

Enjeux à date de la recherche en cancérologie 2022

*Réflexion conjointe
ITMO Cancer d'Aviesan/Pôle recherche et innovation de l'INCa*

Ce document a été coordonné et rédigé par Muriel Altabef

Ont contribué à l'élaboration de ce document : C. Albiges-Rizo, M. Altabef, M. Audebert, I. Baldi, F. Bellevert, S. Benzekry, D. Bernard, P. Bertolino, M. Billaud, N. Boucherit, K. Bourougaa, I. Buvat, I. Cervenka, T. Chardès, E. Chollet, J. Clavel, G. Cristofari, V. Dardalhon, P.A. Defossez, L. Delva, L. Djerroudi, Y. de Launoit, G. Eberl, A. Eychène, M.O. Fauvarque, F. Fenaille, J. Foucaud, J.L. Galzi, A.F. Guillemain, L. Gorvel, F. Haddad, A. Harel-Bellan, A.L. Huber, F. Hubert, S. Incerti, P. Juin, C. Junot, V. Lallemand-Breitenbach, L. Le Cam, S. Le Ricousse, F. Lévi, F. Lopez, B. Malissen, P. Martineau, F. Mechta-Grigoriou, T. Mercher, D. Olive, A. Ouadi, K. Padmanabhan, M.C. Parrini, A. Pélegrin, J.C. Portais, N. Poupin, B. Quesnel, N. Rasclé, J.E. Ricci, D. Rognan, S. Roman, R. Rossignol, C. Sardet, J.E. Sarry, A. Scalbert, E. Solary, S. Tartare-Deckert, C. Theillet, S. Vasseur, A. Vincent-Salomon, A. Zarubica.

SOMMAIRE

GLOSSAIRE	4
DU NORMAL AU PATHOLOGIQUE : DES TRAJECTOIRES COMPLEXES.....	5
UNE TRANSITION DIFFICILE A ANALYSER DANS UN ENVIRONNEMENT COMPLEXE ET HETEROGENE	5
COMPRENDRE ET MODELISER AU-DELA DE LA GENETIQUE.....	6
<i>Le poids croissant du métabolisme.....</i>	6
<i>L'impact de l'horloge circadienne.....</i>	8
<i>Le rôle-clé du vieillissement</i>	8
<i>L'effet déterminant de l'exposition et du style de vie.....</i>	9
<i>Les réponses rares, une clé de compréhension difficile à mettre en œuvre</i>	10
PRENDRE LA MESURE DE LA COMPLEXITE ET AFFINER LA STRATIFICATION	11
<i>De « omiques » à « multiomiques », les « omiques » toujours d'actualité.....</i>	11
<i>La fonction pivot des biomarqueurs au-delà de la médecine de précision</i>	12
L'IMPERATIF DE NOUVEAUX MODELES.....	13
BESOINS ET ECUEILS.....	14
LA MODELISATION 3D : EN QUETE DU MODELE ABSOLU DE « L'AVATAR »	15
<i>Plusieurs types de modèles.....</i>	16
<i>Différentes approches d'analyse</i>	18
<i>Limites des modèles 3D</i>	18
L'AVENIR DES MODELES IN VIVO.....	19
<i>L'enjeu des 3R.....</i>	19
<i>Quel usage pour les modèles in vivo ?.....</i>	19
<i>Améliorations à apporter</i>	20
<i>Alternatives au modèle murin</i>	21
DES INNOVATIONS THERAPEUTIQUES MULTIFORMES	22
PRENDRE LA MALADIE DE VITESSE	22
<i>Prévention et interception</i>	22
<i>L'innovation en prévention : la culture sanitaire pour une prévention de précision</i>	23
<i>L'enjeu de l'équité des soins</i>	23
LE RENOUVEAU DE L'IMMUNOTHERAPIE	24
<i>L'espoir de l'ingénierie cellulaire</i>	25
<i>Vaccins thérapeutiques : une perspective de plus en plus concrète</i>	26
COMBINAISONS DE TRAITEMENTS INNOVANTS	27
IMPACTS ET EFFETS SECONDAIRES DES TRAITEMENTS INNOVANTS.....	28
<i>L'évaluation du poids de la maladie sur le long terme</i>	28
<i>L'accompagnement aux thérapies ciblées.....</i>	29
<i>Le cas particulier des cancers pédiatriques</i>	29
NOUVELLES MOLECULES BIOACTIVES ET REPOSITIONNEMENT D'ANCIENNES	29
<i>Une double finalité et deux types de cibles</i>	29
<i>La nécessité de tests automatisés robustes et de modèles cellulaires adaptés.....</i>	30
<i>Le repositionnement de molécules pour de nouveaux traitements rapidement utilisables</i>	30
<i>La nécessaire optimisation des approches et des collections</i>	31

LE CHANGEMENT DE PARADIGME EN RADIOTHERAPIE	31
<i>Promesses de l'hadronthérapie</i>	31
<i>La radiothérapie interne vectorisée</i>	33
<i>La théranostique : des liens de plus en plus étroits entre radiothérapie et imagerie</i>	34
INNOVATIONS THERAPEUTIQUES NON MEDICAMENTEUSES	35
QUELS FREINS A L'INNOVATION THERAPEUTIQUE ?	36
<i>Le problème de la valorisation</i>	36
<i>Le manque de cliniciennes et cliniciens scientifiques</i>	37
<i>Le besoin d'encadrer l'évaluation des modèles numériques</i>	38
DE L'UTILISATION DU NUMERIQUE DANS LA RECHERCHE SUR LES CANCERS.....	39
L'EXPLOSION DU RECOURS A L'INTELLIGENCE ARTIFICIELLE ET AUX SCIENCES DES DONNEES	40
<i>Bien appréhender la terminologie</i>	40
<i>Au cœur du numérique : les données</i>	41
<i>Adéquation, conditions et coût de l'IA</i>	42
<i>L'IA explicable, car IA ne rime pas seulement avec boîte noire</i>	44
<i>L'indispensable dialogue entre expérimentation et résultats des sciences numériques</i>	45
<i>Un nouveau carrefour interdisciplinaire à soutenir</i>	46
<i>Vers une représentation numérique de la maladie</i>	47
<i>Une approche holistique encore en devenir</i>	48
<i>Le cas particulier de la modélisation en petite dimension</i>	49
LA DIFFICILE MAIS INDISPENSABLE STRUCTURATION DES DONNEES DE SANTE	50
LE POINT CLE DU PARTAGE DES DONNEES ET DES CODES SOURCES	51
L'INCONTOURNABLE QUALITE DES DONNEES.....	52
<i>La prise en compte des écueils</i>	52
<i>La reconnaissance de la constitution d'une base de données de qualité</i>	53
LE DEFI DE LA PROTECTION DES DONNEES PERSONNELLES ET DE L'HARMONISATION DES PRATIQUES	54
<i>Des contraintes multiples</i>	54
<i>Des freins multiples</i>	55
<i>La nécessité d'une harmonisation</i>	55

GLOSSAIRE

3R (règle) : remplacer, réduire, raffiner (Replace, Reduce, Refine)
5-FU : 5-fluoro-uracile
AMM : autorisation de mise sur le marché
ANSM : Agence nationale de sécurité du médicament et des produits de santé
ATR/CHK1 : ATM and Rad3 Related/Checkpoint Kinase 1
BCMA : B Cells Maturation Antigen
BRCA1 : Breast Cancer 1
CAR-T Cells : Chimeric Antigenic Receptor-T Cells
CAS9 : CRISPR Associated Protein 9
CLIP² : centre labellisé de phase précoce
COSMIC : Catalogue of Somatic Mutations in Cancer
CRISPR : Clustered Regularly Interspaced Short Palindromic Repeats
CyTOF : Cytometry by Time of Flight
DM : dispositif médical
DORA : San Francisco Declaration on Research Assessment
FAIR : faisable, accessible, interopérable, réutilisable (Findable, Accessible, Interoperable, Reusable)
FDA : Food and Drug Administration (USA)
GATE : GEANT4 Application for Emission Tomography
GDPR : General Data Protection Regulation
GEANT4 : GEometry ANd Tracking 4 ;
GMED : Groupement pour l'évaluation des dispositifs médicaux
HAS : Haute autorité de santé
HCERES : Haut conseil de l'évaluation de la recherche et de l'enseignement supérieur
HLA : Human Leukocyte Antigen
HT(C)S : High Throughput (or Content) Screening
IA : intelligence artificielle
ICCC : International Classification of Childhood Cancer
Ion CT : Ion Computed Tomography
IRM : imagerie par résonance magnétique
LAG-3 : Lymphocyte-Activation Gene 3
Linac : Linear Accelerator
MALDI-TOF : Matrix Assisted Laser Desorption Ionization-Time of Flight
PARP-1 : poly [ADP-ribose] polymérase 1
PD1 : Programmed cell Death protein 1
PDX : Patient-Derived Xenografts
PEPR : programmes et équipements prioritaires de recherche
PSMA : Prostate-Specific Membrane Antigen
PTCH1 : Patched 1
RGPD : Règlement général sur la protection des données
RIV : radiothérapie interne vectorisée
SATT : société d'accélération du transfert de technologies
scRNAseq : séquençage ARN en cellule unique
SHH : Sonic Hedgehog
SNDS : Système national des données de santé
TCGA : The Cancer Genome Atlas
TCIA : The Cancer Imaging Archive
TEMP : tomographie d'émission monophotonique
TEP : tomographie par émission de positons
XIA : intelligence artificielle explicable

Du normal au pathologique : des trajectoires complexes

Aujourd'hui, il n'existe pas de cartographie des différents événements moléculaires impliqués dans le passage d'une cellule normale à une cellule tumorale, et la modélisation de l'évolution d'une tumeur reste difficile d'accès. Comprendre et prédire la trajectoire cellulaire d'un état normal à un état tumoral nécessite une oncologie de précision dans le temps et l'espace, ainsi qu'à différents niveaux : génétique, épigénétique, phénotypique, signalisation et microenvironnement. L'objectif, à terme, est de développer des stratégies d'interception, adapter l'escalade ou la désescalade des traitements en fonction du moment ou de l'évolution de la maladie.

UNE TRANSITION DIFFICILE A ANALYSER DANS UN ENVIRONNEMENT COMPLEXE ET HETEROGENE

Plusieurs éléments font obstacle à la compréhension du passage d'une cellule normale à une cellule maligne :

- la complexité même de l'organisation tissulaire ;
- l'hétérogénéité tumorale, et donc l'hétérogénéité et la plasticité des réponses cellulaires ;
- le rôle des communications entre différents types ou états cellulaires au sein de leur environnement.

Ces éléments ont des conséquences en termes de complexité expérimentale, de coût et de difficultés d'analyse et de modélisation, qui expliquent le manque actuel de connaissances sur les transitions phénotypiques, génétiques, épigénétiques et microenvironnementales. Les approches permettant de caractériser la signalisation spatiotemporelle sur les modèles et sur les échantillons de malades sont limitées.

Des recherches faisant appel aux concepts de « cellules d'origine », stades « précancéreux » ou « cellules souches cancéreuses » tentent de définir les points de départ et les signaux déterminant les trajectoires cellulaires du normal vers le pathologique. Quelques unes des multiples trajectoires évolutives au sein d'une tumeur établie commencent à être reconstituées, notamment grâce à l'étude de prédispositions génétiques et de lésions prénéoplasiques. Cependant, il reste très difficile d'accéder aux étapes les plus précoces de la tumorigenèse et d'en décrypter les mécanismes.

Les progrès des méthodes d'analyse d'image et des analyses moléculaires à l'échelle de la cellule unique, dans un environnement cellulaire complexe (tumoral et tissulaire), devraient aider à identifier les événements caractérisant la transition entre cellule normale et cellule tumorale, ainsi que les interactions avec les cellules du microenvironnement immun et stromal, ce dernier ayant été longtemps sous-estimé. Cependant, des obstacles persistent à différents niveaux :

- non-possibilité de suivi spatiotemporel et de combinaison de l'analyse d'image et de l'analyse moléculaire ;

Actions

- Encourager les recherches mettant en œuvre ou contribuant au développement de modèles expérimentaux adaptés à partir de cellules humaines normales, notamment en termes de robustesse, reproductibilité, dynamique avec une validation sur échantillons de personnes malades
- Obtenir des analyses dynamiques à l'échelle de la cellule unique (incluant omique et imagerie)
- Encourager le développement de techniques d'analyses permettant de quantifier plusieurs aspects de la biologie au sein d'une même cellule (génétique, épigénétique, signalisation, métabolisme, etc.)
- Favoriser les interactions interdisciplinaires (technologie, biologie cellulaire, « omiques », génétique, épigénétique, métabolisme, mathématiques, physique)

- manque de reproductibilité des modèles expérimentaux 3D, notamment liée à la difficulté de mimer l'hétérogénéité des cellules tumorales et la complexité de leur microenvironnement cellulaire ;
- inadéquation des modèles expérimentaux 3D actuels aux analyses à haut débit ;
- rareté de certaines populations cellulaires ;
- manque de corrélation ou de validation entre modèles tumoraux et échantillons cliniques.

COMPRENDRE ET MODELISER AU-DELA DE LA GENETIQUE

Après la découverte du chromosome Philadelphie dans la leucémie myéloïde chronique, une part importante des recherches a été consacrée à la caractérisation de la composante génétique des cancers, dans une vision réductionniste selon laquelle le cancer serait provoqué par un événement génétique somatique. Il est vrai que les cellules tumorales présentent toujours une accumulation très importante de mutations en comparaison de cellules saines. Cependant, si une composante génétique somatique est indispensable dans la plupart des cancers, elle n'est le plus souvent pas suffisante¹.

Ainsi, des mutations oncogéniques à l'origine de multiples clones de petite taille peuvent être mises en évidence dans une majorité de tissus sains, mais la plupart de ces clones (dont le nombre augmente avec l'âge) ne constituent pas une première étape vers le développement d'une tumeur maligne, et pourraient même avoir un effet protecteur. D'ailleurs, si tous les clones comportant des événements génétiques somatiques affectant des oncogènes conduisaient à des lésions précancéreuses, le nombre de cancers développés serait bien supérieur à ce qui est déjà observé.

Dans ce contexte, plusieurs questions émergent :

- quelles sont les trajectoires possibles d'une cellule normale à une cellule tumorale ? Qu'est-ce qui différencie un clone d'un cancer établi ?
- la réactivation de la télomérase, l'échappement à la sénescence, l'inhibition de la surveillance immunitaire, la stimulation par des cytokines inflammatoires, les modifications épigénétiques, le microenvironnement tumoral (immun et stromal) contribuent-ils indépendamment ou simultanément à cette transition ?
- toutes les cellules peuvent-elles être transformées par une altération ou une combinaison d'altérations génétiques (incluant génome codant et non codant) ?
- quelles sont les cellules cibles ?
- comment le développement, le métabolisme et le vieillissement contribuent-ils à fournir des contextes cellulaires plus permissifs à la transformation par certaines mutations ?

Le poids croissant du métabolisme

Le rôle d'une réorganisation des voies métaboliques affectant à la fois la cellule tumorale, son microenvironnement (cellules stromales diverses, dont cellules immunitaires, fibroblastes, cellules endothéliales, etc.) et l'homéostasie globale de l'organe, ou de l'organisme entier, peut être invoqué à toutes les étapes de la cancérogenèse (des lésions préneoplasiques aux métastases, avant et après traitement). Ainsi, des liens ont d'ores et déjà été établis :

- entre métabolisme, transcription et épigénétique : contrôle de l'expression de gènes impliqués dans la cancérogenèse par le métabolisme énergétique (par exemple, pseudo-hypoxie liée à des mutations de la succinate déshydrogénase, SDH), contrôle de l'épigénétique par la production d'acétyl-CoA, d' α -cétoglutarate et d'autres métabolites mitochondriaux et glycolytiques ;

¹ Le chromosome Philadelphie est à ce titre un contre-exemple, comme le sont de rares événements moléculaires oncogéniques très puissants.

- entre nutrition, microbiote, exercice physique, exposome (p.ex., rôle des pesticides) et survie des personnes malades (cachexie, réponse/résistance aux traitements de radiothérapie et de chimiothérapie, échappement immunitaire et participation à la réponse à l'immunothérapie, etc.), ouvrant la voie à des interventions préventives ou thérapeutiques non médicamenteuses (nutrition, exercice physique, etc.).

Pour autant, les spécificités métaboliques des cancers, par rapport à la situation physiologique ou dans le contexte d'autres maladies chroniques affectant le métabolisme, restent peu connues, en raison :

- de l'hétérogénéité métabolique au sein des tumeurs et de leur microenvironnement, ainsi qu'entre les organes d'un même malade et entre malades ;
- du peu de connaissances disponibles sur l'état métabolique des patientes et patients : statut métabolique pas toujours renseigné dans les dossiers médicaux ; manque de biomarqueurs, tissulaires ou sanguins, du métabolisme ; pas d'études longitudinales de suivi métabolique, ni de cohortes de malades cachectiques à différents stades de la maladie et pour différents types de cancer ;
- d'un déficit de répliquabilité des études, *in vitro* comme *in vivo*, liée notamment à l'interférence de facteurs confondants : composition des milieux, oxygénation des cellules, utilisation d'antioxydants, nutrition, aspects comportementaux, syndromes métaboliques éventuels des patientes ou patients, etc. ;
- d'un manque d'extrapolation des études précliniques aux études métaboliques chez les patientes et patients ;
- du peu d'efficacité des ciblage thérapeutiques du métabolisme déjà tentés ;
- d'un déficit de référentiel sur l'homéostasie énergétique et redox en physiologie humaine.

Relier mécanismes métaboliques et phénotypes, et parvenir ainsi à une classification métabolique des cancers, est compliqué par un accès difficile aux études métaboliques :

- spectrométrie de masse et résonance magnétique nucléaire peu disponibles en routine ;
- métabolomique en cellule unique et métabolomique par imagerie spatiale au niveau cellulaire et tumoral peu accessibles (CyTOF-Hyperion, MALDI-TOF², transcriptomique spatiale) ;
- imagerie métabolique de la personne malade (autre que la TEP (tomographie par émission de positons) cartographiant la consommation de glucose, l'hypoxie, la choline) exceptionnelle et très coûteuse actuellement ;
- échelle insuffisante du profilage métabolique, couramment réalisé sur 100 à 300 métabolites, loin des 5 000 à 10 000 nécessaires : il faudrait désormais tendre vers une métabolomique quantitative, qui nécessite des développements informatiques spécifiques.

Actions

- Clarifier le terme de « métabolisme » qui évoque des sujets différents selon les personnes
- Approfondir les recherches dans cette discipline et lui donner de la visibilité (réseau, institut dédié, formation spécifique, formation des doctorants ou intégration d'unités d'enseignement dédiées dans les parcours M2, école d'été, actions envers les industriels et le monde médical)
- Favoriser le maillage national de plateformes technologiques dédiées ainsi que le développement de bases de données métaboliques tumorales (animaux et malades) à l'échelle nationale et internationale
- Favoriser les interactions interdisciplinaires (biologie, chimie, informatique, clinique)
- Constituer des cohortes humaines ou de modèles animaux dédiées et adaptées
- Développer la métabolomique et fluxomique précliniques et cliniques (mesures *in vivo* des perturbations métaboliques – distribution des flux

² CyTOF : *Cytometry by Time of Flight* ; MALDI-TOF : *Matrix Assisted Laser Desorption Ionization-Time of Flight*

D'un point de vue structurel, ce domaine de recherche reste cloisonné, avec un manque de connexion entre les laboratoires experts en recherche métabolique, notamment mitochondriale, clinique ou fondamentale, les laboratoires experts sur le cancer et son traitement, ainsi que les industries pharmaceutiques actives dans ces deux domaines. Les innovations thérapeutiques et industrielles pour prendre en compte ces éléments se développent donc timidement.

L'impact de l'horloge circadienne

Plusieurs arguments soutiennent le rôle de l'horloge circadienne dans la physiopathologie humaine, notamment les relations :

- entre dysfonctionnement de l'horloge circadienne, oncogenèse et syndromes métaboliques ;
- entre fonctionnement de l'horloge circadienne, processus épigénétiques, métaboliques et microbiote ;
- entre décalage horaire, travail de nuit, changements de modes de vie et augmentation de l'incidence de différents types de cancer³.

Bien que les relations entre horloge circadienne, cancer et chronothérapie soient explorées depuis des années, les dérégulations de l'horloge aux différentes étapes de l'oncogenèse sont encore mal connues : peu de biomarqueurs ont été validés pour évaluer en routine les perturbations de l'horloge circadienne dans les cancers et, d'un point de vue biologique, la rythmicité des processus biologiques dans les processus de transformation maligne est encore peu étudiée. Ce type d'études est par essence transdisciplinaire et requiert notamment une expertise en modélisation mathématique pour intégrer les horloges biologiques dans la cancérologie de précision. Ce recours à la modélisation est également indispensable sur le plan thérapeutique, afin d'optimiser l'heure d'administration des traitements aux personnes malades en tenant compte des rythmes biologiques (chronothérapie).

