjamini–Hochberg method. For RNA-seq data gene expression analysis, the edgeR[66] R package was used, followed by limma-voom[71] and standard differential expression (DE) analysis. Gene Ontology (GO) enrichment analysis was performed with the R packages clusterProfiler[72] and enrichR[73]. Ontologies with a P_{adj} <0.01 and gene counts of more than five were examined. The Kyoto Encyclopedia of Genes and Genomes (KEGG), Molecular Signatures Database (MSigDB), and REACTOME pathway databases were searched for enriched terms with the R package msigdbr.

Analysis of genomic alterations

Somatic mutations were analyzed with the GBM-TCGA dataset. The TCGAbiolinks R package was used to download mutation annotation files (MAF) aligned with the hg38 sequence, and analysis was performed with the R package maftools.

Deconvolution of cellular components and tumor cell purity scores

The TCGA-GBM dataset was used to determine the tumor cell purity of each sample and the abundance of cell populations with respect to the GBM-cRegMap classification. The PUREE[36] (pan-cancer tumor purity estimation) web server was used to calculate the purity score for each sample of the dataset and to display it as a boxplot for each class. The GBMdeconvoluteR[24] web server was then used to estimate the abundance of various immune and stromal cell populations in each sample, and the resulting values were averaged by GBM-cRegMap subclass.

Survival analysis

Survival analyses were performed, and survival curves were displayed following the Kaplan–Meier method with the R package survminer. Survival curves were compared by the log-rank test and subsequent survival rate comparison versus overall metacohort survival were done with a univariate Cox proportional hazard model.

Classification of cell lines

CoRegQuery assigned a specific class with a high degree of confidence (SVM posterior probability >0.75) to 33 of 42 GBM cancer cell lines (GBCCL) (Fig. 4 and Supplementary Data 7), revealing a heterogeneous pattern of subclasses across the GBM cell-line panel. We assessed the robustness of our method by analyzing multiple independent datasets for a single-cell line. These datasets were generated with various transcriptomic profiling techniques, including microarrays and RNA-seq, and were sourced from seven different datasets (Supplementary Data 7). The classification outcomes were consistent across the different profiling techniques and datasets, demonstrating the reliability of our classification approach. An exploration of the PubMed citation status of the studied GBCCL (Supplementary Data 7) revealed a bias, with only five (150> citations each) cell lines used by the scientific community. Our classification system was consistent with published results for these five cell lines. For instance, U87MG, a well-known mesenchymal GBCCL, was assigned to the MES subclass, LN229 to the PN subclass, and T98G, which is known to be dependent on oxidative phosphorylation[74], to the CL-B class. Normal brain samples (GSE15824, $n = 2$; GSE15209, $n = 6$) were categorized as NL.

Identification and validation of subclass-specific regulators

We used Wilcoxon rank-sum tests to identify differentially Influencing regulators (DIRs) with the Seurat::FindAllMarkers function (min.pct = 0.25 and logFC.threshold = 0.20) and *P* value adjustment by Bonferroni correction. Wilcoxon–Mann–Whitney tests between the mean CERES dependency score of the cell lines assigned to the subclass and the mean dependency score of the rest to rank the DIRs further. The CERES dependency score of the cell lines was acquired with the R package depmap[38].

Cell lines, culture, and conditions

The U87MG GBM cell line was obtained from the American Type Culture Collection (LGC Standards Sarl, Molsheim, France). Two TMZ-resistant clonal lines were established in our laboratory by exposing the parental U87MG line to 50 µM TMZ for extended periods: U87MG-R50 and U87MG-OFF. U87MG-R50 cells were continuously cultured in a medium supplemented with 50 µM TMZ, whereas U87MG R50 OFF cells were treated with TMZ for two months, after which they were grown in TMZ-free medium. TMZ (cat #T2577; Sigma-Aldrich) was prepared as a 100 mM stock solution in DMSO and stored at 4 °C until use. These cell lines were routinely cultured in Eagle's minimum essential medium (EMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS), 1% sodium pyruvate, and 1% non-essential amino acids, at 37 °C, under an atmosphere containing 5% CO_2.

Immunofluorescence assays

Cells were fixed by incubation with ice-cold methanol for 10 min. They were then incubated overnight at 4 °C with the PAX8 Proteintech #10336-1-AP primary antibody (1:50). The cells were then incubated with a goat anti-rabbit AF488 secondary antibody (1:500) at room temperature for 2 h. The expression of PAX8 was assessed in U87MG, U87MG R50, and U87MG R50 OFF cell lines and analyzed under a fluorescence microscope.

Seahorse metabolic flux analysis

Real-time analysis of oxygen consumption rates (OCR) and extracellular acidification rates (ECAR) was conducted using the Seahorse XFp analyzer (Agilent). Prior to analysis, XFp flux cartridges were pre-hydrated with XF Calibrant overnight at 37 °C. Microplates were coated with Cell-Tak at a concentration of 22.4 µg/mL (Corning®) to ensure cell adherence. Cells were seeded at a density of 60,000 cells per well in lineage-specific medium and both seeding, and analysis were performed on the same day. ATP production rates were determined using the Seahorse XF Real-Time ATP Rate Assay Kit, following the manufacturer's recommended protocol using the medium recommended by the manufacturer (DMEM, 10 mM glucose, 2 mM glutamine, and 1 mM pyruvate, pH 7.4). OCR and ECAR were assessed under basal conditions and in response to sequential injections of Oligomycin (1.5 µM) and Rotenone/Antimycin A (0.5 µM each).

Western blotting analysis

Cells were lysed in Laemmli sample buffer (Biorad) supplemented with 5% β-mercaptoethanol, and proteins were fractionated by SDS-PAGE (20%), transferred to PVDF membrane (GE Healthcare, Velizy, France), which were blocked with 5% (w/v) nonfat milk in PBST for 1 h at room temperature, probed with the primary antibody in blocking solution overnight at 4 °C, incubated with horseradish peroxidase (HRP)-conjugated secondary antibodies (GE Healthcare) for 1 h at room temperature and the chemiluminescent signal was detected using enhanced chemiluminescence (ECL) system (ECL™ Prime Western Blotting System, GE Healthcare Bioscience) with a ImageQuant™ LAS 4000 analyser (GE Healthcare). Quantification of non-saturated images was done using ImageJ software (National Institutes of Health, Bethesda, MD, USA). For each experiment, 3 lysates from different cell cultures were used. Tubulin or GAPDH was used as the loading control for all samples. The primary antibodies used were antibodies for NKX2.5 (Cell Signaling, cat #8792, 1:500), HK2 (Proteintech, cat #22029-1-AP, 1:5000), PAX8 (Proeintech, cat#10336-1-AP; 1:10,000), α-Tubulin (cat #T9026, Sigma-Aldrich, 1:1000 dilution) and GAPDH (cat #MAB374, Sigma-Aldrich, 1:500). The secondary antibodies used were anti-mouse IgG-HRP (cat #W4028, Promega, 1:10,000 dilution), anti-rabbit IgG-HRP (cat#W4018, Promega, 1:10,000 dilution).

RT-qPCR analysis

U87MG/U87MG R50/U87MG OFF were seeded in a 6-well plate at a density of 0.5 million cells per well and cultured in pure OptiMEM medium, or supplemented with Lipofectamine 2000 (ThermoFisher), negative control plasmid (5 µg), or NKX2.5 plasmid (5 µg) for a period of 24 h. The cells were then washed with 1X PBS and lysed with 350 µL of RLT buffer from the extraction kit (RNEasy Plus Mini Kit, Qiagen) according to the manufacturer's instructions. RNA concentrations were measured using an ND-1000 NanoDrop spectrophotometer (Labtech, Palaiseau, France). 1 µg of extracted RNA was used for cDNA synthesis using iScript Reverse Transcription SuperMix kit (BioRad). cDNA was analyzed by Fast SYBR™ Green (Applied Biosystem, ThermoFisher Scientific, MA, USA) in duplicate, using the StepOne real-time PCR system (Applied Biosystem). qRT-PCR data were analyzed using StepOne Plus software. The relative expression of each target was calculated using the relative quantification method ($2^{-\Delta\Delta CT}$) with RNA18S (5'-TGTGGTGTTGAGGAAAGCAG-3' and 3'-TCCAGA CCATTGGCTAGGAC-5'; Invitrogen) as internal control. GAPDH primers were 5'-GTCACCAGGGCTGCTTTTAACTCT-3' and 3'-GGAAT CATATTGGAACATGTAAACCAT-5'; Invitrogen). The following primer pairs were from Qiagen (RT2 qPCR primer assay): GPI (PPH00897C), HK2 (PPH00983B), LDHA (PPH02047H).

Data availability

Public data used in this article were acquired from GEO (https://www.ncbi.nlm.nih.gov/geo/), DepMap (https://depmap.org/portal/download/all/), GlioVis (http://gliovis.bioinfo.cnio.es), and TCGA (https://portal.gdc.cancer.gov/). The in-house dataset of the U87MG cell line and its TMZ-resistant variants, which we made available, can be found in the GEO repository under accession number GSE253458.

Received: 22 September 2024; Accepted: 21 March 2025;
Published online: 15 April 2025

References

1. Ostrom, Q. T., Cioffi, G., Waite, K., Kruchko, C. & Barnholtz-Sloan, J. S. CBTRUS statistical report: primary brain and other central nervous system tumors diagnosed in the United States in 2014–2018. *Neuro-Oncology* **23**, iii1–iii105 (2021).
2. Gilbertson, R. J. High-grade glioma: can we teach an old dogma new tricks? *Cancer Cell* **9**, 147–148 (2006).
3. Verhaak, R. G. W. et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**, 98–110 (2010).
4. Wang, Q. et al. Tumor evolution of glioma-intrinsic gene expression subtypes associates with immunological changes in the microenvironment. *Cancer Cell* **32**, 42–56.e6 (2017).
5. Neftel, C. et al. An integrative model of cellular states, plasticity, and genetics for glioblastoma. *Cell* **178**, 835–849.e21 (2019).
6. Lasorella, A. & Iavarone, A. The making of the glioblastoma classification. *Br. J. Cancer* **125**, 4–6 (2021).
7. Garofano, L. et al. Pathway-based classification of glioblastoma uncovers a mitochondrial subtype with therapeutic vulnerabilities. *Nat. Cancer* **2**, 141–156 (2021).
8. Margolin, A. A. et al. ARACNE: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context. *BMC Bioinforma.* **7**, S7 (2006).
9. Elati, M. et al. LICORN: learning cooperative regulation networks from gene expression data. *Bioinformatics* **23**, 2407–2414 (2007).
10. Chebil, I., Nicolle, R., Santini, G., Rouveirol, C. & Elati, M. Hybrid method inference for the construction of cooperative regulatory network in human. *IEEE Trans. Nanobiosci.* **13**, 97–103 (2014).

11. Aibar, S. et al. SCENIC: single-cell regulatory network inference and clustering. *Nat. Methods* **14**, 1083–1086 (2017).
12. Nicolle, R., Elati, M. & Radvanyi, F. Network transformation of gene expression for feature extraction. In *2012 11th International Conference on Machine Learning and Applications* (IEEE) 108–113 (IEEE, 2012).
13. Nicolle, R., Radvanyi, F. & Elati, M. CoRegNet: reconstruction and integrated analysis of co-regulatory networks. *Bioinformatics* **31**, 3066–3068 (2015).
14. Alvarez, M. J. et al. Functional characterization of somatic mutations in cancer using network-based inference of protein activity. *Nat. Genet.* **48**, 838–847 (2016).
15. Xiong, A., Zhang, J., Chen, Y., Zhang, Y. & Yang, F. Integrated single-cell transcriptomic analyses reveal that GPNMB-high macrophages promote PN-MES transition and impede T cell activation in GBM. *EBioMedicine* **83**, 104239 (2022).
16. Carro, M. S. et al. The transcriptional network for mesenchymal transformation of brain tumours. *Nature* **463**, 318–325 (2010).
17. Becht, E. et al. Dimensionality reduction for visualizing single-cell data using UMAP. *Nat. Biotechnol.* **37**, 38–44 (2019).
18. McInnes, L., Healy, J., Saul, N. & Großberger, L. UMAP: uniform manifold approximation and projection. *J. Open Source Softw.* **3**, 861 (2018).
19. B.V., A. & Contributors, T.B. visNetwork. Network Visualization using « vis.js, R package version 2.1.1. https://github.com/datastorm-open/visNetwork (2022).
20. Kerry, S. M. & Bland, J. M. Sample size in cluster randomisation. *BMJ* **316**, 549 (1998).
21. Nguyen, H. D. et al. A machine learning analysis of a "normal-like" IDH-WT diffuse glioma transcriptomic subgroup associated with prolonged survival reveals novel immune and neurotransmitter-related actionable targets. *BMC Med.* **18**, 280 (2020).
22. Bailey, M. L. et al. Glioblastoma cells containing mutations in the cohesin component STAG2 are sensitive to PARP inhibition. *Mol. Cancer Ther.* **13**, 724–732 (2014).
23. Xu, W. et al. STAG2 mutations regulate 3D genome organization, chromatin loops, and Polycomb signaling in glioblastoma multiforme. *J. Biol. Chem.* **300**, 107341 (2024).
24. Ajaib, S. et al. GBMdeconvoluteR accurately infers proportions of neoplastic and immune cell populations from bulk glioblastoma transcriptomics data. *Neuro-Oncology* **25**, 1236–1248 (2023).
25. Kawai, S. et al. Generation of a lung squamous cell carcinoma three-dimensional culture model with keratinizing structures. *Sci. Rep.* **11**, 24305 (2021).
26. Saidak, Z. et al. Squamous cell carcinoma antigen-encoding genes SERPINB3/B4 as potentially useful markers for the stratification of HNSCC tumours. *Anticancer Res* **38**, 1343–1352 (2018).
27. Magnus, N., Gerges, N., Jabado, N. & Rak, J. Coagulation-related gene expression profile in glioblastoma is defined by molecular disease subtype. *J. Thromb. Haemost.* **11**, 1197–1200 (2013).
28. Rodriguez, F. J., Scheithauer, B. W., Giannini, C., Bryant, S. C. & Jenkins, R. B. Epithelial and pseudoepithelial differentiation in glioblastoma and gliosarcoma: a comparative morphologic and molecular genetic study. *Cancer* **113**, 2779–2789 (2008).
29. Guigni, B. A. et al. Skeletal muscle atrophy and dysfunction in breast cancer patients: role for chemotherapy-derived oxidant stress. *Am. J. Physiol. Cell Physiol.* **315**, C744–C756 (2018).
30. Maghe, C. et al. The paracaspase MALT1 controls cholesterol homeostasis in glioblastoma stem-like cells through lysosome proteome shaping. *Cell Rep.* **43**, 113631 (2024).
31. Ho, K. -H. et al. A key role of DNA damage-inducible transcript 4 (DDIT4) connects autophagy and GLUT3-mediated stemness to desensitize temozolomide efficacy in glioblastomas. *Neurotherapeutics* **17**, 1212–1227 (2020).
32. Cheng, Y. et al. LGR6 promotes glioblastoma malignancy and chemoresistance by activating the Akt signaling pathway. *Exp. Ther. Med.* **22**, 1364 (2021).
33. Leventoux, N. et al. Transformation foci in IDH1-mutated gliomas show STAT3 phosphorylation and downregulate the metabolic enzyme ETNPPL, a negative regulator of glioma growth. *Sci. Rep.* **10**, 5504 (2020).
34. Goranci-Buzhala, G. et al. Cilium induction triggers differentiation of glioma stem cells. *Cell Rep.* **36**, 109656 (2021).
35. Laws, M. T. et al. Glioblastoma may evade immune surveillance through primary cilia-dependent signaling in an IL-6 dependent manner. *Front. Oncol.* **13**, 1279923 (2023).
36. Revkov, E., Kulshrestha, T., Sung, K. W. -K. & Skanderup, A. J. PUREE: accurate pan-cancer tumor purity estimation from gene expression data. *Commun. Biol.* **6**, 394 (2023).
37. Hao, Y. et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. *Nat. Biotechnol.* https://doi.org/10.1038/s41587-023-01767-y (2023).
38. Tsherniak A. et al. Defining a Cancer Dependency Map. *Cell*. **170**, 564–576 (2017).
39. Li, X. et al. NCBP3/SNHG6 inhibits GBX2 transcription in a histone modification manner to facilitate the malignant biological behaviour of glioma cells. *RNA Biol.* **18**, 47–63 (2021).
40. Chuikov, S., Levi, B. P., Smith, M. L. & Morrison, S. J. Prdm16 promotes stem cell maintenance in multiple tissues, partly by regulating oxidative stress. *Nat. Cell Biol.* **12**, 999–1006 (2010).
41. Held-Feindt, J., Hattermann, K., Knerlich-Lukoschus, F., Mehdorn, H. M. & Mentlein, R. SP100 reduces malignancy of human glioma cells. *Int. J. Oncol.* **38**, 1023–1030 (2011).
42. Song, Y. -C. et al. Proteogenomic characterization and integrative analysis of glioblastoma multiforme. *Oncotarget* **8**, 97304–97312 (2017).
43. Chakraborty, C. et al. Rewiring of the promoter-enhancer interactome and regulatory landscape in glioblastoma orchestrates gene expression underlying neuroglioma synaptic communication. *Nat. Commun.* **14**, 6446 (2023).
44. Spinelli, C. et al. Mesenchymal glioma stem cells trigger vasectasia—distinct neovascularization process stimulated by extracellular vesicles carrying EGFR. *Nat. Commun.* **15**, 2865 (2024).
45. Shi, Z. -F. et al. EGFRvIII promotes the proneural–mesenchymal transition of glioblastoma multiforme and reduces its sensitivity to temozolomide by regulating the NF-κB/ALDH1A3 Axis. *Genes* **14**, 651 (2023).
46. Mao, X. et al. CEBPD is a master transcriptional factor for hypoxia regulated proteins in glioblastoma and augments hypoxia induced invasion through extracellular matrix-integrin mediated EGFR/PI3K pathway. *Cell Death Dis.* **14**, 269 (2023).
47. Renfrow, J. J. et al. Hypoxia-inducible factor 2α: a novel target in gliomas. *Future Med. Chem.* **10**, 2227–2236 (2018).
48. Komiya, Y. et al. The Rho guanine nucleotide exchange factor ARHGEF5 promotes tumor malignancy via epithelial-mesenchymal transition. *Oncogenesis* **5**, e258 (2016).
49. Cheng, H. S., Yip, Y. S., Lim, E. K. Y., Wahli, W. & Tan, N. S. PPARs and tumor microenvironment: the emerging roles of the metabolic master regulators in tumor stromal-epithelial crosstalk and carcinogenesis. *Cancers* https://doi.org/10.3390/cancers13092153 (2021).
50. Li, J. et al. Clinicopathological and prognostic significance of PRAME overexpression in human cancer: a meta-analysis. *BioMed. Res. Int.* **2020**, 8828579 (2020).
51. Lopez-Bertoni, H. et al. Sox2 induces glioblastoma cell stemness and tumor propagation by repressing TET2 and deregulating 5hmC and 5mC DNA modifications. *Signal Transduct. Target. Ther.* **7**, 37 (2022).
52. Suvà, M. L. et al. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**, 580–594 (2014).

53. Wang, L. et al. A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. *Nat. Cancer* **3**, 1534–1552 (2022).
54. Park, J. H. et al. Gene regulatory network topology governs resistance and treatment escape in glioma stem-like cells. *Sci. Adv.* **10**, eadj7706 (2024).
55. Sani, S. et al. Temozolomide-acquired resistance is associated with modulation of the integrin repertoire in glioblastoma, impact of α5β1 integrin. *Cancers* **14**, 369 (2022).
56. Hung, N. et al. Increased paired box transcription factor 8 has a survival function in glioma. *BMC Cancer* **14**, 159 (2014).
57. Di Palma, T., Lucci, V., De Cristofaro, T., Filippone, M. G. & Zannini, M. A role for PAX8 in the tumorigenic phenotype of ovarian cancer cells. *BMC Cancer* **14**, 292 (2014).
58. Cao, C. et al. Nkx2.5: a crucial regulator of cardiac development, regeneration and diseases. *Front. Cardiovasc. Med.* **10**, 1270951 (2023).
59. Zhao, Z. et al. Chinese Glioma Genome Atlas (CGGA): a comprehensive resource with functional genomic data from chinese glioma patients. *Genomics Proteom. Bioinforma.* **19**, 1–12 (2021).
60. Dhifli, W., Puig, J., Dispot, A. & Elati, M. Latent network-based representations for large-scale gene expression data analysis. *BMC Bioinformatics* **19**, 81–90 (2019).
61. Jetten, A. M. GLIS1-3 transcription factors: critical roles in the regulation of multiple physiological processes and diseases. *Cell. Mol. Life Sci. CMLS* **75**, 3473–3494 (2018).
62. Qin, Z. et al. ZNF536, a novel zinc finger protein specifically expressed in the brain, negatively regulates neuron differentiation by repressing retinoic acid-induced gene transcription. *Mol. Cell. Biol.* **29**, 3633–3643 (2009).
63. Sachamitr, P. et al. PRMT5 inhibition disrupts splicing and stemness in glioblastoma. *Nat. Commun.* **12**, 979 (2021).
64. Santamarina-Ojeda, P. et al. Multi-omic integration of DNA methylation and gene expression data reveals molecular vulnerabilities in glioblastoma. *Mol. Oncol.* **17**, 1726–1743 (2023).
65. Bowman, R. L., Wang, Q., Carro, A., Verhaak, R. G. W. & Squatrito, M. GlioVis data portal for visualization and analysis of brain tumor expression datasets. *Neuro-Oncology* **19**, 139–141 (2017).
66. 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).
67. Ritchie, M. E. et al. limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* **43**, e47–e47 (2015).
68. Lambert, S. A. et al. The human transcription factors. *Cell* **172**, 650–665 (2018).
69. Schmeier, S., Alam, T., Essack, M. & Bajic, V. B. TcoF-DB v2: update of the database of human and mouse transcription co-factors and transcription factor interactions. *Nucleic Acids Res.* **45**, D145–D150 (2017).
70. Meyer, D. et al. e1071: misc functions of the department of statistics, probability theory group (formerly: E1071), TU Wien. *Version 1.7-13* (2022).
71. 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).
72. Wu, T. et al. clusterProfiler 4.0: a universal enrichment tool for interpreting omics data. *Innovation* **2**, 100141 (2021).
73. Kuleshov, M. V. et al. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* **44**, W90–W97 (2016).
74. Oppermann, H., Schnabel, L., Meixensberger, J. & Gaunitz, F. Pyruvate attenuates the anti-neoplastic effect of carnosine independently from oxidative phosphorylation. *Oncotarget* **7**, 85848–85860 (2016).

Acknowledgements

This work was supported by grants from the Institut National du Cancer (INCa) (PL-Bio project no. 2017-145, SIRIC EN-HOPE SMART4CBT), the French National Research Agency (ANR) (ANR-20-CE20-0028), and the Ligue contre le Cancer CCIR EST. C.B. was supported by a doctoral fellowship from the French Ministry of Higher Education, Research, and Innovation (MESRI). GP received a doctoral fellowship from the French National Research Agency (ANR) (ANR-20-THIA-0014, AI_PhD@Lille program). M.M. was supported by a doctoral fellowship from the CNRS (MITI 80 PRIME program).

Author contributions

Biological investigation: C.B., M.K., D.R., S.A.C., M.D., and M.E.; software development and bioinformatics analysis: G.P., K.G., W.D., A.D., J.D., and M.E.; writing–review and editing: C.B., K.G., G.P., S.A.C., M.D., and M.E.; conceptualization, supervision, and funding acquisition: M.D. and M.E. All authors provided critical feedback, helped shape the research, discussed the results, and commented on the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41698-025-00890-0.

Correspondence and requests for materials should be addressed to Mohamed Elati.

Reprints and permissions information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

© The Author(s) 2025

Discussion

1. Réseau de régulation génique

1.1 Approche multiomique

Notre approche par détermination du réseau de régulation génique a montré son efficacité et fait ses preuves dans l'étude du réseau de régulation du récepteur METex14del. Le comportement du réseau correspondait à nos attentes et nous avons proposé une autre voie de signalisation activée par METex14del exposé à l'HGF.

Les algorithmes h-LICORN et CoRegNet s'appuient sur des informations transcriptionnelles et sur des preuves d'interactions régulatrices issues de la littérature pour inférer le GRN et calculer l'influence des régulateurs. Cependant, la qualité du GRN inféré dépend fortement du contexte spécifique étudié, et le recours à des bases de données généralistes permet certes d'obtenir des relations de régulation éprouvées dans de nombreux contextes, mais n'offre pas la possibilité de pondérer ces interactions en fonction, par exemple, de l'accessibilité de la chromatine aux facteurs de transcription.

Ainsi, l'adjonction d'autres types de données omiques lors de l'inférence du GRN pourrait améliorer la concordance du modèle avec son contexte. Des tests tels que l'immunoprécipitation de la chromatine suivie d'un séquençage (ChIP-seq) permettent de mesurer la liaison des facteurs de transcription (TF) sur l'ensemble du génome, et ces informations peuvent être utilisées pour construire des GRN, notamment en assignant des sites de liaison de TF à des gènes cibles. Toutefois, le profilage de la liaison des TF reste coûteux et limité aux facteurs de transcription pour lesquels des anticorps sont disponibles (262).

Une alternative à cette approche est la méthodologie ATAC-seq (394), qui mesure les régions de chromatine ouverte et identifie, dans les régions régulatrices, les positions avec ou sans nucléosomes, afin d'analyser l'occupation des facteurs de transcription. Par l'ajout de ce type de données, nous pourrons renforcer l'explicabilité de nos conclusions.

Depuis la fin des années 1990, des bases de données rassemblent des informations sur les voies de signalisation (395). Ces bases de données fournissent une source unifiée d'informations dans des formats que les utilisateurs peuvent parcourir, récupérer et traiter.

Elles ont été complétées par des données sur les interactions ligand-récepteur (396), les éléments de régulation transcriptionnelle (397–399), les interactions protéines-protéines (256,396,400,401), les interactions ligand-récepteur (402), les annotations fonctionnelles (403–405), ainsi que les informations structurelles et mécanistiques (406–408).

Cependant, les différentes bases de données restent dispersées. Pour combler cette lacune, plus de 100 ressources ont été combinées au sein de la base de données OmniPath, couvrant les interactions et les rôles des protéines dans la signalisation inter- et intracellulaire, ainsi que la régulation transcriptionnelle et post-transcriptionnelle, afin de constituer un point d'accès unifié aux processus intra- et intercellulaires (409). Ce type

de base de données peut permettre de raffiner les GRN et de les rapprocher du contexte cellulaire, améliorant ainsi la pertinence des conclusions.

Dans une étude, pour s'affranchir de la dépendance aux tissus, Ding et al. ont développé metaVIPER, qui agrège des centaines d'interactomes dérivés de jeux multiomiques afin d'inférer l'activité de protéines régulatrices dans des échantillons orphelins, tels que des cancers rares ou mal caractérisés, ou des cellules uniques.

Plus de la moitié des gènes porteurs de mutations récurrentes convergent vers un petit ensemble de maîtres régulateurs, illustrant la convergence fonctionnelle des altérations et offrant une base mécanistique solide pour la priorisation thérapeutique (410). Le modèle LINGER pousse l'intégration plus loin : un réseau de neurones pré-entraîné sur des atlas bulk publics, formant un foundation model, est affiné sur des couples scRNA-seq / scATAC-seq tumoraux. Le gain de précision permet de distinguer les gènes « trans-dominants » (contrôlés par des facteurs de transcription) et « cis-dominants » (dépendants de l'accessibilité chromatinienne), révélant de nouveaux régulateurs trans-cancers, tels qu'IKZF1 dans les leucémies ou TFEB dans le myélome multiple (411).

Ces informations pourraient être intégrées à notre modèle, soit en filtrant la liste des régulateurs pour ne conserver que ceux dont l'expression est possible, soit dès les premières étapes d'inférence du réseau de régulation, en restreignant les liens corégulateurs–gènes cibles à ceux effectivement plausibles au regard des informations a priori.

1.2 Mise à l'échelle des GRN

Les analyses présentées dans ce manuscrit ont été effectuées sur des données issues du séquençage de l'ARNm, par technologie microarray ou RNA-seq, autrement appelé « RNA-seq bulk », car la mesure de l'expression génique y est réalisée sur une masse de cellules.

Néanmoins, ce mode de séquençage produit un signal moyen de l'expression génique pour un échantillon. L'hétérogénéité intra-échantillon est ainsi fortement atténuée dans ce signal global.

1.2.1 Augmenter la résolution : le single-cell

L'essor du séquençage de l'ARN en cellule unique (scRNA-seq) a bouleversé notre capacité à cartographier l'expression génique, en rendant possible l'exploration de l'hétérogénéité cellulaire intratumorale — phénomène central en oncologie — ou des étapes précoces du développement embryonnaire, caractérisées par de très petits contingents cellulaires. Contrairement au RNA-seq bulk, qui ne fournit qu'une expression moyenne issue de milliers de cellules, le scRNA-seq saisit la variabilité transcriptionnelle propre à chaque cellule, ouvrant ainsi la voie à une compréhension plus fine des dynamiques cellulaires et de leurs dérégulations tumorales. L'inférence des réseaux de régulation génique à partir de données scRNA-seq, associée à des informations de phénotypage des cellules, permet de construire des réseaux spécifiques à un type ou un état cellulaire, en capturant l'hétérogénéité cellulaire intra-échantillon. Aujourd'hui, les technologies de séquençage permettent de quantifier d'autres modalités : l'accessibilité des régions chromatiniennes via scATAC-seq, la conformation de la chromatine via scHi-C, et les interactions entre protéines et ADN via scChIP-seq, facilitant ainsi une capture complète de la dynamique intermoléculaire au sein des cellule (412).

Par exemple, CellOracle, MICA et IReNA utilisent scRNA-seq et scATAC-seq séparément, en deux étapes, ce qui implique de filtrer les liens de régulation putatifs, puis de construire le GRN final, ou inversement (413–415). 'ajout de données multimodales (ex. scRNA-seq + scATAC-seq) permet de lier l'expression génique à l'accessibilité de l'ADN et à la présence de motifs de liaison de facteurs de transcription (TF), améliorant l'interprétabilité et la directionnalité des régulations, tout en limitant les faux positifs.

Cependant, l'intégration de ces données augmente fortement les coûts en temps de calcul et en ressources mémoire. De plus, la nature creuse des données single-cell (fort taux de zéros dans les matrices d'expression) les rend particulièrement sensibles au bruit technique et à la variabilité biologique, ce qui peut biaiser l'inférence (262,412). Enfin, les opportunités offertes par les approches en cellules uniques ont provoqué une explosion des méthodes bio-informatiques d'analyse et de la quantité de données associées.

Quelques études parues entre 2024 et 2025 démontrent la maturité des approches d'inférence de GRN à partir de données transcriptomiques en cellule unique. Green et al. combinent scRNA-seq et scATAC-seq multi-cancer sur plus de 100 000 cellules avec SCENIC, enrichi de liens enhancer-gène (ChromVAR/ABC), et révèlent un régulon BATF–RUNX3 pilotant la transition des lymphocytes résidents mémoire vers des lymphocytes épuisés, dont la suppression réactive les lymphocytes $CD8^+$ infiltrés dans les tumeurs (416). Iglesia et al. intègrent snATAC-seq et snRNA-seq dans 61 cancers du sein. Via le couplage SCENIC + Link-Peaks, ils identifient BHLHE40 (LumA/B) et KLF5 (Basal-like) comme nœuds maîtres, et relient les duos SOX6/KCNQ3 (basal) et FAM155A/LRP1B (luminal) aux phénotypes cliniques (417). Park et al. ont analysé 150 000 cellules de gliome soumises au pitavastatin et, grâce à SCINGE, révèlent une boucle feed-forward STAT3 → CEBPβ → MCL1 qui s'amplifie sous traitement. À partir de ce résultat, ils ont démontré que le knock-out de CEBPβ retarde la rechute chez la souris (418).

Ces travaux illustrent la convergence de données multiomiques, de la dynamique et des cribles fonctionnels pour transformer des réseaux inférés en cibles testables, hissant les GRN au cœur de la prochaine médecine de précision.

1.2.2 Une dimension supérieure : Transcriptomique spatiale

Récemment, des techniques d'analyse omiques spatiales ont été développées pour ajouter une dimension spatiale aux analyses bio-informatiques. La transcriptomique spatiale permet d'ancrer les interactions régulateurs-gènes cibles dans des régions tissulaires précises et d'obtenir des GRN spatialement restreints (419). Dans la même approche, les analyses spatiales peuvent révéler les micro-environnements fonctionnels, là où, par exemple, la régulation inflammatoire est plus importante au contact des macrophages, ou là où la réponse hypoxique semble plus active autour des capillaires (420,421).

Néanmoins, ce genre d'approche souffre de limitations. Premièrement, la résolution est encore limitée : la technologie 10x Visium permet d'analyser des coupes histologiques carrées de 6,5 mm de côté, pour un total de 4 992 sondes, et chaque spot de capture a une longueur de 55 µm, ce qui est suffisant pour inclure plusieurs cellules. Même si la solution de Nanostring offre une sélection de la région d'intérêt pouvant être aussi petite que 10 µm, le débit du GeoMx n'est pas propice à l'analyse d'une seule cellule : il faut au moins 20 à 300 cellules par ROI pour une quantification fiable des transcrits et des protéines.

Concrètement, on bénéficierait de l'information de coexpression spatiale des gènes, permettant d'ajouter un degré de précision supplémentaire aux couples coactivateur (ou corépresseur) – gène cible.

D'une autre façon, il serait possible d'inférer des réseaux de corégulation spécifiques à certaines parties des données spatiales en appliquant d'abord des solutions de détection automatique de domaines cellulaires(422,423). Plutôt que d'obtenir un seul réseau de régulation moyen à la fin de l'analyse, nous aurions plusieurs réseaux de régulation spécifiques reflétant l'hétérogénéité spatiale tissulaire, nous donnant la capacité d'étudier les perturbations spécifiques d'une population.

1.2.3 Représenter les trajectoires de variations phénotypiques

L'hétérogénéité intratumorale rend compte du fait que des variations génétiques et non génétiques sont à l'œuvre au sein d'une même tumeur. Ces variations engendrent une plasticité phénotypique de la tumeur, notamment via l'impact d'une variation du microenvironnement tumoral (424), pouvant conduire à un changement d'état tumoral. Cette variation d'état tumoral a un impact sur la prise en charge du patient et sur son pronostic, car elle est associée à une diminution de la survie globale et à une augmentation du risque de progression (425). Elle favorise également l'émergence de mécanismes d'échappement thérapeutique via une adaptation rapide des cellules tumorales au stress (426).

Il semble donc pertinent de vouloir tracer ces variations d'état tumoral en identifiant les trajectoires empruntées par les cellules cancéreuses. L'onglet CoRegQuery, associé à la projection de nouvelles données sur notre carte de tumeurs pulmonaires, permet de visualiser ces transitions phénotypiques. Par exemple, Simard et al. ont étudié la variation du microenvironnement tumoral de NSCLC présentant l'altération METex14Del après traitement par tépotinib, un inhibiteur de MET. Le microenvironnement est passé d'un état immunosuppressif à un paysage immuno-permissif, avec une augmentation des cellules présentatrices d'antigènes et des cellules pro-inflammatoires, ainsi qu'une réduction notable des molécules de checkpoint immunitaire, dont PD-L1.

Projetées sur notre carte 2D des tumeurs pulmonaires, ces données révèlent une différence de zone de projection suivant le design avant/après traitement, laissant supposer une variation de l'état tumoral (Figure 41).

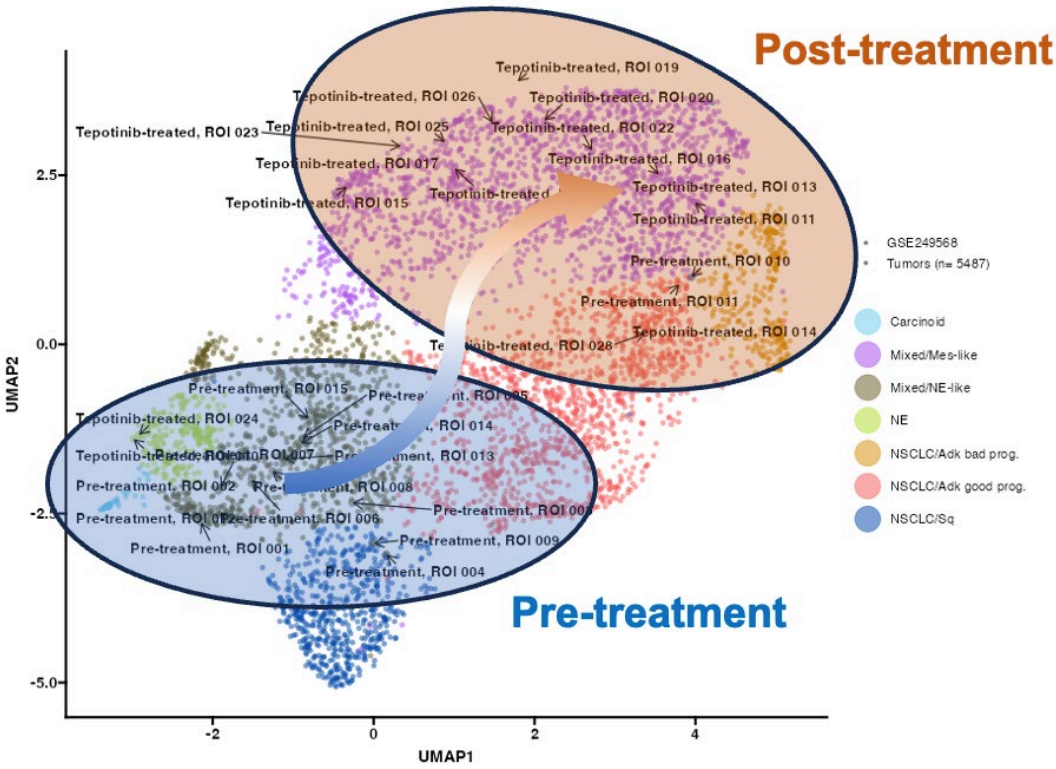


Figure 41 : Projection du jeu de données GSE249568.

On peut facilement représenter une bascule phénotypique des tumeurs après traitement.

L'une des méthodes utilisables pour étudier cette variation d'état cellulaire est l'analyse de trajectoires pseudo-temporelles à partir de données transcriptomiques. L'analyse pseudo-temporelle ordonne les cellules tumorales le long de trajectoires transcriptionnelles et reconstruit des transitions évolutives intra-tumorales. Parmi les principaux algorithmes utilisés, on peut citer Monocle 3 et Slingshot (427–429). Appliquée en oncologie, cette approche a mis en évidence des états cellulaires hétérogènes liés au cycle, à la résistance pharmacologique et à la maladie résiduelle minimale après double inhibition RAF/MEK dans le mélanome (430,431) et a révélé un programme de p-EMT prédictif de métastases dans le carcinome épidermoïde de la tête et du cou, affiné par l'intégration de données unicellulaires et bulk (432).

Une autre méthode d'analyse pseudo-temporelle est l'analyse de vélocité RNA-seq, qui estime la dérivée temporelle de l'expression génique en exploitant les ratios ARN non épissés/épissés, et permet de prédire la trajectoire transcriptionnelle à court terme des cellules tumorales (433). Des approches comme CellRank et sa version multivue, ou encore cell2fate, transforment ces vecteurs de vélocité en probabilités de destin, identifient les états initiaux/terminaux et modélisent des réseaux biophysiques plus réalistes, améliorant la robustesse des inférences en contexte cancéreux (434,435). Appliquée à l'oncologie, la vélocité a permis de reconstruire des trajectoires multiples et indépendantes menant à la tolérance ou à la résistance aux thérapies ciblées dans le mélanome, révélant des sous-populations persistantes se reprogrammant métaboliquement sous inhibition RAF/MEK ou tyrosine kinase (436,437). Étendue à d'autres données multiomiques couplant snRNA-seq, snATAC-seq et transcriptomique spatiale, des études montrent que l'information directionnelle de la vélocité se

combine utilement aux changements d'accessibilité chromatinienne et à la localisation tissulaire pour contextualiser l'émergence d'états résistants (438,439).

Cependant, Bergen et al. (440) puis Zheng et al.(441) nt mis en lumière plusieurs dangers concernant ce type d'analyse. La vélocité repose sur des hypothèses (état quasi-stable, taux d'épissage homogènes parmi l'ensemble des gènes), reste sensible aux erreurs de comptage des transcrits, exige des modèles dynamiques et des contrôles rigoureux, et l'existence de différences dans les dynamiques de régulation des gènes peut engendrer des évolutions de transition d'état cellulaire contrefactuelles.

Au total, la vélocité RNA-seq peut s'avérer utile pour prédire les trajectoires tumorales, anticiper la résistance, stratifier les patients et concevoir des thérapies adaptatives, comme le confirment les synthèses récentes en glioblastome, sein ou vessie (442–444).

1.2.4 Meta-GRN

Que ce soit via les données en cellule unique ou la transcriptomique spatiale, ces approches permettent d'intégrer les analyses en réseau de régulation génique dans l'étude des comportements multicellulaires à l'aide de la modélisation basée sur les agents. Ce genre d'analyse met en œuvre une modélisation mathématique des cellules individuelles ou des régions cellulaires pour répondre à des questions portant sur la complexité des systèmes biologiques. Chaque cellule individuelle ou type cellulaire est considéré comme un agent doté de son propre état (réseau), évoluant au sein d'un environnement continu, où peuvent être prises en considération la présence de cytokines ou de médicaments, et constitué d'une matrice extracellulaire avec ses propriétés mécaniques ou topologiques. Pas à pas, tous les agents et l'environnement évoluent.

Les comportements cellulaires reposent sur des règles déterministes ou stochastiques, et ces règles peuvent être pilotées par les GRN inférés pour chaque agent (445). Par exemple, la méthode PhysiBoSS (446) associe l'inférence d'un réseau booléen stochastique pour la signalisation intracellulaire d'une cellule, MaBoSS (447). PhysiCell gère la représentation de la mécanique des cellules et des comportements phénotypiques clés. Une cellule est représentée comme une sphère molle avec deux rayons : cellulaire et nucléaire. Elle peut se déplacer et interagir avec les objets voisins, se diviser et changer ses propriétés en fonction de conditions spécifiques (448). BioFVM, quant à lui, gère la simulation d'une ou plusieurs entités environnementales diffusantes. Il simule la diffusion, la dégradation et la libération d'entités diffusibles dans l'espace extracellulaire, y compris la matrice extracellulaire (449). PhysiBoSS a montré sa capacité à relier la signalisation intracellulaire et les interactions mécaniques qui gouvernent l'invasion tumorale. En modulant simultanément les paramètres du modèle et les nœuds du réseau de signalisation, le modèle reproduit les principaux modes d'invasion et peut servir de banc d'essai virtuel pour identifier des cibles pharmacologiques ou des altérations génétiques capables de les neutraliser (450).

2. Portail cRegMap

Le portail interactif en ligne cRegMap permet à la communauté scientifique d'explorer la projection d'une cohorte de patients à travers un réseau de corégulation génique. Actuellement, l'accent a été mis sur les capacités de l'application à caractériser les sous-groupes déterminés par notre procédure d'analyse.

Notre module de projection de données externes, CoRegQuery, permet d'ajouter de nouvelles modalités à nos sous-groupes, notamment des données de perturbations pharmacologiques.

La projection des profils transcriptomiques perturbés pharmacologiquement permet de cartographier les régulateurs et les voies de signalisation affectées (451) u d'identifier les régulateurs médiateurs de la réponse pharmacologique, de la même manière que nous l'avons fait chez les cellules METex14Del, où nous avons identifié un module de régulateurs clés médiateurs de la réponse à l'exposition à l'HGF. La projection de nombreuses données de perturbation permet d'identifier les groupes de molécules qui provoquent des perturbations du GRN, des changements de sous-groupes d'origine similaire ou des susceptibilités pharmacologiques analogues. Dans cet effort d'étude de perturbation pharmacologique, des bases de données ont été constituées et peuvent servir de matière première à nos analyses. On peut citer la plateforme L1000, un projet à faible coût de séquençage ciblé de 1000 gènes, produisant 1 319 138 profils L1000 pour 42 080 perturbations génétiques (452), a base de données Genomics of Drug Sensitivity in Cancer (453) (453) qui comprend les analyses RNA-seq de plus de 1000 lignées cellulaires et 600 composés, ou la récente étude Tahoe-100M, qui a produit un atlas à l'échelle de la cellule unique à partir de 100 millions de profils transcriptomiques, mesurant l'impact de chacune des 1 100 petites molécules sur les cellules de 50 lignées cellulaires cancéreuses. Cependant, une telle approche n'est pas sans limitations. Premièrement, la perturbation pharmacologique agit à plusieurs niveaux, comme le protéome, le métabolome ou la signalisation cellulaire, alors que le GRN se base sur la transcriptomique. En ce sens, il se pourrait que la projection fournie par le GRN ne capture qu'une partie de l'effet. Ensuite, les analyses par co-activation de certains régulateurs ou de coexpression de gènes cibles peuvent entraîner un risque de surinterprétation, d'où l'intérêt d'un filtrage causal sur le GRN.