Actions

- Référencer les communautés académiques françaises sur le territoire et à l'étranger
- Donner de la visibilité à la discipline (colloques, formation, actions envers les industriels et le monde médical en particulier le personnel soignant)
- Favoriser les interactions interdisciplinaires (oncologues, biologistes, épidémiologistes, pharmacologistes, spécialistes des données, personnels soignants, autorités de régulation des médicaments et de sécurité sanitaire, industriels)
- Mettre en place les conditions de recueil des données concernant les heures d'administration des traitements dans les dossiers des malades pour une évolution vers une chronothérapie personnalisée en cancérologie
- Prendre en compte les perturbations des rythmes de vie

Le rôle-clé du vieillissement

Le risque de cancer augmente avec le vieillissement, la principale signature moléculaire identifiée dans les cancers étant, d'ailleurs, une signature du vieillissement. Le traitement d'un cancer intervient donc souvent dans un contexte de comorbidité, du fait de la présence d'autres maladies liées au vieillissement. De plus, la tumeur proprement dite se développe dans un organe ou un environnement altérés par le vieillissement (fibrose, altérations hormonales, dysfonctionnement des mécanismes de détoxification du foie et des reins, immunosénescence/immunodépression, inflammation, etc.).

³ Le travail de nuit a ainsi été reconnu comme « probablement cancérigène » par le CIRC en 2007 et confirmé comme tel en 2019. Cette question, qui touche aussi celle des cancers professionnels, est donc un enjeu de santé publique.

Malgré l'impact du vieillissement sur le développement et la prise en charge des cancers, des questions demeurent sur les relations entre cancer et vieillissement :

- spécificités de l'oncogenèse liée au vieillissement ;
- impact du vieillissement sur la résistance thérapeutique ;
- effets secondaires des médicaments chez les sujets âgés, voire accélération du vieillissement de l'organisme sous l'effet du traitement et risque de second cancer induit par le traitement d'une première tumeur.

Par ailleurs, la plupart des essais cliniques excluant les malades âgés de plus de 75 ans, les possibilités thérapeutiques diminuent avec l'âge.

La recherche sur la biologie du vieillissement a progressé au cours des dernières décennies, mais trop peu d'études s'intéressent à la biologie et aux caractéristiques des cancers des sujets âgés et/ou souffrant de maladies, métaboliques ou autres, liées au vieillissement (diabète, BPCO (bronchopneumopathie chronique obstructive), etc.). D'un point de vue clinique, il est en effet plus complexe de développer des essais chez les personnes souffrant de cancer de plus de 75 ans, compte-tenu notamment du poids de la prise en charge multidisciplinaire (gériatrie, endocrinologie, oncologie, etc.). Du côté de la recherche expérimentale, il existe là aussi peu de cohortes humaines ou animales (> 30 mois chez la souris) portant sur le vieillissement, ou de modèles animaux adaptés à ces études. Ces travaux sont par essence longs et coûteux, et seraient par ailleurs difficilement extrapolables à l'être humain.

L'effet déterminant de l'exposition et du style de vie

L'environnement, au sens large du terme, est en cause dans les différentes étapes de la cancérogenèse, y compris dans les phases précoces (dont la période prénatale). Si plusieurs cancers ont pu être attribués à des expositions chimiques ou physiques uniques, en situation professionnelle ou accidentelle, le rôle des multiexpositions (expositions conjointes ou successives) est avéré dans plusieurs cas et doit être mieux apprécié en contexte professionnel et domestique. La reconnaissance de la grande diversité des facteurs de risque pour les différents types de cancer, ainsi que de leurs nombreuses interactions, a conduit au développement du concept « d'exposome », défini comme la totalité des expositions auxquelles chaque individu est soumis tout au long de sa vie. « L'exposomique » est donc une nouvelle manière d'aborder les questions liant facteurs d'exposition environnementaux, styles de vie et cancers, qui nécessite des techniques d'exploration particulières, et notamment des acquisitions dynamiques et exhaustives tout au long de la vie.

Les périmètres d'exploration sont multiples. Les liens entre modifications du microbiote, habitudes de vie, nutrition et cancers constituent notamment un champ de recherche en pleine expansion. La caractérisation des signatures mutationnelles et épigénétiques corrélées avec l'exposition à des composés/cancérogènes ouvre également la perspective de disposer de biomarqueurs d'exposition

Actions

- Référencer les communautés académique, médicale, industrielle sur le sujet
- Donner de la visibilité à la discipline (colloques, formation, actions envers les industriels et le monde médical)
- Favoriser les interactions interdisciplinaires (biologie, biologie des systèmes, informatique, SHS, clinique, gériatrie, maladies dégénératives)
- Constituer des cohortes et collections biologiques dédiées et adaptées
- Encourager les études épidémiologiques intégrées dans les projets expérimentaux
- Prendre en compte les effets secondaires des traitements ainsi que l'âge biologique (versus âge civil)

dans les approches rétrospectives. Ces approches concernent aussi bien les cancers d'origine professionnelle que ceux liés aux expositions dans l'environnement domestique et la vie quotidienne (produits chimiques, polluants atmosphériques, alimentation, styles de vie, microbiote, etc.). Elles ont des prolongements en sciences humaines et sociales, sur des aspects portant sur l'association entre inégalités environnementales ou socioéconomiques, modes de vie et cancer, qui sont des questions majeures en santé publique (cf. infra).

D'un point de vue méthodologique, les méthodes d'estimation rétrospective des expositions et du style de vie à partir d'interviews, bases de données, matrices emplois-exposition, matrices cultures-exposition, cartographies, algorithmes et modélisations de nature diverse (géographique, toxicocinétique, etc.) restent indispensables pour retracer l'historique des expositions passées en lien avec les caractéristiques socioéconomiques. De leur côté, les cohortes prospectives de grande taille en population générale, ou dans des populations porteuses de prédispositions génétiques, permettent de suivre plusieurs années d'expositions, d'infections, d'alimentation et d'habitudes de vie précédant la survenue d'un cancer.

Elles ne peuvent cependant pas retracer la vie entière et sont adaptées à des cancers relativement fréquents, après de longues périodes de suivi. Les cancers plus rares (notamment tous les cancers des enfants, adolescents et jeunes adultes, mais aussi certaines localisations rares de l'adulte plus âgé) ne peuvent que très difficilement être explorés prospectivement. Or en approche cas-témoins (une fois le cancer présent), on ne dispose, sauf exception, d'aucun marqueur biologique pouvant retracer l'historique des expositions des personnes cas et témoins. Afin de mieux circonscrire le rôle précoce de l'exposome, il reste nécessaire d'identifier des données d'expositions historiques pertinentes et de disposer de matériel biologique stocké suffisamment longtemps avant la survenue du cancer. La constitution et le maintien de cohortes de très grande taille impliquent une organisation et un financement important en personnel, ainsi que la mise en place de biobanques et de suivis longitudinaux durables. Leur soutien, dans le cadre juridique, réglementaire et financier actuel, est très difficile. Elles requièrent en parallèle la constitution de collections biologiques de grande taille associées à des informations cliniques et environnementales pertinentes sur une base d'échantillonnage contrôlée.

Ces approches exposomiques (à partir de biomarqueurs ou d'autres données) et de signatures mutationnelles peuvent permettre aux études épidémiologiques d'identifier les liens qui existent entre polyexposition à des substances chimiques (ainsi qu'agents physiques ou pathogènes) et cancer. Elles nécessitent le développement de plateformes technologiques *ad hoc* pour étudier l'exposomique, notamment de spectrométrie de masse à haute résolution et d'analyses bioinformatiques.

Les réponses rares, une clé de compréhension difficile à mettre en œuvre

Le développement des thérapeutiques ciblées, puis des inhibiteurs de points de contrôle, a réactivé la notion de personnes répondeuses aux traitements de façon rare ou exceptionnelle. Ces réponses concernent des personnes atteintes de tumeurs de très mauvais pronostic qui, néanmoins, répondent au traitement et survivent à 5 ans. Contrairement aux recherches sur les mécanismes de

Actions

- Encourager le développement de nouvelles méthodes d'exploration des modes d'exposition sur une longue durée utilisant de nouvelles approches technologiques (SMHR) ne se limitant pas aux échantillons biologiques classiques
- Promouvoir le développement de techniques permettant d'identifier de nouveaux xénobiotiques associés à des cancers
- Soutenir la mise en commun d'informations cartographiques pour l'évaluation de l'exposome environnemental
- Renforcer l'interopérabilité et le partage des données d'exposition
- Organiser le soutien aux biobanques constitutionnelles et tumorales adossées aux grandes enquêtes épidémiologiques en complément de France-cohortes

résistance thérapeutique, les recherches sur les mécanismes de ces réponses inattendues sont peu nombreuses. De fait, les études se heurtent à la difficulté de constituer des cohortes de taille suffisante (ces personnes étant, par définition, peu nombreuses) et au manque de disponibilité des matériels tumoral et non tumoral (dont l'analyse est nécessaire pour tenter d'identifier les mécanismes biologiques de la réponse rare). Pourtant, l'intérêt d'étudier ces réponses rares pourrait permettre de mieux comprendre les réponses non rares et les optimiser, voire de mettre au jour des mécanismes oncologiques insoupçonnés.

PRENDRE LA MESURE DE LA COMPLEXITE ET AFFINER LA STRATIFICATION

La compréhension de la complexité des tumeurs implique d'intégrer différents types de données (génomiques, épigénomiques, transcriptomiques, métabolomiques, etc.), afin de saisir l'interrelation entre ces processus biologiques au cours de la progression tumorale et de la résistance aux traitements, et d'identifier de nouvelles approches thérapeutiques. En corollaire, la stratification adéquate des individus malades est un enjeu constant de la cancérologie, afin d'administrer le bon médicament au bon moment et à la bonne personne.

De « omiques » à « multiomiques », les « omiques » toujours d'actualité

Les méthodes « omiques » permettent d'explorer et cataloguer l'ensemble des molécules bioactives, leurs caractéristiques fonctionnelles et leurs interactions. En clinique, les données omiques sont de plus en plus exploitées pour caractériser les spécificités de la maladie de chaque patiente ou patient et, si possible, prédire et suivre son évolution et sa réponse aux traitements. Il s'agit d'optimiser la prise en charge, notamment en permettant l'identification de mécanismes moléculaires et de vulnérabilité pouvant être ciblés par de petites molécules ou d'autres interventions. En recherche, les données omiques permettent de sonder la complexité des tumeurs en identifiant les composantes des tumeurs et de leur microenvironnement soit de façon globale, soit au niveau de la cellule unique, et peuvent désormais prendre en compte la dimension spatiale des tumeurs (p. ex., transcriptomique spatiale).

Néanmoins, la prise en compte d'un seul type de données omiques est souvent insuffisante pour caractériser le cancer d'une personne malade dans toute sa complexité. Il est ainsi nécessaire de combiner plusieurs méthodes et plusieurs échelles d'observation, regroupées sous le terme générique d'approches « multiomiques ».

En clinique, les données recueillies pour une même personne couvrent des aspects de plus en plus vastes de la biologie tumorale et peuvent inclure (de façon non exhaustive) : séquençage complet du génome ou de l'exome ; méthylome ; transcriptome ; protéome ou phosphoprotéome ; métabolome ; données en cellules uniques ou spatiales ; tests fonctionnels ou de sensibilité aux drogues sur des modèles avatars. En outre, afin de caractériser la trajectoire et la plasticité des cellules tumorales et de leur microenvironnement, et décrypter les mécanismes associés, il est nécessaire de passer d'une vision statique à une vision dynamique. Il faut donc capter des données temporelles et spatiales pour évaluer l'évolution longitudinale (pré/post-traitement, rechute, etc.) de la maladie. Des méthodes informatiques commencent à permettre d'intégrer et analyser ces grands volumes de données, qui sont hétérogènes par nature (omiques, imagerie, cliniques, biologiques, anatomopathologiques, etc.), mais aussi dans l'échantillonnage, la qualité - les données

Actions

- Créer un registre (à terme international) en source ouverte dédié aux réponses rares (en vie à 5 ans en dépit d'une tumeur de très mauvais pronostic)
- Mettre en place les conditions de la conservation prospective (y compris sous forme de cellules viables) de matériel tumoral et non tumoral en vue de l'analyse des réponses exceptionnelles
- Structurer un programme dédié d'analyse multidimensionnelle

étant souvent obtenues sur de très petites quantités de matériel -, les protocoles ou les technologies d'acquisition, les méthodes d'analyse et d'intégration.

Des interfaces simples d'utilisation, accessibles aux biologistes, sont disponibles *via* des portails du type de ceux de l'ICGC (*International Cancer Genome Consortium*), notamment. L'accès aux données brutes est une source d'information très riche pour tester de nouveaux algorithmes, réutiliser les données et répondre à de nouvelles questions. Il faut donc stocker à la fois les données brutes et les données prétraitées. Les données brutes de séquençage, notamment des tissus non tumoraux, sont identifiantes, même sur des échantillons anonymisés, et sont donc soumises à la législation sur les données personnelles⁴.

Aujourd'hui, les bases de données multiomiques correctement annotées et mises à jour régulièrement sont rares, car les procédures pour accéder et traiter les données brutes sont complexes à mettre en œuvre, et ce pour plusieurs raisons :

- la lourdeur des lois sur les données personnelles ;
- le coût et la complexité du recueil des données ;
- le recours nécessaire à des personnels hautement qualifiés pour accéder à des méthodes performantes et pertinentes d'intégration de données hétérogènes ;
- l'absence même d'une « meilleure manière » d'agrèger des données hétérogènes ;
- le gigantisme des capacités nécessaires pour le stockage (plusieurs téra (10^{12}), voire peta (10^{15}) octets), et même le téléchargement, de données multiomiques.

Actions

- Faciliter l'accès aux données recueillies chez les malades à des fins de recherche
- Mettre en place un portail unique d'aide au montage de projets sur les aspects « données personnelles »
- Soutenir la création d'entrepôts de données et d'infrastructures de calcul en santé interopérables
- Mettre en relation le monde médical avec les plateformes ou des équipes développant des projets technologiques multiomiques
- Développer des appels d'offres couvrant le coût de l'accès aux plateformes technologiques ou aux infrastructures de recherche
- Stimuler un développement technologique par les entreprises françaises en partenariat avec les laboratoires publics
- Mettre en place une base de données multiomiques « cancer » commune en France
- Favoriser les interactions interdisciplinaires (sciences des données, biologie, clinique) pour le développement de méthodes d'analyses adaptées aux données complexes
- Former au recueil de données multiomiques et favoriser le couplage d'approches technologiques permettant d'étudier plusieurs caractéristiques cellulaires, idéalement dans une même cellule
- Généraliser la diffusion des technologies *long-read* avec séquençage direct (ADN et ARN)
- Favoriser les études longitudinales afin d'identifier les trajectoires

La fonction pivot des biomarqueurs au-delà de la médecine de précision

La stratification des cancers s'est faite par étapes, définissant les maladies d'abord en fonction de l'organe d'origine, puis de l'aspect anatomopathologique des tumeurs, puis de marqueurs détectables par des approches immunologiques. Les progrès récents de la génomique ont considérablement affiné cette stratification en définissant de nouveaux sous-groupes de cancers, et donc de personnes qui en sont atteintes. Cette dynamique se poursuit avec l'arrivée de nouvelles approches de stratification.

Cette évolution, combinée au développement de thérapies ciblées dont l'usage est guidé par l'analyse moléculaire des tumeurs, a conduit au développement du concept de médecine de

⁴ Règlement général sur la protection des données : règlement (UE) 2016/679 du Parlement européen et du Conseil du 27 avril 2016.

précision dans laquelle une cible, dite « actionnable », est visée par des traitements spécifiques. Cette médecine repose sur l'analyse de biomarqueurs non seulement au moment du diagnostic, mais aussi au fil du temps, le niveau et le type de biomarqueurs pouvant évoluer sous l'effet d'un traitement. Cette médecine de précision nécessite également une approche solide en termes de contrôle standardisé des variations de biomarqueurs « naturelles », c.-à-d. non liées à la présence d'un cancer (nutrition, cycle circadien, états de fatigue, inflammation, infections inconnues, activité physique, etc.), qui sont génératrices de bruit de fond. L'identification d'un biomarqueur et son implémentation en routine clinique sont donc des étapes très longues, d'autant que la validation d'un biomarqueur est délicate lorsqu'il définit une population restreinte. En outre, certaines études publiées peuvent ne pas être reproductibles du fait d'une approche insuffisamment rigoureuse sur ces aspects.

En complément des données moléculaires utilisant les outils de l'intelligence artificielle (cf. infra), le développement de l'analyse des images (cytologiques, histologiques, radiologiques) faisant aussi appel à des outils numériques d'analyse et de modélisation devrait permettre de définir de nouveaux niveaux de stratification et de nouvelles possibilités d'évaluation longitudinale des tumeurs. Une approche de stratification supplémentaire consisterait à distinguer l'âge biologique d'une personne de son âge civil, en analysant une signature épigénétique, protéique ou métabolique dans le sang, cette évaluation pouvant ensuite guider, en sus de l'âge civil, la prise en charge thérapeutique d'un cancer ou sa prévention. La profondeur des données considérées pour caractériser la maladie cancéreuse d'une personne va ainsi croissant, tendant vers une prise en compte de son état physiologique complet (métabolique, immunologique, psychique, etc.) afin d'individualiser toujours plus sa prise en charge. Cependant, cette approche se heurte à la difficulté de valider les résultats : plus une stratification est précise et fine, moins le nombre de personnes concernées est grand. En dehors des cancers les plus fréquents, les cohortes constituées par sous-groupes ont donc une taille limitée.

Au-delà de leur intérêt dans le diagnostic, dans la mise en place des lignes de traitements et du suivi de leur efficacité, les marqueurs moléculaires ont un intérêt croissant dans l'évaluation et la surveillance de la maladie résiduelle après traitement, en particulier lorsqu'ils sont accessibles à une biopsie liquide. Utilisée depuis des décennies dans certaines hémopathies malignes, cette mesure de la maladie résiduelle occupe une place croissante dans la surveillance des tumeurs solides. De surcroît, au-delà de la surveillance des tumeurs déclarées, l'approche « biomarqueur » pourrait être utilisée pour le dépistage précoce des tumeurs, à condition que le dosage dispose d'une sensibilité suffisante.

L'impératif de nouveaux modèles

Comprendre les propriétés des tumeurs et anticiper les récurrences ou la résistance aux traitements nécessitent d'évaluer précisément la complexité et l'hétérogénéité tumorale dans son intégralité, au

Actions

- Promouvoir le développement et la validation de biomarqueurs associant les outils numériques et prenant en compte la variabilité « normale » du biomarqueur en fonction de situations physiopathologiques (rythme circadien, nutrition, état physique, exercice, autres maladies, infections, etc.)
- Encourager les analyses longitudinales évaluant la dynamique d'un biomarqueur dans différents contextes (évolution spontanée, réponse au traitement, prédiction de la rechute, voire prédiction de la maladie chez des sujets identifiés « à risque »)
- Poursuivre les efforts de recherche de biomarqueurs susceptibles de mieux guider le recours aux immunothérapies et de prévoir leur tolérance
- Promouvoir le partage des bases de données pour constituer des cohortes de malades stratifiés de taille suffisante

plus près de la réalité clinique. Idéalement, il s'agirait de reproduire au mieux l'écosystème tumoral et de mettre en place des micro-systèmes fiables, robustes et standardisés pour travailler avec des prélèvements de personnes malades.