Enfin, certaines bases de données proposent des profils transcriptomiques restreints (L1000) ou ne comportent que des données RNA-seq. Les autres dimensions cellulaires ne sont pas explorées, ce qui peut entraver la validation des conclusions du modèle par des analyses orthogonales.

2.1 Affiner les prédictions

La collaboration avec l'équipe du Dr Tulasne, dans le cadre de l'étude de la stimulation du récepteur METex14Del par l'HGF, a conduit à l'identification d'un nœud de régulation composé de trois facteurs de transcription interconnectés : ETS1, FOSL1 et SMAD3, tous associés à la voie de signalisation RAS-ERK. Dans ce travail, l'une de nos contributions a été de fournir les gènes cibles de ces régulateurs. FOSL1, ETS1 et SMAD3 possèdent respectivement 431, 440 et 756 gènes cibles, dont 174, 191 et 364 activés, et 257, 249 et 392 réprimés. Le nombre de gènes cibles associés à un régulateur peut facilement être très élevé. Durant cette collaboration, nous avons classé les associations régulateurs–gènes cibles selon la corrélation entre l'expression de chaque gène cible et l'influence de son régulateur, ainsi que la présence d'interactions de régulation connues. Sur ce point, des améliorations peuvent être apportées à notre approche afin d'optimiser la sélection expérimentale et de réduire le ratio entre cibles à tester et cibles initiales. La visualisation du GRN via le graphe des régulateurs constitue un

point de départ, mais cette vue centrée sur les corégulateurs ne répond pas aux questions biologiques fondamentales. Les phénotypes tumoraux dépendent de gènes cibles spécifiques. Cette vue condensée ne permet ni d'identifier le gène médiateur, ni de savoir s'il est régulé de manière unique ou redondante. Par exemple, il est difficile de déterminer si une fonction pro-migratoire attribuée à un régulateur R1 est due à l'activation de MMP2 ou à un ensemble de gènes également régulés par R2. Deux régulateurs peuvent paraître proches du fait de cibles communes, alors qu'ils possèdent aussi un répertoire privé de cibles critiques. La suppression combinée de R1 et R2 pourrait n'être létale que via ces cibles exclusives — une interdépendance que la visualisation classique sous-estime. Pour chaque gène cible GG, hLICORN génère plusieurs réseaux locaux (coactivateurs et corépresseurs), puis ajuste un modèle de régression linéaire intégrant l'expression des régulateurs et leurs interactions. Le meilleur modèle, basé sur le R^2 ajusté, est conservé. À ce score prédictif est associé un score hybride intégrant les preuves expérimentales (ChIP-seq, motifs PWM, ATAC-seq). Une solution consiste à attribuer à chaque lien régulateur–cible un score agrégé :

$$S = \omega_1(Score\ hybride) + \sum_{i=1}^{n} \omega_i(Preuves\ expérimentales) + \omega_{n+1}(Corrélation).$$

Par exemple, ce score agrégé pourrait combiner le score CoRegNet, les preuves expérimentales d'interaction régulateur–gène cible (ChIP-seq, motifs de liaison, ATAC-seq), ainsi que la corrélation entre l'influence du régulateur et l'expression de son gène cible. De cette manière, l'ensemble des gènes cibles associés à un régulateur ne serait plus simplement réparti en deux listes distinctes (gènes activés et gènes réprimés) aux éléments équivalents, mais structuré en deux listes hiérarchisées, où chaque interaction est pondérée selon sa plausibilité fonctionnelle et sa validité expérimentale.

2.2 Agrégation modulaire et simplification visuelle

Lors de la visualisation du réseau de corégulation, il est possible de régler un paramètre permettant d'afficher tous les régulateurs du réseau ou seulement ceux qui reviennent le plus fréquemment dans les GRN locaux. Un seuil bas affiche l'intégralité des régulateurs et génère une représentation visuelle fidèle du GRN, mais provoque un effet « pelote de laine » rendant son exploration et son interprétation visuelle difficiles. Un seuil plus élevé génère un réseau plus simple et plus facile à interpréter, mais masque l'essentialité des interactions de corégulation. Pour conserver l'information issue de l'ensemble des régulateurs du réseau tout en préservant la lisibilité de la représentation visuelle, un compromis est possible. À partir d'un réseau fortement connecté, on peut appliquer un algorithme de détection de communautés, c'est-à-dire un algorithme capable d'identifier, dans un graphe complexe, des groupes de nœuds présentant plus de connexions internes qu'avec le reste du réseau (357,454,455). Appliqués aux GRN, ces algorithmes réduisent la complexité visuelle en regroupant les nœuds en modules, lesquels peuvent ensuite être analysés pour mettre en évidence des programmes transcriptionnels particuliers associés à des fonctions biologiques, en lien avec des bases telles que Gene Ontology (408), Reactome (456) ou les *Hallmarks* du cancer (457). Le passage à une représentation simplifiée — mais non simpliste — du réseau de corégulation s'inscrit également dans une démarche de priorisation des validations expérimentales.

Plutôt que de tester arête par arête, on cible d'abord un module entier afin de maximiser la probabilité d'observer un phénotype pertinent.

Un autre axe de développement et d'amélioration consiste à permettre la détection automatique des modules fonctionnels de notre réseau de régulation génique.

Par module fonctionnel, j'entends l'enrichissement du GRN dans le but d'annoter, regrouper ou construire des sous-réseaux associés à des processus biologiques spécifiques. Grâce à plusieurs avancées méthodologiques, l'oncologie a démontré que l'enrichissement des GRN avec des référentiels de signalisation (KEGG, Reactome ou MSigDB) permet d'identifier automatiquement des modules fonctionnels cohérents, souvent plus informatifs que les voies canoniques considérées isolément. Après avoir reconstruit des GRN spécifiques à plus de 30 cohortes tumorales à l'aide d'ARACNe, l'algorithme réévalue l'activité de chaque facteur régulateur, puis projette ces scores sur les modules de Reactome/MSigDB pour révéler des régulateurs cachés. Appliqué au neuroblastome et au médulloblastome, NetBID2 met en évidence des noyaux kinasiques (CDK-MYC, WNT-β-caténine) dont la surexpression est parfois absente aux niveaux génomique ou transcriptionnel, mais dont l'activité prédit la sensibilité à des inhibiteurs ciblés (458). Utilisé sur des jeux RNA-seq appariés diagnostic/rechute dans la leucémie aiguë lymphoïde, NetBID2 a identifié le module de polyglutamylation du folate comme principal régulateur caché de la rechute, lié au métabolisme du méthotrexate. Il a également détecté une activation compensatoire du nœud DHFR-TYMS, signalant une dépendance métabolique exploitable : les cellules FPGS-basses deviennent vulnérables au trimetrexate, prédiction confirmée in vitro et dans des modèles murins de xénogreffe (459). Une autre étude, centrée sur le glioblastome, utilise PANDA/LIONESS pour générer des réseaux individualisés, puis évalue la centralité des gènes du module PD-1 extrait de KEGG. Les auteurs montrent qu'une dérégulation du module PD-1, plutôt que son expression absolue, distingue les tumeurs au pronostic défavorable, indépendamment de la charge mutationnelle ou de l'infiltration immunitaire. Cette signature topologique, validée dans la cohorte German Glioma Network, positionne le statut de régulation de PD-1 comme un biomarqueur pertinent pour orienter les stratégies d'immunothérapie dans le GBM(460).

3. Étude du cancer du poumon

3.1 NSCLC vs SCLC

Notre étude du réseau de corégulation du cancer du poumon s'est appuyée sur des données transcriptomiques issues de lignées cellulaires de NSCLC et de SCLC. Nous avons délibérément inclus des lignées cellulaires représentant deux types de cancers du poumon, histologiquement très différents et historiquement étudiés séparément, afin d'enrichir le contexte de l'inférence du GRN, de démontrer sa capacité à discriminer les SCLC des NSCLC, et de maximiser le nombre d'échantillons pour détecter des événements rares. Cependant, la projection 2D des tumeurs pulmonaires selon notre modèle UMAP révèle trois bassins distincts : un bassin dense en tumeurs neuroendocrines, un bassin de carcinomes épidermoïdes, et un vaste bassin d'adénocarcinomes (Figure 42).

Une approche spécifique à chaque sous-type de cancer du poumon impliquerait l'inférence de réseaux de régulation distincts pour les NSCLC et les SCLC.

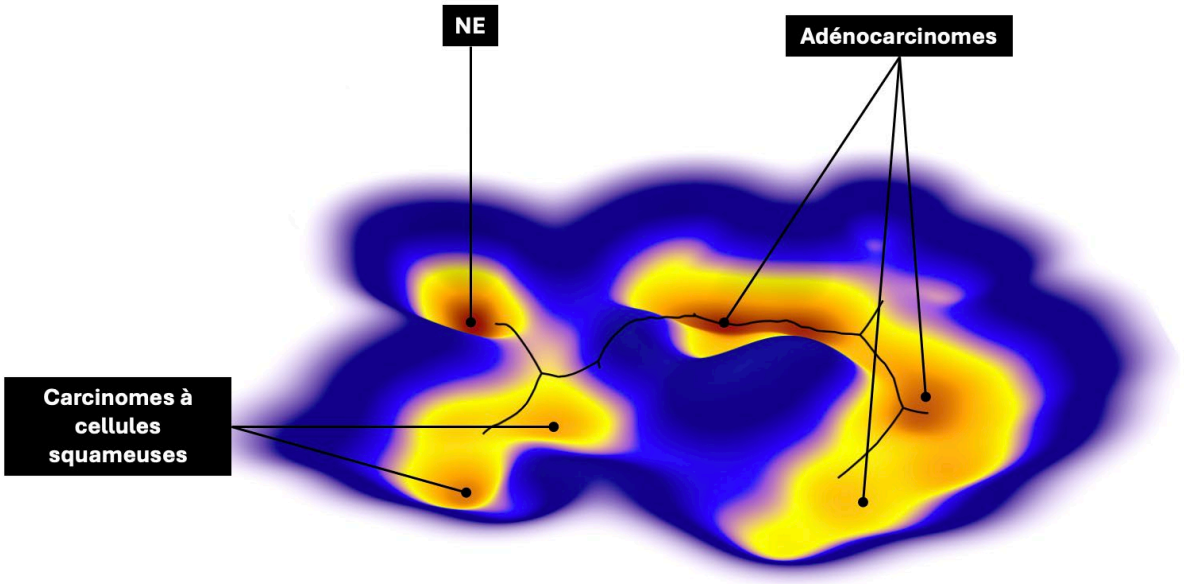


Figure 42 : Carte de densité des tumeurs pulmonaires sur la projection 2D du modèle UMAP

La carte de densité des tumeurs de cancer du poumon montre trois bassins de tumeurs, un bassin neuroendocrine (NE), un bassin de carcinomes épidermoïdes et un grand bassin d'adénocarcinomes.

L'inférence du GRN dans le cancer du poumon menée dans notre étude a permis de révéler la mixité de sous-types histologiques pouvant coexister au sein d'une même tumeur. Certaines tumeurs pulmonaires présentent simultanément des cellules évoquant un adénocarcinome et d'autres un carcinome épidermoïde (461). Notre approche globale des tumeurs pulmonaires a mis en évidence que la discrimination tumorale repose non seulement sur l'homogénéité histologique, mais également sur d'autres composantes telles que la composition du microenvironnement tumoral et les processus d'EMT en cours. L'inférence de GRN spécifiques aux NSCLC, et plus particulièrement aux adénocarcinomes ou aux carcinomes épidermoïdes, pourrait offrir une opportunité d'exploration approfondie de ces processus. À titre d'exemple, l'analyse d'un GRN inféré à partir de données transcriptomiques de lignées cellulaires pulmonaires soumises à un processus d'EMT a permis d'identifier de nouveaux gènes signatures de la dynamique de l'EMT, présentant des profils d'expression spécifiques au cours de cette transition(462). La même approche, appliquée à des données multiomiques, a permis d'identifier un ensemble de gènes associés à la progression tumorale (463). En 2022, la reconstruction de réseaux de coexpression génique et de méthylation à partir de carcinomes épidermoïdes pulmonaires, suivie de leur fusion, a révélé des interactions étroites entre les gènes impliqués dans le cycle cellulaire, la kératinisation (différenciation épithéliale), le système immunitaire et le métabolisme des xénobiotiques (464). La même année, une étude portant sur sept jeux de données d'expression publics a identifié des régulateurs clés récurrents associés à la progression, aux rechutes et à la résistance aux traitements dans l'adénocarcinome pulmonaire.

Ces résultats montrent qu'il est possible d'analyser plus finement un sous-type tumoral particulier afin de mieux comprendre les processus biologiques spécifiques en jeu. On s'éloigne ainsi d'une simple comparaison de sous-groupes définis par algorithmes pour approfondir les mécanismes propres à un groupe donné.

Ainsi, une analyse ciblée permet d'accéder à une granularité plus fine, révélant à la fois les voies communes en place et les mécanismes moléculaires spécifiques.

3.2 Transdifférenciation NSCLC vers SCLC

Au sein du pôle neuroendocrine regroupant les clusters NE, Carcinoid et Mixed/NE-like il est possible de faire une analyse plus profonde et de déterminer un cluster riche en adénocarcinome (Figure 43).

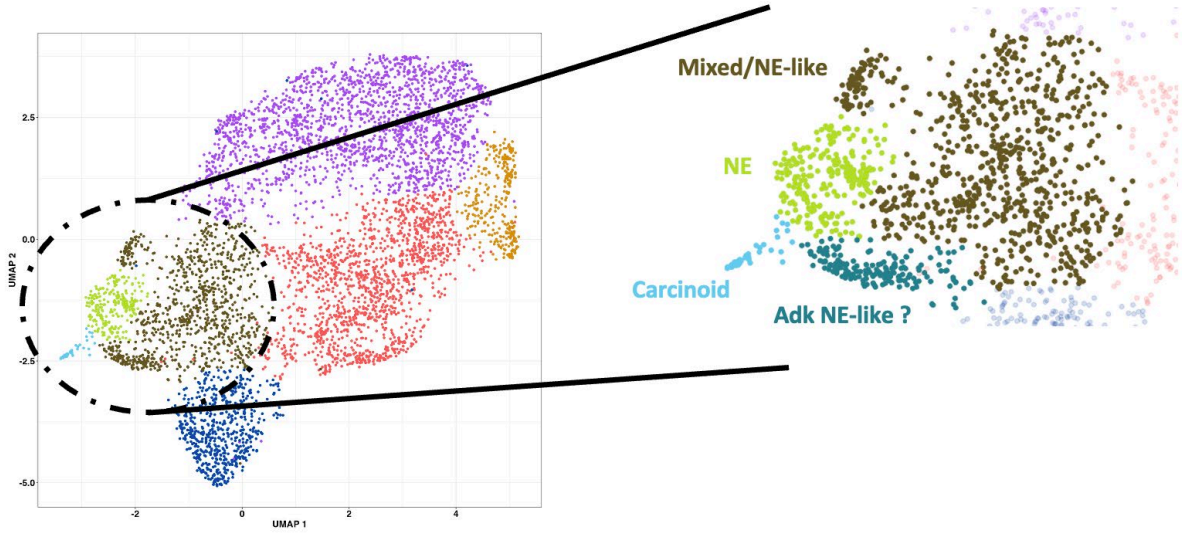


Figure 43 : Sous-cluster NE riche en adénocarcinomes.

L'étude fine du groupes de clusters à signature neuroendocrine révèle un sous-cluster riche en adénocarcinomes.

Les adénocarcinomes sont pourtant très différents des tumeurs neuroendocrines. Cependant, un phénomène de transdifférenciation a été décrit, dans lequel certaines tumeurs initialement classées comme adénocarcinomes acquièrent progressivement un phénotype de SCLC. Ce phénomène, rapporté dans 3 à 14 % des cas d'adénocarcinomes (465,466) constitue un mécanisme majeur d'échappement thérapeutique, observé principalement lors des traitements ciblant les altérations oncogéniques de type EGFR ou ALK. Cette transformation implique une reprogrammation cellulaire, au cours de laquelle les cellules tumorales perdent leur identité épithéliale pour acquérir un phénotype neuroendocrine. Ce processus repose sur l'inactivation des gènes suppresseurs de tumeurs RB1 et TP53, ainsi que sur des modifications épigénétiques globales modulant l'expression de facteurs de transcription neuroendocriniens tels qu'ASCL1, NEUROD1 ou POU2F3 (48). Plusieurs études récentes soulignent l'implication précoce d'une plasticité cellulaire intrinsèque à certaines sous-populations tumorales épithéliales, caractérisées par une transition épithélio-neuroendocrine partielle précédant l'initiation des thérapies ciblées (467,468). ne étude récente, combinant transcriptomique classique, transcriptomique spatiale et immunofluorescence, a montré qu'avant leur transformation en SCLC, les NSCLC présentent un paysage transcriptomique les rendant plus susceptibles à cette transition, notamment via des

altérations épigénétiques spécifiques des cellules tumorales. Les auteurs ont également mis en évidence des différences moléculaires marquées entre les SCLC primaires et ceux dérivés d'une transformation, ainsi qu'un état d'épuisement immunitaire corrélé à la durée du traitement par EGFR-TKI (469). Compte tenu des preuves apportées par cette étude, mettant en évidence une signature transcriptionnelle associée à une prédisposition à la transformation des adénocarcinomes en SCLC, ainsi que de l'enjeu thérapeutique et de l'intérêt pronostique majeurs que représente cette problématique, il apparaît pertinent d'étudier plus en détail ce sous-cluster d'adénocarcinomes par une analyse ciblée de son réseau de régulation génique.

3.3 Anticiper les expérimentations

Une autre difficulté rencontrée a concerné la validation, par siRNA, des gènes cibles de leurs corégulateurs. L'inhibition conjointe de la triade de régulateurs FOSL1, ETS1 et SMAD3 a induit une inhibition partielle et cumulative de la réponse de type wound healing. Cependant, les mesures d'expression de certains gènes cibles communs à ces régulateurs révèlent des profils contrastés : certains, tels que VIM, NOG et SERPINA1, présentent une diminution plus marquée de leur expression lors de l'inhibition complète de la triade régulatrice, tandis que d'autres, comme IRS1, ABCA1 ou LETM2, ne montrent pas d'effet cumulatif. Avant de valider expérimentalement la robustesse des liens régulateur–gène cible, une estimation in silico de ces perturbations peut être envisagée. Il s'agirait de simuler l'inactivation ou la surexpression de chaque régulateur, puis de mesurer la variation d'expression des cibles partagées.

Des outils comme dyngen (470) ou CellOracle (415) permettent de mettre en place ce type de simulation. Ils offrent la possibilité de quantifier l'impact effectif d'un facteur de transcription sur ses gènes cibles et de sélectionner uniquement les liens biologiquement pertinents à tester expérimentalement.

Conclusion

La biologie des systèmes vise à étudier les interactions entre réseaux de gènes, de protéines et de métabolites au sein d'un organisme, en mobilisant à la fois des modèles mathématiques et des simulations informatiques. La révolution technologique du XX^{e} siècle a permis une étude fine des tumeurs et a généré une masse considérable de données expérimentales, inaugurant ainsi l'ère de la biologie des systèmes expérimentale. Cette profusion de données a motivé le développement d'algorithmes de reconstruction de réseaux de régulation génique. Ces derniers visent à modéliser les mécanismes moléculaires régissant le comportement cellulaire, en identifiant les interactions régulatrices à partir de profils d'expression génique. Depuis la découverte des premiers sous-types de cancers du poumon, leur classification a évolué au gré des avancées technologiques : d'abord fondée sur des critères microscopiques, puis immunohistochimiques, et enfin moléculaires. Cette spécialisation progressive vise à optimiser la prise en charge des patients.

L'évolution des options thérapeutiques a notamment conduit au développement des thérapies ciblées, véritable atout dans l'arsenal thérapeutique lorsque la tumeur exprime un marqueur oncogénique à l'origine d'une addiction oncogénique. Parmi l'ensemble des marqueurs oncogéniques, les mutations METex14Del dans le cancer du poumon représentent environ 3 % des cas de NSCLC. Sur les 2,1 millions de cas annuels dans le monde, cela correspondrait à environ 63 000 patients porteurs d'une mutation METex14Del.

Cette découverte a conduit au développement de thérapies ciblées contre METex14Del. Les résultats des essais cliniques de phase I/II, GEOMETRY mono-1 et VISION, ont démontré une activité clinique significative chez les patients traités par capmatinib ou tepotinib. Cependant, environ 50 % des patients porteurs de la mutation METex14Del répondent au traitement, ce qui suggère l'existence de mécanismes additionnels responsables de l'addiction oncogénique. Par ailleurs, les travaux de l'équipe du Dr Tulasne ont montré que l'activation et l'oncogénicité du récepteur METex14Del nécessitent la présence de son ligand, l'HGF. Contrairement à d'autres oncogènes drivers, les altérations conduisant à METex14Del ne suffisent pas à induire une activation constitutive du récepteur.

À l'intersection de ces deux domaines, l'application des réseaux de régulation génique à l'étude des mécanismes intrinsèques du cancer du poumon offre la possibilité d'agréger, au sein d'un même cadre de comparabilité, de nombreuses données expérimentales, notamment les profils transcriptomiques de patients atteints de cette pathologie. Nous avons inféré un réseau de régulation génique spécifique au cancer du poumon, constitué de 502 régulateurs, 18 053 interactions régulatrices et 4 143 gènes cibles, à partir de profils transcriptomiques de lignées cellulaires tumorales. La plateforme lung.cRegMap a été développée pour intégrer ce réseau ainsi que les profils d'expression de lignées cellulaires et de plusieurs cohortes, totalisant environ 5 487 patients atteints de cancer du poumon. Ce type de cancer, caractérisé par une grande hétérogénéité et des mécanismes de régulation complexes, a fait l'objet d'une analyse transcriptomique permettant d'inférer son réseau de régulation génique. Cette analyse à grande échelle a permis de définir sept sous-groupes distincts de

cancers du poumon, chacun caractérisé par un ensemble spécifique de régulateurs, fournissant ainsi un éclairage sur les processus moléculaires sous-jacents. Cette classification en sous-groupes affine les typologies existantes, en s'appuyant sur des différences fonctionnelles dans les réseaux de régulation génique plutôt que sur les seuls critères histopathologiques ou mutationnels.

Le réseau de signalisation du cancer du poumon a été étudié dynamiquement par l'analyse de données transcriptomiques issues de cellules porteuses de la mutation METex14Del, exposées ou non au ligand HGF. L'analyse de ce réseau a mis en évidence un rôle central de la voie RAS-ERK : la mutation METex14Del induit un nœud régulateur majeur constitué des facteurs de transcription ETS1, FOSL1 et SMAD3, ainsi qu'une activation soutenue de RAS-ERK, strictement dépendante de la présence du ligand HGF. Ce résultat souligne la prédominance de la voie RAS-ERK dans l'addiction oncogénique induite par METex14Del, en accord avec des observations antérieures d'activation prolongée de cette cascade dans ce contexte. Cette étude confirme l'intérêt de l'approche par réseau de régulation génique comme stratégie originale pour identifier des nœuds transcriptionnels clés, associés à des voies de signalisation spécifiques, susceptibles de constituer de nouvelles cibles thérapeutiques.

L'application des méthodes de biologie des systèmes et d'intelligence artificielle aux données transcriptomiques pulmonaires a permis d'approfondir la compréhension de l'addiction oncogénique et de raffiner la classification des tumeurs. Les sept sous-groupes identifiés reflètent une diversité de profils régulateurs et complètent les classifications histopathologiques et moléculaires traditionnelles. La mise en évidence du rôle pivot de la voie RAS-ERK dans la réponse METex14Del a contribué à expliquer la dépendance de ces tumeurs au récepteur MET. Plus largement, ces travaux confirment qu'une analyse à grande échelle des réseaux de régulation génique constitue un outil puissant et robuste pour décrire les circuits moléculaires essentiels à la survie tumorale.

La biologie des cancers n'a pas fini de révéler ses mécanismes les plus complexes.

Références

1.Bray F, Laversanne M, Sung H, Ferlay J, Siegel RL, Soerjomataram I, et al. Global cancer statistics 2022: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA Cancer J Clin. 2024;74(3):229-63.

2.Lortet-Tieulent J, Renteria E, Sharp L, Weiderpass E, Comber H, Baas P, et al. Convergence of decreasing male and increasing female incidence rates in major tobacco-related cancers in Europe in 1988-2010. Eur J Cancer Oxf Engl 1990. juin 2015;51(9):1144-63.

3.Laennec R. A Treatise on the Diseases of the Chest and on Mediate Auscultation [Internet]. New York : Samuel S. and William Wood; 1838. Disponible sur: http://resource.nlm.nih.gov/9308216

4.van Meerbeeck JP, Fennell DA, De Ruysscher DKM. Small-cell lung cancer. Lancet Lond Engl. 12 nov 2011;378(9804):1741-55.

5.Nicholson AG, Tsao MS, Beasley MB, Borczuk AC, Brambilla E, Cooper WA, et al. The 2021 WHO Classification of Lung Tumors: Impact of advances since 2015. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. 19 nov 2021;S1556-0864(21)03316-5.

6.Antoine M, Moroch J, Kebir F, Georges M, Ruppert AM, Wislez M, et al. Classification histomoléculaire des cancers pulmonaires. Rev Mal Respir Actual. 1 oct 2018;10(3):236-47.

7.Prinzi N, Rossi RE, Proto C, Leuzzi G, Raimondi A, Torchio M, et al. Recent Advances in the Management of Typical and Atypical Lung Carcinoids. Clin Lung Cancer. 1 mai 2021;22(3):161-9.

8.Bischoff P, Trinks A, Wiederspahn J, Obermayer B, Pett JP, Jurmeister P, et al. The single-cell transcriptional landscape of lung carcinoid tumors. Int J Cancer. 2022;150(12):2058-71.

9.Lantuejoul S, Fernandez-Cuesta L, Damiola F, Girard N, McLeer A. New molecular classification of large cell neuroendocrine carcinoma and small cell lung carcinoma with potential therapeutic impacts. Transl Lung Cancer Res. oct 2020;9(5):2233-44.

10.Williams JF, Vivero M. Diagnostic criteria and evolving molecular characterisation of pulmonary neuroendocrine carcinomas. Histopathology. 2022;81(5):556-68.

11.Meng L, Cao B, Ji R, Chen DT, Laber DA, Shafique M. Enhanced Efficacy of Chemotherapy by Addition of Immune Checkpoint Inhibitors in Stage IV Large Cell Neuroendocrine Carcinoma of the Lung: A Real-World Analysis. J Cancer. 25 sept 2023;14(17):3169-75.

12.Vendramin R, Litchfield K, Swanton C. Cancer evolution: Darwin and beyond. EMBO J. 15 sept 2021;40(18):e108389.

13.Visser KE de, Joyce JA. The evolving tumor microenvironment: From cancer initiation to metastatic outgrowth. Cancer Cell. 13 mars 2023;41(3):374-403.

14.Wang Q, Shao X, Zhang Y, Zhu M, Wang FXC, Mu J, et al. Role of tumor microenvironment in cancer progression and therapeutic strategy. Cancer Med. 2023;12(10):11149-65.

15.Ostroverkhova D, Przytycka TM, Panchenko AR. Cancer driver mutations: predictions and reality. Trends Mol Med. 1 juill 2023;29(7):554-66.

16.Iranzo J, Martincorena I, Koonin EV. Cancer-mutation network and the number and specificity of driver mutations. Proc Natl Acad Sci. 26 juin 2018;115(26):E6010-9.

17.McGranahan N, Favero F, de Bruin EC, Birkbak NJ, Szallasi Z, Swanton C. Clonal status of actionable driver events and the timing of mutational processes in cancer evolution. Sci Transl Med. 15 avr 2015;7(283):283ra54-283ra54.

18.Carney DN, Gazdar AF, Bepler G, Guccion JG, Marangos PJ, Moody TW, et al. Establishment and identification of small cell lung cancer cell lines having classic and variant features. Cancer Res. juin 1985;45(6):2913-23.

19.Augert A, Zhang Q, Bates B, Cui M, Wang X, Wildey G, et al. Small Cell Lung Cancer Exhibits Frequent Inactivating Mutations in the Histone Methyltransferase *KMT2D/MLL2*: CALGB 151111 (Alliance). J Thorac Oncol. 1 avr 2017;12(4):704-13.

20.Poirier JT, Dobromilskaya I, Moriarty WF, Peacock CD, Hann CL, Rudin CM. Selective tropism of Seneca Valley virus for variant subtype small cell lung cancer. J Natl Cancer Inst. 17 juill 2013;105(14):1059-65.

21.Borromeo MD, Savage TK, Kollipara RK, He M, Augustyn A, Osborne JK, et al. ASCL1 and NEUROD1 Reveal Heterogeneity in Pulmonary Neuroendocrine Tumors and Regulate Distinct Genetic Programs. Cell Rep. 2 août 2016;16(5):1259-72.

22.Ireland AS, Micinski AM, Kastner DW, Guo B, Wait SJ, Spainhower KB, et al. MYC Drives Temporal Evolution of Small Cell Lung Cancer Subtypes by Reprogramming Neuroendocrine Fate. Cancer Cell. 13 juill 2020;38(1):60-78.e12.

23.McColl K, Wildey G, Sakre N, Lipka MB, Behtaj M, Kresak A, et al. Reciprocal expression of INSM1 and YAP1 defines subgroups in small cell lung cancer. Oncotarget. 26 sept 2017;8(43):73745-56.

24.Huang YH, Klingbeil O, He XY, Wu XS, Arun G, Lu B, et al. POU2F3 is a master regulator of a tuft cell-like variant of small cell lung cancer. Genes Dev. 7 janv 2018;32(13-14):915-28.

25.Rudin CM, Poirier JT, Byers LA, Dive C, Dowlati A, George J, et al. Molecular subtypes of small cell lung cancer: a synthesis of human and mouse model data. Nat Rev Cancer. mai 2019;19(5):289-97.

26.Baine MK, Hsieh MS, Lai WV, Egger JV, Jungbluth AA, Daneshbod Y, et al. SCLC Subtypes Defined by ASCL1, NEUROD1, POU2F3, and YAP1: A Comprehensive Immunohistochemical and Histopathologic Characterization. J Thorac Oncol. 1 déc 2020;15(12):1823-35.

27.Shue YT, Drainas AP, Li NY, Pearsall SM, Morgan D, Sinnott-Armstrong N, et al. A conserved YAP/Notch/REST network controls the neuroendocrine cell fate in the lungs. Nat Commun. 16 mai 2022;13(1):2690.

28.Horn L, Mansfield AS, Szczęsna A, Havel L, Krzakowski M, Hochmair MJ, et al. First-Line Atezolizumab plus Chemotherapy

in Extensive-Stage Small-Cell Lung Cancer. N Engl J Med. 6 déc 2018;379(23):2220-9.

29.Gay CM, Stewart CA, Park EM, Diao L, Groves SM, Heeke S, et al. Patterns of transcription factor programs and immune pathway activation define four major subtypes of SCLC with distinct therapeutic vulnerabilities. Cancer Cell. 8 mars 2021;39(3):346-360.e7.

30.Nabet BY, Hamidi H, Lee MC, Banchereau R, Morris S, Adler L, et al. Immune heterogeneity in small-cell lung cancer and vulnerability to immune checkpoint blockade. Cancer Cell. 11 mars 2024;42(3):429-443.e4.

31.Zhang J, Zeng X, Guo Q, Sheng Z, Chen Y, Wan S, et al. Small cell lung cancer: emerging subtypes, signaling pathways, and therapeutic vulnerabilities. Exp Hematol Oncol. 5 août 2024;13(1):78.

32.Schwendenwein A, Megyesfalvi Z, Barany N, Valko Z, Bugyik E, Lang C, et al. Molecular profiles of small cell lung cancer subtypes: Therapeutic implications. Mol Ther - Oncolytics. 26 mars 2021;20:470-83.

33.Ahn MJ, Cho BC, Felip E, Korantzis I, Ohashi K, Majem M, et al. Tarlatamab for Patients with Previously Treated Small-Cell Lung Cancer. N Engl J Med. 29 nov 2023;389(22):2063-75.

34.George J, Lim JS, Jang SJ, Cun Y, Ozretić L, Kong G, et al. Comprehensive genomic profiles of small cell lung cancer. Nature. 6 août 2015;524(7563):47-53.

35.Park S, Hong TH, Hwang S, Heeke S, Gay CM, Kim J, et al. Comprehensive analysis of transcription factor-based molecular subtypes and their correlation to clinical outcomes in small-cell lung cancer. eBioMedicine [Internet]. 1 avr 2024 [cité 3 avr 2024];102. Disponible sur: https://www.thelancet.com/journals/ebiom/article/PIIS2352-3964(24)00097-5/fulltext

36.Sen T, Takahashi N, Chakraborty S, Takebe N, Nassar AH, Karim NA, et al. Emerging advances in defining the molecular and therapeutic landscape of small-cell lung cancer. Nat Rev Clin Oncol. août 2024;21(8):610-27.

37.Gotink KJ, Verheul HMW. Anti-angiogenic tyrosine kinase inhibitors: what is their mechanism of action? Angiogenesis. 1 mars 2010;13(1):1-14.

38.Qin S, Li A, Yi M, Yu S, Zhang M, Wu K. Recent advances on anti-angiogenesis receptor tyrosine kinase inhibitors in cancer therapy. J Hematol OncolJ Hematol Oncol. 12 mars 2019;12(1):27.

39.Pierotti MA, Borrello MG. Receptor Tyrosine Kinases. In: Schwab M, éditeur. Encyclopedia of Cancer [Internet]. Berlin, Heidelberg: Springer; 2011 [cité 6 févr 2025]. p. 3198-203. Disponible sur: https://doi.org/10.1007/978-3-642-16483-5_4982

40.Sigismund S, Avanzato D, Lanzetti L. Emerging functions of the EGFR in cancer. Mol Oncol. 2018;12(1):3-20.

41.Lemmon MA, Schlessinger J. Cell signaling by receptor-tyrosine kinases. Cell. 25 juin 2010;141(7):1117-34.

42.Midha A, Dearden S, McCormack R. EGFR mutation incidence in non-small-cell lung cancer of adenocarcinoma histology: a systematic review and global map by ethnicity (mutMapII). Am J Cancer Res. 2015;5(9):2892-911.

43.Passaro A, Jänne PA, Mok T, Peters S. Overcoming therapy resistance in EGFR-mutant lung cancer. Nat Cancer. avr 2021;2(4):377-91.

44.Tomuleasa C, Tigu AB, Munteanu R, Moldovan CS, Kegyes D, Onaciu A, et al. Therapeutic advances of targeting receptor tyrosine kinases in cancer. Signal Transduct Target Ther. 14 août 2024;9(1):1-51.

45.Reinmuth N, Meister M, Muley T, Steins M, Kreuter M, Herth FJF, et al. Molecular determinants of response to RTK-targeting agents in nonsmall cell lung cancer. Int J Cancer. 2006;119(4):727-34.

46.Hojjat-Farsangi M, Moshfegh A, Daneshmanesh AH, Khan AS, Mikaelsson E, Österborg A, et al. The receptor tyrosine kinase ROR1 – An oncofetal antigen for targeted cancer therapy. Semin Cancer Biol. 1 déc 2014;29:21-31.

47.Almutairy BS, Andey T. Camptothecin Sensitizes Refractory Lung Cancer Cells to Receptor Kinase Inhibitor. FASEB J. 2017;31(S1):996.9-996.9.

48.Niederst MJ, Engelman JA. Bypass Mechanisms of Resistance to Receptor Tyrosine Kinase Inhibition in Lung Cancer. Sci Signal. 24 sept 2013;6(294):re6-re6.

49.Maulik G, Kijima T, Salgia R. Role of Receptor Tyrosine Kinases in Lung Cancer. In: Driscoll B, éditeur. Lung Cancer: Volume 1: Molecular Pathology Methods and Reviews [Internet]. Totowa, NJ: Humana Press; 2003 [cité 30 juin 2025]. p. 113-25. Disponible sur: https://doi.org/10.1385/1-59259-323-2:113

50.Herrera-Juárez M, Serrano-Gómez C, Bote-de-Cabo H, Paz-Ares L. Targeted therapy for lung cancer: Beyond EGFR and ALK. Cancer. 2023;129(12):1803-20.

51.Soda M, Choi YL, Enomoto M, Takada S, Yamashita Y, Ishikawa S, et al. Identification of the transforming EML4-ALK fusion gene in non-small-cell lung cancer. Nature. 2 août 2007;448(7153):561-6.

52.Camidge DR, Doebele RC. Treating ALK-positive lung cancer—early successes and future challenges. Nat Rev Clin Oncol. mai 2012;9(5):268-77.

53.Chin LP, Soo RA, Soong R, Ou SHI. Targeting ROS1 with Anaplastic Lymphoma Kinase Inhibitors: A Promising Therapeutic Strategy for a Newly Defined Molecular Subset of Non–Small-Cell Lung Cancer. J Thorac Oncol. 1 nov 2012;7(11):1625-30.

54.Gendarme S, Bylicki O, Chouaid C, Guisier F. ROS-1 Fusions in Non-Small-Cell Lung Cancer: Evidence to Date. Curr Oncol. févr 2022;29(2):641-58.

55.Ou SHI, Zhu VW. Catalog of 5' fusion partners in RET+ NSCLC Circa 2020. JTO Clin Res Rep. juin 2020;1(2):100037.

56.Chevallier M, Borgeaud M, Addeo A, Friedlaender A. Oncogenic driver mutations in non-small cell lung cancer: Past, present and future. World J Clin Oncol. 24 avr 2021;12(4):217-37.

57.Davies KD, Doebele RC. Molecular Pathways: ROS1 Fusion Proteins in Cancer. Clin Cancer Res. 31 juill 2013;19(15):4040-5.

58.Jain P, Iyer S, Straka J, Surrey LF, Pogoriler J, Han H, et al. Discovery and functional characterization of the oncogenicity and targetability of a novel NOTCH1–ROS1 gene fusion in pediatric angiosarcoma. Mol Case Stud. 10 janv 2022;8(6):a006222.

59.Liu Y, Liu T, Li N, Wang T, Pu Y, Lin R. Identification of a novel *WNK1–ROS1* fusion in a lung adenocarcinoma sensitive to crizotinib. Lung Cancer. 1 mars 2019;129:92-4.

60.Ferrara R, Auger N, Auclin E, Besse B. Clinical and Translational Implications of RET Rearrangements in Non-Small Cell Lung Cancer. J Thorac Oncol Off Publ Int Assoc Study Lung Cancer. janv 2018;13(1):27-45.

61.Cancer Genome Atlas Research Network. Comprehensive molecular profiling of lung adenocarcinoma. Nature. 31 juill 2014;511(7511):543-50.

62.Bronte G, Ulivi P, Verlicchi A, Cravero P, Delmonte A, Crinò L. Targeting RET-rearranged non-small-cell lung cancer: future prospects. Lung Cancer Targets Ther. 20 mars 2019;10:27-36.

63.Li S, Qi X, Huang Y, Liu D, Zhou F, Zhou C. Ceritinib (LDK378): A Potent Alternative to Crizotinib for ALK-Rearranged Non–Small-Cell Lung Cancer. Clin Lung Cancer. 1 mars 2015;16(2):86-91.

64.Casaluce F, Sgambato A, Maione P, Rossi A, Ferrara C, Napolitano A, et al. ALK inhibitors: a new targeted therapy in the treatment of advanced NSCLC. Target Oncol. 1 mars 2013;8(1):55-67.

65.Azelby CM, Sakamoto MR, Bowles DW. ROS1 Targeted Therapies: Current Status. Curr Oncol Rep. 14 juin 2021;23(8):94.

66.Drilon A. Targeted therapy outcomes in RET-rearranged lung cancers: drug or driver? Lancet Respir Med. 1 janv 2017;5(1):5-6.

67.Gautschi O, Milia J, Filleron T, Wolf J, Carbone DP, Owen D, et al. Targeting RET in Patients With RET-Rearranged Lung Cancers: Results From the Global, Multicenter RET Registry. J Clin Oncol. mai 2017;35(13):1403-10.

68.Karnoub AE, Weinberg RA. Ras oncogenes: split personalities. Nat Rev Mol Cell Biol. juill 2008;9(7):517-31.

69.Tulpule A, Guan J, Neel DS, Allegakoen HR, Lin YP, Brown D, et al. Kinase-mediated RAS signaling via membraneless cytoplasmic protein granules. Cell. 13 mai 2021;184(10):2649-2664.e18.

70.Huang L, Guo Z, Wang F, Fu L. KRAS mutation: from undruggable to druggable in cancer. Signal Transduct Target Ther. 15 nov 2021;6(1):1-20.

71.Hoxhaj G, Manning BD. The PI3K-AKT network at the interface of oncogenic signalling and cancer metabolism. Nat Rev Cancer. févr 2020;20(2):74-88.

72.Simanshu DK, Nissley DV, McCormick F. RAS Proteins and Their Regulators in Human Disease. Cell. 29 juin 2017;170(1):17-33.

73.Friedlaender A, Drilon A, Weiss GJ, Banna GL, Addeo A. KRAS as a druggable target in NSCLC: Rising like a phoenix after decades of development failures. Cancer Treat Rev. avr 2020;85:101978.

74.Lee JK, Sivakumar S, Schrock AB, Madison R, Fabrizio D, Gjoerup O, et al. Comprehensive pan-cancer genomic landscape of KRAS altered cancers and real-world outcomes in solid tumors. NPJ Precis Oncol. 9 déc 2022;6(1):91.

75.Wood K, Hensing T, Malik R, Salgia R. Prognostic and Predictive Value in KRAS in Non–Small-Cell Lung Cancer: A Review. JAMA Oncol. 1 juin 2016;2(6):805-12.

76.Dearden S, Stevens J, Wu YL, Blowers D. Mutation incidence and coincidence in non small-cell lung cancer: meta-analyses by ethnicity and histology (mutMap). Ann Oncol. 1 sept 2013;24(9):2371-6.

77.Rabara D, Tran TH, Dharmaiah S, Stephens RM, McCormick F, Simanshu DK, et al. KRAS G13D sensitivity to neurofibromin-mediated GTP hydrolysis. Proc Natl Acad Sci. 29 oct 2019;116(44):22122-31.

78.Yun J, Nakagawa R, Tham K. KRAS-targeted therapy in the treatment of non-small cell lung cancer. J Oncol Pharm Pract. 1 mars 2023;29(2):422-30.

79.Rohatgi A, Govindan R. Targeting KRAS G12C mutation in lung adenocarcinoma. Lung Cancer. 1 mars 2022;165:28-33.

80.Reck M, Carbone DP, Garassino M, Barlesi F. Targeting KRAS in non-small-cell lung cancer: recent progress and new approaches. Ann Oncol. 1 sept 2021;32(9):1101-10.

81.Salgia R, Pharaon R, Mambetsariev I, Nam A, Sattler M. The improbable targeted therapy: KRAS as an emerging target in non-small cell lung cancer (NSCLC). Cell Rep Med [Internet]. 19 janv 2021 [cité 30 juin 2025];2(1). Disponible sur: https://www.cell.com/cell-reports-medicine/abstract/S2666-3791(20)30241-X

82.Leonetti A, Facchinetti F, Rossi G, Minari R, Conti A, Friboulet L, et al. BRAF in non-small cell lung cancer (NSCLC): Pickaxing another brick in the wall. Cancer Treat Rev. mai 2018;66:82-94.

83.Tabbò F, Pisano C, Mazieres J, Mezquita L, Nadal E, Planchard D, et al. How far we have come targeting BRAF-mutant non-small cell lung cancer (NSCLC). Cancer Treat Rev. 1 févr 2022;103:102335.

84.Sforza V, Palumbo G, Cascetta P, Carillio G, Manzo A, Montanino A, et al. BRAF Inhibitors in Non-Small Cell Lung Cancer. Cancers. janv 2022;14(19):4863.

85.Roviello G, D'Angelo A, Sirico M, Pittacolo M, Conter FU, Sobhani N. Advances in anti-BRAF therapies for lung cancer. Invest New Drugs. 1 juin 2021;39(3):879-90.

86.Baik CS, Myall NJ, Wakelee HA. Targeting BRAF-Mutant Non-Small Cell Lung Cancer: From Molecular Profiling to Rationally Designed Therapy. The Oncologist. 1 juill 2017;22(7):786-96.

87.Gautschi O, Milia J, Cabarrou B, Bluthgen MV, Besse B, Smit EF, et al. Targeted Therapy for Patients with BRAF-Mutant Lung Cancer Results from the European EURAF Cohort. J Thorac Oncol. 1 oct 2015;10(10):1451-7.

88.Kropf-Sanchen C, Rasokat A, Christopoulos P, Saalfeld FC, Wehler T, Rost M, et al. Response to targeted and immune therapy in patients with NSCLC and BRAF non-V600E mutations: A retrospective, multicenter analysis within the national Network Genomic Medicine Lung Cancer (nNGM) in Germany. J Clin Oncol. juin 2024;42(16_suppl):8591-8591.

89.Liu D, Ding K, Yin K, Peng Z, Li X, Pan Y, et al. A real world analysis of secondary BRAF variations after targeted therapy

resistance in driver gene positive NSCLC. Sci Rep. 1 sept 2024;14(1):20302.