BESOINS ET ECUEILS

Cette quête de nouveaux modèles a pour objectif de répondre à plusieurs nécessités :

- prendre en compte les variabilités interindividuelles : l'hétérogénéité des tumeurs et la variabilité interindividuelle constituent un problème majeur dans le traitement du cancer, chaque individu réagissant différemment aux thérapies médicamenteuses actuelles. Il est nécessaire de prendre en compte la complexité et la diversité des tumeurs selon les personnes, notamment l'environnement immunologique, métabolique, stromal et vasculaire, ainsi que l'architecture matricielle ;
- analyser la dynamique spatiotemporelle des tumeurs : des études longitudinales sont indispensables pour comprendre la dynamique et la plasticité des tumeurs dans leur écosystème complexe, notamment au cours des traitements.
- analyser la dynamique spatiotemporelle des processus biologiques : la transduction du signal est au cœur de l'interaction de la cellule cancéreuse avec son microenvironnement cellulaire et physicochimique. Les signaux extérieurs et leur transduction contrôlent le destin des cellules en voie de cancérisation en régulant les options qui se présentent à elles : prolifération, différenciation ou mort cellulaire. L'étude des cascades de signalisation a apporté une vision fonctionnelle de la transduction du signal, mais peu de travaux ont abordé la dynamique de ces cascades en corrélation avec l'évolution d'une tumeur, sa capacité à métastaser ou à résister aux traitements, et avec l'hétérogénéité intratumorale. La propagation du signal à l'échelle d'une cellule et la vitesse d'adaptation au signal extérieur restent également sous-explorées ;
- relier signalisation et hétérogénéité intratumorale : la transduction du signal est également au cœur des interactions entre les différentes sous-populations tumorales de leur environnement, et notamment entre les cellules souches cancéreuses et leurs cellules filles. Dans un contexte d'hétérogénéité intratumorale, il est important d'identifier les différentes populations par analyse d'image sur des modèles expérimentaux, pour repositionner et valider ces connaissances sur les tissus de patientes et patients. Il est nécessaire d'établir un lien entre l'analyse phénotypique des tumeurs et la dynamique de signalisation ;
- avoir un usage raisonné des modèles animaux : dans le cadre de l'application de la règle des 3R (réduire, remplacer et raffiner les modèles animaux), il est indispensable de créer des modèles alternatifs permettant d'étudier l'hétérogénéité tumorale et les interactions des cellules tumorales avec leur environnement immédiat ;
- disposer de plateformes de tests thérapeutiques fidèles : le criblage et l'analyse de nouvelles molécules thérapeutiques (y compris en immunothérapie) et de leurs combinaisons nécessitent des modèles fidèles, reproductibles et fiables (notamment pour minimiser l'usage de modèles *in vivo*), afin de maximiser les chances de réussite en phase clinique. En outre, les études portant

Actions

- Favoriser le développement d'approches pour la détection de la signalisation dans l'espace et le temps et de nouveaux outils, et de technologies d'observation et d'acquisition des données à l'échelle subcellulaire et cellulaire en 2D et 3D
- Encourager la modélisation concernant la dynamique de la signalisation, les comportements collectifs
- Soutenir le développement d'approches permettant de suivre, mimer ou contrôler la signalisation dans l'espace et le temps
- Favoriser les développements technologiques mimant l'hétérogénéité tumorale et du micro-environnement, notamment en réponse aux traitements

sur la biodistribution de nouveaux composés à visée pharmaceutique sont capitales pour évaluer la pertinence de cibles spécifiques ;

- disposer de méthodes économes en échantillons de tissus cancéreux et adaptables à la routine clinique : les échantillons de personnes malades étant rares (souvent restreints à l'étape du diagnostic) et généralement de petite taille, il est indispensable de développer des modèles miniaturisés afin de les économiser. Par ailleurs, les nouvelles méthodes développées auront plus de chance de conduire à de véritables changements de pratiques cliniques si elles sont compatibles avec un usage en routine.

Les obstacles à surmonter pour développer des modèles plus performants sont pluriels. Pour les modèles cellulaires, les difficultés sont notamment liées à la diversité phénotypique, à la multiplicité des voies de signalisation intracellulaires et à la difficulté à identifier les différentes cellules constituant une tumeur (cellules tumorales et du microenvironnement) et leurs interactions. L'hétérogénéité intra- et intertumorale *ex vivo* reste difficile à reproduire, avec une vision parfois réductrice de la complexité tumorale (résolution au niveau de la cellule unique). Au niveau technique, les difficultés rencontrées concernent la détection et l'analyse en 3D, ainsi que la complexité des systèmes d'études (transcriptomique spatiale, Hyperion/MACSima/DSP (*Digital Spatial Profiling*), MALDI-TOF, etc.) et leur coût à l'achat, à l'entretien et à l'usage (suites logicielles, kits et personnels dédiés) complexes et onéreux. Au niveau structurel, l'établissement de nouveaux modèles requiert une grande interdisciplinarité (médecine, biologie, bioinformatique, biophysique, mathématiques, statistiques, micro-ingénierie, physique, imagerie, informatique, etc.), avec des compétences souvent difficiles à combiner en un même lieu.

LA MODELISATION 3D : EN QUETE DU MODELE ABSOLU DE « L'AVATAR »

Les « avatars » représentent un rêve inachevé des oncologues, dont l'objectif est d'établir, à partir d'échantillons tumoraux d'une personne malade donnée, des modèles expérimentaux *in vitro/ex vivo* qui mimeraient fidèlement la biologie de son cancer.

Théoriquement plus faciles à manipuler et plus fidèles aux conditions pathologiques réelles, les avatars permettraient de modéliser la maladie sur du matériel humain plutôt que sur un modèle distant d'un point de vue évolutif et, dans le même temps, de répondre à la demande de limitation du recours à l'expérimentation animale. Ces modèles seraient précieux pour remplir les objectifs de médecine personnalisée en permettant de tester/prédire les réponses ou résistances aux traitements, et de suivre l'évolution de la maladie pour anticiper une éventuelle récurrence. Ils pourraient également être utilisés pour des études longitudinales précliniques et cliniques ou de toxicité, à moindre coût et sans risques pour les malades. Pour être pertinent, l'avatar devrait être le plus proche possible histologiquement et génétiquement de la tumeur à traiter. Idéalement, il devrait aussi maintenir l'hétérogénéité des populations cellulaires tumorales (hétérogénéités clonale et cellulaire) et du microenvironnement cellulaire et acellulaire (p.ex., métabolites, facteurs de croissance et cytokines sécrétés), ainsi que l'hétérogénéité du pseudo-organe tumoral associée à une résistance aux traitements. Enfin, ils devraient répondre aux critères de manipulation au laboratoire : croissance/expansion, survie, reproductibilité et, si possible, conservation à moyen/long terme.

Un panel d'approches *ex vivo* (systèmes microphysiologiques, MPS) complété par des approches *in silico* peut être développé. L'enjeu est de modéliser de façon fiable, reproductible et aussi complète que possible les interactions des cellules tumorales et de leur microenvironnement, l'organisation spatiale des tumeurs, ainsi que leur hétérogénéité et leur plasticité en 3 dimensions. L'ingénierie *ex vivo* des milieux tumoraux est toutefois complexe, car la progression du cancer et les métastases impliquent :

- des interactions cellule-matrice évoluant dans l'espace et dans le temps ;
- une myriade d'interactions entre cellules tumorales et cellules auxiliaires ;
- des environnements hypoxiques ;

- une vascularisation non organisée ;
- une multitude de molécules de signalisation.

À l'avenir, les avatars pourraient permettre d'identifier/anticiper la ligne de traitement suivante en cas d'échec thérapeutique ou de récurrence-récidive systémique.

Plusieurs types de modèles

Il existe de multiples types de systèmes microphysiologiques : sphéroïdes, organoïdes, culture organotypique, tumeur-sur-puce, reprogrammation des cellules de personnes malades en cellules souches pluripotentes (iPS)⁵, etc. L'analyse de ces modèles peut être réalisée à l'échelle unicellulaire et spatiale, et comparée aux données collectées par l'analyse directe des échantillons issus de malades.

Les tumeurs-sur-puces sont des cultures en 3D de cellules tumorales et de leur microenvironnement dans un système miniaturisé microfluidique⁶ transparent. Elles peuvent recueillir différents types cellulaires, éventuellement de façon compartimentée. L'objectif est de récapituler les différents types de signaux cellulaires et extracellulaires (moléculaires, structurels et physiques) présents *in vivo*, afin de reproduire au mieux le contexte physiologique d'origine (architecture, flux, nutriments, conditions hypoxiques, etc.).

Les xénogreffes issues de patientes et patients (Patient-Derived Xenografts, PDX) peuvent être implantées sur de nombreux modèles animaux (souris, poisson-zèbre, œuf de poule, etc.). Elles présentent de nombreux avantages, notamment celui d'étudier le cancer *in vivo* en situation orthotopique, en interaction avec un réseau vasculaire, un stroma fibroblastique, etc., d'origine animale, toutefois. Il est théoriquement possible de suivre les métastases et l'évolution des tumeurs résiduelles, d'étudier les récurrences et les mécanismes de résistance acquise. Cette technique permet une bonne préservation génétique et histologique du matériel tumoral (hors stroma), l'amplification du matériel tumoral, une imagerie et des essais précliniques proches de la situation chez les malades. Cependant :

- chez la souris, le taux de succès de la xénotransplantation des tumeurs humaines reste assez faible et très variable d'un type, ou d'un sous-type, tumoral à l'autre (avec un biais en faveur des tumeurs les plus agressives). Par ailleurs, le stroma étant murin, les interactions entre tumeur et système immunitaire/microenvironnement cellulaire humain ne peuvent être approchées qu'à condition d'utiliser des souris partiellement « humanisées », mais là encore, de façon très limitée : il est, notamment, très difficile d'étudier une réponse immune « complète » (innée et adaptative). La probabilité de succès des PDX avatars murins n'est actuellement suffisante que pour des tumeurs avancées et récidivantes. D'un point de vue économique, la création d'avatars murins coûte cher et nécessite une infrastructure dédiée (animalerie SOPF, *Specific and Opportunistic Pathogen Free*) et réglementée (3R, éthique, saisine, etc.) ;
- chez les non-mammifères (poisson-zèbre, poulet, etc.), les problèmes de prise de greffes interspèces et d'interprétation sont encore plus accentués que chez la souris. De plus, ces PDX ont une fenêtre d'utilisation limitée à une quinzaine de jours maximum (durée de vie du poisson ou de développement de l'embryon de poulet) : ce sont donc des systèmes modèles applicables une fois. En revanche, ils présentent plusieurs avantages, notamment celui d'une diminution des coûts et des contraintes éthiques ;

⁵ Cette reprogrammation est une approche utilisée pour évaluer l'impact des altérations génomiques. Elle est cependant complexe à intégrer à un modèle 3D, qui nécessite des cellules de différentes nature et origine.

⁶ D'un point de vue général, les systèmes de culture microfluidique ont montré leur intérêt en permettant de recréer *in vitro* des architectures complexes mimant l'architecture physiologique. De plus, il est désormais possible d'exploiter les avantages conférés par la microfluidique en gouttelettes par rapport aux technologies conventionnelles et aux autres techniques microfluidiques (en termes d'automatisation, de débit et de puissance combinatoire) pour manipuler et analyser des cellules uniques. Il est d'ailleurs possible de combiner la microfluidique en gouttelettes et des techniques moléculaires de pointe, comme le séquençage de nouvelle génération, pour effectuer le profilage génomique des tissus à une résolution unicellulaire.

- il faut noter le regain d'attention porté au test CAM (choroallantoïde membrane), qui permet la greffe de tissus au niveau de la membrane chorioallantoïde sans rejets, l'embryon étant naturellement dépourvu de système immunitaire. De surcroît, l'angiogenèse est possible et la tumeur transplantée est donc irriguée. Selon la réglementation européenne, ce modèle CAM correspond à un stade intermédiaire entre les cellules en culture et l'animal, d'où l'absence de problèmes éthiques et légaux.

Les organoïdes/tumoroïdes/sphéroïdes constituent un autre groupe de modèles :

- les organoïdes sont des cocultures 3D de microtissus multicellulaires, généralement issus de tissus sains, qui présentent un ordre supérieur d'autoassemblage en organe, mais qui perdent rapidement les populations issues du microenvironnement (notamment les cellules immunitaires) ;
- les tumoroïdes sont un type particulier d'organoïdes dérivés de cellules tumorales primaires prélevées chez des personnes atteintes de cancer ;
- les sphéroïdes, formes plus basiques que les organoïdes ou tumoroïdes, sont des agrégats cellulaires en 3D réalisés à partir de cellules issues de culture de lignées cellulaires tumorales ou non (cellules tumorales seules ou en contact avec des cellules stromales).

Ces formes de modèles sont propices au criblage de molécules thérapeutiques *ex vivo*, dans certains cas à haut débit, pour une meilleure prise en charge des cancers en clinique, mais également pour le développement de nouveaux candidats médicaments en recherche fondamentale. Des temps d'incubation plus longs qu'avec les modèles 2D couramment utilisés peuvent être mis en œuvre. Ces modèles permettent de mimer des traitements de chimiothérapie, mais aussi l'effet de composés cancérigènes (exposome) avant l'utilisation de modèles animaux, en toute fin de démonstration de concept. Il devrait être possible d'obtenir une meilleure prédiction des effets des traitements testés et de préciser leur éventuelle utilité en clinique en apportant une information spécifique à du matériel humain. La mise en place du modèle et l'analyse phénotypique/omique (imagerie, omiques, etc.) sont relativement simples au laboratoire.

Il existe une complémentarité entre PDX et tumoroïdes. Les modèles PDX déjà établis et caractérisés peuvent jouer un rôle essentiel pour l'amplification du matériel tumoral *in vivo* chez la souris. Cette étape est nécessaire pour standardiser/caractériser/mettre en place de nouvelles cultures de tumoroïdes,

Actions

- Encourager la création et la structuration d'un réseau unique dédié aux technologies et modèles 3D, en particulier tumeurs sur puce (intégration groupements de recherche Micronit, Microfluidique, Organoïdes, etc.)
- Favoriser le développement de différents types d'avatars (culture d'échantillons de malades, PDX, modèles 3D, tumeurs sur puce, etc.), leur caractérisation complète, leur comparaison et leur validation, et soutenir la mise en place de plateformes technologiques ouvertes permettant d'y avoir accès
- Améliorer l'accès aux échantillons frais, à l'exemple du CHTN (*Cooperative Human Tissue Network*) du NCI (*National Cancer Institute*) ou du programme *Rapid Autopsy* de l'université Johns Hopkins (USA)
- Promouvoir leur utilisation au niveau clinique
- Encourager la formation (cours, ateliers, écoles thématiques, etc.)
- Soutenir l'acquisition d'équipements dédiés
- Favoriser l'interdisciplinarité (clinique, biologie, ingénierie et bioinformatique)
- Promouvoir l'utilisation des systèmes de modélisation 3D des tumeurs pour la réalisation d'études longitudinales évaluant la plasticité de la composition des tumeurs, notamment dans l'espace, au cours de l'évolution de la maladie, spontanée ou sous l'influence d'un traitement
- Promouvoir la constitution d'atlas des modèles 3D de tumeurs humaines et des résultats des recherches conduites avec ces modèles
- Favoriser les approches combinant bioimpression et microfluidique

notamment pour les tumeurs rares. Le développement de tumeurs sur puce diagnostiques, compatibles avec la prise de décision clinique en médecine personnalisée, est en cours et devrait permettre d'améliorer les traitements. Il est cependant probable que ces modèles *ex vivo* avantagent la croissance ou l'expression de phénotypes ou de clones cellulaires spécifiques, et préservent mal l'hétérogénéité clonale.

L'impression 3D permet de placer avec précision des cellules, des facteurs bioactifs et des biomatériaux à une position donnée dans une construction tissulaire, afin de reproduire le plus fidèlement possible les conditions réelles de la tumeur ou plusieurs caractéristiques associées au microenvironnement tumoral *in vivo*. Ce type de modèles peut servir de plateformes robustes pour étudier les mécanismes de progression de la maladie, permettre le criblage à haut débit de médicaments et contribuer au développement de thérapies moléculaires de nouvelle génération et de thérapies ciblées. Il existe certaines applications de l'impression 3D dans le domaine du cancer, tant dans le développement de modèles de cancer *in vitro* que dans celui de médicaments personnalisés anticancéreux.

D'autres pistes sont explorées, telles que les cultures organotypiques ou les explants de tumeur pour les tumeurs sur coupe. Il s'agit de modèles plus proches de la tumeur d'origine, qui prennent en compte un microenvironnement acellulaire « autochtone ». Cependant, leur durée de vie est courte et leur obtention nécessite de travailler à partir d'une pièce opératoire de taille conséquente.

Différentes approches d'analyse

Les analyses multiomiques spatiales sont en pleine expansion, dans la logique d'acquérir la connaissance spatiotemporelle la plus complète possible des tumeurs. Ces méthodes s'inscrivent dans la continuité des analyses à haut débit cellulaires (CyTOF, cytométrie en flux spectrale) ou transcriptomiques (scRNAseq, Nanostring, etc.) en combinant ces approches, tout en y intégrant une dimension spatiale, cruciale à la compréhension de la complexité du microenvironnement tumoral. Ce type d'analyses permet également de réaliser des analyses fines en utilisant un matériel biologique disponible en anatomopathologie (donc de s'implémenter en clinique), et d'effectuer des cribles fonctionnels (technologie CRISPR). En comparaison d'autres techniques, ces approches sont très précises et très économes en matériel tumoral.

Limites des modèles 3D

Il existe encore de nombreuses limites au développement des modèles 3D :

- le coût, l'accès aux échantillons et les taux de réussite sont souvent relativement faibles, et très variables d'un cancer à l'autre. Cela nécessite la manipulation de cellules primaires issues de pièces opératoires incluant cellules tumorales et cellules environnantes ;
- les modèles ont une durée de vie limitée et une reproductibilité variable ;
- un véritable « contrôle qualité » de chaque avatar nécessiterait de vérifier par séquençage que le modèle mis en culture reste identique à la tumeur d'origine et qu'il n'y a pas de dérive clonale ;
- l'accès aux biopsies et pièces opératoires reste souvent compliqué ;
- pour les tumeurs solides, il est difficile d'obtenir des échantillons de tumeurs récurrentes et métastatiques permettant un suivi longitudinal ;
- la reconstitution 3D d'une tumeur reste un modèle déconnecté de l'organisme dans son ensemble. Le comportement cellulaire peut être différent dans ce contexte de ce qu'il est *in vivo*, particulièrement lorsqu'une cellule reprogrammée est isolée de son contexte. Le choix du modèle doit intégrer ces limites et s'adapter à la question posée ;
- la reproduction de l'organisation spatiale des tumeurs reste largement sous-estimée dans les systèmes actuels ;
- la composition du microbiote symbiotique est peu prise en considération jusqu'à présent dans ces modélisations ;
- l'exploration du rôle de l'âge ou du sexe dans l'évolution tumorale n'est pas possible.

- la complexité de la réponse immune, composante essentielle du microenvironnement tumoral de nombreux cancers, ne peut être actuellement appréhendée par un modèle 3D. Une réponse immune efficace, ou inefficace, se constitue *via* des éléments à la fois périphériques et centraux. On peut citer la production de cellules T naïves, qui requiert une fonction thymique affectée chez certains malades ; la reconnaissance d'un antigène tumoral, qui nécessite une présentation par les cellules dendritiques et se déroule souvent dans les ganglions de drainage ; la synthèse d'anticorps facilitateurs, ou bloquants, qui peut se situer au niveau médullaire. L'échange d'information s'effectue *via* de nombreux médiateurs systémiques, cytokines, exosomes, microvésicules, médiateurs chimiques. Un système immunitaire est ainsi constitué d'éléments complexes et variés sur le plan fonctionnel et anatomique qu'il semble illusoire de reproduire par un modèle 3D. Seules des étapes précises, locales, de la réponse antitumorale peuvent être évaluées, avec beaucoup de prudence, *via* ces systèmes.

Compte tenu de ces limites, les modèles *in vivo* restent importants en complément de ces modèles 3D.

L'AVENIR DES MODELES *IN VIVO*

Le choix du modèle animal pertinent en recherche en cancérologie est un enjeu majeur afin de disposer d'outils les plus proches possible de la maladie humaine. Cependant, l'expérimentation animale fait l'objet d'une attention soutenue de la part des pouvoirs publics et de la société.

L'enjeu des 3R

Au cours des dernières décennies, des efforts très importants ont été consacrés au respect d'une éthique de l'expérimentation animale. Le contexte principal correspond à la réduction, autant que possible, de la douleur, de l'angoisse et de la souffrance des animaux lorsque leur utilisation se révèle indispensable⁷. Dans le contexte d'une réglementation de plus en plus stricte sur l'usage des modèles animaux, de nouveaux modèles d'études émergent, comme les reconstitutions 3D de tumeurs (cf. supra). Pourtant, certains modèles *in vivo* sont et resteront indispensables pour progresser dans la compréhension, le diagnostic et le traitement des cancers sur l'individu entier, dans la mesure où certains paramètres biologiques sont difficilement évaluables autrement (cf. infra).

Quel usage pour les modèles *in vivo* ?

L'expérimentation animale est encore très utilisée par la recherche sur le cancer pour trois raisons principales⁸ :

- les progrès de la génomique ont permis de recenser les événements génétiques somatiques impliqués en cancérologie, mais des travaux de

Actions

- Recenser et favoriser la constitution de biobanques (animaux, échantillons de malades, services centralisés d'analyse des modèles en anatomopathologie, histologie, physiologie)
- Favoriser la mise en place d'outils de gestion et de professionnels dédiés pour communiquer sur les collections publiques interrogeables de modèles murins et d'échantillons issus de ces modèles (« blocothèques ») déjà marqués de façon standardisée et selon de nombreux critères, ou congelés
- Encourager et clarifier la large utilisation des ressources du consortium EuroPDX ou d'autres banques de souris européennes (Phenomin/EMMA)
- Soutenir l'usage de modèles alternatifs (p.ex., poisson zèbre) au travers de plateformes mutualisées

⁷ Le terme « indispensable » est très relatif et évolue au cours du temps.