90.Timmermann C. A History of Lung Cancer [Internet]. London: Palgrave Macmillan UK; 2014 [cité 2 sept 2024]. Disponible sur: http://link.springer.com/10.1057/9781137384232

91.Karnofsky DA, Abelmann WH, Craver LF, Burchenal JH. The use of the nitrogen mustards in the palliative treatment of carcinoma. With particular reference to bronchogenic carcinoma. Cancer. 1948;1(4):634-56.

92.Green RA, Humphrey E, Close H, Patno ME. Alkylating agents in bronchogenic carcinoma. Am J Med. avr 1969;46(4):516-25.

93.Rosenberg B, Vancamp L, Krigas T. Inhibition of cell division in escherichia coli by electrolysis products from a platinum electrode. Nature. 13 févr 1965;205:698-9.

94.Zhang C, Xu C, Gao X, Yao Q. Platinum-based drugs for cancer therapy and anti-tumor strategies. Theranostics. 7 févr 2022;12(5):2115-32.

95.Faller BA, Pandit TN. Safety and Efficacy of Vinorelbine in the Treatment of Non-Small Cell Lung Cancer. Clin Med Insights Oncol. 1 janv 2011;5:CMO.S5074.

96.Salmaninejad A, Khoramshahi V, Azani A, Soltaninejad E, Aslani S, Zamani MR, et al. PD-1 and cancer: molecular mechanisms and polymorphisms. Immunogenetics. 1 févr 2018;70(2):73-86.

97.Borghaei H, Paz-Ares L, Horn L, Spigel DR, Steins M, Ready NE, et al. Nivolumab versus Docetaxel in Advanced Nonsquamous Non–Small-Cell Lung Cancer. N Engl J Med. 22 oct 2015;373(17):1627-39.

98.Garon EB, Rizvi NA, Hui R, Leighl N, Balmanoukian AS, Eder JP, et al. Pembrolizumab for the Treatment of Non–Small-Cell Lung Cancer. N Engl J Med. 21 mai 2015;372(21):2018-28.

99.Chen J, Wang J, Xu H. Comparison of atezolizumab, durvalumab, pembrolizumab, and nivolumab as first-line treatment in patients with extensive-stage small cell lung cancer: A systematic review and network meta-analysis. Medicine (Baltimore). 16 avr 2021;100(15):e25180.

100.Gogishvili M, Melkadze T, Makharadze T, Giorgadze D, Dvorkin M, Penkov K, et al. Cemiplimab plus chemotherapy versus chemotherapy alone in non-small cell lung cancer: a randomized, controlled, double-blind phase 3 trial. Nat Med. nov 2022;28(11):2374-80.

101.Brahmer JR, Lee JS, Ciuleanu TE, Bernabe Caro R, Nishio M, Urban L, et al. Five-Year Survival Outcomes With Nivolumab Plus Ipilimumab Versus Chemotherapy as First-Line Treatment for Metastatic Non–Small-Cell Lung Cancer in CheckMate 227. J Clin Oncol. 20 févr 2023;41(6):1200-12.

102.Schuler M, Cuppens K, Plönes T, Wiesweg M, Du Pont B, Hegedus B, et al. Neoadjuvant nivolumab with or without relatlimab in resectable non-small-cell lung cancer: a randomized phase 2 trial. Nat Med. juin 2024;30(6):1602-11.

103.Hellmann MD, Paz-Ares L, Caro RB, Zurawski B, Kim SW, Costa EC, et al. Nivolumab plus Ipilimumab in Advanced Non–Small-Cell Lung Cancer. N Engl J Med. 21 nov 2019;381(21):2020-31.

104.Paez JG, Jänne PA, Lee JC, Tracy S, Greulich H, Gabriel S, et al. EGFR Mutations in Lung Cancer: Correlation with Clinical Response to Gefitinib Therapy. Science. 4 juin 2004;304(5676):1497-500.

105.Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, et al. Non–Small Cell Lung Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 1 mai 2022;20(5):497-530.

106.Tovar EA, Graveel CR. MET in human cancer: germline and somatic mutations. Ann Transl Med. mai 2017;5(10):205.

107.Cooper CS, Park M, Blair DG, Tainsky MA, Huebner K, Croce CM, et al. Molecular cloning of a new transforming gene from a chemically transformed human cell line. Nature. sept 1984;311(5981):29-33.

108.Uchikawa E, Chen Z, Xiao GY, Zhang X, Bai X chen. Structural basis of the activation of c-MET receptor. Nat Commun. 1 juill 2021;12(1):4074.

109.Rodrigues GA, Park M. Isoforms of the met receptor tyrosine kinase. EXS. 1 janv 1993;65:167-79.

110.Bottaro DP, Rubin JS, Faletto DL, Chan AM, Kmiecik TE, Vande Woude GF, et al. Identification of the hepatocyte growth factor receptor as the c-met proto-oncogene product. Science. 15 févr 1991;251(4995):802-4.

111.Stoker M, Gherardi E, Perryman M, Gray J. Scatter factor is a fibroblast-derived modulator of epithelial cell mobility. Nature. mai 1987;327(6119):239-42.

112.Weidner KM, Arakaki N, Hartmann G, Vandekerckhove J, Weingart S, Rieder H, et al. Evidence for the identity of human scatter factor and human hepatocyte growth factor. Proc Natl Acad Sci U S A. 15 août 1991;88(16):7001-5.

113.Fajardo-Puerta AB, Prado MM, Frampton AE, Jiao LR. Gene of the month: HGF. J Clin Pathol. 1 juill 2016;69(7):575-9.

114.Zhao Y, Ye W, Wang YD, Chen WD. HGF/c-Met: A Key Promoter in Liver Regeneration. Front Pharmacol [Internet]. 17 mars 2022 [cité 3 avr 2025];13. Disponible sur: https://www.frontiersin.org/journals/pharmacology/articles/10.3389/fphar.2022.808855/full

115.Miyazawa K, Tsubouchi H, Naka D, Takahashi K, Okigaki M, Arakaki N, et al. Molecular cloning and sequence analysis of cDNA for human hepatocyte growth factor. Biochem Biophys Res Commun. 15 sept 1989;163(2):967-73.

116.Holmes O, Pillozzi S, Deakin JA, Carafoli F, Kemp L, Butler PJG, et al. Insights into the Structure/Function of Hepatocyte Growth Factor/Scatter Factor from Studies with Individual Domains. J Mol Biol. 23 mars 2007;367(2):395-408.

117.Kong-Beltran M, Stamos J, Wickramasinghe D. The Sema domain of Met is necessary for receptor dimerization and activation. Cancer Cell. 1 juill 2004;6(1):75-84.

118.Lyon M, Deakin JA, Lietha D, Gherardi E, Gallagher JT. The interactions of hepatocyte growth factor/scatter factor and its NK1 and NK2 variants with glycosaminoglycans using a modified gel mobility shift assay. Elucidation of the minimal size of binding

and activatory oligosaccharides. J Biol Chem. 15 oct 2004;279(42):43560-7.

119.Rubin JS, Day RM, Breckenridge D, Atabey N, Taylor WG, Stahl SJ, et al. Dissociation of heparan sulfate and receptor binding domains of hepatocyte growth factor reveals that heparan sulfate-c-met interaction facilitates signaling. J Biol Chem. 31 août 2001;276(35):32977-83.

120.Sakata H, Stahl SJ, Taylor WG, Rosenberg JM, Sakaguchi K, Wingfield PT, et al. Heparin binding and oligomerization of hepatocyte growth factor/scatter factor isoforms. Heparan sulfate glycosaminoglycan requirement for Met binding and signaling. J Biol Chem. 4 avr 1997;272(14):9457-63.

121.Simonneau C, Leclercq B, Mougel A, Adriaenssens E, Paquet C, Raibaut L, et al. Semi-synthesis of a HGF/SF kringle one (K1) domain scaffold generates a potent in vivo MET receptor agonist. Chem Sci. 2015;6(3):2110-21.

122.Nola G de, Leclercq B, Mougel A, Taront S, Simonneau C, Forneris F, et al. Dimerization of kringle 1 domain from hepatocyte growth factor/scatter factor provides a potent MET receptor agonist. Life Sci Alliance [Internet]. 1 déc 2022 [cité 4 avr 2025];5(12). Disponible sur: https://www.life-science-alliance.org/content/5/12/e202201424

123.Ponzetto C, Bardelli A, Zhen Z, Maina F, dalla Zonca P, Giordano S, et al. A multifunctional docking site mediates signaling and transformation by the hepatocyte growth factor/scatter factor receptor family. Cell. 22 avr 1994;77(2):261-71.

124.Huang X, Li E, Shen H, Wang X, Tang T, Zhang X, et al. Targeting the HGF/MET Axis in Cancer Therapy: Challenges in Resistance and Opportunities for Improvement. Front Cell Dev Biol. 6 mai 2020;8:152.

125.Rhoades Smith KE, Bilen MA. A Review of Papillary Renal Cell Carcinoma and MET Inhibitors. Kidney Cancer. 1 nov 2019;3(3):151-61.

126.Hynes RO. Integrins: Bidirectional, Allosteric Signaling Machines. Cell. 20 sept 2002;110(6):673-87.

127.Jahangiri A, Nguyen A, Chandra A, Sidorov MK, Yagnik G, Rick J, et al. Cross-activating c-Met/β1 integrin complex drives metastasis and invasive resistance in cancer. Proc Natl Acad Sci. 10 oct 2017;114(41):E8685-94.

128.Singh R, Peng S, Viswanath P, Sambandam V, Shen L, Rao X, et al. Non-canonical cMet regulation by vimentin mediates Plk1 inhibitor–induced apoptosis. EMBO Mol Med. mai 2019;11(5):e9960.

129.Hui AY, Meens JA, Schick C, Organ SL, Qiao H, Tremblay EA, et al. Src and FAK mediate cell–matrix adhesion-dependent activation of met during transformation of breast epithelial cells. J Cell Biochem. 2009;107(6):1168-81.

130.Wang R, Ferrell LD, Faouzi S, Maher JJ, Bishop JM. Activation of the Met Receptor by Cell Attachment Induces and Sustains Hepatocellular Carcinomas in Transgenic Mice. J Cell Biol. 29 mai 2001;153(5):1023-34.

131.Huang X, Ye Q, Chen M, Li A, Mi W, Fang Y, et al. N-glycosylation-defective splice variants of neuropilin-1 promote metastasis by activating endosomal signals. Nat Commun. 16 août 2019;10(1):3708.

132.Yoshioka T, Otero J, Chen Y, Kim YM, Koutcher JA, Satagopan J, et al. β4 Integrin signaling induces expansion of prostate tumor progenitors. J Clin Invest. 1 févr 2013;123(2):682-99.

133.Barrow-McGee R, Kishi N, Joffre C, Ménard L, Hervieu A, Bakhouche BA, et al. Beta 1-integrin–c-Met cooperation reveals an inside-in survival signalling on autophagy-related endomembranes. Nat Commun. 23 juin 2016;7(1):11942.

134.Bertotti A, Comoglio PM, Trusolino L. β4 Integrin Is a Transforming Molecule that Unleashes Met Tyrosine Kinase Tumorigenesis. Cancer Res. 1 déc 2005;65(23):10674-9.

135.Bertotti A, Comoglio PM, Trusolino L. β4 integrin activates a Shp2–Src signaling pathway that sustains HGF-induced anchorage-independent growth. J Cell Biol. 11 déc 2006;175(6):993-1003.

136.Qamar T, Misra DP, Kar S. Semaphorins and its receptors: Emerging cellular biomarkers and therapeutic targets in autoimmune and inflammatory disorders. Life Sci. 15 janv 2025;361:123281.

137.Khan A, Sharma P, Dahiya S, Sharma B. Plexins: Navigating through the neural regulation and brain pathology. Neurosci Biobehav Rev. févr 2025;169:105999.

138.Conrotto P, Corso S, Gamberini S, Comoglio PM, Giordano S. Interplay between scatter factor receptors and B plexins controls invasive growth. Oncogene. juill 2004;23(30):5131-7.

139.Swiercz JM, Worzfeld T, Offermanns S. ErbB-2 and Met Reciprocally Regulate Cellular Signaling via Plexin-B1 *. J Biol Chem. 25 janv 2008;283(4):1893-901.

140.Soong J, Scott G. Plexin B1 inhibits MET through direct association and regulates Shp2 expression in melanocytes. J Cell Sci. 15 janv 2013;126(2):688-95.

141.Peacock JW, Takeuchi A, Hayashi N, Liu L, Tam KJ, Al Nakouzi N, et al. SEMA3C drives cancer growth by transactivating multiple receptor tyrosine kinases via Plexin B1. EMBO Mol Med. févr 2018;10(2):219-38.

142.Yamaguchi T, Lu C, Ida L, Yanagisawa K, Usukura J, Cheng J, et al. ROR1 sustains caveolae and survival signalling as a scaffold of cavin-1 and caveolin-1. Nat Commun. 4 janv 2016;7(1):10060.

143.Schmid S, Früh M, Peters S. Targeting MET in EGFR resistance in non-small-cell lung cancer—ready for daily practice? Lancet Oncol. 1 mars 2020;21(3):320-2.

144.Wang Q, Yang S, Wang K, Sun SY. MET inhibitors for targeted therapy of EGFR TKI-resistant lung cancer. J Hematol OncolJ Hematol Oncol. 21 juin 2019;12(1):63.

145.Guo A, Villén J, Kornhauser J, Lee KA, Stokes MP, Rikova K, et al. Signaling networks assembled by oncogenic EGFR and c-Met. Proc Natl Acad Sci. 15 janv 2008;105(2):692-7.

146.Rikova K, Guo A, Zeng Q, Possemato A, Yu J, Haack H, et al. Global Survey of Phosphotyrosine Signaling Identifies Oncogenic Kinases in Lung Cancer. Cell. 14 déc 2007;131(6):1190-203.

147.Lai AZ, Abella JV, Park M. Crosstalk in Met receptor oncogenesis. Trends Cell Biol. 1 oct 2009;19(10):542-51.

148.Mueller KL, Hunter LA, Ethier SP, Boerner JL. Met and c-Src cooperate to compensate for loss of epidermal growth factor receptor kinase activity in breast cancer cells. Cancer Res. 1 mai 2008;68(9):3314-22.

149.Breindel JL, Haskins JW, Cowell EP, Zhao M, Nguyen DX, Stern DF. EGF Receptor Activates MET through MAPK to Enhance Non–Small Cell Lung Carcinoma Invasion and Brain Metastasis. Cancer Res. 13 août 2013;73(16):5053-65.

150.Engelman JA, Zejnullahu K, Mitsudomi T, Song Y, Hyland C, Park JO, et al. MET Amplification Leads to Gefitinib Resistance in Lung Cancer by Activating ERBB3 Signaling. Science. 18 mai 2007;316(5827):1039-43.

151.Bardelli A, Ponzetto C, Comoglio PM. Identification of functional domains in the hepatocyte growth factor and its receptor by molecular engineering. J Biotechnol. 30 sept 1994;37(2):109-22.

152.Paumelle R, Tulashe D, Kherrouche Z, Plaza S, Leroy C, Reveneau S, et al. Hepatocyte growth factor/scatter factor activates the ETS1 transcription factor by a RAS-RAF-MEK-ERK signaling pathway. Oncogene. avr 2002;21(15):2309-19.

153.Nazir SU, Kumar R, Singh A, Khan A, Tanwar P, Tripathi R, et al. Breast cancer invasion and progression by MMP-9 through Ets-1 transcription factor. Gene. 30 août 2019;711:143952.

154.Ozaki I, Mizuta T, Zhao G, Yotsumoto H, Hara T, Kajihara S, et al. Involvement of the Ets-1 Gene in Overexpression of Matrilysin in Human Hepatocellular Carcinoma1. Cancer Res. 15 nov 2000;60(22):6519-25.

155.Park YH, Jung HH, Ahn JS, Im YH. Ets-1 upregulates HER2-induced MMP-1 expression in breast cancer cells. Biochem Biophys Res Commun. 12 déc 2008;377(2):389-94.

156.Liebmann C. Regulation of MAP kinase activity by peptide receptor signalling pathway: paradigms of multiplicity. Cell Signal. nov 2001;13(11):777-85.

157.Kherrouche Z, Monte D, Werkmeister E, Stoven L, De Launoit Y, Cortot AB, et al. PEA3 transcription factors are downstream effectors of Met signaling involved in migration and invasiveness of Met-addicted tumor cells. Mol Oncol. 1 nov 2015;9(9):1852-67.

158.Rosário M, Birchmeier W. How to make tubes: signaling by the Met receptor tyrosine kinase. Trends Cell Biol. juin 2003;13(6):328-35.

159.Zhang Z, Yao L, Yang J, Wang Z, Du G. PI3K/Akt and HIF-1 signaling pathway in hypoxia-ischemia (Review). Mol Med Rep. 1 oct 2018;18(4):3547-54.

160.Carnero A, Blanco-Aparicio C, Renner O, Link W, Leal JFM. The PTEN/PI3K/AKT signalling pathway in cancer, therapeutic implications. Curr Cancer Drug Targets. mai 2008;8(3):187-98.

161.Boccaccio C, Andò M, Tamagnone L, Bardelli A, Michieli P, Battistini C, et al. Induction of epithelial tubules by growth factor HGF depends on the STAT pathway. Nature. janv 1998;391(6664):285-8.

162.Zhang YW, Wang LM, Jove R, Vande Woude GF. Requirement of Stat3 signaling for HGF/SF-Met mediated tumorigenesis. Oncogene. janv 2002;21(2):217-26.

163.Zinatizadeh MR, Schock B, Chalbatani GM, Zarandi PK, Jalali SA, Miri SR. The Nuclear Factor Kappa B (NF-kB) signaling in cancer development and immune diseases. Genes Dis. mai 2021;8(3):287-97.

164.Chaturvedi M, Sung B, Yadav V, Kannappan R, Aggarwal B. NF-κB addiction and its role in cancer: 'one size does not fit all'. Oncogene. 7 avr 2011;30(14):1615-30.

165.Luo JL, Kamata H, Karin M. IKK/NF-κB signaling: balancing life and death – a new approach to cancer therapy. J Clin Invest. 1 oct 2005;115(10):2625-32.

166.Villa-Moruzzi E, Puntoni F, Bardelli A, Vigna E, Rosa SD, Comoglio PM. Protein tyrosine phosphatase PTP-S binds to the juxtamembrane region of the hepatocyte growth factor receptor Met. Biochem J. 15 nov 1998;336(1):235-9.

167.Lu KV, Chang JP, Parachoniak CA, Pandika MM, Aghi MK, Meyronet D, et al. VEGF inhibits tumor cell invasion and mesenchymal transition through a MET/VEGFR2 complex. Cancer Cell. 10 juill 2012;22(1):21-35.

168.Sangwan V, Paliouras GN, Abella JV, Dubé N, Monast A, Tremblay ML, et al. Regulation of the Met Receptor-tyrosine Kinase by the Protein-tyrosine Phosphatase 1B and T-cell Phosphatase. J Biol Chem. 5 déc 2008;283(49):34374-83.

169.Stoker AW. Protein tyrosine phosphatases and signalling. J Endocrinol. avr 2005;185(1):19-33.

170.Huang WC, Yen JH, Sung YW, Tung SL, Chen PM, Chu PY, et al. Novel function of THEMIS2 in the enhancement of cancer stemness and chemoresistance by releasing PTP1B from MET. Oncogene. févr 2022;41(7):997-1010.

171.Xu Y, Xia W, Baker D, Zhou J, Cha HC, Voorhees JJ, et al. Receptor-type Protein Tyrosine Phosphatase β (RPTP-β) Directly Dephosphorylates and Regulates Hepatocyte Growth Factor Receptor (HGFR/Met) Function*. J Biol Chem. 6 mai 2011;286(18):15980-8.

172.D'Agostino S, Lanzillotta D, Varano M, Botta C, Baldrini A, Bilotta A, et al. The receptor protein tyrosine phosphatase PTPRJ negatively modulates the CD98hc oncoprotein in lung cancer cells. Oncotarget. 4 mai 2018;9(34):23334-48.

173.Miller MA, Meyer AS, Beste MT, Lasisi Z, Reddy S, Jeng KW, et al. ADAM-10 and -17 regulate endometriotic cell migration via concerted ligand and receptor shedding feedback on kinase signaling. Proc Natl Acad Sci U S A. 28 mai 2013;110(22):E2074-2083.

174.Ancot F, Leroy C, Muharram G, Lefebvre J, Vicogne J, Lemiere A, et al. Shedding-generated Met receptor fragments can be routed to either the proteasomal or the lysosomal degradation pathway. Traffic Cph Den. sept 2012;13(9):1261-72.

175.Schelter F, Kobuch J, Moss ML, Becherer JD, Comoglio PM, Boccaccio C, et al. A Disintegrin and Metalloproteinase-10 (ADAM-10) Mediates DN30 Antibody-induced Shedding of the Met Surface Receptor *. J Biol Chem. 20 août 2010;285(34):26335-40.

176.Miller MA, Oudin MJ, Sullivan RJ, Wang SJ, Meyer AS, Im H, et al. Reduced Proteolytic Shedding of Receptor Tyrosine Kinases Is a Post-Translational Mechanism of Kinase Inhibitor Resistance. Cancer Discov. 3 avr 2016;6(4):382-99.

177.Foveau B, Leroy C, Ancot F, Deheuninck J, Ji Z, Fafeur V, et al. Amplification of apoptosis through sequential caspase

cleavage of the MET tyrosine kinase receptor. Cell Death Differ. avr 2007;14(4):752-64.

178.Tulasne D, Deheuninck J, Lourenço FC, Lamballe F, Ji Z, Leroy C, et al. Proapoptotic Function of the MET Tyrosine Kinase Receptor through Caspase Cleavage. Mol Cell Biol. 1 déc 2004;24(23):10328-39.

179.Lefebvre J, Muharram G, Leroy C, Kherrouche Z, Montagne R, Ichim G, et al. Caspase-generated fragment of the Met receptor favors apoptosis via the intrinsic pathway independently of its tyrosine kinase activity. Cell Death Dis. oct 2013;4(10):e871-e871.

180.Duplaquet L, Leroy C, Vinchent A, Paget S, Lefebvre J, Vanden Abeele F, et al. Control of cell death/survival balance by the MET dependence receptor. Espinosa JM, Cooper JA, éditeurs. eLife. 24 févr 2020;9:e50041.

181.Mebratu YA, Leyva-Baca I, Wathelet MG, Lacey N, Chand HS, Choi AMK, et al. Bik reduces hyperplastic cells by increasing Bak and activating DAPk1 to juxtapose ER and mitochondria. Nat Commun. 6 oct 2017;8(1):803.

182.Giordano S, Di Renzo MF, Narsimhan RP, Cooper CS, Rosa C, Comoglio PM. Biosynthesis of the protein encoded by the c-met proto-oncogene. Oncogene. nov 1989;4(11):1383-8.