⁸ Les protocoles d'expérimentation animale en cancérologie requièrent depuis de nombreuses années une formation *ad hoc* et une autorisation délivrée par une commission de spécialistes.

généétique et de génomique fonctionnelle sont encore nécessaires pour comprendre leur contribution ;

- le développement des médicaments anticancéreux passe encore le plus souvent par une phase préclinique de preuve de concept chez l'animal ;
- les modèles animaux sont encore à l'origine de progrès significatifs dans la compréhension des cancers humains, par exemple dans l'étude du rôle de l'hôte dans le développement des tumeurs.

Remplacer les tests sur modèle animal par des tests *ex vivo* et *in vitro* n'est pas toujours aisé, et il est nécessaire de déterminer la faisabilité du passage du modèle *in vivo* aux modèles alternatifs. Certains contextes ou questions de recherche et d'innovation impliquent de maintenir l'utilisation de modèles animaux : PDX, *knock-in* permettant plus spécifiquement de travailler sur la notion de cellule à l'origine du développement pathologique, cellules souches cancéreuses, modèles précancéreux et métastatiques, métabolisation hépatique, toxicité, vaccination, vectorisation, accessibilité des traitements à travers la barrière hémato-méningée, etc.

Dans le même temps, la généralisation de l'usage de la technologie CRISPR/CAS9, et son efficacité à produire des *knock-out* et des *knock-in*, va amplifier l'établissement de nouvelles lignées de souris et la production d'échantillons biologiques connexes (issus de ces modèles). L'absence d'outils d'information, de stockage (biobanque) et de partage de l'existant en termes de modèles, murins notamment, et d'échantillons biologiques connexes déjà bien caractérisés, est responsable d'une redondance dans la production des modèles par différents laboratoires : le coût pour la recherche, le temps passé et le nombre d'animaux utilisés s'en trouvent inutilement augmentés.

Améliorations à apporter

Mimer le contexte immunitaire humain de façon optimale est un objectif majeur du développement de nouveaux modèles *in vivo*. De fait, compte tenu du rôle majeur du système immunitaire dans la réponse antitumorale, le modèle murin, bien qu'important, est encore insatisfaisant. Les travaux en immuno-oncologie nécessitent donc le recours à des souris « humanisées », mais peu de structures académiques sont en capacité de les produire et de les maintenir en élevage prolongé, et leur coût est prohibitif. Dans le même temps, les modèles « humanisés » commerciaux n'abordent pas l'ensemble des populations immunes dans leur diversité⁹. Par ailleurs, il est possible que les conditions très strictes d'environnement dans lesquelles ces animaux sont maintenus aient des conséquences sur les résultats des études d'impact du métabolisme, du microbiote, etc., qu'il conviendrait de documenter.

La prise en compte du sexe est peu considérée dans les projets de recherche en cancérologie. Les études d'impact du contexte hormonal sur l'activité médicamenteuse ne sont pas suffisantes. Or les hormones, et plus généralement la différenciation sexuelle, incluant l'inactivation du chromosome X, jouent un rôle important en cancérologie, de la biologie des tumeurs à la pharmacologie des agents anticancéreux. Ainsi, les doses de référence des médicaments sont le plus souvent calculées pour un

Actions

- Mener une réflexion pour généraliser et améliorer l'usage d'animaux « humanisés » à un coût raisonnable
- Encourager la création de modèles plus complets, d'un point de vue immunologique notamment
- Intégrer la notion de sexe à la recherche clinique, translationnelle et cognitive en cancérologie
- Intégrer la notion de sexe aux procédures d'élaboration des nouveaux médicaments
- Surveiller l'incidence des cancers chez les personnes suivant un traitement hormonal
- Soutenir le développement de modèles d'étude de l'impact du sexe sur le développement de cancers

⁹ On sait maintenant que les immunothérapies n'affectent pas seulement les CD8, mais un ensemble de populations innées et acquises.

homme/adulte/de corpulence moyenne. Même dans les modèles animaux, le sexe est rarement considéré en tant que paramètre susceptible d'influencer le développement des tumeurs et, pour des raisons de facilité et de coût, les animaux sont principalement mâles. Il est indispensable que les paramètres de sexe soient clairement identifiés dans les études, dans les modèles expérimentaux comme chez l'humain, pour évaluer leur impact potentiel sur le développement des cellules tumorales, par un effet direct ou indirect (microenvironnement tumoral dans sa globalité). Les bases de données mériteraient d'être enrichies en indications concernant l'exposition naturelle, ou pharmacologique, aux hormones sexuelles (en référant le sexe, les traitements hormonaux pris dans le cadre d'une contraception ou d'un changement de genre, etc.).

La prise en compte du microbiote dans la constitution de modèles, ou dans la prise en charge thérapeutique, est encore faible, mais tend à se développer. L'importance d'analyser le microbiote intratumoral a récemment pris un essor important : de plus en plus de travaux mettent en avant l'impact du microbiote (et des dysbioses) sur le développement tumoral, mais également sur l'efficacité du système immunitaire et sur l'effet de nombreux médicaments. Ces travaux montrent notamment une interconnexion entre métabolisme/adaptation métabolique et tumeur/microenvironnement, ce qui suggère un lien potentiel avec la nutrition. Le développement d'applications thérapeutiques impliquant la manipulation du microbiote émerge, même s'il y a encore peu d'applications concrètes dans les traitements standards.

Il est donc nécessaire de caractériser de façon plus systématique les microbiotes (en association avec le statut clinique, le contexte physiopathologique, les informations pertinentes du dossier médical) et d'identifier de façon précise la composition du microbiote dans le tissu des personnes atteintes de cancer. Pour comprendre les chaînes de causalité, la prise en compte de ce paramètre dans le développement de modèles *in vivo* et *ex vivo* 2D et 3D est cruciale pour étudier de manière adaptée le développement tumoral, la réponse immunitaire, l'interaction avec la thérapie utilisée en combinaison avec d'autres paramètres tels que l'adaptation métabolique ou le statut hormonal. C'est un point particulièrement compliqué qui implique de prendre en compte le microbiote propre de l'animal et son fond génétique (et la clonalité des modèles murins). Le développement en cours de méthodes de métagénomique devrait contribuer à assurer une comparaison adéquate du microbiote *in situ* (tissu du malade) et de la modélisation *ex vivo* : conditions de maintien en culture (milieu *ad hoc*), accès à des métabolites variés, structure 3D, impact de différents antibiotiques.

Actions

- Développer et améliorer l'accessibilité de plateformes – notamment en -omics (incluant métagénomique et métabolomique)
- Soutenir le développement de bases de données associant microbiote et contexte physio-pathologique à l'échelle nationale et internationale
- Favoriser le développement de modèles *ex vivo* mimant la complexité du microenvironnement tumoral (immun, stromal, vasculaire) en prenant en considération le microbiote intratumoral / symbiotique
- Encourager la compréhension fonctionnelle du microbiote

Alternatives au modèle murin

Bien que leur pertinence soit souvent critiquée, 75 % des modèles animaux sont des rongeurs (souris ou rats). Des investissements intellectuels (maîtrise) et matériels (réactifs, animalerie) importants sont nécessaires pour introduire de nouvelles espèces modèles, telles que :

- le poisson-zèbre, utilisé depuis le début des années 2000 : la réglementation est en effet plus souple avec ce modèle, car les larves du poisson-zèbre ne sont considérées comme des animaux qu'à partir du moment où elles se nourrissent (6 jours après la fertilisation) ;

- la mouche drosophile, intéressante car 75 % des gènes impliqués dans les cancers humains ont un orthologue chez la mouche : ce modèle est utilisé notamment pour étudier les cancers cérébraux ;
- le ver nématode, doté de plusieurs avantages : structures cellulaire et génétique simples, reproduction rapide, organes transparents, manipulation génétique aisée ;
- le killifish, qui ne vit que quelques mois, et le rat-taupe, qui, au contraire, vit une trentaine d'années : ces deux modèles tiennent une place croissante dans l'étude des liens entre vieillissement et cancer.

Des innovations thérapeutiques multifformes

PRENDRE LA MALADIE DE VITESSE

Prévention et interception

La prévention est sans doute la stratégie la plus efficace pour réduire la mortalité associée aux cancers. Avec son approche politique (lutte contre le tabagisme et la consommation d'alcool, alimentation équilibrée, protection solaire, amélioration de l'environnement, hygiène du travail, identification et réduction des polluants, etc.), la prévention collective a un bénéfice individuel faible, mais un bénéfice collectif important. À l'inverse, la prévention personnalisée et son approche médicale, ciblée sur les vulnérabilités de chaque personne (profil génétique, expositions (tabac, alcool, amiante, etc.), hygiène de vie, etc.) a un bénéfice individuel fort, mais un bénéfice collectif plus limité (mise en œuvre difficile, voire impossible, à l'échelle de la totalité d'une population).

La notion d'« interception » a été introduite par Elisabeth Blackburn en 2011 dans un éditorial de *Cancer Research*. Il s'agit d'éliminer les cellules cancéreuses très précocement, avant même qu'elles ne soient détectables, en s'appuyant sur la connaissance des facteurs de risque et sur des présupposés. Il peut s'agir, par exemple, d'utiliser : des inhibiteurs de SHH pour prévenir l'émergence de carcinomes basocellulaires chez des sujets porteurs d'une mutation du gène *PTCH1* (syndrome de Gorlin) ; des inhibiteurs de PARP-1 pour prévenir des cancers du sein ou de l'ovaire chez les sujets portant une mutation de *BRCA1* ; des sénolytiques (traitements qui éliminent les cellules sénescents) comme le Venetoclax chez des sujets vieillissants, ou ayant eu des expositions particulières responsables de l'accumulation de cellules sénescents dans leurs tissus.

L'utilisation de l'aspirine pour prévenir les cancers colorectaux est sans doute la démonstration la plus aboutie d'une stratégie d'interception, avec des limites

(notamment l'âge) connues. Un autre exemple concerne le traitement contre des agents pathogènes prédisposant aux cancers (éradication des virus de l'hépatite B ou C, vaccination des jeunes contre certains papillomavirus). Cependant, la notion d'interception reste le plus souvent très théorique. Le temps nécessaire à l'évaluation de stratégies d'interception, comme de prévention, est long, et leur niveau de preuve d'efficacité est pour l'instant faible : il doit être mis en balance avec les risques éventuels d'une prise médicamenteuse sur le long terme. Les modèles restent ainsi encore largement à développer. Leur utilisation/conception est compliquée, notamment, par la faible pénétrance des

Actions

- Encourager l'identification des populations pouvant bénéficier de stratégies d'interception*, le développement de nouvelles classes de molécules d'interception et la pharmacogénomique dans une approche de prévention personnalisée des cancers
- Favoriser les stratégies d'interception des cellules tumorales à des stades précoces
- Prendre en compte les risques individuels vis-à-vis des traitements anti-cancéreux

* pour les cancers clairement associés à une prédisposition génétique ou à une exposition accidentelle

expositions dans l'apparition des cancers et la difficulté d'établir des preuves de concept ou de monter des essais chez l'homme à des stades précoces.

L'innovation en prévention : la culture sanitaire pour une prévention de précision

Les récents progrès de la médecine, des approches psychosociales, de la gestion et de l'organisation des soins, ainsi que des actions de promotion et d'éducation à la santé ont contribué à réduire l'impact des cancers. Ce résultat est également rendu possible par des actions ciblées – notamment auprès de populations vulnérables – ainsi que par une implication de plus en plus importante de leurs bénéficiaires – personnes malades et leur entourage – dans les soins et, plus largement, dans la santé.

La prévention de précision est centrée sur l'évaluation personnalisée du risque. Or les ressources des individus, leur environnement social, leur accès à la culture sanitaire impactent leur mode de vie et donc directement leur état de santé. La culture sanitaire est donc un levier important de la prévention, du rétablissement et de la qualité de vie, en agissant sur les déterminants sociaux de santé (meilleure connaissance des facteurs de risque) et le développement de comportements et environnements favorables à la santé.

Cependant, le « Baromètre cancer » de l'INCa montre que la société civile a beaucoup de mal à appréhender l'ampleur des progrès en cancérologie, mais aussi le poids des facteurs de risques. En outre, les progrès scientifiques permettent le développement de prestations, traitements, produits et dispositifs d'accompagnement pour les malades et leur entourage qui rendent le paysage complexe. Ces constats concourent à rendre nécessaire le développement d'une culture sanitaire et donc un certain niveau de littératie en santé¹⁰. En effet, quels que soient les progrès en termes de soin ou de prévention, ils ne seront pleinement bénéfiques que s'ils remportent l'adhésion d'une population (malades et société civile) qui comprend et s'approprie la stratégie proposée. Cette approche nécessite aussi la prise en compte de savoirs profanes, ou expérientiels, des populations, ainsi que la réduction des inégalités sociales et territoriales de santé (ISS).

Actions

- Favoriser la représentation des malades, des parents de malades ou des citoyen.nes dans les Conseils scientifiques des instituts de recherche en cancérologie
- Encourager l'intervention de malades et de leurs proches dans les cursus universitaires des médecins et des scientifiques
- Promouvoir la formation de la société civile à la recherche, et en particulier à la recherche sur le cancer

De nombreux freins entravent le développement de la culture sanitaire :

- les inégalités socioéconomiques, qui creusent la complexité d'accès aux soins ou aux programmes d'intervention de promotion de la santé ;
- la méconnaissance des enjeux de la culture sanitaire chez les professionnels de santé et de l'éducation ;
- la complexité des messages d'information et de promotion de la santé ;
- la méfiance à l'égard des institutions publiques sanitaires ; le nombre important de sources d'information non fiables.

L'enjeu de l'équité des soins

L'équité des soins est une problématique majeure en France, où d'importantes inégalités territoriales et sociodémographiques se traduisent par la disparité des facteurs de risque et de l'incidence des cancers. Elle est à distinguer de l'égalité des chances face au cancer, qui commence en amont,

¹⁰ La littératie en santé représente la capacité d'accéder à l'information, de la comprendre, de l'évaluer et de la communiquer, afin de promouvoir, maintenir et améliorer sa santé dans divers milieux et tout au long de sa vie.

notamment par l'égalité d'accès à la prévention primaire. L'équité des soins se pose à toutes les étapes de la maladie, mais particulièrement au moment du dépistage, du traitement, de la reconnaissance éventuelle de maladie professionnelle ou de handicap, et de l'après-cancer.

Les inégalités comportent une dimension territoriale : les 30 % de la population française résidant dans un désert médical sont plus sujets à un déficit de dépistage, à des retards de diagnostic et à un manque d'accompagnement en relais des structures hospitalières ou dans l'après-cancer. Ces inégalités concernent également des territoires à plus haut risque de cancers liés à l'alcool, au tabac, à l'obésité ou à des facteurs environnementaux particuliers (p. ex., mercure en Guyane, chlordécone aux Antilles), ainsi que des territoires où l'accès de la population au dépistage et à la prise en charge est rendu difficile par des spécificités de topographie, de multiculturalisme, de multilinguisme et de complexités sociales. Enfin, les questions d'accès aux soins présentent des spécificités dans certaines populations, notamment dans les situations de handicap, d'extrême pauvreté, d'exclusion sociale (p. ex., milieu carcéral) ou, encore, de prédispositions génétiques aux cancers.

L'enjeu de l'équité des soins porte sur une offre et une organisation de soins adaptées et de qualité sur l'ensemble du territoire français, proposant un parcours de santé/soin optimal. Il englobe de nombreux aspects : littératie en santé, promotion de la santé, dimensions politiques et économiques, dimensions sociales et de croyances religieuses, pratiques coutumières, comportements communautaires, influence des réseaux sociaux, etc.

Pour la recherche, l'enjeu est d'appréhender et développer les connaissances sur ces aspects, afin de proposer des interventions efficaces aux différentes étapes. Il s'agit également de déterminer les indicateurs individuels et contextuels concernant l'offre et la qualité des soins à disposition, ainsi que la mesure de la défaveur territoriale, socioéconomique ou culturelle. Le choix et la finesse de ces indicateurs sont décisifs pour cerner les inégalités en fonction du type de soins, des âges et des populations. L'évaluation épidémiologique des inégalités d'accès aux soins, de leur évolution et de l'efficacité des interventions suppose de disposer d'indicateurs de référence stables dans le temps et au niveau géographique, ainsi qu'une meilleure prise en compte des disparités populationnelles sur le territoire, en particulier par le système d'organisation des soins et d'accès à l'information, très centralisé en France. Il s'agit là d'une thématique transversale, qui nécessite un décroisement disciplinaire et une approche inclusive à toutes les étapes.

Actions

- Développer les réseaux avec la société civile
- Étendre la communauté de recherche en proposant des programmes dédiés aux jeunes scientifiques
- Réfléchir avec les épidémiologistes, les économistes de la santé et des spécialistes opérationnels (Insee (Institut national de la statistique et des études économiques), Drees (Direction de la recherche, des études, de l'évaluation et des statistiques), Observatoire des inégalités, etc.) aux indicateurs existants et à leur périmètre d'utilisation
- Encourager des recherches interventionnelles systémiques
- Diffuser la connaissance auprès des structures de décision, collectivités, acteurs locaux, etc.

LE RENOUVEAU DE L'IMMUNOTHERAPIE

La découverte des points de contrôle et de leur inhibition par des anticorps monoclonaux a permis de redonner une seconde vie à une piste qui, jusque-là, n'avait pas rencontré de succès : miser sur le système immunitaire pour contenir, voire éradiquer la tumeur. Bien que les biomarqueurs susceptibles de guider l'utilisation de tels anticorps restent insuffisants, cette approche est incontestablement source de succès impressionnants et d'espoirs en mono- et, surtout, combothérapie.

Le second grand progrès en matière d'immunothérapie a été la mise au point de nouvelles thérapies cellulaires fondées sur la manipulation génétique des lymphocytes T du malade, afin de les cibler spécifiquement vers sa propre tumeur (*CAR-T Cells*). Cette approche est en plein développement, mais à un coût qui reste encore prohibitif. De plus, elle n'est, pour le moment, efficace que vis-à-vis un nombre très limité de cibles (CD19, BCMA), et uniquement dans les hémopathies lymphoïdes. Par ailleurs, une approche très prometteuse sur le plan logistique, voire économique, les CAR-T allogéniques, reste encore au stade du développement malgré certains succès cliniques. Enfin, seuls quelques signaux d'activités ont été obtenus dans quelques tumeurs solides, sans amélioration de la survie démontrée pour l'instant. La recherche dans ce domaine devra donc être soutenue, afin d'optimiser la technologie et la rendre accessible à toutes et tous, sur un large éventail de maladies.

L'espoir de l'ingénierie cellulaire

La génération post-CAR-T : l'ingénierie cellulaire connaît un important essor depuis quelques années, y compris en phases translationnelle et clinique. Afin de surmonter les limitations des CAR-T décrites plus haut, il convient d'explorer l'hypothèse que d'autres effecteurs immuns moins connus, ou d'autres types cellulaires, puissent servir de vecteurs pour des facteurs de croissance et des médicaments. Ainsi, de très nombreuses nouvelles approches sont explorées : utilisation d'autres effecteurs immuns comme les CAR-K, CAR-B, etc. ; modification massive de CAR-T allogéniques *via* les technologies CRISPR/CAS9 ; doubles CAR avec une logique booléenne afin d'améliorer la sélectivité de la reconnaissance de la cellule tumorale ; CD28 ou 41BB mutants pour optimiser la réponse. Le potentiel de développement des CAR est, en théorie, infini, puisque toutes les modifications des effecteurs sont possibles. Une difficulté majeure est toutefois représentée par la nécessité du développement clinique, indispensable pour tester ces approches en l'absence de modèles précliniques réellement pertinents.

Les nouveaux anticorps constituent un domaine stratégique au cœur des nouveaux traitements, avec des enjeux médicoéconomiques majeurs. En 2022, une centaine d'anticorps thérapeutiques ont été approuvés dans le monde, dont 50 % en oncologie :

- les anticorps monospécifiques non modifiés voient leur développement continuer (p.ex , anti-LAG-3 en combinaison avec les anti-PD1). Ils peuvent également être couplés à des toxines ou des médicaments cytotoxiques (*Antibody-Drug Conjugates*) : le spectre des toxines ou des nouvelles molécules chimiques conjuguées s'élargit constamment (p.ex., inhibiteurs de topoisomérase comme pour l'exatecan et ses dérivés). En complément, leur biodisponibilité, leur stabilité et leur ciblage sont optimisés (p.ex., anticorps fonctionnant à pH acide).
- les anticorps bispécifiques et multispécifiques sont encore peu nombreux à être approuvés en Europe dans le cancer. En revanche, une centaine sont aujourd'hui en essai clinique, toutes indications confondues, dont 18 en phase III pour une indication cancer. On peut donc s'attendre à une forte croissance du domaine des anticorps bi/multispécifiques dans les dix

Actions

- Favoriser la mise en place de plateformes académiques (développement et sélection d'anticorps, analyse de structure, affinité et spécificité, production, etc.)
- Soutenir les approches de type intelligence artificielle dans ce domaine
- Encourager les interactions entre scientifiques et médecins sur ces sujets (nouvelles cibles pertinentes, spécificité, biodistribution, balance efficacité/toxicité, galénique etc.)
- Renforcer l'enseignement ainsi que la formation académique et techniques appliquées sur les anticorps, notamment en tant que biomédicament (pharmacie, médecine, écoles d'ingénieur.e.s, etc.)
- Favoriser les interactions académiques/industriels pour la maîtrise des procédés de production qui sont de haute technicité
- Valoriser l'apport intellectuel, le savoir-faire et l'expertise du monde académique

prochaines années, d'autant que ce type de molécules représente un tiers du portefeuille des grandes industries pharmaceutiques. Un des grands intérêts des anticorps bi/multispécifiques étant de créer des molécules inédites dans la nature, le domaine est ouvert à des idées originales. Aujourd'hui, deux grandes familles de mécanismes d'actions sont mises à profit en oncologie : le cociblage tumoral de 2 récepteurs sur une même cellule, de façon à augmenter l'efficacité du ciblage, diminuer la toxicité sur le tissu sain et limiter à la baisse les régulations d'échappement (p.ex., amivantamab), donc agir sur la récurrence ; le ciblage de 2 récepteurs sur 2 cellules différentes, permettant leur rapprochement. Des anticorps bispécifiques ciblant certaines mutations commencent à être développés et sont très prometteurs (p.ex., anticorps bipécifique dirigé contre un complexe p53 mutée (néoantigène)/allèle HLA à la surface des cellules tumorales). En parallèle, ce concept est aussi actuellement développé pour les cellules CAR-T.

En complément des techniques d'ingénierie cellulaire, de nombreux travaux sont consacrés à de nouveaux vecteurs de communication intercellulaire tels que les vésicules, les microparticules, les nanotubules, les exosomes, etc. Des outils et technologies d'étude *ex vivo* et *in vivo* de ces micro- et nanostructures cellulaires nécessitent d'être développés et optimisés pour soutenir ce qui s'annonce comme un axe de recherche prometteur.

Vaccins thérapeutiques : une perspective de plus en plus concrète

L'immunothérapie a certainement marqué un tournant décisif dans le traitement des cancers, notamment ceux pour lesquels les options thérapeutiques conventionnelles n'étaient pas suffisantes. Outre l'utilisation d'anticorps bloquants contre les inhibiteurs de points de contrôle et l'ingénierie cellulaire des cellules du système immunitaire, une autre voie très prometteuse concerne la mise au point de vaccins thérapeutiques. La pandémie de Covid-19 a démontré que les approches vaccinales préventives fondées sur l'utilisation de l'ARN messager pouvaient se révéler d'une efficacité rarement atteinte en matière de vaccination. De surcroît, les principales firmes pharmaceutiques qui ont mis à disposition de la population ces vaccins en un temps record ont pu le faire grâce aux recherches menées ces dernières années en matière de vaccins contre le cancer.

L'idée qu'un vaccin thérapeutique à ARN puisse être efficace contre les cancers a mûri en parallèle de notre meilleure compréhension de l'hétérogénéité et de la plasticité tumorale. En effet, celles-ci se matérialisent par l'accumulation d'anomalies génétiques de plus en plus nombreuses (mutations, translocations, épissages alternatifs aberrants, etc.), parallèlement à l'acquisition de variations épigénétiques au cours des traitements (par chimiothérapie, notamment) responsables de mécanismes de résistance. Ces modifications de l'identité cellulaire pourraient se révéler être un talon d'Achille des cellules cancéreuses, par l'expression de néoantigènes spécifiques représentant des cibles potentielles pour un vaccin curatif ou préventif de récurrences.

L'enjeu, qui nécessite un effort de recherche important, réside dans l'identification la plus rapide et précoce possible, chez les personnes malades, des néoantigènes susceptibles d'être exprimés par les ARN messagers encapsulés dans les vaccins. Un même vaccin pourrait contenir plusieurs molécules d'ARN exprimant plusieurs néoantigènes différents, diminuant d'autant les risques de résistance. Cette approche sera plus difficile à mettre en œuvre pour les tumeurs « froides », résistantes aux autres approches d'immunothérapie, pour lesquelles il faudra développer des approches moléculaires permettant d'induire efficacement l'expression de néoantigènes pour rendre ceux-ci visibles au système immunitaire. À plus long terme, cette approche pourrait être combinée avec les traitements conventionnels de référence (chimiothérapie, radiothérapie, inhibiteurs de point de contrôle, etc.).

COMBINAISONS DE TRAITEMENTS INNOVANTS

Combiner des drogues pour le traitement des cancers est un principe accepté depuis de nombreuses années (p.ex., combinaison cytarabine et anthracyclines pour les leucémies aiguës myéloïdes), et constitue même, parfois, la seule approche curative dont l'efficacité est démontrée (p.ex., combinaison acide rétinoïque et arsenic dans le cas des leucémies promyélocytaïres). La nécessité de combiner des molécules répond à la biologie de la plupart des cancers, dont le développement résulte d'altérations affectant plusieurs processus biologiques indépendants ou redondants, d'une façon souvent spécifique de la personne malade et à l'origine de résistance à la chimiothérapie. Ces résistances sont fréquemment la cause de rechutes incurables : les connaître et prévenir leur survenue par l'utilisation de combinaisons thérapeutiques revêt donc un intérêt clinique majeur.

Les traitements actuels sont souvent séquentiels, avec en première ligne un traitement conventionnel de référence, puis des traitements utilisant des combinaisons de molécules développées plus récemment, afin de cibler les résistances. L'enjeu est de mieux anticiper les résistances par des combinaisons thérapeutiques et déterminer à quelle phase du traitement les combinaisons seront les plus efficaces. Pour cela, il est nécessaire de prendre en compte la biologie des tumeurs et leurs interactions avec le système immunitaire pour définir comment combiner, de façon la plus efficace possible, les approches pharmacologiques (molécules chimiques), les approches utilisant des produits biologiques (p.ex., anticorps ou cellules immunitaires) ou des stratégies de radiothérapie ciblée.

Les freins au développement des combinaisons de molécules sont de plusieurs ordres :

- le manque de connaissance des mécanismes de résistance et de définition des doses efficaces pour chaque molécule (multipliant les conditions à étudier), en excluant d'éventuelles toxicités additives. En revanche, l'hypothèse que les combinaisons, par leur synergie, pourraient permettre d'utiliser chacune des drogues à des doses plus faibles est prometteuse pour limiter leur toxicité ;
- la difficulté de documenter le compromis bénéfice-risk en faveur d'une combinaison thérapeutique nuit à la possibilité de s'écarter de la première ligne de traitement classiquement utilisée, même dans les cas où les sous-types de cancers sont connus pour être presque systématiquement associés à des rechutes incurables. L'exploitation d'effets synergiques de combinaison de drogues dans le contexte spécifique de certains profils génétiques/phénotypiques a été démontrée dans des études de preuve de concept prometteuses. Mais ces résultats sont encore sous-exploités en pratique clinique (p.ex., combinaison d'inhibiteurs de PARP et drogues endommageant l'ADN dans un contexte de tumeurs déficientes pour BRCA,

Actions

- Soutenir les approches multidisciplinaires englobant la caractérisation moléculaire fine des tumeurs, la chimie moléculaire et l'immunologie pour une vision globale des mécanismes de résistance
- Promouvoir des études expérimentales à grande échelle sur les effets synergiques des drogues déjà existantes, dans tous les contextes mutationnels connus liés à la cancérogenèse
- Promouvoir des études précliniques standardisées de validation et de toxicité des combinaisons identifiées
- Ne pas réserver les combinaisons de molécules innovantes aux seules personnes malades résistantes à la première ligne de traitement conventionnel
- Encourager la définition de critères cliniques permettant de préconiser d'emblée une combinaison thérapeutique selon les profils moléculaires (systématiques) des tumeurs
- Soutenir le développement d'essais cliniques nationaux de faible nombre de malades possédant un profil moléculaire compatible avec des combinaisons thérapeutiques ayant fait l'objet de preuves de concept
- Convaincre les industriels d'étudier et de consentir à la combinaison de molécules issues d'entreprises différentes

etc. ; association d'inhibiteurs d'ATR/CHK1 et d'agents bloquant la réplication pour gemcitabine, 5-FU, etc.). Cette approche nécessite une bonne connaissance du profil phénotypique/mutationnel des tumeurs, données souvent absentes en pratique clinique quotidienne ;

- le manque de collaboration entre industriels commercialisant les molécules impliquées, et le coût de ces combinaisons au regard des possibilités de prise en charge par les systèmes de santé ;
- la difficulté de réaliser des essais cliniques pour justifier l'utilisation d'une combinaison thérapeutique en 1^{re} intention.

IMPACTS ET EFFETS SECONDAIRES DES TRAITEMENTS INNOVANTS

Bien que l'efficacité des traitements des cancers ait progressé dans une majorité de cas depuis 20 ans, leur impact sur la vie des malades et de leur entourage reste très prégnant : il est d'usage de parler dans le champ du cancer de « rétablissement » et non de « guérison », avec plusieurs enjeux à la clé.

L'évaluation du poids de la maladie sur le long terme

L'impact des traitements et du cancer sur la qualité de vie dans toutes ses dimensions (physique, sociale, psychologique, économique) a fait l'objet de nombreuses études dans la dernière décennie, mais la plupart d'entre elles s'arrêtent à la fin des traitements. Or des questions à long terme se posent, soulevant des points méthodologiques :

- peur de la récurrence dans les mécanismes d'immunodépression ;
- impacts des troubles dépressifs non diagnostiqués et non traités ;
- conséquences physiques et cognitives, non prises en charge, des traitements ;
- non-adhésion à la surveillance post-traitement.

La question se pose également des représentations collectives associées à ces besoins d'accompagnement et de veille sur le long terme :

- place et légitimité des médecins généralistes ;
- besoin en formation des professions médicales ;
- types d'accompagnement, notamment pour le retour à l'emploi ;
- évaluation, par les malades, des dispositifs mis en place.

Le poids de la maladie pour les personnes seniors soulève des points spécifiques dans la mesure où elles constituent une population :

- sans accès aux essais cliniques, voire aux nouveaux médicaments ;
- avec une prise en charge très complexe du fait des comorbidités ;
- avec une exposition accrue à la perte d'autonomie. L'étude des mécanismes psychosociaux associés à cette vulnérabilité est cruciale pour une meilleure prise en charge.

La prise en compte du critère « qualité de vie » dans les essais cliniques n'est pas systématique, et les outils utilisés ne sont pas toujours adaptés. Dans le contexte particulier de l'innovation thérapeutique, des outils standardisés ne sont pas disponibles. De même, chez les sujets les plus âgés, qui sont aussi les plus nombreux, la question de la présentation des choix thérapeutiques en fonction des objectifs attendus de la personne malade – durée de survie *versus* qualité de vie – n'est pas encore standardisée.

L'accompagnement aux thérapies ciblées

Si l'accès aux thérapies ciblées est très encadré dans la trajectoire de soins des personnes malades et connaît des améliorations vers plus d'équité sociale et territoriale, le retentissement de ces nouvelles modalités thérapeutiques doit être encore documenté, notamment en ce qui concerne :

- l'observance des traitements, surtout lorsque la thérapie est orale (77 % des thérapies ciblées) ;
- la perception des immunothérapies, dont les effets secondaires, parfois graves, peuvent laisser penser aux malades, notamment celles et ceux ayant préalablement bénéficié d'un traitement en chimiothérapie, que le traitement n'est pas efficace.

Il est important de connaître les déterminants individuels et environnementaux ainsi que les besoins en accompagnement pour ces thérapies innovantes. En parallèle, il convient de développer des outils permettant d'évaluer la qualité de vie des malades subissant ces traitements.

Le cas particulier des cancers pédiatriques

Les conséquences à long terme, et particulièrement après la fin des traitements, restent importants pour les enfants et les adolescents ou adolescentes atteints de cancers. Cela affecte leur réintégration dans la vie courante sur les plans scolaire, affectif et social. Il existe aussi de nombreuses disparités socio-économiques et territoriales en matière de type et d'accès aux soins de support, malgré les bénéfices connus de ces dispositifs.

NOUVELLES MOLECULES BIOACTIVES ET REPOSITIONNEMENT D'ANCIENNES

Les cribles de collections de molécules chimiques ou biologiques effectués dans les laboratoires académiques, ou dans les sociétés pharmaceutiques, sont à l'origine de l'identification de molécules bioactives qui représentent une ressource considérable en nouveaux principes actifs pour le traitement des cancers. Obtenir des molécules chimiques capables d'interférer sur une cible biologique constitue la première preuve de concept que cette dernière est bien une cible thérapeutique pour de futures recherches de médicaments.

Une double finalité et deux types de cibles

Ces molécules bioactives constituent, en premier lieu, de nouveaux médicaments en clinique. Par ailleurs, dans le milieu académique, elles servent aussi des objectifs de recherche fondamentale et translationnelle, pour l'identification ou l'étude de nouvelles cibles thérapeutiques par des approches de génétique chimique, et pour la validation de stratégies thérapeutiques.

Les cibles biologiques peuvent être :

- des protéines purifiées ou des interfaces protéine-protéine (*target-based screen*) ;
- des modèles cellulaires plus ou moins complexes (*phenotypic-based screen*).

La connaissance du rôle d'une protéine, ou d'un mécanisme oncogénique, est rarement suffisante pour entreprendre des cribles à l'échelle industrielle sur cette cible. Or les biologistes ont peu conscience de l'importance qu'il y a à développer un programme de chimie médicinale à partir d'une famille de touches (*hits*), une molécule unique ne suffisant ni à développer un médicament, ni à intéresser un industriel.

Actions

- Favoriser l'utilisation de molécules bioactives pour la recherche fondamentale et la validation de cibles thérapeutiques
- Favoriser la mise au point de tests préindustriels en finançant cette phase dans les AAP académiques et en organisant des transferts industriels de ces tests à l'issue de cribles pilotes académiques
- Encourager la recherche de modèles pertinents pour évaluer le potentiel anti-oncogénique d'un candidat-médicament à des stades précoces

La nécessité de tests automatisés robustes et de modèles cellulaires adaptés

La mise au point de tests *statistiquement robustes* sur des cibles purifiées ou sur des modèles cellulaires pertinents est une étape cruciale pour la réalisation et le succès de campagnes de criblage à petite, moyenne ou grande échelle. Cet aspect constitue certainement un aspect méconnu – mais aussi l’une des forces majeures – du monde académique, qui possède un trésor d’expertises et de connaissances sur les mécanismes moléculaires et cellulaires du cancer. Ces connaissances permettent la mise au point de tests préindustriels innovants nécessaires à la mise en œuvre de cribles pilotes, ou à grande échelle, dans le monde académique ou industriel. Cependant, les phases de mise au point de tests automatisables (HT(C)S) à partir des essais de paillasse peuvent être longues et difficiles, et ne sont ni financées ni valorisées.

La mise au point de modèles multicellulaires adaptés au criblage ou au test de candidat-médicaments est indispensable. Que ce soit pour des criblages sur des lignées cellulaires, ou dans des études cliniques menées chez des personnes malades, les tests et essais cliniques s’effectuent le plus souvent, sinon exclusivement, sur des modèles de cancers avancés : de fait, les lignées cellulaires utilisées classiquement correspondent généralement à des stades de cancer avancés, les malades inclus dans les essais étant souvent en échec de premières lignes de traitements. Or ces approches ne permettent pas de révéler le potentiel anti-oncogénique d’un candidat-médicament qui pourrait se révéler efficace à des stades précoces, ce qui est pourtant d’une importance cruciale.

Ce défaut est peut-être en passe d’être contourné avec les modèles 3D *ex vivo* (cf. supra), mais reste encore plus vrai et difficile à éviter pour les essais cliniques. De façon générale, la pertinence des modèles cellulaires utilisés est déterminante pour éviter un taux d’attrition élevé et la poursuite inopportune vers le développement de médicaments inefficaces.

Le repositionnement de molécules pour de nouveaux traitements rapidement utilisables

Le repositionnement repose sur au moins deux stratégies :

- l’une reposant sur des observations cliniques mettant en évidence des cas de succès (stratification des personnes malades) ;
- l’autre appliquant une démarche rationnelle de nouveaux cribles sur des médicaments déjà très ciblés, mais avec encore peu de succès. Cette stratégie permet de bénéficier d’études de toxicité, pharmacocinétiques et cliniques déjà réalisées.

Deux types de méthodologies peuvent être utilisées :

- l’utilisation de nouveaux médicaments pour une nouvelle indication sans optimisation chimique ;
- l’optimisation sélective d’activités secondaires (SOSA). Cette dernière stratégie est souhaitable pour optimiser la sélectivité vers la nouvelle cible, mais peut rendre nécessaire une nouvelle réalisation de phases cliniques, avec le risque de perdre certains bénéfices du « repositionnement ».

L’approche reposant sur les données cliniques et de pharmacovigilance est difficile à mettre en œuvre car ces données ne sont pas toujours FAIR, et reposent sur des constats de personnels soignants qui font des signalements à l’ANSM. Les critères de stratification des bénéficiaires peuvent aussi être difficiles à anticiper. Enfin, les industriels

Actions

- Favoriser les cribles de repositionnement, seul ou en combinaison avec des molécules originales (nouvelles) pour détecter les synergies (léthalité synthétique) en vue de découvrir des mécanismes de redondance menant à l’identification de combinaisons thérapeutiques (molécule connue / molécule originale)
- Réfléchir avec les industriels et le législateur (État) sur les conditions juridiques de l’utilisation d’un médicament pour de nouvelles indications (rôle particulier de la puissance publique dans le repositionnement)

peuvent freiner ce type de développement qui produit peu, ou pas, de propriété intellectuelle, surtout si la molécule est déjà dans le domaine public.

Du point de vue fondamental, le repositionnement permet de « découvrir » de nouveaux mécanismes oncologiques à partir de l'effet anticancéreux de médicaments prescrits pour d'autres indications, et de mettre à jour des mécanismes de létalité synthétique insoupçonnés. Les risques d'effets secondaires peuvent être plus élevés qu'avec les autres molécules-médicaments puisque, par définition, ce sont des molécules multicibles.

La nécessaire optimisation des approches et des collections

Afin d'accélérer la découverte de nouveaux candidats-médicaments en cancérologie – en diminuant les coûts et le temps de criblage par la baisse de la quantité de molécules à cribler –, il faut améliorer les collections en utilisant des approches variées et complémentaires reposant sur la prédiction *in silico*, les cribles à haut débit non orientés et l'utilisation de molécules bioactives connues. Il peut être très profitable d'adapter l'espace chimique utilisé pour le problème posé. L'optimisation doit reposer sur les connaissances académiques des cibles biologiques, en particulier les données de biologie structurale, ou de bioactivité, pour identifier des candidats-médicaments et bénéficier de l'émergence des interfaces protéines-protéines comme nouvelles cibles thérapeutiques. Il faut veiller à ne pas perdre d'opportunités par la seule application de démarches avec *a priori*, et donc soutenir également des stratégies sans *a priori* sur l'espace chimique.

LE CHANGEMENT DE PARADIGME EN RADIOTHERAPIE

Les techniques de radiothérapie, externe et interne (dont la radiothérapie interne vectorisée) évoluent rapidement (p.ex., Flash, minifaisceaux, etc.) et offrent des options de plus en plus variées pour délivrer très précisément la dose prescrite à la tumeur, tout en épargnant au mieux les tissus sains. En complément, les techniques de radiothérapie sont de plus en plus combinées à d'autres modalités de traitement (dont l'immunothérapie). L'évolution des techniques d'irradiation et de leur usage pose de nouveaux défis : ainsi, par exemple, la délivrance de dose pouvant être spatialement très précise, il devient décisif de définir exactement la région à cibler. Dans la mesure où différentes modalités de traitement par radiothérapie existent, le choix de l'approche à privilégier pour chaque personne à traiter peut se révéler complexe. La toxicité des différents modes de traitement envers les tissus sains reste aussi à mieux caractériser.