183.Wu YM, Liu CH, Huang MJ, Lai HS, Lee PH, Hu RH, et al. C1GALT1 enhances proliferation of hepatocellular carcinoma cells via modulating MET glycosylation and dimerization. Cancer Res. 1 sept 2013;73(17):5580-90.

184.Qian J, Zhu C hua, Tang S, Shen A jun, Ai J, Li J, et al. α2,6-hyposialylation of c-Met abolishes cell motility of ST6Gal-I-knockdown HCT116 cells. Acta Pharmacol Sin. juill 2009;30(7):1039-45.

185.Chen R, Li J, Feng CH, Chen SK, Liu YP, Duan CY, et al. c-Met function requires N-linked glycosylation modification of pro-Met. J Cell Biochem. avr 2013;114(4):816-22.

186.Hu X, Tang F, Liu P, Zhong T, Yuan F, He Q, et al. Structural and Functional Insight Into the Glycosylation Impact Upon the HGF/c-Met Signaling Pathway. Front Cell Dev Biol. 18 juin 2020;8:490.

187.Joffre C, Barrow R, Ménard L, Calleja V, Hart IR, Kermorgant S. A direct role for Met endocytosis in tumorigenesis. Nat Cell Biol. 5 juin 2011;13(7):827-37.

188.Kermorgant S, Parker PJ. Receptor trafficking controls weak signal delivery: a strategy used by c-Met for STAT3 nuclear accumulation. J Cell Biol. 8 sept 2008;182(5):855-63.

189.Peschard P, Ishiyama N, Lin T, Lipkowitz S, Park M. A Conserved DpYR Motif in the Juxtamembrane Domain of the Met Receptor Family Forms an Atypical c-Cbl/Cbl-b Tyrosine Kinase Binding Domain Binding Site Required for Suppression of Oncogenic Activation *. J Biol Chem. 9 juill 2004;279(28):29565-71.

190.Gui Y, Yeganeh M, Ramanathan S, Leblanc C, Pomerleau V, Ferbeyre G, et al. SOCS1 controls liver regeneration by regulating HGF signaling in hepatocytes. J Hepatol. déc 2011;55(6):1300-8.

191.Saigi M, Alburquerque-Bejar JJ, Mc Leer-Florin A, Pereira C, Pros E, Romero OA, et al. MET-Oncogenic and JAK2-Inactivating Alterations Are Independent Factors That Affect Regulation of PD-L1 Expression in Lung Cancer. Clin Cancer Res Off J Am Assoc Cancer Res. 15 sept 2018;24(18):4579-87.

192.Hu CT, Cheng CC, Wu JR, Pan SM, Wu WS. PKCε-mediated c-Met endosomal processing directs fluctuant c-Met-JNK-paxillin signaling for tumor progression of HepG2. Cell Signal. juill 2015;27(7):1544-55.

193.Muharram G, Sahgal P, Korpela T, De Franceschi N, Kaukonen R, Clark K, et al. Tensin-4-Dependent MET Stabilization Is Essential for Survival and Proliferation in Carcinoma Cells. Dev Cell. 9 juin 2014;29(5):629-30.

194.Parachoniak CA, Luo Y, Abella JV, Keen JH, Park M. GGA3 Functions as a Switch to Promote Met Receptor Recycling, Essential for Sustained ERK and Cell Migration. Dev Cell. 14 juin 2011;20(6):751-63.

195.Duplaquet L, Kherrouche Z, Baldacci S, Jamme P, Cortot AB, Copin MC, et al. The multiple paths towards MET receptor addiction in cancer. Oncogene. juin 2018;37(24):3200-15.

196.Cancer Genome Atlas Research Network. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. Nature. 23 oct 2008;455(7216):1061-8.

197.Umeki K, Shiota G, Kawasaki H. Clinical Significance of c-met Oncogene Alterations in Human Colorectal Cancer. Oncology. 21 mai 1999;56(4):314-21.

198.Ujiie H, Tomida M, Akiyama H, Nakajima Y, Okada D, Yoshino N, et al. Serum hepatocyte growth factor and interleukin-6 are effective prognostic markers for non-small cell lung cancer. Anticancer Res. août 2012;32(8):3251-8.

199.Stransky N, Cerami E, Schalm S, Kim JL, Lengauer C. The landscape of kinase fusions in cancer. Nat Commun. 10 sept 2014;5(1):4846.

200.Guérin C, Tulasne D. Recording and classifying MET receptor mutations in cancers. Dötsch V, éditeur. eLife. 23 avr 2024;13:e92762.

201.Schmidt L, Duh FM, Chen F, Kishida T, Glenn G, Choyke P, et al. Germline and somatic mutations in the tyrosine kinase domain of the MET proto-oncogene in papillary renal carcinomas. Nat Genet. mai 1997;16(1):68-73.

202.Cortot AB, Kherrouche Z, Descarpentries C, Wislez M, Baldacci S, Furlan A, et al. Exon 14 Deleted MET Receptor as a New Biomarker and Target in Cancers. J Natl Cancer Inst. 1 mai 2017;109(5).

203.Pecci F, Ricciuti B, Alessi JVM, Barrichello AP de C, Vaz VR, Lamberti G, et al. Activating MET kinase domain mutations define a novel molecular subtype of non–small cell lung cancer that is clinically targetable with the MET inhibitor elzovantinib (TPX-0022). J Clin Oncol. juin 2022;40(16_suppl):9124-9124.

204.Kalmuk L, Descarpentries C, Mascaux C, Munck C, Kherrouche Z, Wasielewski E, et al. MET kinase mutations in lung cancer patients. J Clin Oncol. juin 2022;40(16_suppl):e21014-e21014.

205.Estevam GO, Linossi EM, Macdonald CB, Espinoza CA, Michaud JM, Coyote-Maestas W, et al. Conserved regulatory

motifs in the juxtamembrane domain and kinase N-lobe revealed through deep mutational scanning of the MET receptor tyrosine kinase domain. Dötsch V, éditeur. eLife. 13 sept 2024;12:RP91619.

206.Bardelli A, Pugliese L, Comoglio PM. “Invasive-growth” signaling by the Met/HGF receptor. Biochim Biophys Acta BBA - Rev Cancer. 9 déc 1997;1333(3):M41-51.

207.Jeffers M, Schmidt L, Nakaigawa N, Webb CP, Weirich G, Kishida T, et al. Activating mutations for the met tyrosine kinase receptor in human cancer. Proc Natl Acad Sci U S A. 14 oct 1997;94(21):11445-50.

208.Jeffers M, Fiscella M, Webb CP, Anver M, Koochekpour S, Vande Woude GF. The mutationally activated Met receptor mediates motility and metastasis. Proc Natl Acad Sci. 24 nov 1998;95(24):14417-22.

209.Chiara F, Michieli P, Pugliese L, Comoglio PM. Mutations in the met Oncogene Unveil a “Dual Switch” Mechanism Controlling Tyrosine Kinase Activity *. J Biol Chem. 1 août 2003;278(31):29352-8.

210.Michieli P, Basilico C, Pennacchietti S, Maffè A, Tamagnone L, Giordano S, et al. Mutant Met-mediated transformation is ligand-dependent and can be inhibited by HGF antagonists. Oncogene. sept 1999;18(37):5221-31.

211.Graveel C, Su Y, Koeman J, Wang LM, Tessarollo L, Fiscella M, et al. Activating Met mutations produce unique tumor profiles in mice with selective duplication of the mutant allele. Proc Natl Acad Sci U S A. 7 déc 2004;101(49):17198-203.

212.Graveel CR, London ,Cheryl A., and Woude GFV. A Mouse Model of Activating Met Mutations. Cell Cycle. 28 avr 2005;4(4):518-20.

213.Seiwert TY, Jagadeeswaran R, Faoro L, Janamanchi V, Nallasura V, El Dinali M, et al. The MET Receptor Tyrosine Kinase Is a Potential Novel Therapeutic Target for Head and Neck Squamous Cell Carcinoma. Cancer Res. 1 avr 2009;69(7):3021-31.

214.Lee JH, Han SU, Cho H, Jennings B, Gerrard B, Dean M, et al. A novel germ line juxtamembrane Met mutation in human gastric cancer. Oncogene. 12 oct 2000;19(43):4947-53.

215.Krishnaswamy S, Kanteti R, Duke-Cohan JS, Loganathan S, Liu W, Ma PC, et al. Ethnic Differences and Functional Analysis of MET Mutations in Lung Cancer. Clin Cancer Res. 14 sept 2009;15(18):5714-23.

216.Jagadeeswaran R, Jagadeeswaran S, Bindokas VP, Salgia R. Activation of HGF/c-Met pathway contributes to the reactive oxygen species generation and motility of small cell lung cancer cells. Am J Physiol-Lung Cell Mol Physiol. juin 2007;292(6):L1488-94.

217.Ma PC. MET Receptor Juxtamembrane Exon 14 Alternative Spliced Variant: Novel Cancer Genomic Predictive Biomarker. Cancer Discov. 3 août 2015;5(8):802-5.

218.Tyner JW, Fletcher LB, Wang EQ, Yang WF, Rutenberg-Schoenberg ML, Beadling C, et al. Met receptor sequence variants R970C and T992I lack transforming capacity. Cancer Res. 1 août 2010;70(15):6233-7.

219.Montagne R, Baranzelli A, Muharram G, Catherine L, Lesaffre M, Vinchent A, et al. MET receptor variant R970C favors calpain-dependent generation of a fragment promoting epithelial cell scattering. Oncotarget. 14 févr 2017;8(7):11268-83.

220.Fernandes M, Duplaquet L, Tulasne and D. Proteolytic cleavages of MET: the divide-and-conquer strategy of a receptor tyrosine kinase. BMB Rep. 30 avr 2019;52(4):239-49.

221.Tengs T, Lee JC, Guillermo Paez J, Zhao X, LaFramboise T, Giannoukos G, et al. A transforming MET mutation discovered in non-small cell lung cancer using microarray-based resequencing. Cancer Lett. 8 août 2006;239(2):227-33.

222.Kong LR, Mohamed Salleh NAB, Ong RW, Tan TZ, Syn NL, Goh RM, et al. A common MET polymorphism harnesses HER2 signaling to drive aggressive squamous cell carcinoma. Nat Commun. 25 mars 2020;11(1):1556.

223.Shieh JM, Tang YA, Yang TH, Chen CY, Hsu HS, Tan YHC, et al. Lack of Association of C-Met-N375S Sequence Variant with Lung Cancer Susceptibility and Prognosis. Int J Med Sci. 14 juin 2013;10(8):988-94.

224.Yeh YM, Wu PY, Lin PC, Su PF, Hsu YT, Hsu KF, et al. MET Mutation Is a Potential Therapeutic Target for Advanced Endometrial Cancer. Cancers. janv 2021;13(16):4231.

225.Fujino T, Mitsudomi T. Acquired Resistance Mechanism for MET Tyrosine Kinase Inhibitor. JTO Clin Res Rep. 16 mars 2021;2(3):100134.

226.Choueiri TK, Vaishampayan U, Rosenberg JE, Logan TF, Harzstark AL, Bukowski RM, et al. Phase II and Biomarker Study of the Dual MET/VEGFR2 Inhibitor Foretinib in Patients With Papillary Renal Cell Carcinoma. J Clin Oncol. 10 janv 2013;31(2):181-6.

227.Schöffski P, Wozniak A, Escudier B, Rutkowski P, Anthoney A, Bauer S, et al. Crizotinib achieves long-lasting disease control in advanced papillary renal-cell carcinoma type 1 patients with MET mutations or amplification. EORTC 90101 CREATE trial. Eur J Cancer. 1 déc 2017;87:147-63.

228.Choueiri TK, Heng DYC, Lee JL, Cancel M, Verheijen RB, Mellemgaard A, et al. Efficacy of Savolitinib vs Sunitinib in Patients With MET-Driven Papillary Renal Cell Carcinoma: The SAVOIR Phase 3 Randomized Clinical Trial. JAMA Oncol. 1 août 2020;6(8):1247-55.

229.Suárez C, Larkin JMG, Patel P, Valderrama BP, Rodriguez-Vida A, Glen H, et al. Phase II Study Investigating the Safety and Efficacy of Savolitinib and Durvalumab in Metastatic Papillary Renal Cancer (CALYPSO). J Clin Oncol. 10 mai 2023;41(14):2493-502.

230.Estevam GO, Linossi E, Rao J, Macdonald CB, Ravikumar A, Chrispens KM, et al. Mapping kinase domain resistance mechanisms for the MET receptor tyrosine kinase via deep mutational scanning. bioRxiv [Internet]. 2024 [cité 30 juin 2025]; Disponible sur: https://consensus.app/papers/mapping-kinase-domain-resistance-mechanisms-for-the-met-macdonald-estevam/3ab1d7076da05523b0ecc23115dd771d/

231.Cepero V, Sierra JR, Corso S, Ghiso E, Casorzo L, Perera T, et al. MET and KRAS Gene Amplification Mediates Acquired Resistance to MET Tyrosine Kinase Inhibitors. Cancer Res. 29 sept 2010;70(19):7580-90.

232.Goltsov AA, Fang B, Pandita TK, Maru DM, Swisher SG, Hofstetter WL. HER2 Confers Resistance to Foretinib Inhibition of MET-Amplified Esophageal Adenocarcinoma Cells. Ann Thorac Surg. 1 févr 2018;105(2):363-70.

233.Bahcall M, Awad MM, Sholl LM, Wilson FH, Xu M, Wang S, et al. Amplification of Wild-type KRAS Imparts Resistance to Crizotinib in MET Exon 14 Mutant Non–Small Cell Lung Cancer. Clin Cancer Res. 3 déc 2018;24(23):5963-76.

234.Mazieres J, Vioix H, Pfeiffer BM, Campden RI, Chen Z, Heeg B, et al. *MET* Exon 14 Skipping in NSCLC: A Systematic Literature Review of Epidemiology, Clinical Characteristics, and Outcomes. Clin Lung Cancer. 1 sept 2023;24(6):483-97.

235.Lee CC, Yamada KM. Identification of a novel type of alternative splicing of a tyrosine kinase receptor. Juxtamembrane deletion of the c-met protein kinase C serine phosphorylation regulatory site. J Biol Chem. 29 juill 1994;269(30):19457-61.

236.Ma PC, Jagadeeswaran R, Jagadeesh S, Tretiakova MS, Nallasura V, Fox EA, et al. Functional expression and mutations of c-Met and its therapeutic inhibition with SU11274 and small interfering RNA in non-small cell lung cancer. Cancer Res. 15 févr 2005;65(4):1479-88.

237.Lee JK, Madison R, Classon A, Gjoerup O, Rosenzweig M, Frampton GM, et al. Characterization of Non–Small-Cell Lung Cancers With MET Exon 14 Skipping Alterations Detected in Tissue or Liquid: Clinicogenomics and Real-World Treatment Patterns. JCO Precis Oncol. août 2021;(5):1354-76.

238.Heist RS, Sequist LV, Borger D, Gainor JF, Arellano RS, Le LP, et al. Acquired Resistance to Crizotinib in NSCLC with MET Exon 14 Skipping. J Thorac Oncol. 1 août 2016;11(8):1242-5.

239.Paik PK, Felip E, Veillon R, Sakai H, Cortot AB, Garassino MC, et al. Tepotinib in Non–Small-Cell Lung Cancer with MET Exon 14 Skipping Mutations. N Engl J Med. 2 sept 2020;383(10):931-43.

240.Wolf J, Seto T, Han JY, Reguart N, Garon EB, Groen HJM, et al. Capmatinib in MET Exon 14–Mutated or MET-Amplified Non–Small-Cell Lung Cancer. N Engl J Med. 2 sept 2020;383(10):944-57.

241.Fernandes M, Hoggard B, Jamme P, Paget S, Truong MJ, Grégoire V, et al. MET exon 14 skipping mutation is a hepatocyte growth factor (HGF)-dependent oncogenic driver in vitro and in humanised HGF knock-in mice. Mol Oncol. 2023;17(11):2257-74.

242.Lee GD, Lee SE, Oh DY, Yu D bi, Jeong HM, Kim J, et al. MET Exon 14 Skipping Mutations in Lung Adenocarcinoma: Clinicopathologic Implications and Prognostic Values. J Thorac Oncol. 1 août 2017;12(8):1233-46.

243.Drusbosky LM, Dawar R, Rodriguez E, Ikpeazu CV. Therapeutic strategies in METex14 skipping mutated non-small cell lung cancer. J Hematol OncolJ Hematol Oncol. 23 août 2021;14(1):129.

244.Hashigasako A, Machide M, Nakamura T, Matsumoto K, Nakamura T. Bi-directional regulation of Ser-985 phosphorylation of c-met via protein kinase C and protein phosphatase 2A involves c-Met activation and cellular responsiveness to hepatocyte growth factor. J Biol Chem. 18 juin 2004;279(25):26445-52.

245.Nakayama M, Sakai K, Yamashita A, Nakamura T, Suzuki Y, Matsumoto K. Met/HGF receptor activation is regulated by juxtamembrane Ser985 phosphorylation in hepatocytes. Cytokine. 1 juin 2013;62(3):446-52.

246.Peschard P, Fournier TM, Lamorte L, Naujokas MA, Band H, Langdon WY, et al. Mutation of the c-Cbl TKB domain binding site on the Met receptor tyrosine kinase converts it into a transforming protein. Mol Cell. nov 2001;8(5):995-1004.

247.Abella JV, Peschard ,Pascal, Naujokas ,Monica A., Lin ,Tong, Saucier ,Caroline, Urbé ,Sylvie, et al. Met/Hepatocyte Growth Factor Receptor Ubiquitination Suppresses Transformation and Is Required for Hrs Phosphorylation. Mol Cell Biol. 1 nov 2005;25(21):9632-45.

248.Fernandes M, Paget S, Kherrouche Z, Truong MJ, Vinchent A, Meneboo JP, et al. Transforming properties of MET receptor exon 14 skipping can be recapitulated by loss of the CBL ubiquitin ligase binding site. FEBS Lett. 2023;597(18):2301-15.

249.Truong MJ, Pawlak G, Meneboo JP, Sebda S, Fernandes M, Figeac M, et al. Transcriptional program-based deciphering of the MET exon 14 skipping regulation network [Internet]. bioRxiv; 2024 [cité 27 févr 2025]. p. 2024.09.13.612820. Disponible sur: https://www.biorxiv.org/content/10.1101/2024.09.13.612820v2

250.Bravo González-Blas C, Matetovici I, Hillen H, Taskiran II, Vandepoel R, Christiaens V, et al. Single-cell spatial multi-omics and deep learning dissect enhancer-driven gene regulatory networks in liver zonation. Nat Cell Biol. janv 2024;26(1):153-67.

251.Vitale I, Sistigu A, Manic G, Rudqvist NP, Trajanoski Z, Galluzzi L. Mutational and Antigenic Landscape in Tumor Progression and Cancer Immunotherapy. Trends Cell Biol. 1 mai 2019;29(5):396-416.

252.Waddington CH. The Strategy of the Genes. London: Routledge; 1957. 274 p.

253.Huang S, Soto AM, Sonnenschein C. The end of the genetic paradigm of cancer. PLOS Biol. 18 mars 2025;23(3):e3003052.

254.Alon U. Network motifs: theory and experimental approaches. Nat Rev Genet. juin 2007;8(6):450-61.

255.Shin D, Cho KH. Critical transition and reversion of tumorigenesis. Exp Mol Med. avr 2023;55(4):692-705.

256.Alanis-Lobato G, Andrade-Navarro MA, Schaefer MH. HIPPIE v2.0: enhancing meaningfulness and reliability of protein–protein interaction networks. Nucleic Acids Res. 4 janv 2017;45(D1):D408-14.

257.Oughtred R, Stark C, Breitkreutz BJ, Rust J, Boucher L, Chang C, et al. The BioGRID interaction database: 2019 update. Nucleic Acids Res. 8 janv 2019;47(D1):D529-41.

258.Szklarczyk D, Kirsch R, Koutrouli M, Nastou K, Mehryary F, Hachilif R, et al. The STRING database in 2023: protein-protein association networks and functional enrichment analyses for any sequenced genome of interest. Nucleic Acids Res. 6 janv 2023;51(D1):D638-46.

259.Huang J, Zheng W, Zhang P, Lin Q, Chen Z, Xuan J, et al. ChIPBase v3.0: the encyclopedia of transcriptional regulations of non-coding RNAs and protein-coding genes. Nucleic Acids Res. 6 janv 2023;51(D1):D46-56.

260.Keenan AB, Torre D, Lachmann A, Leong AK, Wojciechowicz ML, Utti V, et al. ChEA3: transcription factor enrichment analysis by orthogonal omics integration. Nucleic Acids Res. 2 juill 2019;47(W1):W212-24.

261.Rauluseviciute I, Riudavets-Puig R, Blanc-Mathieu R, Castro-Mondragon JA, Ferenc K, Kumar V, et al. JASPAR 2024: 20th anniversary of the open-access database of transcription factor binding profiles. Nucleic Acids Res. 5 janv 2024;52(D1):D174-82.

262.Badia-i-Mompel P, Wessels L, Müller-Dott S, Trimbour R, Ramirez Flores RO, Argelaguet R, et al. Gene regulatory network inference in the era of single-cell multi-omics. Nat Rev Genet. nov 2023;24(11):739-54.

263.Li X, Li M, Huang M, Lin Q, Fang Q, Liu J, et al. The multi-molecular mechanisms of tumor-targeted drug resistance in precision medicine. Biomed Pharmacother. 1 juin 2022;150:113064.

264.Heath AP, Ferretti V, Agrawal S, An M, Angelakos JC, Arya R, et al. The NCI Genomic Data Commons. Nat Genet. mars 2021;53(3):257-62.

265.Tsherniak A, Vazquez F, Montgomery PG, Weir BA, Kryukov G, Cowley GS, et al. Defining a Cancer Dependency Map. Cell. 27 juill 2017;170(3):564-576.e16.

266.Liang S, Fuhrman S, Somogyi R. Reveal, a general reverse engineering algorithm for inference of genetic network architectures. Pac Symp Biocomput Pac Symp Biocomput. 1998;18-29.

267.Butte AJ, Kohane IS. Mutual information relevance networks: functional genomic clustering using pairwise entropy measurements. Pac Symp Biocomput Pac Symp Biocomput. 2000;418-29.

268.Margolin AA, Nemenman I, Basso K, Wiggins C, Stolovitzky G, Favera RD, et al. ARACNE: An Algorithm for the Reconstruction of Gene Regulatory Networks in a Mammalian Cellular Context. BMC Bioinformatics. 20 mars 2006;7(1):S7.

269.Chávez Montes RA, Coello G, González-Aguilera KL, Marsch-Martínez N, de Folter S, Alvarez-Buylla ER. ARACNe-based inference, using curated microarray data, of Arabidopsis thaliana root transcriptional regulatory networks. BMC Plant Biol. 16 avr 2014;14:97.