Promesses de l'hadronthérapie

Les besoins cliniques ne peuvent être remplis par une seule modalité de radiothérapie, et les carbone-, hélium- ou protonthérapies peuvent être des solutions prometteuses pour augmenter l'efficacité du traitement des tumeurs radiorésistantes tout en limitant la toxicité des radiations. Les ions permettent une grande précision balistique grâce au « pic de Bragg », et divers facteurs peuvent être utilisés pour augmenter le dépôt de dose local :

- particules à fort transfert d'énergie linéique (ions plus lourds que les protons, produits de capture neutronique par le bore) ;

Actions

- Soutenir la production de collections intelligentes par de multiples approches (prédiction *in silico*, démarche sans *a priori*, à partir de substances bioactives connues)
- Rendre accessible les données de criblages expérimentaux en source ouverte (après levée d'embargo) pour l'apprentissage machine, dont apprentissage de signatures phénotypiques (cribles par imagerie) avec les critères FAIR
- Favoriser les interactions interdisciplinaires entre chimistes, biologistes, chimio-informaticiens, ainsi qu'entre expérimentalistes et modélisateurs

- éléments et facteurs radiosensibilisants ;
- fractionnement spatial par mini- ou microfaisceaux.

Par ailleurs, les effets associés au fractionnement temporel (hypofractionnement, effet Flash, etc.) sont également à appréhender pour l'amélioration du gain thérapeutique de l'hadronthérapie. Pour ce faire, il est nécessaire de comprendre et modéliser les mécanismes physiques, chimiques et biologiques de chaque type de rayonnement, et ce aux échelles spatiale et temporelle, afin d'interpréter et prédire l'efficacité d'un traitement et sa radiotoxicité éventuelle. D'un point de vue expérimental, plusieurs difficultés sont à prendre en compte :

- les irradiations *in vitro* ne permettent pas de répondre à l'ensemble des questions ;
- les modèles animaux et les données des personnes malades (dont les questions éthiques) n'offrent pas, ou peu, la possibilité de contrôler certains paramètres ;
- l'accès aux différentes plateformes d'irradiation (temps et coûts) peut être un frein aux avancées expérimentales ;
- le développement et le maintien des outils de calcul pertinents (GEANT4, GATE, GEANT4-DNA, etc.) nécessitent un soutien à long terme.

L'utilisation en routine de l'hadronthérapie requiert la maîtrise des incertitudes, le contrôle des conditions du vivant, l'asservissement du faisceau et la prédiction fiable des effets, afin :

- de caractériser les tissus qui vont être irradiés et de limiter le volume correspondant (multiénergie et *Ion Computer Tomography*, contrôle en ligne du parcours, etc.) ;
- d'utiliser la distalité (irradiation avec un organe à risque localisé dans l'axe du faisceau en aval du volume planifié) ;
- de réduire la pénombre latérale ;
- d'optimiser le fractionnement et viser si possible l'hypofractionnement, y compris par l'utilisation du mode Flash avec des hadrons ;
- d'aller vers l'hadronthérapie stéréotaxique (minifaisceaux) ;
- de traiter des tumeurs en évolution et/ou mouvement (inter- et infrafraction) ;
- d'utiliser des techniques combinées : incertitudes additionnelles de cumul de doses et d'effets ;
- de modéliser les effets locaux et systémiques à court et long terme des traitements radiothérapeutiques et médicamenteux combinés (chimio-, immuno-, hormonothérapies, etc.) en intégrant aussi les caractéristiques des malades, des tumeurs et du microenvironnement tumoral ;
- d'aller vers des modalités d'imagerie de guidage non ionisantes, non interférentielles sur le vivant.

Enfin, il est nécessaire d'accompagner le corps médical pour le choix de l'orientation thérapeutique et l'optimisation des traitements en termes de modélisation/prédiction des bénéfices

Actions

- Soutenir le développement de nouvelles plateformes/dispositifs d'irradiation permettant de répondre de façon spécifique aux questions en cours
- Encourager le développement des outils dosimétriques pour permettre une évaluation standardisée des résultats
- Promouvoir une utilisation optimisée des équipements (énergie minimale, accessoires, modalités d'imagerie embarquée)
- Promouvoir les collaborations interdisciplinaires entre physiciens, chimistes, oncologues et biologistes
- Contribuer à développer les infrastructures de recherche clinique dans les établissements de soin
- Promouvoir les recherches interdisciplinaires et translationnelles, en s'appuyant sur une animation scientifique nationale de la communauté (GDR MI2B, Outils et méthodes nucléaires pour la lutte contre le cancer)

thérapeutiques, de coût associé à la délivrance, de planification incluant le guidage par l'image, d'assurance qualité, d'organisation/gestion des données pertinentes et d'infrastructures de recherche clinique.

Exploiter le plein potentiel de l'hadronthérapie nécessite non seulement d'en maîtriser les effets biologiques et les incertitudes physiques – trop longtemps négligées pour les protons, mais aussi d'équiper les salles avec des systèmes de contrôle et d'asservissement (imagerie / suivi) au moins équivalents à ceux utilisés avec les techniques avancées de photons. Or, ce dernier point est très en retard en hadronthérapie. D'une façon générale, l'approche hadronthérapie est donc très fortement pluridisciplinaire.

La radiothérapie interne vectorisée

Également appelée radiothérapie moléculaire ou métabolique, la RIV consiste à administrer par voie interne, le plus souvent intraveineuse, un vecteur porteur d'un élément radioactif ciblant de façon la plus spécifique possible l'organe à traiter. C'est une méthode de thérapie utilisée avec succès depuis de très nombreuses années pour les cancers de la thyroïde, traités à l'iode 131 (^{131}I , émetteur β^-). Plus récemment, en 2013, le chlorure de radium a été autorisé par les autorités sanitaires pour le traitement palliatif des métastases osseuses de certains cancers de la prostate métastatiques. Ce procédé est rarement utilisé seul (sauf pour des tumeurs de très petite taille) à cause de la faible pénétration des rayonnements.

La RIV est en pleine expansion, mais nécessite l'identification de cibles biologiques précises et de vecteurs spécifiques adaptés. Il est également important de comprendre les mécanismes de radiobiologie à l'œuvre pour aller vers des traitements plus efficaces et moins toxiques. Plusieurs sociétés pharmaceutiques majeures sont aujourd'hui engagées dans le développement de ce type de produits (p. ex., Novartis, Bayer), qui montrent des effets spectaculaires dans la prise en charge de certains cancers. Depuis 2017, un radiopharmaceutique utilisant le Lutécium 177 (^{177}Lu) émetteur β^- , le Lutathera®, est disponible en routine clinique pour le traitement des tumeurs neuroendocrines. En mars 2022, la FDA a autorisé le PSMA-617-Lu177 pour le traitement des cancers de la prostate.

Les particules émettrices β^- ne sont pas les seules d'intérêt : de premières données cliniques montrent un fort potentiel des

Actions

- Soutenir les projets innovants à la croisée de l'imagerie et de la radiothérapie, et le développement de la radiothérapie interne vectorisée guidée par l'imagerie
- Encourager les équipes de radiobiologie à interagir avec les autres spécialités (imagerie, immunologie, etc.)
- Investir sur la valorisation des innovations en imagerie et en radiothérapie, et soutenir les sociétés impliquées dans ces développements
- Soutenir financièrement les sites de production de radionucléides innovants
- Favoriser les collaborations entre les équipes développant des vecteurs ou travaillant sur des cibles biologiques et celles travaillant en médecine nucléaire
- Encourager et soutenir des études dosimétriques dans les essais cliniques afin de permettre de démontrer l'apport de la dosimétrie
- Encourager le développement d'outils de simulation multi-échelle (espace, temps) et modèles biophysiques adaptés aux particules de haut TEL pour prédire la réponse thérapeutique en RIV
- Promouvoir les études translationnelles de désescalade thérapeutique (suivi de la réponse au traitement, évaluation des effets systémiques, etc.)
- S'appuyer sur des structures transversales et multi-organismes pour animer la communauté scientifique nationale
- Soutenir la filière industrielle française désireuse de s'investir dans l'usage de radionucléides à des fins thérapeutiques
- Simplifier les procédures réglementaires et accélérer les délais nécessaires aux obtentions des autorisations nécessaires aux travaux de recherche

traitements utilisant des émetteurs α (^{211}At , $^{212}\text{Pb}/^{212}\text{Bi}$, ^{225}Ac , ^{213}Bi , ^{227}Th , ^{223}Ra , ^{149}Tb) grâce au transfert d'énergie linéique (TEL) très élevé de ces particules. On peut imaginer également un impact important des émetteurs « Auger », qui sont aussi des particules de haut TEL (^{195}mPt , ^{193}mPt , ^{111}Ln , ^{103}Rh , ^{97}Ru)¹¹.

Il est indispensable que les effets systémiques de l'ensemble de ces radiations ionisantes à faibles doses soient évalués et modélisés à court et long terme (inflammation, effets sur les lignées sanguines, cancers secondaires, etc.). En parallèle, l'utilisation de la dosimétrie doit être accélérée dans les essais cliniques.

La théranostique : des liens de plus en plus étroits entre radiothérapie et imagerie

L'imagerie est largement employée pour caractériser l'efficacité d'un traitement et ses toxicités induites, ainsi que pour le suivi des personnes malades et la détection des récidives. Des systèmes d'imagerie fonctionnelle et moléculaire embarqués sur les équipements de radiothérapie (type Linac, *Linear Accelerator*) sont de plus en plus développés (IRM- et TEP-Linac), qui devraient permettre une meilleure intégration des nouvelles méthodologies d'imagerie au service de la radiothérapie.

Pourtant, les avancées récentes en radiothérapie et en imagerie se font surtout de manière parallèle, alors que la fertilisation croisée entre les deux disciplines servirait considérablement chacune d'elles. Ainsi, les équipes et laboratoires spécialisés en imagerie ne sont souvent pas les mêmes que ceux spécialisés en radiothérapie. De plus, le recueil de données nécessaires à la recherche croisant imagerie et radiothérapie est encore plus complexe que lorsque la recherche n'implique « que » l'imagerie ou « que » la radiothérapie, puisqu'elle nécessite de créer des bases de données incluant l'imagerie amont, les plans de traitement, l'imagerie embarquée le cas échéant, puis l'imagerie post-traitement. Les données issues des dernières innovations technologiques (traitement Flash, RIV) ou les équipements implémentant ces innovations (IRM-Linac, TEP-Linac) sont encore rares, et l'implémentation expérimentale de ces innovations est souvent complexe et sujette à une réglementation lourde. De fait, les innovations françaises dans le domaine de la radiothérapie rencontrent des difficultés dans leur phase de valorisation, qui passe généralement par des partenariats privé-public. Ainsi en est-il de la radiothérapie Flash qui, inventée en France, a été valorisée en Suisse.

Les propriétés d'émissions photoniques des radionucléides permettent de faire de l'imagerie fonctionnelle (TEP ; TEMP). Les imageries anatomique, fonctionnelle et moléculaire jouent un rôle crucial pour la planification des traitements en radiothérapie, puisqu'elles servent notamment à délimiter les zones à irradier. La médecine nucléaire, qui est à la fois en charge des examens d'imagerie TEMP et TEP et de l'administration de la RIV, offre donc de nombreuses possibilités d'approches théranostiques (développement coordonné de l'imagerie et de la thérapie), car un même ligand peut être utilisé pour l'imagerie et pour la RIV, sous réserve de modifier le radionucléide associé au ligand (émetteur γ ou de positons pour l'imagerie, émetteur β^- ou α pour la RIV). De nombreuses possibilités existent, combinant $^{44}\text{Sc}/^{47}\text{Sc}$, $^{64}\text{Cu}/^{67}\text{Cu}$, $^{124}\text{I}/^{131}\text{I}$, les isotopes du Tb, $^{68}\text{Ga}/^{177}\text{Lu}$ ¹¹, etc. Ces produits doivent répondre aux normes de bonnes pratiques de fabrication, et leur disponibilité peut être un obstacle important à leur utilisation.

D'une façon générale, l'approche théranostique permet notamment de :

- sélectionner les personnes malades qui répondront à une thérapie donnée, ce qui a pour avantages d'éviter les thérapies inutiles, leurs éventuels effets secondaires, une perte de temps pour la mise en œuvre d'une autre piste thérapeutique et des dépenses inutiles pour le système de santé ;

¹¹ At : Astate ; Pb : Plomb, Bi : Bismuth ; Ac : Actinium ; Th : Thorium ; Ra : Radium ; Tb : Terbium ; Pt : Platine ; Ln : Lanthanide ; Rh : Rhodium ; Ru : Ruthenium ; Sc : Scandium ; Cu : Cuivre ; Ga : Gallium ; Lu : Lutécium

- réaliser les calculs dosimétriques afin d'optimiser le traitement en anticipant l'impact sur les organes à risque et la réponse tumorale ;
- suivre l'évolution du traitement après son administration.

INNOVATIONS THERAPEUTIQUES NON MEDICAMENTEUSES

L'innovation thérapeutique est très souvent entendue comme médicamenteuse. Or le traitement des cancers fait traditionnellement appel à l'association d'approches de chirurgie, radiothérapie, chimiothérapie, médicaments ciblés et immunothérapie (le plus souvent par anticorps inhibiteurs de points contrôlés). Dans cet ensemble, l'innovation thérapeutique non médicamenteuse concerne notamment la chirurgie et sa robotisation, la radiothérapie et son couplage à l'analyse d'image, la radiologie interventionnelle, la thérapie de la douleur par laser, les thérapies et les instruments/applications développés pour guider la distribution des médicaments ou assurer le suivi clinicobiologique. Elle est moins développée et, surtout, moins valorisée que l'innovation médicamenteuse au sens large.

La validation de l'apport thérapeutique d'une innovation technologique n'obéit pas toujours aux mêmes règles que celle d'un médicament : plus que sur la validation du bénéfice apporté aux personnes malades, elle repose souvent sur la démonstration de l'absence de danger immédiat et sur la capacité des professionnels à s'approprier l'innovation, ce qui est problématique. Ainsi :

- les innovations chirurgicales font rarement l'objet d'une analyse randomisée les comparant à la technique de référence ;
- des traitements relativement rares et très coûteux, comme la greffe de cellules souches hématopoïétiques, n'obéissent que rarement aux règles de validation d'une innovation médicamenteuse classique ;
- plusieurs thérapies de soins de support (p. ex., activité physique adaptée dans la prévention de la sarcopénie, traitement laser antalgique de la mucite, etc.) sont, d'une manière générale, difficiles à valider selon les critères utilisés pour les soins médicamenteux.

En revanche :

- la validation des innovations en radiothérapie se rapproche de celle d'une innovation médicamenteuse ;
- la thérapie cellulaire se situe à la frontière entre non médicamenteux et médicamenteux, puisque les cellules CAR-T sont désormais assimilées à des médicaments.

Il existe par ailleurs une vaste innovation non thérapeutique, qui concerne les outils de dépistage, de diagnostic et de suivi, incluant les biomarqueurs de toute nature et les outils d'imagerie (à l'échelle moléculaire, cellulaire, tissulaire, de l'organisme) et d'analyse d'image. Du fait de leur couplage fréquent au développement médicamenteux, les biomarqueurs sont validés de façon plus rigoureuse.

Actions

- Soutenir les projets portant sur la validation du bénéfice clinique de thérapies non médicamenteuses et une validation rigoureuse des innovations technologiques thérapeutiques, diagnostiques ou de suivi
- Encourager la recherche en thérapie chirurgicale (robotisation, chirurgie non délabrante, réparation tissulaire, etc.)
- Soutenir la recherche en physique médicale (distribution et débit de dose, prise en compte des mouvements respiratoires, tissus sains environnants, etc.)
- Encourager la recherche sur les soins de support, incluant la prévention des séquelles de la maladie et du traitement

QUELS FREINS A L'INNOVATION THERAPEUTIQUE ?

L'activité de mise sur le marché de médicaments de biothérapies issues de filières de recherche et développement principalement françaises est aujourd'hui inexistante. Les perspectives pour les 5, et même 10, prochaines années ne laissent pas entrevoir d'amélioration. Ce constat peut d'ailleurs être fait pour l'ensemble de la filière française du médicament en oncologie, où les AMM obtenues au cours des 40 dernières années sont très rares, et plutôt sur des marchés de niche.

Le problème de la valorisation

Les portefeuilles de développement des quelques sociétés pharmaceutiques nationales comprennent des candidats-médicaments issus, pour l'essentiel, d'équipes de recherche d'autres pays. En ce qui concerne plus spécifiquement les biothérapies, la capacité des structures de recherche et développement françaises à amener un produit jusqu'au marché (et à son remboursement) reste à construire. À l'heure actuelle, par exemple, les thérapies CAR-T ayant une AMM – ou sur le point de l'obtenir – sont issues d'équipes académiques aux États-Unis ou en Chine et développées par Novartis, Gilead, Janssen ou Bristol-Myers-Squibb.

Une structuration de la transition entre preuve de concept et développement industriel a été initiée, mais un soutien plus efficace aux stades de maturation semble encore nécessaire. De fait, même si différentes structures de transfert de technologies françaises (SATT, Inserm Transfert, CNRS Innovation, ou propres à chaque Centre de lutte contre le cancer, etc.) sont maintenant bien établies pour permettre les stades de prématuration (p. ex., dépôt de brevet), leur dispersion ne permet pas d'avoir des apports conséquents en termes de maturation de projets. De plus, bien que des Centres labellisés de phase précoce (CLIP²) aient été créés en 2010 par l'INCa, il semble important de réaliser un bilan de l'origine des thérapies utilisées dans ces structures, afin de déterminer si leur vocation est l'utilisation de thérapies développées par de grands groupes internationaux, ou si une prise en compte d'un soutien particulier aux thérapies émergentes françaises est souhaitable. Il existe donc au moins deux « vallées de la mort » à traverser pour le transfert d'une découverte à une application : entre la prématuration et la création d'une start-up, et à l'entrée en phase clinique I, voire II.

Si l'analyse montre qu'il existe une abondance de financements publics pour de la prématuration (Inserm Transfert, Prémat CNRS, etc.) et pour la création de start-up et de spin-off (SATT, etc.), ces financements sont accordés avec trop peu de contraintes et d'évaluation, notamment faute d'expertise dans le domaine d'intérêt. Or ce saupoudrage conduit à la création de multiples start-up qui ne seront probablement jamais profitables.

La mise en place d'une commission nationale experte, commune à toutes les structures de valorisation, pourrait pallier ce manque d'expertise dans les domaines d'intérêt. Dans le même ordre d'idée, l'évaluation des unités par l'HCERES (ou les EPST) sur le critère des brevets les incite à déposer trop de brevets qui ne seront jamais transformés en licence. C'est donc un leurre aux conséquences néfastes : de très nombreuses découvertes ne conduiront à aucune valorisation marchande.

Actions

- Encourager une recherche d'amont non ciblée permettant l'émergence de preuves de concepts dans différents domaines des biothérapies
- Favoriser l'obtention de preuves de concepts *in vivo* utilisant des modèles adéquats
- Promouvoir l'émergence d'une structure de production de biothérapies issues de la recherche académique
- Poursuivre la structuration des organismes de transfert de technologies pour réduire le nombre d'interlocuteurs et simplifier le parcours
- Encourager la prise d'initiatives et de risques par les acteurs de la valorisation en favorisant les liens académiques/clinique/industrie avec des appels d'offre plus orientés
- Promouvoir l'investissement sur les projets présentant la plus forte expertise scientifique et industrielle en France.
- Dimensionner le financement de façon réaliste en fonction de l'enjeu

Les difficultés à l'entrée en phase I d'essai clinique sont d'un autre ordre. En France, environ 3 000 molécules sont en phases cliniques, ce qui pose clairement le problème du recrutement de cohortes valables. De plus, le coût de la mise sur le marché d'un médicament est d'un ordre de grandeur très différent de celui nécessaire pour obtenir une preuve de concept (environ 300 M€ pour une phase III, contre environ 15 M€ pour une phase I). Le développement d'un anticorps monoclonal jusqu'à l'obtention d'un premier lot clinique incluant la toxicologie coûte près de 5 M€, et les chiffres sont encore supérieurs pour les anticorps couplés à des toxines ou les anticorps multispécifiques. Dans ces conditions, le dimensionnement des financements actuels relève fréquemment du complément de recherche académique plutôt que de celui nécessaire aux premières étapes permettant un développement industriel. Un soutien efficace à l'étape de maturation nécessiterait donc un calibrage réaliste des financements alloués, adapté aux coûts des projets mais aussi au nombre de projets aboutis souhaités.

Enfin, la perspective d'accélérer les phases cliniques en cancérologie pour faire baisser les coûts n'est pas envisageable car, à la différence de ce qui s'est passé pour les vaccins anti-Covid, la durée de l'évaluation des effets des molécules anticancéreuses se compte en années. Le modèle des « oncoclusters » sur des sites identifiés est quant à lui séduisant, mais se heurte aux pratiques industrielles en cours : de fait, lorsqu'elles souhaitent développer une molécule en clinique sur une cohorte identifiée de malades, les industries choisissent la meilleure proposition permettant une étude multicentrique (4-6 centres), rapide et impliquant donc des centres gros recruteurs.

Le manque de cliniciennes et cliniciens scientifiques

Le transfert des avancées de la recherche dans la routine clinique repose sur le trio médecin/scientifique/malade. Pour répondre aux nouveaux enjeux de la recherche en cancérologie (médecine de précision, approches expérimentales pluridisciplinaires et combinées, exploitation et méta-analyse de données, intelligence artificielle, etc.), l'analyse des échantillons issus de personnes atteintes de cancer par des équipes de recherche et le transfert des résultats de la recherche cognitive vers les malades constituent un pilier de l'innovation thérapeutique. Ce pilier implique notamment un dialogue de qualité qui concerne tous les pans de la cancérologie : pratique clinique, recherche clinique et autres formes de recherche – biologie translationnelle et cognitive, mais aussi sciences humaines et sociales, biophysique, biostatistiques, etc.

Or le nombre de médecins (et de pharmaciennes et pharmaciens) participant activement à la recherche non clinique reste modeste. La filière est relativement peu attractive, et la perspective d'une carrière hospitalo-universitaire est trop souvent la première – et parfois la seule – motivation pour suivre une filière MD-PhD ou PharmaD-PhD. Assurer le renouvellement des générations de médecins-chercheurs est pourtant un enjeu majeur en cancérologie : il en va du maintien, dans les départements cliniques, d'une expérience des exigences, contraintes et apports de la recherche

Actions

- Repenser la filière de médecine-scientifique (y compris dans la nature des enseignements dispensés), indépendamment du statut d'enseignant ou enseignante (MCU-PH, PU-PH)
- S'assurer de l'existence de financements adaptés pour les médecins souhaitant se former à la recherche à différents niveaux (MD, PhD, post-doctorat)
- Aider au maintien d'une activité de recherche protégée des MD-PhD recrutés dans les hôpitaux en compensant le temps médical dans leur service d'origine
- Proposer des pédagogies et des formations innovantes : classes virtuelles, écoles d'été, immersions (dans les laboratoires pour les médecins, dans les services cliniques pour les scientifiques)
- Promouvoir la formation continue aux nouvelles technologies, à l'interdisciplinarité et à l'intégrité scientifique

translationnelle et cognitive, afin de prévenir le développement d'une recherche clinique découplée de la recherche translationnelle et cognitive académique, et entièrement au service de l'industrie du médicament.

Il est par ailleurs indispensable de favoriser le dialogue entre médecins et scientifiques pour informer ces derniers des contraintes de la prise en charge clinique et thérapeutique des personnes malades et, ainsi, disposer dans les laboratoires de recherche de scientifiques ayant une connaissance approfondie des maladies humaines, des pratiques médicales et de leurs limites. Amener des personnes malades, voire non malades, au cœur de la recherche cognitive participe aussi de ce processus de transversalité visant à accélérer le transfert au bénéfice des patientes et des patients.

Il est donc nécessaire d'adapter l'offre de formation des médecins à la recherche fondamentale ou translationnelle :

- en tenant compte de leur charge clinique, qui ne cesse d'augmenter, de la fragilité de la démographie médicale, du décalage salarial important entre les postes de médecin et de scientifique ;
- en les sensibilisant très précocement dans leur cursus médical à l'importance de la recherche scientifique ;
- en présentant les enjeux de la science ouverte, de l'éthique et de l'intégrité scientifique ;
- en mettant en place des mécanismes de compensation de « temps médecin » indépendamment de l'enseignement universitaire (création de filière *ad hoc* « médecin-scientifique ») ;
- en favorisant les interactions et les cursus communs entre scientifiques et médecins, afin de familiariser les scientifiques à la pratique clinique et implanter la démarche scientifique dans les études médicales.

Le besoin d'encadrer l'évaluation des modèles numériques

De nombreux travaux ont pour objectif de développer et valider des outils numériques de médecine personnalisée. Des outils numériques sont déjà utilisés pour adapter les posologies en fonction du suivi thérapeutique pharmacologique. Pour aboutir à une utilisation clinique, les outils numériques doivent démontrer le bénéfice qu'ils apportent aux patientes et patients : il est donc indispensable de poser un encadrement juridique à leur évaluation.

La législation européenne définit la notion de dispositif médical (DM) comme « *tout instrument, appareil, équipement, matière, produit, à l'exception des produits d'origine humaine, ou autre article utilisé seul ou en*

Actions

- Encourager les essais cliniques prospectifs, randomisés et multicentriques pour évaluer le bénéfice d'un outil numérique donné : un bras contrôle où la décision thérapeutique est prise selon le standard de traitement et le choix de l'oncologue, un bras où la décision thérapeutique est prise en utilisant l'outil numérique
- Sensibiliser la communauté à la nécessité d'établir la non-nocivité de l'utilisation d'un outil numérique et leur validation de manière prospective
- Créer un registre public des outils numériques entrés dans le processus d'évaluation et d'homologation permettant de connaître à tout moment les informations sur sa conception, son implémentation, la vérification de son code, l'état du processus d'évaluation, le nombre de malades sur lesquels un outil a été évalué, le type d'évaluation en cours, le type d'homologation (CE, remboursement, etc.), les données de vie réelle, les études prospectives, etc.
- Adapter les processus d'homologation aux outils numériques évoluant par apprentissage continu
- Instaurer des méthodes de contrôle qualité périodique (à fréquence élevée) afin de détecter précocement les effets de dérive susceptibles de dégrader significativement les performances de l'algorithme
- Soutenir les recherches sur la caractérisation longitudinale des performances des outils numériques

association, y compris les accessoires et logiciels nécessaires au bon fonctionnement de celui-ci, destiné par le fabricant à être utilisé chez l'homme à des fins médicales et dont l'action principale voulue n'est pas obtenue par des moyens pharmacologiques ou immunologiques ni par métabolisme, mais dont la fonction peut être assistée par de tels moyens. Constitue également un dispositif médical le logiciel destiné par le fabricant à être utilisé spécifiquement à des fins diagnostiques ou thérapeutiques »¹². Cette notion inclut donc l'utilisation de logiciel. De son côté, le Parlement européen définit un « système d'IA [Intelligence artificielle] » comme « un système qui est soit fondé sur des logiciels, soit intégré dans des dispositifs matériels, et qui affiche un comportement simulant l'intelligence, notamment en collectant et traitant des données, en analysant et en interprétant son environnement et en agissant, avec un certain degré d'autonomie, pour atteindre des objectifs spécifiques ». Par ailleurs, il existe un marquage « CE » établi par un organisme notifié – en France, le Groupement pour l'évaluation des dispositifs médicaux (GMED) – pour garantir une conformité en matière de sécurité et de performances. À l'issue du marquage CE, le DM peut être mis sur le marché.

Concernant le possible remboursement des DM en France, c'est la commission nationale d'évaluation des dispositifs médicaux et des technologies de santé (CNEDiMTS) de la HAS qui en évalue la pertinence. Cette instance a récemment étendu son champ aux DM embarquant de l'IA en instituant que « tout dispositif médical (DM), dispositif médical implantable actif (DMIA) ou dispositif médical de diagnostic in vitro (DMDIV) dont la composante numérique représente la part prépondérante de l'actif, ou la part prépondérante de la proposition de valeur » est un dispositif médical numérique. Elle a aussi établi une classification des DM numériques, dont les catégories suivantes sont pertinentes pour la cancérologie clinique : C5 = aide au dépistage sans autonomie de la solution numérique ; C6 = aide au diagnostic sans autonomie de la solution numérique ; C7 = aide au traitement sans autonomie de la solution numérique ; C8 = aide à la décision thérapeutique sans autonomie de la solution numérique ; D1 = système qui utilise des données de la personne malade afin de diagnostiquer une situation pathologique et, ensuite, de modifier automatiquement, sans intervention humaine préalable, les paramètres d'un traitement.

L'ensemble de ces notions pourraient être utilisées pour établir le cadre réglementaire permettant d'évaluer le bénéfice et le remboursement du recours aux technologies numériques dans le diagnostic ou le traitement des maladies cancéreuses. Cependant, la classification fonctionnelle actuelle des DM numériques n'est pas entièrement adaptée à des outils d'analyse de données pour l'aide à la décision en cancérologie. De plus, les algorithmes reposant sur l'apprentissage peuvent généralement être raffinés et voir leurs performances améliorées à mesure qu'ils analysent davantage de cas (notion d'apprentissage continu) : cette spécificité complexifie les principes d'homologation. À l'inverse, un algorithme fondé sur l'apprentissage peut voir ses performances se dégrader progressivement si les données qu'il analyse connaissent une dérive dans le temps par rapport aux données utilisées pour l'apprentissage de l'algorithme (p. ex., rajeunissement d'une population de malades consécutive à l'évolution d'un bassin de population).

De l'utilisation du numérique dans la recherche sur les cancers

Il est désormais acté que la compréhension des cancers et leur traitement nécessitent une approche intégrant des informations de nature variée et hétérogène (moléculaires, environnementales, antécédents, etc.), et acquises à différentes échelles spatiales (de la génomique jusqu'à l'imagerie corps entier) et temporelles (prise en compte de données acquises tout au long du parcours du malade, ou même de la vie d'une personne). Des modèles permettant de prendre en compte l'ensemble des informations potentiellement pertinentes, toutes complexes, et d'identifier les informations utiles sont nécessaires pour identifier les rôles respectifs de différents facteurs (par

¹² [Code de la santé publique \(article L.5211-1\)](#)

exemple à des fins de prévention), définir les éléments à mesurer pour optimiser la prise en charge des personnes malades et pouvoir prédire l'évolution de la maladie, afin de déterminer le parcours de soin de plus adapté.

L'EXPLOSION DU RECOURS A L'INTELLIGENCE ARTIFICIELLE ET AUX SCIENCES DES DONNEES

Le développement de modèles d'IA, en particulier de modèles d'apprentissage automatique, ainsi que le recours aux sciences des données pour la cancérologie clinique ont pris beaucoup d'ampleur dans la dernière décennie. Des succès importants dans l'aide à la détection (diagnostic) et la caractérisation (classification de la maladie et de son extension) des cancers ont été rapportés.

L'IA est définie comme « *tout algorithme statistique dont les paramètres optimaux ont été entraînés à partir d'un jeu de données représentatif de la population cible* ». Elle a désormais un rôle omniprésent dans toute la biologie expérimentale et, dans une moindre mesure, dans la routine clinique en oncologie. De plus en plus de modèles mathématiques et d'intelligence artificielle sont développés dans le domaine de la cancérologie, dont la vocation est cognitive (compréhension des mécanismes biologiques, identification des groupes d'observations présentant des similarités, etc.) ou prédictive (de l'évolution d'une maladie, de la réponse d'une tumeur à un traitement, pour l'optimisation d'une posologie afin de maximiser l'efficacité thérapeutique, etc.). L'IA est devenue un outil incontournable pour l'analyse, l'intégration et l'inférence de données, qu'elles soient moléculaires, de séquençage à haut débit (cellule unique et transcriptomique spatiale), de cytométrie ou d'imagerie.

Les performances de l'IA, restées modestes jusque récemment, ont progressé notamment grâce à :

- des avancées théoriques remarquables, avec l'invention des réseaux de neurones convolutifs et de l'apprentissage profond ;
- une augmentation massive des données numériques disponibles ;
- une augmentation continue de la puissance de calcul des ordinateurs.

L'IA a le potentiel d'accélérer des processus (p. ex., segmentation d'images) et de classer les personnes malades, conduisant ainsi à une amélioration de la détection, du diagnostic et de la compréhension de la maladie, voire à une meilleure anticipation de la réponse aux traitements, devenant alors un véritable outil d'aide à la décision thérapeutique. L'IA permet également de suggérer de nouvelles hypothèses biologiques, qui doivent être validées ultérieurement par l'expérience.

Pour autant, l'engouement pour l'IA et les sciences des données ne doit pas masquer le fait que :

- l'IA et les méthodes d'apprentissage ne sont pas indispensables à tous les projets de biologie et de médecine ;
- la mise en œuvre de méthodes d'IA nécessite des conditions particulières (notamment des données en quantité suffisante et correctement annotées, cf. infra) qui ne sont pas toujours remplies ;
- des analyses statistiques sans apprentissage (analyse en composantes principales ou en composantes indépendantes, classification hiérarchique, etc.) ont fait leurs preuves pendant des années et restent extrêmement pertinentes dans de nombreuses situations ;
- l'utilisation des termes « IA » et « sciences des données » n'est en aucun cas garante de la qualité d'une étude ou d'un projet.

Bien appréhender la terminologie

Du fait de l'explosion soudaine et exponentielle du recours aux techniques numériques, les termes « IA » et « sciences des données » ne sont pas toujours utilisés à bon escient, les frontières entre sciences des données, apprentissage statistique, apprentissage machine et intelligence artificielle n'étant pas toujours bien appréhendées. Même si certaines terminologies se recouvrent

partiellement (p. ex., l'apprentissage automatique est une technique d'intelligence artificielle), ces termes sont souvent considérés, à tort, comme interchangeables.

L'intelligence artificielle est une sous-discipline de l'informatique qui a émergé dans les années 50, avec l'apparition des premiers ordinateurs. En utilisant des algorithmes, l'IA effectue des tâches « intelligentes/cognitives » requérant normalement l'intelligence humaine, telles que la vision (des données d'imagerie, par exemple, mais pas des données omiques), la reconnaissance de la parole, la prise de décision ou la traduction (p.ex., extraction de rapports médicaux).

L'apprentissage machine (*machine learning*) est un sous-domaine de l'IA, à l'intersection des mathématiques statistiques et de l'informatique. L'objectif est d'établir un algorithme prédictif amélioré, pour une tâche donnée, à partir de l'expérience issue de l'analyse d'exemples. Le modèle sous-jacent n'est pas nécessairement bien caractérisé mathématiquement, mais plutôt de manière algorithmique à partir des seules données. À partir d'un jeu de données dans lequel entrées et sortie sont connues, l'algorithme cherche à prévoir une variable de sortie : il s'agit donc de prédiction.

L'apprentissage statistique a pour objectif de découvrir des relations entre les variables considérées, afin de comprendre les mécanismes sous-jacents. Il repose sur un cadre mathématique spécifié et quantifie souvent l'incertitude sur les paramètres du modèle, qui sont estimés à partir des données : il s'agit donc d'inférence.

L'apprentissage supervisé a pour objectif de déterminer la relation entre une variable de sortie et des variables d'entrée, à partir d'un jeu de données dans lequel entrées et sortie sont connues (apprentissage statistique). Il se décline en deux sous-branches : la régression (la variable de sortie est continue) et la classification (la variable de sortie est catégorielle, donc discrète).

L'apprentissage non supervisé a pour objectif de découvrir des similitudes entre des données d'entrée en l'absence de donnée de sortie (p. ex, regrouper des malades partageant les mêmes profils dans les données d'entrée).

Les « sciences des données » regroupent des méthodes issues des statistiques, de l'analyse de données et de l'informatique pour analyser des phénomènes, les comprendre et effectuer des prédictions à partir de données.

Actions

- Mettre en place des formations pour la communauté scientifique en biologie et en clinique pour la compréhension et l'utilisation correcte des termes
- Faire évaluer les projets incluant ces termes et leurs résultats par des spécialistes d'IA et de sciences des données, qui sauront déterminer la pertinence méthodologique des propositions et distinguer ce qui est du domaine du verbiage
- Instaurer des recommandations précisant les définitions des différents termes
- Promouvoir la création de départements de sciences des données au sein des centres de lutte contre le cancer

Au cœur du numérique : les données

Au cours de deux dernières décennies, les informations collectées chez les personnes atteintes de cancer et dans les modèles expérimentaux ont considérablement évolué. À côté des traditionnelles données, principalement qualitatives, permettant de caractériser la maladie et d'évaluer l'efficacité des traitements utilisés, des données quantitatives massives sont produites par l'analyse des images (cellules, tissus, organes, individus) et des prélèvements biologiques (analyses omiques multidimensionnelles des tissus malades et des tissus sains). L'imagerie et la biologie sont devenues quantitatives, du génotype au phénotype. La compréhension de la physiologie et de la physiopathologie est devenue accessible à la modélisation des systèmes complexes. Cette évolution produit une « biologie digitale » dont l'impact attendu sur les pratiques médicales est considérable. Cette évolution s'est accompagnée de nouveaux besoins d'interdisciplinarité, d'innovation et d'intégration (cf. infra).

Les données collectées sont de nature, d'origine et de destination différentes :

- une part très importante des données réunies actuellement est constituée de données ponctuelles collectées au diagnostic, voire en cours d'évolution d'une maladie constituée, et parfois traitée, mais les données longitudinales sont encore trop rares ;
- les collections de données constituées par les épidémiologistes dans le cadre de cohortes et de registres en population ne cessent de croître en volume et en précision ;
- les collections de données en vie réelle sur des populations larges et non sélectionnées apparaissent de plus en plus pertinentes pour évaluer les pratiques de soin, leur efficacité à long terme et les effets secondaires des traitements. L'ouverture des accès aux bases médicoadministratives de l'assurance maladie et des séjours hospitaliers (SNDS¹³) contribue aux capacités de suivi longitudinal de personnes à grande échelle.

Les enjeux de la gestion de ces données sont particulièrement nombreux en termes de stockage, partage et confidentialité :

- la quantité d'information collectée augmente exponentiellement au cours du temps, de même que le nombre de cas nécessaires pour accroître l'efficacité des analyses réalisées ;
- la production des données, leur contrôle qualité, leur maintenance et leur enrichissement, de même que les multiples analyses qu'elles engendrent, nécessitent des savoir-faire spécifiques et évolutifs ;
- il est indispensable de structurer ces données et de les partager au sein de la communauté scientifique, tout en respectant leur confidentialité. Cependant, les données de sources multiples et à des temps répétés sont nécessairement nominatives. Elles sont donc soumises à des règles précises de confidentialité et de sécurité. Le respect de ces règles (RGPD et compléments) nécessite des supports informatiques et juridiques rarement disponibles au niveau d'un laboratoire, ou même d'un centre de recherche. Une harmonisation de l'interprétation des règlements est indispensable ;
- les données doivent subir des contrôles qualité, être interopérables et standardisées pour être accessibles et utilisables (cf. infra).

Adéquation, conditions et coût de l'IA

L'IA représente un arsenal de méthodes – certes très puissantes – au même titre que des tests statistiques pour vérifier des hypothèses, ou des supercalculateurs pour effectuer des calculs longs et

Actions

- Faciliter l'accès aux données des personnes malades et leur standardisation
- Encourager le développement de méthodes d'estimation capables de gérer les données manquantes et les données hétérogènes
- Promouvoir la formation de la jeune génération à l'analyse bioinformatique et biostatistique des données massives, à leur intégration (dimensions, temps, espace) et à la modélisation des processus biologiques
- Soutenir la formation continue aux nouvelles ressources et techniques
- Soutenir l'investissement dans les ressources technologiques permettant la production, l'analyse, le stockage et le partage des données massives
- Promouvoir l'adaptation des infrastructures aux règles FAIR
- Favoriser l'organisation, la standardisation, l'interopérabilité, le financement, la pérennité et l'indépendance des bases de données partagées, nationales ou internationales
- Mettre en place les ressources juridiques permettant d'encadrer les partages de données en accord avec les règles (notamment RGPD)
- Encourager et structurer la collecte de données longitudinales ordonnées et structurées à différentes échelles de temps sur le long terme

¹³ Le Système national des données de santé comporte notamment les données de l'Assurance maladie (base SNIIRAM, les données des hôpitaux (base PMSI), les causes médicales de décès (base du CépiDC de l'Inserm) et les données relatives au handicap (en provenance des MDPH).

complexes. L'IA est trop souvent perçue comme « la » solution à un problème non encore résolu, sans que les conditions de mise en œuvre d'approches relevant de l'IA soient même examinées. Or l'IA n'a pas l'exclusivité de l'innovation : parfois, celle-ci vient de l'adéquation d'une méthode non innovante à une question pertinente restée jusque-là sans réponse, souvent à cause d'un manque de données pour réaliser l'étude.

Le choix du modèle adéquat doit dépendre du problème à résoudre, des connaissances *a priori* sur le problème et sa solution, et des données (nature, quantité) disponibles pour optimiser le modèle. Il existe de très nombreux modèles d'IA, dont les variations sont quasiment infinies, et de nouveaux modèles sont perpétuellement proposés. Ainsi, toutes les méthodes et modèles d'IA n'ont pas les mêmes exigences et ne produisent pas les résultats sous la même forme : certains modèles reposent sur peu de paramètres, tandis que d'autres comptent des millions de paramètres ; certains permettent de prendre en compte des connaissances existantes, tandis que d'autres sont construits uniquement à partir des données d'entraînement ; certains fournissent un résultat unique, tandis que d'autres associent une incertitude au résultat fourni. Il est impossible de tester toutes les options. De plus, actuellement, beaucoup de choix relatifs à la mise au point d'un modèle reposant sur l'IA sont empiriques.