270.Madar A, Greenfield A, Vanden-Eijnden E, Bonneau R. DREAM3: Network Inference Using Dynamic Context Likelihood of Relatedness and the Inferelator. PLOS ONE. 22 mars 2010;5(3):e9803.

271.Olsen C, Meyer PE, Bontempi G. On the impact of entropy estimation on transcriptional regulatory network inference based on mutual information. EURASIP J Bioinforma Syst Biol. 2009;2009(1):308959.

272.Gómez-Vela F, Barranco CD, Díaz-Díaz N. Incorporating biological knowledge for construction of fuzzy networks of gene associations. Appl Soft Comput. 1 mai 2016;42:144-55.

273.Trescher S, Münchmeyer J, Leser U. Estimating genome-wide regulatory activity from multi-omics data sets using mathematical optimization. BMC Syst Biol. 27 mars 2017;11(1):41.

274.Lu J, Dumitrascu B, McDowell IC, Jo B, Barrera A, Hong LK, et al. Causal network inference from gene transcriptional time-series response to glucocorticoids. PLOS Comput Biol. 29 janv 2021;17(1):e1008223.

275.Pušnik Ž, Mraz M, Zimic N, Moškon M. Review and assessment of Boolean approaches for inference of gene regulatory networks. Heliyon [Internet]. 1 août 2022 [cité 10 sept 2024];8(8). Disponible sur: https://www.cell.com/heliyon/abstract/S2405-8440(22)01510-9

276.Cao J, Qi X, Zhao H. Modeling Gene Regulation Networks Using Ordinary Differential Equations. In: Wang J, Tan AC, Tian T, éditeurs. Next Generation Microarray Bioinformatics: Methods and Protocols [Internet]. Totowa, NJ: Humana Press; 2012 [cité 26 juin 2025]. p. 185-97. Disponible sur: https://doi.org/10.1007/978-1-61779-400-1_12

277.Seaton DD. ODE-Based Modeling of Complex Regulatory Circuits. In: Kaufmann K, Mueller-Roeber B, éditeurs. Plant Gene Regulatory Networks: Methods and Protocols [Internet]. New York, NY: Springer; 2017 [cité 26 juin 2025]. p. 317-30. Disponible sur: https://doi.org/10.1007/978-1-4939-7125-1_20

278.Kang X, Hajek B, Hanzawa Y. From graph topology to ODE models for gene regulatory networks. PLOS ONE. 30 juin 2020;15(6):e0235070.

279.Hossain I, Fanfani V, Fischer J, Quackenbush J, Burkholz R. Biologically informed NeuralODEs for genome-wide regulatory dynamics. Genome Biol. 21 mai 2024;25(1):127.

280.Ma B, Fang M, Jiao X. Inference of gene regulatory networks based on nonlinear ordinary differential equations. Bioinformatics. 8 déc 2020;36(19):4885-93.

281.Liu F, Zhang SW, Guo WF, Wei ZG, Chen L. Inference of Gene Regulatory Network Based on Local Bayesian Networks. PLOS Comput Biol. 1 août 2016;12(8):e1005024.

282.Zhao M, He W, Tang J, Zou Q, Guo F. A comprehensive overview and critical evaluation of gene regulatory network inference technologies. Brief Bioinform. 1 sept 2021;22(5):bbab009.

283.Chai LE, Loh SK, Low ST, Mohamad MS, Deris S, Zakaria Z. A review on the computational approaches for gene regulatory network construction. Comput Biol Med. 1 mai 2014;48:55-65.

284.Brandman O, Meyer T. Feedback Loops Shape Cellular Signals in Space and Time. Science. 17 oct 2008;322(5900):390-5.

285.Suter P, Kuipers J, Beerenwinkel N. Discovering gene regulatory networks of multiple phenotypic groups using dynamic Bayesian networks. Brief Bioinform. 1 juill 2022;23(4):bbac219.

286.Rumelhart DE, Hinton GE, Williams RJ. Learning representations by back-propagating errors. Nature. oct 1986;323(6088):533-6.

287.LeCun Y, Bengio Y, Hinton G. Deep learning. Nature. mai 2015;521(7553):436-44.

288.Camacho DM, Collins KM, Powers RK, Costello JC, Collins JJ. Next-Generation Machine Learning for Biological Networks. Cell. 14 juin 2018;173(7):1581-92.

289.Cao Y, Geddes TA, Yang JYH, Yang P. Ensemble deep learning in bioinformatics. Nat Mach Intell. sept 2020;2(9):500-8.

290.Hwang H, Jeon H, Yeo N, Baek D. Big data and deep learning for RNA biology. Exp Mol Med. juin 2024;56(6):1293-321.

291.Li H, Tian S, Li Y, Fang Q, Tan R, Pan Y, et al. Modern deep learning in bioinformatics. J Mol Cell Biol. 1 nov 2020;12(11):823-7.

292.Li Y, Huang C, Ding L, Li Z, Pan Y, Gao X. Deep learning in bioinformatics: Introduction, application, and perspective in the big data era. Methods. 15 août 2019;166:4-21.

293.Mahmud M, Kaiser MS, McGinnity TM, Hussain A. Deep Learning in Mining Biological Data. Cogn Comput. 1 janv 2021;13(1):1-33.

294.Mikolov T, Sutskever I, Chen K, Corrado G, Dean J. Distributed representations of words and phrases and their compositionality. In: Proceedings of the 26th International Conference on Neural Information Processing Systems - Volume 2. Red Hook, NY, USA: Curran Associates Inc.; 2013. p. 3111-9. (NIPS'13).

295.Perozzi B, Al-Rfou R, Skiena S. DeepWalk: online learning of social representations. In: Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining [Internet]. New York, NY, USA: Association for Computing Machinery; 2014 [cité 18 sept 2024]. p. 701-10. (KDD '14). Disponible sur: https://doi.org/10.1145/2623330.2623732

296.Gligorijevic V, Barot M, Bonneau R. deepNF: deep network fusion for protein function prediction. Bioinforma Oxf Engl. 15 nov 2018;34(22):3873-81.

297.Grover A, Leskovec J. node2vec: Scalable Feature Learning for Networks. KDD Proc Int Conf Knowl Discov Data Min. août 2016;2016:855-64.

298.Yue X, Wang Z, Huang J, Parthasarathy S, Moosavinasab S, Huang Y, et al. Graph embedding on biomedical networks: methods, applications and evaluations. Bioinformatics. 15 févr 2020;36(4):1241-51.

299.Yang K, Swanson K, Jin W, Coley C, Eiden P, Gao H, et al. Analyzing Learned Molecular Representations for Property Prediction. J Chem Inf Model. 26 août 2019;59(8):3370-88.

300.Scarselli F, Gori M, Tsoi AC, Hagenbuchner M, Monfardini G. The graph neural network model. IEEE Trans Neural Netw. janv 2009;20(1):61-80.

301.Wang D, Cui P, Zhu W. Structural Deep Network Embedding. In: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining [Internet]. New York, NY, USA: Association for Computing Machinery; 2016 [cité 19 sept 2024]. p. 1225-34. (KDD '16). Disponible sur: https://dl.acm.org/doi/10.1145/2939672.2939753

302.Kipf TN, Welling M. Semi-Supervised Classification with Graph Convolutional Networks [Internet]. arXiv; 2017 [cité 19 sept 2024]. Disponible sur: http://arxiv.org/abs/1609.02907

303.You J, Ying R, Ren X, Hamilton WL, Leskovec J. GraphRNN: Generating Realistic Graphs with Deep Auto-regressive Models [Internet]. arXiv; 2018 [cité 19 sept 2024]. Disponible sur: http://arxiv.org/abs/1802.08773

304.Ma J, Li C, Zhang Y, Wang Z, Li S, Guo Y, et al. MULGA, a unified multi-view graph autoencoder-based approach for identifying drug–protein interaction and drug repositioning. Bioinformatics. 1 sept 2023;39(9):btad524.

305.Wang J, Qiao TJ, Zheng CH, Liu JX, Shang JL. A New Graph Autoencoder-Based Multi-level Kernel Subspace Fusion Framework for Single-cell Type Identification. IEEE/ACM Trans Comput Biol Bioinform. 2024;1-12.

306.Luo X, Wang L, Hu P, Hu L. Predicting Protein-Protein Interactions Using Sequence and Network Information via Variational Graph Autoencoder. IEEE/ACM Trans Comput Biol Bioinform. sept 2023;20(5):3182-94.

307.Li G, Lin Y, Luo J, Xiao Q, Liang C. GGAECDA: Predicting circRNA-disease associations using graph autoencoder based on graph representation learning. Comput Biol Chem. 1 août 2022;99:107722.

308.Ganeshamoorthy S, Roden L, Klepl D, He F. Gene Regulatory Network Inference through Link Prediction using Graph Neural Network. In: 2022 IEEE Signal Processing in Medicine and Biology Symposium (SPMB) [Internet]. 2022 [cité 19 sept 2024]. p. 1-5. Disponible sur: https://ieeexplore.ieee.org/abstract/document/10014835

309.Harada S, Akita H, Tsubaki M, Baba Y, Takigawa I, Yamanishi Y, et al. Dual graph convolutional neural network for predicting chemical networks. BMC Bioinformatics. 23 avr 2020;21(3):94.

310.Ma X, Wang Z, Li H. Application of an Improved Graph Neural Network for Drug Property Prediction. IEEE Access. 2024;12:46812-20.

311.Shishir FS, Hasib KMd, Sakib S, Maitra S, Shah FM. De Novo Drug Property Prediction using Graph Convolutional Neural Networks. In: 2021 IEEE 9th Region 10 Humanitarian Technology Conference (R10-HTC) [Internet]. 2021 [cité 19 sept 2024]. p. 01-6. Disponible sur: https://ieeexplore.ieee.org/abstract/document/9641611

312.Zhao H, Li Y, Wang J. A convolutional neural network and graph convolutional network-based method for predicting the classification of anatomical therapeutic chemicals. Bioinformatics. 29 sept 2021;37(18):2841-7.

313.Liu Q, Hu Z, Jiang R, Zhou M. DeepCDR: a hybrid graph convolutional network for predicting cancer drug response. Bioinforma Oxf Engl. 30 déc 2020;36(Suppl_2):i911-8.

314.Ma T, Liu Q, Li H, Zhou M, Jiang R, Zhang X. DualGCN: a dual graph convolutional network model to predict cancer drug response. BMC Bioinformatics. 15 avr 2022;23(4):129.

315.Nguyen TT, Nguyen GTT, Nguyen T, Le DH. Graph convolutional networks for drug response prediction. IEEE/ACM Trans Comput Biol Bioinform. 19 févr 2021;PP.

316.Wang H, Dai C, Wen Y, Wang X, Liu W, He S, et al. GADRP: graph convolutional networks and autoencoders for cancer drug response prediction. Brief Bioinform. 1 janv 2023;24(1):bbac501.

317.Cai L, Lu C, Xu J, Meng Y, Wang P, Fu X, et al. Drug repositioning based on the heterogeneous information fusion graph convolutional network. Brief Bioinform. 1 nov 2021;22(6):bbab319.

318.Du X, Sun X, Li M. Knowledge Graph Convolutional Network with Heuristic Search for Drug Repositioning. J Chem Inf Model. 24 juin 2024;64(12):4928-37.

319.Sun X, Wang B, Zhang J, Li M. Partner-Specific Drug Repositioning Approach Based on Graph Convolutional Network.

IEEE J Biomed Health Inform. nov 2022;26(11):5757-65.

320.Sun X, Jia X, Lu Z, Tang J, Li M. Drug repositioning with adaptive graph convolutional networks. Bioinformatics. 1 janv 2024;40(1):btad748.

321.Shen C, Ding P, Wee J, Bi J, Luo J, Xia K. Curvature-enhanced graph convolutional network for biomolecular interaction prediction. Comput Struct Biotechnol J. 1 déc 2024;23:1016-25.

322.Wang F, Lei X, Liao B, Wu FX. Predicting drug–drug interactions by graph convolutional network with multi-kernel. Brief Bioinform. 1 janv 2022;23(1):bbab511.

323.Wang W, Yang X, Wu C, Yang C. CGINet: graph convolutional network-based model for identifying chemical-gene interaction in an integrated multi-relational graph. BMC Bioinformatics. 26 nov 2020;21(1):544.

324.Yuan Q, Chen J, Zhao H, Zhou Y, Yang Y. Structure-aware protein–protein interaction site prediction using deep graph convolutional network. Bioinformatics. 22 déc 2021;38(1):125-32.

325.Zhao T, Hu Y, Valsdottir LR, Zang T, Peng J. Identifying drug–target interactions based on graph convolutional network and deep neural network. Brief Bioinform. 1 mars 2021;22(2):2141-50.

326.Ditzler G, Polikar R, Rosen G. Multi-Layer and Recursive Neural Networks for Metagenomic Classification. IEEE Trans NanoBioscience. sept 2015;14(6):608-16.

327.Yue Y, Ye C, Peng PY, Zhai HX, Ahmad I, Xia C, et al. A deep learning framework for identifying essential proteins based on multiple biological information. BMC Bioinformatics. 4 août 2022;23(1):318.

328.Zarayeneh N, Ko E, Oh JH, Suh S, Liu C, Gao J, et al. Integration of multi-omics data for integrative gene regulatory network inference. Int J Data Min Bioinforma. janv 2017;18(3):223-39.

329.Petralia F, Wang P, Yang J, Tu Z. Integrative random forest for gene regulatory network inference. Bioinformatics. 15 juin 2015;31(12):i197-205.

330.Wani N, Raza K. iMTF-GRN: Integrative Matrix Tri-Factorization for Inference of Gene Regulatory Networks. IEEE Access. 2019;7:126154-63.

331.Siahpirani AF, Chasman D, Roy S. Integrative Approaches for Inference of Genome-Scale Gene Regulatory Networks. In: Sanguinetti G, Huynh-Thu VA, éditeurs. Gene Regulatory Networks: Methods and Protocols [Internet]. New York, NY: Springer; 2019 [cité 27 juin 2025]. p. 161-94. Disponible sur: https://doi.org/10.1007/978-1-4939-8882-2_7

332.Glass K, Huttenhower C, Quackenbush J, Yuan GC. Passing Messages between Biological Networks to Refine Predicted Interactions. PLOS ONE. 31 mai 2013;8(5):e64832.

333.Kamal A, Arnold C, Claringbould A, Moussa R, Servaas NH, Kholmatov M, et al. GRaNIE and GRaNPA: inference and evaluation of enhancer-mediated gene regulatory networks. Mol Syst Biol. 12 juin 2023;19(6):e11627.

334.Lehnertz K, Bröhl T, Rings T. The Human Organism as an Integrated Interaction Network: Recent Conceptual and Methodological Challenges. Front Physiol. 2020;11:598694.

335.Wilkinson AC, Nakauchi H, Göttgens B. Mammalian Transcription Factor Networks: Recent Advances in Interrogating Biological Complexity. Cell Syst. 25 oct 2017;5(4):319-31.

336.Elati M, Neuvial P, Bolotin-Fukuhara M, Barillot E, Radvanyi F, Rouveirol C. LICORN: learning cooperative regulation networks from gene expression data. Bioinforma Oxf Engl. 15 sept 2007;23(18):2407-14.

337.Chebil I, Nicolle R, Santini G, Rouveirol C, Elati M. Hybrid method inference for the construction of cooperative regulatory network in human. IEEE Trans Nanobioscience. juin 2014;13(2):97-103.

338.Marbach D, Costello JC, Küffner R, Vega NM, Prill RJ, Camacho DM, et al. Wisdom of crowds for robust gene network inference. Nat Methods. 15 juill 2012;9(8):796-804.

339.Huynh-Thu VA, Irrthum A, Wehenkel L, Geurts P. Inferring Regulatory Networks from Expression Data Using Tree-Based Methods. PLOS ONE. 28 sept 2010;5(9):e12776.

340.Lachmann A, Giorgi FM, Lopez G, Califano A. ARACNe-AP: gene network reverse engineering through adaptive partitioning inference of mutual information. Bioinforma Oxf Engl. 15 juill 2016;32(14):2233-5.

341.Saha A, Kim Y, Gewirtz ADH, Jo B, Gao C, McDowell IC, et al. Co-expression networks reveal the tissue-specific regulation of transcription and splicing. Genome Res. 11 janv 2017;27(11):1843-58.

342.Faith JJ, Hayete B, Thaden JT, Mogno I, Wierzbowski J, Cottarel G, et al. Large-scale mapping and validation of Escherichia coli transcriptional regulation from a compendium of expression profiles. PLoS Biol. janv 2007;5(1):e8.

343.Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 29 déc 2008;9:559.

344.Nicolle R, Radvanyi F, Elati M. CoRegNet: reconstruction and integrated analysis of co-regulatory networks. Bioinforma Oxf Engl. 15 sept 2015;31(18):3066-8.

345.Gatto L, Killian T. depmap: Cancer Dependency Map Data Package. 2022.

346.Fay, Colin, Vincent Guyader, Sébastien Rochette, and Cervan Girard. Golem: A Framework for Robust Shiny Applications. 2021.

347.Merkel D. Docker: lightweight Linux containers for consistent development and deployment. Linux J. 1 mars 2014;2014(239):2:2.

348.Shannon P, Richards M. MotifDb: An Annotated Collection of Protein-DNA Binding Sequence Motifs. 2024.

349.Stojnic R, Diez D. PWMEnrich: PWM enrichment analysis. 2022.

350.Kou Y, Chen EY, Clark NR, Duan Q, Tan CM, Ma'ayan A. ChEA2: Gene-Set Libraries from ChIP-X Experiments to Decode

the Transcription Regulome. In: Cuzzocrea A, Kittl C, Simos DE, Weippl E, Xu L, éditeurs. Availability, Reliability, and Security in Information Systems and HCI. Berlin, Heidelberg: Springer; 2013. p. 416-30.

351.Mora A, Donaldson IM. iRefR: an R package to manipulate the iRefIndex consolidated protein interaction database. BMC Bioinformatics. 24 nov 2011;12:455.

352.Almende B.V. and Contributors, Thieurmel B. visNetwork: Network Visualization using « vis.js » Library. 2022.

353.Maaten L van der, Hinton G. Visualizing Data using t-SNE. J Mach Learn Res. 2008;9(86):2579-605.

354.Newman MEJ, Girvan M. Finding and evaluating community structure in networks. Phys Rev E. 26 févr 2004;69(2):026113.

355.Waltman L, van Eck NJ. A smart local moving algorithm for large-scale modularity-based community detection. Eur Phys J B. 13 nov 2013;86(11):471.

356.Zhang J, Fei J, Song X, Feng J. An Improved Louvain Algorithm for Community Detection. Math Probl Eng. 2021;2021(1):1485592.

357.Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. Sci Rep. 26 mars 2019;9(1):5233.

358.H. Wickham. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York. 2016.

359.Meyers RM, Bryan JG, McFarland JM, Weir BA, Sizemore AE, Xu H, et al. Computational correction of copy number effect improves specificity of CRISPR–Cas9 essentiality screens in cancer cells. Nat Genet. déc 2017;49(12):1779-84.

360.Clough E, Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, et al. NCBI GEO: archive for gene expression and epigenomics data sets: 23-year update. Nucleic Acids Res. 2 nov 2023;52(D1):D138.

361.Hunt GP, Grassi L, Henkin R, Smeraldi F, Spargo TP, Kabiljo R, et al. GEOexplorer: a webserver for gene expression analysis and visualisation. Nucleic Acids Res. 5 juill 2022;50(W1):W367-74.

362.Hao Y, Stuart T, Kowalski MH, Choudhary S, Hoffman P, Hartman A, et al. Dictionary learning for integrative, multimodal and scalable single-cell analysis. Nat Biotechnol. févr 2024;42(2):293-304.

363.Vapnik V, Golowich SE, Smola A. Support vector method for function approximation, regression estimation and signal processing. In: Proceedings of the 9th International Conference on Neural Information Processing Systems. Cambridge, MA, USA: MIT Press; 1996. p. 281-7. (NIPS'96).

364.Oyebode O, Ighravwe DE. Urban Water Demand Forecasting: A Comparative Evaluation of Conventional and Soft Computing Techniques. Resources. sept 2019;8(3):156.

365.Siegel RL, Kratzer TB, Giaquinto AN, Sung H, Jemal A. Cancer statistics, 2025. CA Cancer J Clin. 2025;75(1):10-45.

366.Siegel RL, Miller KD, Jemal A. Cancer statistics, 2020. CA Cancer J Clin. 2020;70(1):7-30.

367.Ding MQ, Chen L, Cooper GF, Young JD, Lu X. Precision Oncology beyond Targeted Therapy: Combining Omics Data with Machine Learning Matches the Majority of Cancer Cells to Effective Therapeutics. Mol Cancer Res. 1 févr 2018;16(2):269-78.

368.Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. janv 2018;553(7689):446-54.

369.Dlamini Z, Francies FZ, Hull R, Marima R. Artificial intelligence (AI) and big data in cancer and precision oncology. Comput Struct Biotechnol J. 1 janv 2020;18:2300-11.

370.Pich O, Bailey C, Watkins TBK, Zaccaria S, Jamal-Hanjani M, Swanton C. The translational challenges of precision oncology. Cancer Cell. 9 mai 2022;40(5):458-78.

371.Garber ME, Troyanskaya OG, Schluens K, Petersen S, Thaesler Z, Pacyna-Gengelbach M, et al. Diversity of gene expression in adenocarcinoma of the lung. Proc Natl Acad Sci. 20 nov 2001;98(24):13784-9.

372.Hayes DN, Monti S, Parmigiani G, Gilks CB, Naoki K, Bhattacharjee A, et al. Gene expression profiling reveals reproducible human lung adenocarcinoma subtypes in multiple independent patient cohorts. J Clin Oncol Off J Am Soc Clin Oncol. 1 nov 2006;24(31):5079-90.

373.Collisson EA, Campbell JD, Brooks AN, Berger AH, Lee W, Chmielecki J, et al. Comprehensive molecular profiling of lung adenocarcinoma. Nature. juill 2014;511(7511):543-50.

374.Fauteux F, Surendra A, McComb S, Pan Y, Hill JJ. Identification of transcriptional subtypes in lung adenocarcinoma and squamous cell carcinoma through integrative analysis of microarray and RNA sequencing data. Sci Rep. 22 avr 2021;11(1):8709.

375.Lehtiö J, Arslan T, Siavelis I, Pan Y, Socciarelli F, Berkovska O, et al. Proteogenomics of non-small cell lung cancer reveals molecular subtypes associated with specific therapeutic targets and immune-evasion mechanisms. Nat Cancer. nov 2021;2(11):1224-42.

376.Inamura K, Fujiwara T, Hoshida Y, Isagawa T, Jones MH, Virtanen C, et al. Two subclasses of lung squamous cell carcinoma with different gene expression profiles and prognosis identified by hierarchical clustering and non-negative matrix factorization. Oncogene. 27 oct 2005;24(47):7105-13.

377.Wilkerson MD, Yin X, Hoadley KA, Liu Y, Hayward MC, Cabanski CR, et al. Lung squamous cell carcinoma mRNA expression subtypes are reproducible, clinically important, and correspond to normal cell types. Clin Cancer Res Off J Am Assoc Cancer Res. 1 oct 2010;16(19):4864-75.

378.Hammerman PS, Lawrence MS, Voet D, Jing R, Cibulskis K, Sivachenko A, et al. Comprehensive genomic characterization of squamous cell lung cancers. Nature. sept 2012;489(7417):519-25.

379.Leiter A, Veluswamy RR, Wisnivesky JP. The global burden of lung cancer: current status and future trends. Nat Rev Clin Oncol. sept 2023;20(9):624-39.

380.Thai AA, Solomon BJ, Sequist LV, Gainor JF, Heist RS. Lung cancer. The Lancet. 7 août 2021;398(10299):535-54.

381.Shintani H, Oura S, Yamaguchi T, Makimoto S. Bone Metastasis of Non-Small-Cell Lung Cancer Showing Pathological Complete Response to Osimertinib Monotherapy. Case Rep Oncol. 30 déc 2021;14(3):1876-81.

382.Lim ZF, Ma PC. Emerging insights of tumor heterogeneity and drug resistance mechanisms in lung cancer targeted therapy. J Hematol OncolJ Hematol Oncol. 9 déc 2019;12(1):134.

383.Willers H, Azzoli CG, Santivasi WL, Xia F. Basic Mechanisms of Therapeutic Resistance to Radiation and Chemotherapy in Lung Cancer. Cancer J. juin 2013;19(3):200.

384.Cani M, Napoli VM, Garbo E, Ferrari G, Del Rio B, Novello S, et al. Targeted Therapies in Small Cell Lung Cancer: From Old Failures to Novel Therapeutic Strategies. Int J Mol Sci. janv 2023;24(10):8883.

385.Tan AC, Tan DSW. Targeted Therapies for Lung Cancer Patients With Oncogenic Driver Molecular Alterations. J Clin Oncol Off J Am Soc Clin Oncol. 20 févr 2022;40(6):611-25.

386.Acunzo M, Romano G, Palmieri D, Laganá A, Garofalo M, Balatti V, et al. Cross-talk between MET and EGFR in non-small cell lung cancer involves miR-27a and Sprouty2. Proc Natl Acad Sci. 21 mai 2013;110(21):8573-8.

387.Lu X, Peled N, Greer J, Wu W, Choi P, Berger AH, et al. MET Exon 14 Mutation Encodes an Actionable Therapeutic Target in Lung Adenocarcinoma. Cancer Res. 14 août 2017;77(16):4498-505.

388.Singh S, Yeat NY, Wang YT, Lin SY, Kuo IY, Wu KP, et al. PTPN23 ubiquitination by WDR4 suppresses EGFR and c-MET degradation to define a lung cancer therapeutic target. Cell Death Dis. 11 oct 2023;14(10):1-17.

389.Awad MM, Oxnard GR, Jackman DM, Savukoski DO, Hall D, Shivdasani P, et al. MET Exon 14 Mutations in Non–Small-Cell Lung Cancer Are Associated With Advanced Age and Stage-Dependent MET Genomic Amplification and c-Met Overexpression. J Clin Oncol. mars 2016;34(7):721-30.

390.Heist RS, Shim HS, Gingipally S, Mino-Kenudson M, Le L, Gainor JF, et al. MET Exon 14 Skipping in Non-Small Cell Lung Cancer. The Oncologist. 1 avr 2016;21(4):481-6.

391.Song Y, Li G, Ju K, Ran W, Zhao H, Liu X, et al. Mesenchymal-Epithelial Transition Exon 14 Skipping Mutation and Amplification in 5,008 Patients With Lung Cancer. Front Oncol [Internet]. 30 sept 2021 [cité 23 sept 2024];11. Disponible sur: https://www.frontiersin.org/journals/oncology/articles/10.3389/fonc.2021.755031/full

392.Fernandes M, Jamme P, Cortot AB, Kherrouche Z, Tulasne D. When the MET receptor kicks in to resist targeted therapies. Oncogene. juin 2021;40(24):4061-78.

393.Jamme P, Fernandes M, Copin MC, Descarpentries C, Escande F, Morabito A, et al. Alterations in the PI3K Pathway Drive Resistance to MET Inhibitors in NSCLC Harboring MET Exon 14 Skipping Mutations. J Thorac Oncol. 1 mai 2020;15(5):741-51.

394.Buenrostro JD, Giresi PG, Zaba LC, Chang HY, Greenleaf WJ. Transposition of native chromatin for fast and sensitive epigenomic profiling of open chromatin, DNA-binding proteins and nucleosome position. Nat Methods. déc 2013;10(12):1213-8.

395.Xenarios I, Rice DW, Salwinski L, Baron MK, Marcotte EM, Eisenberg D. DIP: the database of interacting proteins. Nucleic Acids Res. 1 janv 2000;28(1):289-91.

396.Efremova M, Vento-Tormo M, Teichmann SA, Vento-Tormo R. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. Nat Protoc. avr 2020;15(4):1484-506.

397.ENCODE Project Consortium. An integrated encyclopedia of DNA elements in the human genome. Nature. 6 sept 2012;489(7414):57-74.

398.Garcia-Alonso L, Holland CH, Ibrahim MM, Turei D, Saez-Rodriguez J. Benchmark and integration of resources for the estimation of human transcription factor activities. Genome Res. août 2019;29(8):1363-75.

399.Lo Surdo P, Iannuccelli M, Contino S, Castagnoli L, Licata L, Cesareni G, et al. SIGNOR 3.0, the SIGnaling network open resource 3.0: 2022 update. Nucleic Acids Res. 6 janv 2023;51(D1):D631-7.

400.Thomas PD, Ebert D, Muruganujan A, Mushayahama T, Albou LP, Mi H. PANTHER: Making genome-scale phylogenetics accessible to all. Protein Sci. 2022;31(1):8-22.

401.Winograd-Katz SE, Fässler R, Geiger B, Legate KR. The integrin adhesome: from genes and proteins to human disease. Nat Rev Mol Cell Biol. avr 2014;15(4):273-88.

402.Ramilowski JA, Goldberg T, Harshbarger J, Kloppmann E, Lizio M, Satagopam VP, et al. A draft network of ligand–receptor-mediated multicellular signalling in human. Nat Commun. 22 juill 2015;6(1):7866.

403.Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 1 janv 2000;28(1):27-30.

404.Liberzon A, Birger C, Thorvaldsdóttir H, Ghandi M, Mesirov JP, Tamayo P. The Molecular Signatures Database (MSigDB) hallmark gene set collection. Cell Syst. 23 déc 2015;1(6):417-25.

405.Seal RL, Braschi B, Gray K, Jones TEM, Tweedie S, Haim-Vilmovsky L, et al. Genenames.org: the HGNC resources in 2023. Nucleic Acids Res. 6 janv 2023;51(D1):D1003-9.

406.Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The Protein Data Bank. Nucleic Acids Res. 1 janv 2000;28(1):235-42.

407.Sprenger J, Lynn Fink J, Karunaratne S, Hanson K, Hamilton NA, Teasdale RD. LOCATE: a mammalian protein subcellular localization database. Nucleic Acids Res. janv 2008;36(Database issue):D230-233.

408.Thomas PD, Hill DP, Mi H, Osumi-Sutherland D, Van Auken K, Carbon S, et al. Gene Ontology Causal Activity Modeling (GO-CAM) moves beyond GO annotations to structured descriptions of biological functions and systems. Nat Genet. oct 2019;51(10):1429-33.

409.Türei D, Valdeolivas A, Gul L, Palacio-Escat N, Klein M, Ivanova O, et al. Integrated intra- and intercellular signaling

knowledge for multicellular omics analysis. Mol Syst Biol. mars 2021;17(3):e9923.

410.Ding H, Douglass EF, Sonabend AM, Mela A, Bose S, Gonzalez C, et al. Quantitative assessment of protein activity in orphan tissues and single cells using the metaVIPER algorithm. Nat Commun. 16 avr 2018;9(1):1471.

411.Yuan Q, Duren Z. Inferring gene regulatory networks from single-cell multiome data using atlas-scale external data. Nat Biotechnol. févr 2025;43(2):247-57.

412.Kim D, Tran A, Kim HJ, Lin Y, Yang JYH, Yang P. Gene regulatory network reconstruction: harnessing the power of single-cell multi-omic data. Npj Syst Biol Appl. 19 oct 2023;9(1):1-13.

413.Alanis-Lobato G, Bartlett TE, Huang Q, Simon CS, McCarthy A, Elder K, et al. MICA: a multi-omics method to predict gene regulatory networks in early human embryos. Life Sci Alliance [Internet]. 1 janv 2024 [cité 23 avr 2025];7(1). Disponible sur: https://www.life-science-alliance.org/content/7/1/e202302415

414.Jiang J, Lyu P, Li J, Huang S, Tao J, Blackshaw S, et al. IReNA: Integrated regulatory network analysis of single-cell transcriptomes and chromatin accessibility profiles. iScience [Internet]. 18 nov 2022 [cité 23 avr 2025];25(11). Disponible sur: https://www.cell.com/iscience/abstract/S2589-0042(22)01631-5

415.Kamimoto K, Stringa B, Hoffmann CM, Jindal K, Solnica-Krezel L, Morris SA. Dissecting cell identity via network inference and in silico gene perturbation. Nature. févr 2023;614(7949):742-51.

416.Green WD, Gomez A, Plotkin AL, Pratt BM, Merritt EF, Mullins GN, et al. Enhancer-driven gene regulatory networks reveal transcription factors governing T cell adaptation and differentiation in the tumor microenvironment. Immunity. 8 juill 2025;58(7):1725-1741.e9.

417.Iglesia MD, Jayasinghe RG, Chen S, Terekhanova NV, Herndon JM, Storrs E, et al. Differential chromatin accessibility and transcriptional dynamics define breast cancer subtypes and their lineages. Nat Cancer. nov 2024;5(11):1713-36.

418.Park JH, Hothi P, de Lomana ALG, Pan M, Calder R, Turkarslan S, et al. Gene regulatory network topology governs resistance and treatment escape in glioma stem-like cells. Sci Adv. 7 juin 2024;10(23):eadj7706.

419.Littman R, Cheng M, Wang N, Peng C, Yang X. SCING: Inference of robust, interpretable gene regulatory networks from single cell and spatial transcriptomics. iScience [Internet]. 21 juill 2023 [cité 23 avr 2025];26(7). Disponible sur: https://www.cell.com/iscience/abstract/S2589-0042(23)01201-4

420.Song X, Zhu Y, Geng W, Jiao J, Liu H, Chen R, et al. Spatial and single-cell transcriptomics reveal cellular heterogeneity and a novel cancer-promoting Treg cell subset in human clear-cell renal cell carcinoma. J Immunother Cancer. 4 janv 2025;13(1):e010183.

421.Sun H, Li Y, Zhang Y, Zhao X, Dong X, Guo Y, et al. The relevance between hypoxia-dependent spatial transcriptomics and the prognosis and efficacy of immunotherapy in claudin-low breast cancer. Front Immunol [Internet]. 4 janv 2023 [cité 23 avr 2025];13. Disponible sur: https://www.frontiersin.orghttps://www.frontiersin.org/journals/immunology/articles/10.3389/fimmu.2022.1042835/full

422.Ruan Z, Zhou W, Liu H, Wei J, Pan Y, Yan C, et al. Precise detection of cell-type-specific domains in spatial transcriptomics. Cell Rep Methods. 19 août 2024;4(8):100841.

423.Shang L, Wu P, Zhou X. Statistical identification of cell type-specific spatially variable genes in spatial transcriptomics. Nat Commun. 26 janv 2025;16(1):1059.

424.Dirkse A, Golebiewska A, Buder T, Nazarov PV, Muller A, Poovathingal S, et al. Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. Nat Commun. 16 avr 2019;10(1):1787.

425.Syga S, Jain HP, Krellner M, Hatzikirou H, Deutsch A. Evolution of phenotypic plasticity leads to tumor heterogeneity with implications for therapy. PLOS Comput Biol. 9 août 2024;20(8):e1012003.

426.Zahir N, Sun R, Gallahan D, Gatenby RA, Curtis C. Characterizing the ecological and evolutionary dynamics of cancer. Nat Genet. août 2020;52(8):759-67.

427.Van den Berge K, Roux de Bézieux H, Street K, Saelens W, Cannoodt R, Saeys Y, et al. Trajectory-based differential expression analysis for single-cell sequencing data. Nat Commun. 5 mars 2020;11(1):1201.

428.Street K, Risso D, Fletcher RB, Das D, Ngai J, Yosef N, et al. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. BMC Genomics. 19 juin 2018;19(1):477.

429.Cao J, Spielmann M, Qiu X, Huang X, Ibrahim DM, Hill AJ, et al. The single-cell transcriptional landscape of mammalian organogenesis. Nature. févr 2019;566(7745):496-502.

430.Rambow F, Rogiers A, Marin-Bejar O, Aibar S, Femel J, Dewaele M, et al. Toward Minimal Residual Disease-Directed Therapy in Melanoma. Cell. 9 août 2018;174(4):843-855.e19.

431.Tirosh I, Izar B, Prakadan SM, Wadsworth MH, Treacy D, Trombetta JJ, et al. Dissecting the multicellular ecosystem of metastatic melanoma by single-cell RNA-seq. Science. 8 avr 2016;352(6282):189-96.

432.Puram SV, Tirosh I, Parikh AS, Patel AP, Yizhak K, Gillespie S, et al. Single-Cell Transcriptomic Analysis of Primary and Metastatic Tumor Ecosystems in Head and Neck Cancer. Cell. 14 déc 2017;171(7):1611-1624.e24.

433.La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, et al. RNA velocity of single cells. Nature. août 2018;560(7719):494-8.

434.Weiler P, Lange M, Klein M, Pe'er D, Theis F. CellRank 2: unified fate mapping in multiview single-cell data. Nat Methods. juill 2024;21(7):1196-205.

435.Aivazidis A, Memi F, Kleshchevnikov V, Er S, Clarke B, Stegle O, et al. Cell2fate infers RNA velocity modules to improve cell fate prediction. Nat Methods. avr 2025;22(4):698-707.

436.Aissa AF, Islam ABMMK, Ariss MM, Go CC, Rader AE, Conrardy RD, et al. Single-cell transcriptional changes associated with drug tolerance and response to combination therapies in cancer. Nat Commun. 12 mars 2021;12(1):1628.

437.Chauvistré H, Shannan B, Daignault-Mill SM, Ju RJ, Picard D, Egetemaier S, et al. Persister state-directed transitioning and vulnerability in melanoma. Nat Commun. 1 juin 2022;13(1):3055.

438.Wang X, Sun Q, Liu T, Lu H, Lin X, Wang W, et al. Single-cell multi-omics sequencing uncovers region-specific plasticity of glioblastoma for complementary therapeutic targeting. Sci Adv. 22 nov 2024;10(47):eadn4306.

439.Wang L, Jung J, Babikir H, Shamardani K, Jain S, Feng X, et al. A single-cell atlas of glioblastoma evolution under therapy reveals cell-intrinsic and cell-extrinsic therapeutic targets. Nat Cancer. déc 2022;3(12):1534-52.

440.Bergen V, Soldatov RA, Kharchenko PV, Theis FJ. RNA velocity—current challenges and future perspectives. Mol Syst Biol. août 2021;17(8):e10282.

441.Zheng SC, Stein-O'Brien G, Boukas L, Goff LA, Hansen KD. Pumping the brakes on RNA velocity by understanding and interpreting RNA velocity estimates. Genome Biol. 26 oct 2023;24(1):246.

442.Liu Z, Mao X, Xie Y, Yan Y, Wang X, Mi J, et al. Single-cell RNA sequencing reveals a fibroblast gene signature that promotes T-cell infiltration in muscle-invasive bladder cancer. Commun Biol. 3 mai 2025;8(1):696.

443.Fang K, Ohihoin AG, Liu T, Choppavarapu L, Nosirov B, Wang Q, et al. Integrated single-cell analysis reveals distinct epigenetic-regulated cancer cell states and a heterogeneity-guided core signature in tamoxifen-resistant breast cancer. Genome Med. 18 nov 2024;16:134.

444.Tirosh I, Suva ML. Cancer cell states: Lessons from ten years of single-cell RNA-sequencing of human tumors. Cancer Cell. 9 sept 2024;42(9):1497-506.

445.Haase M, Comlekoglu T, Petrucciani A, Peirce SM, Blemker SS. Agent-based model demonstrates the impact of nonlinear, complex interactions between cytokines on muscle regeneration. eLife. 2024;13:RP91924.

446.Letort G, Montagud A, Stoll G, Heiland R, Barillot E, Macklin P, et al. PhysiBoSS: a multi-scale agent-based modelling framework integrating physical dimension and cell signalling. Bioinformatics. 1 avr 2019;35(7):1188-96.

447.Stoll G, Viara E, Barillot E, Calzone L. Continuous time boolean modeling for biological signaling: application of Gillespie algorithm. BMC Syst Biol. 29 août 2012;6(1):116.

448.Ghaffarizadeh A, Heiland R, Friedman SH, Mumenthaler SM, Macklin P. PhysiCell: An open source physics-based cell simulator for 3-D multicellular systems. PLOS Comput Biol. 23 févr 2018;14(2):e1005991.

449.Ghaffarizadeh A, Friedman SH, Macklin P. BioFVM: an efficient, parallelized diffusive transport solver for 3-D biological simulations. Bioinformatics. 15 avr 2016;32(8):1256-8.

450.Ruscone M, Montagud A, Chavrier P, Destaing O, Bonnet I, Zinovyev A, et al. Multiscale model of the different modes of cancer cell invasion. Bioinformatics. 1 juin 2023;39(6):btad374.

451.Iorio F, Knijnenburg TA, Vis DJ, Bignell GR, Menden MP, Schubert M, et al. A Landscape of Pharmacogenomic Interactions in Cancer. Cell. 28 juill 2016;166(3):740-54.

452.Subramanian A, Narayan R, Corsello SM, Peck DD, Natoli TE, Lu X, et al. A Next Generation Connectivity Map: L1000 Platform and the First 1,000,000 Profiles. Cell. 30 nov 2017;171(6):1437-1452.e17.

453.Yang W, Soares J, Greninger P, Edelman EJ, Lightfoot H, Forbes S, et al. Genomics of Drug Sensitivity in Cancer (GDSC): a resource for therapeutic biomarker discovery in cancer cells. Nucleic Acids Res. 1 janv 2013;41(D1):D955-61.

454.Blondel VD, Guillaume JL, Lambiotte R, Lefebvre E. Fast unfolding of communities in large networks. J Stat Mech Theory Exp. oct 2008;2008(10):P10008.

455.You T, Cheng HM, Ning YZ, Shia BC, Zhang ZY. Community detection in complex networks using density-based clustering algorithm and manifold learning. Phys Stat Mech Its Appl. 15 déc 2016;464:221-30.

456.Milacic M, Beavers D, Conley P, Gong C, Gillespie M, Griss J, et al. The Reactome Pathway Knowledgebase 2024. Nucleic Acids Res. 5 janv 2024;52(D1):D672-8.

457.Hanahan D. Hallmarks of Cancer: New Dimensions. Cancer Discov. janv 2022;12(1):31-46.

458.Dong X, Ding L, Thrasher A, Wang X, Liu J, Pan Q, et al. NetBID2 provides comprehensive hidden driver analysis. Nat Commun. 4 mai 2023;14(1):2581.

459.Li H, Chen Y, Ding M, Liu J, Sun H, Fang H, et al. Folylpolyglutamate synthetase inactivation in relapsed ALL induces a druggable folate metabolic vulnerability. Drug Resist Updat. 1 nov 2024;77:101141.

460.Lopes-Ramos CM, Belova T, Brunner TH, Ben Guebila M, Osorio D, Quackenbush J, et al. Regulatory Network of PD1 Signaling Is Associated with Prognosis in Glioblastoma Multiforme. Cancer Res. 1 nov 2021;81(21):5401-12.

461.Ruffini E, Rena O, Oliaro A, Filosso PL, Bongiovanni M, Arslanian A, et al. Lung tumors with mixed histologic pattern. Clinico-pathologic characteristics and prognostic significance. Eur J Cardiothorac Surg. 1 nov 2002;22(5):701-7.

462.Wang D, Haley JD, Thompson P. Comparative gene co-expression network analysis of epithelial to mesenchymal transition reveals lung cancer progression stages. BMC Cancer. 6 déc 2017;17(1):830.

463.Tong YF, He QE, Zhu JX, Ding EC, Song K. Multi-omics differential gene regulatory network inference for lung adenocarcinoma tumor progression biomarker discovery. AIChE J. 2022;68(4):e17574.

464.Heryanto YD, Katayama K, Imoto S. Analyzing integrated network of methylation and gene expression profiles in lung squamous cell carcinoma. Sci Rep. 22 sept 2022;12(1):15799.

465.Yu HA, Arcila ME, Rekhtman N, Sima CS, Zakowski MF, Pao W, et al. Analysis of tumor specimens at the time of acquired resistance to EGFR-TKI therapy in 155 patients with EGFR-mutant lung cancers. Clin Cancer Res Off J Am Assoc Cancer Res. 15 avr

2013;19(8):2240-7.

466.Lai L, Meng W, Wei J, Zhang X, Tan Z, Lu Y, et al. Transformation of NSCLC to SCLC after 1st- and 3rd-generation EGFR-TKI resistance and response to EP regimen and erlotinib: 2 CARE-compliant case reports. Medicine (Baltimore). 12 mars 2021;100(10):e25046.

467.Quintanal-Villalonga A, Taniguchi H, Zhan YA, Hasan MM, Chavan SS, Meng F, et al. Multiomic Analysis of Lung Tumors Defines Pathways Activated in Neuroendocrine Transformation. Cancer Discov. 2 déc 2021;11(12):3028-47.

468.Oser MG, Niederst MJ, Sequist LV, Engelman JA. Transformation from non-small-cell lung cancer to small-cell lung cancer: molecular drivers and cells of origin. Lancet Oncol. avr 2015;16(4):e165.

469.Li Y, Xie T, Wang S, Yang L, Hao X, Wang Y, et al. Mechanism exploration and model construction for small cell transformation in EGFR-mutant lung adenocarcinomas. Signal Transduct Target Ther. 2 oct 2024;9(1):261.

470.Cannoodt R, Saelens W, Deconinck L, Saeys Y. Spearheading future omics analyses using dyngen, a multi-modal simulator of single cells. Nat Commun. 24 juin 2021;12(1):3942.