De fait, les méthodes d'IA exigent des quantités de données substantielles, de qualité suffisante, bien annotées et documentées. L'absence de telles données rend bien souvent inopérant le recours à des modèles d'IA. De plus, il existe un très grand nombre de méthodes d'IA (non supervisées, supervisées, faiblement supervisées), et de très nombreuses architectures de modèles d'IA destinées à résoudre différents types de problèmes. Chacune de ces approches a des propriétés particulières, des avantages et des inconvénients. Il est actuellement encore complexe de déterminer, *a priori*, quel modèle est susceptible de répondre au mieux à une question donnée. Si les algorithmes d'IA classiques (UNet, GAN, etc.) sont accessibles aisément, ils nécessitent un niveau d'expertise suffisant pour être correctement sélectionnés, configurés et exploités à bon escient. Or la configuration d'un algorithme d'IA implique encore beaucoup d'approches empiriques et exige de très nombreux tests et calculs. Enfin, il existe beaucoup de questions méthodologiques en IA pour lesquelles il n'y a pas encore de réponse (p. ex., meilleure manière de prendre en compte des données de nature très hétérogène, meilleure façon de gérer des données manquantes, etc.), et des problèmes aujourd'hui non adaptés à l'IA pourraient l'être demain, notamment du fait de la rapidité d'évolution du domaine.

Actions

- Ne pas délaissé des projets ne reposant pas sur l'IA, mais proposant des méthodes plus classiques ayant des fondements mathématiques ou statistiques particulièrement pertinents pour les questions posées, au motif que ces méthodes ne sont pas innovantes
- Soutenir les projets pour lesquels le recours à l'IA est justifié par la nature des questions abordées, qui remplissent les conditions d'applicabilité de l'IA, qui comparent avec les résultats obtenus sans faire appel à l'IA, et qui discutent les avantages et inconvénients des différentes solutions

Actions

- Former les biologistes et les médecins au potentiel, aux prérequis et aux limites de l'intelligence artificielle
- Démultiplier le nombre de spécialistes des données dans les équipes de recherche en biologie/médecine. Ces collègues sont les plus à même d'indiquer si un problème est bien adapté à une solution par IA, ou à proposer des solutions alternatives potentiellement plus pertinentes
- Favoriser le recours à une IA durable, avec partage des codes sources, et partage des poids impliqués dans les modèles pour encourager le transfert d'apprentissage plutôt que de réentraîner *ab initio*
- Soutenir les développements conduisant à des modèles réutilisables
- Former et communiquer sur l'empreinte carbone de l'IA

L'IA a un coût important en termes :

- de données nécessaires pour mettre au point les modèles ou les appliquer ensuite ;
- de ressources humaines sollicitées pour préparer les données, et les annoter le cas échéant ;
- de ressources calculatoires nécessaires pour identifier les paramètres des modèles et évaluer leurs performances.

Il n'existe pas de méthodes simples pour mesurer l'empreinte carbone de développements algorithmiques (c'est un point encore peu discuté), même si des initiatives dans ce sens voient le jour. Cependant, il est certain que l'empreinte de l'IA est lourde : le numérique est responsable aujourd'hui de 4 % des émissions mondiales de gaz à effet de serre, et la forte augmentation des usages laisse présager un doublement de cette empreinte carbone d'ici 2025. L'entraînement d'un modèle peut produire environ 282 tonnes d'équivalent CO₂, soit cinq fois plus qu'une voiture américaine durant tout son cycle de vie. La pollution numérique ne peut donc pas être ignorée, ce qui justifie de privilégier le recours à des méthodes alternatives à l'IA, plus économes en ressources, quand elles sont aussi compétitives.

L'IA explicable, car IA ne rime pas seulement avec boîte noire

Parmi les méthodes relevant de l'intelligence artificielle, certaines produisent des modèles dits explicables, tandis que d'autres fonctionnent davantage comme des « boîtes noires », c'est-à-dire ne permettent pas de comprendre sur quelles informations précises le modèle s'est appuyé pour fournir un résultat. Certaines tâches confiées à des intelligences artificielles (p.ex., segmentation d'une structure dans une image) ne nécessitent pas obligatoirement de faire appel à des modèles explicables. En revanche, pour d'autres tâches (p. ex., prédiction de la réponse à un traitement, ou détermination du schéma optimal de délivrance d'un médicament), l'explicabilité des modèles est un véritable atout, car elle peut permettre :

- de vérifier le résultat produit par l'algorithme ;
- de prédire les cas où l'algorithme ne pourra pas donner un résultat fiable et où il ne faut pas l'utiliser ;
- de comprendre de nouveaux mécanismes révélés par les algorithmes.

L'intelligence artificielle explicable (XIA) est un ensemble de processus et de méthodes permettant aux utilisatrices et utilisateurs humains de comprendre les résultats et les conclusions créés par les algorithmes d'intelligence artificielle, notamment ceux faisant appel à l'apprentissage automatique. Dans le domaine de la santé, l'XIA a de nombreux atouts, puisqu'elle permet :

- de décrire un modèle d'IA de façon intelligible aux médecins, biologistes, personnes atteintes de cancers ;

Actions

- Favoriser dans les AAP le développement de modèles d'XIA pour les applications de cancérologie
- Encourager la fourniture de preuves des explications associées aux modèles d'IA explicables
- Évaluer les modèles d'IA sur la base de leurs résultats (avec une incertitude associée au résultat fourni), de l'explicabilité des modèles ainsi que la validité des explications fournies (explication vérifiable)
- Former les biologistes et les médecins aux grands principes de l'intelligence artificielle explicable, pour encourager l'utilisation de cette approche dans le domaine de la cancérologie et leur donner les moyens d'analyser de manière critique les résultats produits par les spécialistes des données
- Soutenir les équipes alliant spécialistes des données et biologistes/médecins
- Encourager les mobilités de spécialistes des données au sein d'équipes de biologistes/médecins et inversement, pour développer une culture commune nécessaire à l'élaboration de modèles d'IA explicables en biologie-santé

- d'anticiper les situations où le modèle peut être en échec et où il ne faut pas lui faire confiance ;
- de donner aux personnes utilisatrices les moyens de vérifier les conclusions de l'IA si nécessaire ;
- de fiabiliser le modèle en le rendant moins sensible aux fluctuations susceptibles d'affecter les données exploitées par le modèle ;
- de dégager de nouvelles hypothèses biologiques.

Cette approche n'est pas la norme pour l'instant, notamment lorsque les algorithmes reposent sur des méthodes d'apprentissage profond, et parce qu'il n'existe pas une seule manière de produire des modèles d'XIA, mais différentes approches favorisant l'explicabilité des modèles d'IA.

En réalité, le domaine de l'XIA est lui-même un domaine de recherche en pleine évolution. La notion d'explicabilité inclut :

- la compréhension des données sur lesquelles agissent les modèles (nature, distribution, caractéristiques extraites des échantillons), ce qui requiert une forte implication des spécialistes qui les produisent (médecins, biologistes) ;
- la compréhension des modèles, qui est du ressort des spécialistes des données. Elle peut être locale – être capable d'expliquer la décision du modèle pour un cas particulier – ou globale – expliquer le fonctionnement du modèle « en moyenne ». L'explicabilité peut être intrinsèque – construction de modèles directement interprétables par leur architecture – ou *a posteriori* – décodage de l'information utilisée par le modèle après obtention des résultats. Les approches à mettre en œuvre selon ces cas sont différentes.

Ainsi, le travail d'explicabilité ne peut pas être intégralement assuré par les spécialistes de l'IA et des données, il doit être pluridisciplinaire : un dialogue permanent entre spécialistes des données et spécialistes en biologie-santé est nécessaire, pour une compréhension profonde de la nature des données (afin d'interpréter correctement les explications suggérées par les modèles) et des algorithmes (pour comprendre quelles sont les caractéristiques globales et individuelles des données exploitées et comment elles le sont). L'XIA est un enjeu crucial qui contribue fortement à l'adhésion des personnes ayant recours aux algorithmes exploitant l'IA, et augmente la confiance dans les modèles.

L'indispensable dialogue entre expérimentation et résultats des sciences numériques

La biologie du cancer (omiques, cytométrie, imagerie, etc.) et la recherche clinique (radiologie, histopathologie, biomarqueurs, omiques, immuno- et *vasculomonitoring*, etc.) produisent des quantités de données de plus en plus importantes, qu'il est nécessaire d'analyser afin d'acquérir de nouvelles connaissances. Les techniques de production de données massives et complexes (cellule unique, transcriptomique spatiale, etc.) sont en constante et rapide évolution. Le besoin de spécialistes de la mise en œuvre de ces méthodes et de l'interprétation de leurs résultats est donc croissant.

La modélisation de ces données, surtout si elle est faite au moyen de modèles mécanistes ou d'inférence causale, peut permettre de produire des hypothèses biologiques, mais un modèle ne peut remplacer l'expérimentation et, *in fine*, ces hypothèses nécessitent d'être testées expérimentalement. Ainsi, bien que des modèles purement numériques soient développés, il sera toujours indispensable de les confronter à la réalité expérimentale.

Or la publication des travaux de modélisation n'exige bien souvent pas l'étape de validation expérimentale des hypothèses issues des modèles. Les codes de bonne pratique de la méthode scientifique, consistant notamment à dire qu'un modèle doit être falsifiable (K. Popper) pour être qualifié de scientifique, ne sont pas la norme. Les hypothèses restent souvent à l'état de spéculation et sont publiées par de grands journaux généralistes, qui se font ainsi le vecteur d'hypothèses spéculatives présentées comme des faits avérés, alors qu'elles peuvent, à terme, se révéler erronées. Ce phénomène est renforcé par l'isolement les uns des autres des différents acteurs : les spécialistes de la modélisation et de l'IA dans des laboratoires d'informatique et de mathématiques, les

biologistes à même de réaliser les expériences de validation de modèle dans leurs propres laboratoires. En outre, les ressources et les compétences nécessaires pour la modélisation et pour l'expérimentation sont totalement différentes. L'ensemble complexifie la possibilité, pour les spécialistes de la modélisation, de conduire ou, au moins, de mettre en place des expériences pour tester les prédictions de leurs modèles. Enfin, la validation expérimentale des hypothèses produites par les modèles mathématiques est souvent complexe, devant elle-même souvent faire appel à d'autres modèles (p. ex., modèles animaux, tumeur sur puces, malades très rigoureusement sélectionnés, prélèvements invasifs).

D'un point de vue clinique, la grande majorité des études se font dans un cadre rétrospectif, et beaucoup ne sont pas validées sur des cohortes externes, ou le sont avec des effectifs trop faibles pour avoir une puissance statistique suffisante. Il existe peu d'études prospectives permettant d'établir le bénéfice réel des modélisations. De nombreux modèles mathématiques existent, mais peu sont validés sur des données cliniques, certains étant encore trop complexes pour pouvoir l'être avec les données actuelles. De fait, jusqu'à présent, les modèles qui ont fait la preuve d'une utilité clinique sont simples et ne prennent en compte que certains aspects de la maladie (p. ex., modèles de *Tumor growth inhibition*). Ils intègrent encore peu de la complexité biologique, mais l'essor des nouvelles données biologiques longitudinales permet d'envisager des modèles plus évolués, intégrant notamment des composantes de l'hétérogénéité tumorale ou du système immunitaire.

Un nouveau carrefour interdisciplinaire à soutenir

Plusieurs communautés scientifiques distinctes sont concernées par l'analyse des nouvelles données produites : bioinformatique, imagerie, biostatistiques, informatique, biologie des systèmes, etc. L'analyse de données hétérogènes et multiomiques doit nécessairement impliquer les spécialistes de chaque domaine. De plus en plus de formations spécialisées en bioinformatique et sciences numériques se mettent en place au niveau master dans les universités françaises et les écoles d'ingénieurs et d'ingénieurs. Certaines écoles formant des spécialistes en apprentissage, informatique et sciences des données sont d'ailleurs très demandeuses des applications en santé, dont la cancérologie : en témoignent des partenariats stratégiques tels que Centrale SupElec avec Gustave-Roussy, ou Mines ParisTech avec l'Institut Curie. Malgré une très forte expertise française, reconnue à l'échelle internationale, dans les domaines des mathématiques, des statistiques et de l'informatique, les applications en cancérologie sont relativement rares, notamment du fait du cloisonnement des disciplines. Toutefois, des partenariats stratégiques pour des applications en santé se mettent en place entre EPST (Inria, INS2I-CNRS, Inserm, INSB-CNRS, CEA), avec notamment l'implantation d'équipes Inria sur des campus santé, au sein du PEPR santé numérique et de 3 des 4 instituts interdisciplinaires d'intelligence artificielle (3IA) qui affichent une thématique santé.

En dépit de besoins croissants, il est très difficile d'attirer et retenir dans les laboratoires des scientifiques des données, extrêmement prisés par les entreprises privées qui leur proposent des

Actions

- Mettre en place des normes de validation et d'interprétation des résultats d'IA/modélisation, et des recommandations, avec plusieurs niveaux, sur le modèle des « niveaux de preuve » appliqués aux essais cliniques
- Favoriser la mise en place d'essais cliniques prospectifs et multicentriques permettant de valider les hypothèses générées par IA
- Systématiser le dialogue et soutenir les collaborations entre scientifiques produisant les hypothèses et les modélisations (spécialistes des données) et scientifiques aptes à les tester (biologistes et médecins)
- Veiller à faire respecter la distinction entre résultats et interprétation des résultats
- Intégrer l'histoire et la philosophie des sciences aux cursus d'ingénieurs et d'ingénieurs, de mathématiques/informatique pour la biologie et la médecine pour renforcer l'utilisation de la méthode scientifique

salaires très attractifs. La compétition est donc très rude avec les acteurs du privé, que ce soit dans l'industrie pharmaceutique ou bien dans les entreprises d'informatique (p. ex., GAFAM - Google, Apple, Facebook, Amazon, Microsoft). Il existe également des obstacles administratifs à l'interdisciplinarité qui ne favorisent pas l'émergence d'équipes de recherche réellement interdisciplinaires (nécessité d'être rattaché à une seule école doctorale, commission scientifique spécialisée ou commission/section de rattachement à l'Inserm et au CNRS). Bien qu'il existe des Commissions interdisciplinaires (CID) au sein du Comité national du CNRS, les travaux interdisciplinaires pâtissent de façon générale d'une reconnaissance insuffisante dans les autres instances d'évaluation.

Vers une représentation numérique de la maladie

Des avatars numériques, ou « patients et patientes numériques », ont récemment émergé. Ces représentations permettent de visualiser la maladie (p. ex., une représentation en 3D d'une tumeur du sein en réalité virtuelle) afin, par exemple, de préparer et d'améliorer l'acte chirurgical. De manière distincte d'une analyse des données sans *a priori* (analyse statistique, IA), le concept consiste à établir des modèles capables de simuler la maladie, ainsi que la réponse thérapeutique et différents scénarios d'interventions, afin de déterminer la meilleure option thérapeutique. Ces modèles sont une représentation virtuelle de la maladie dont souffre une personne donnée.

À la différence d'autres domaines en santé (p.ex., cardiologie ou rhéologie sanguine), il n'existe pas de lois (bio)physiques fondamentales sur lesquelles s'appuyer pour construire de tels modèles en cancérologie. Formellement, il ne s'agit pas d'approches d'intelligence artificielle, mais de modèles mécanistiques fondés sur une traduction de la biologie des phénomènes impliqués, ou mimant des états ou phénotypes variés d'un même type cellulaire. Cela permet d'inférer des liens de causalité, de tester des hypothèses biologiques et d'apporter de nouvelles connaissances qui prennent en compte l'aspect dynamique des maladies et du contexte dans lesquelles elles se développent.

Cette approche, très séduisante, se heurte à des obstacles. En effet, les données pertinentes pour calibrer et tester ces modèles (p.ex., composition d'une tumeur hétérogène et plastique à différents temps d'observation) sont difficiles à obtenir en raison des difficultés d'accès, de méthodes souvent invasives, de nécessité de récurrence des données. La biologie du cancer fait intervenir des processus complexes, multiéchelles, évolutifs (phénomène de plasticité), augmentant encore la difficulté de

Actions

- Promouvoir la formation de la jeune génération à l'analyse bioinformatique et biostatistique des données massives, à leur intégration (dimensions, temps, espace) et à la modélisation des processus biologiques
- Créer davantage de parcours de master réellement interdisciplinaires, comme cela se fait déjà dans certaines universités ou dans les 3IA
- Soutenir les partenariats locaux laboratoires/écoles d'ingénieurs et d'ingénieurs
- Promouvoir la formation des biologistes et des médecins aux techniques et donner les moyens d'analyse critique des outils numériques à disposition
- Soutenir les mobilités interdisciplinaires et les séjours d'immersion (biologie-santé/sciences du numérique)
- Favoriser la création d'équipes mixtes, la mobilité des scientifiques, les thèses en cotutelle
- Promouvoir la création de départements de sciences des données au sein des centres de lutte contre le cancer
- Soutenir des partenariats académiques-entreprises pour attirer dans les laboratoires des spécialistes des données travaillant dans le privé
- Augmenter fortement et sur plusieurs années le nombre de postes d'ingénieur.es de recherche en bioinformatique/IA dans les EPST, et attribuer des postes fléchés pour les chargées et chargés de recherche

l'exercice de modélisation. La place que vont prendre ces avatars numériques dans la médecine de précision reste donc encore à définir par rapport à d'autres modèles (tumeurs sur puces, PDX, etc.).

Les modèles dynamiques en oncologie s'attèlent à décrire l'évolution temporelle des cancers ainsi que l'impact de différents traitements. L'objectif est de pallier le manque de données longitudinales, sans lesquelles il est difficile de capturer la vitesse de progression de la maladie. Le développement de tels outils doit permettre une meilleure prise en charge thérapeutique individualisée. La dynamique de ces modèles est pilotée par des jeux de paramètres qui interviennent dans l'expression des modèles. À chaque jeu de paramètres peut être associé un sujet malade virtuel. Ces modèles peuvent devenir des outils de prédiction en clinique dès lors qu'il est possible d'estimer à partir des données cliniques de personnes malades un « bon » jeu de paramètres ou, en d'autres termes, d'assimiler une vraie personne malade à son modèle virtuel. Cette phase d'estimation est donc cruciale et requiert un maximum de données sur chaque patiente ou patient. Or ces données sont souvent de nature très différente, et certaines informations manquent. De plus, dans les cas où des données longitudinales sont disponibles, elles sont souvent en faible nombre et il est nécessaire de pallier ce manque par l'utilisation d'informations obtenues sur l'ensemble de la population des patientes et patients (technique de l'imputation).

Plusieurs facteurs contribuent à produire des données longitudinales permettant la calibration et la validation de modèles dynamiques de développement de la maladie et de réponse ou résistance aux traitements :

- les progrès médicaux, et notamment les immunothérapies, ont augmenté l'espérance de vie des personnes malades, donc les données de suivi sont plus nombreuses ;
- le recours de plus en plus important à des traitements néoadjuvants donne accès, de manière personnalisée, à des données longitudinales sur la réponse de la maladie au médicament ;
- de nouvelles méthodes non invasives (biopsie liquide, imagerie fonctionnelle) permettent de suivre l'évolution temporelle de la maladie ;
- de nouvelles données pertinentes peuvent être dérivées de prises de sang (p.ex., *immunomonitoring* ou *vasculomonitoring*).

Une approche holistique encore en devenir

La médecine des systèmes et des réseaux (ou médecine systémique) repose sur une approche holistique selon laquelle il est aussi, voire plus important, de comprendre les interactions entre toutes les composantes d'un système (le système pouvant être la personne malade) que d'élucider les interactions entre quelques éléments à une échelle donnée, comme cela est réalisé dans les approches réductionnistes. Il s'agit d'un concept identique à celui de la biologie des systèmes, mais appliquée à l'étude des maladies, dont les cancers. Elle inclut implicitement la modélisation des interactions entre mécanismes observés à des échelles différentes, ce qui induit des difficultés

Actions

- Établir des recommandations pour le développement, la publication et la validation d'avatars numériques, sur le modèle du *TRIPOD statement* (*Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis*)
- Accorder lors du soutien à la création d'avatars numériques une importance majeure à la validation des modèles et aux méthodes employées pour cette validation
- Encourager la réalisation d'essais cliniques virtuels, à partir d'avatars numériques, en parallèle à des essais cliniques réels, pour produire des résultats qui faciliteront le positionnement de l'usage des avatars numériques en recherche (clinique)
- Promouvoir les projets de comparaison de différents types de modèles susceptibles de servir la médecine de précision (tumeur sur puce, PDX, avatars numériques, etc.)

importantes et nécessite des développements méthodologiques particuliers par rapport à la seule intégration de données, prônée par la médecine de précision.

Ce domaine d'étude pluridisciplinaire consiste à caractériser les maladies en analysant les interactions entre différentes composantes observées à un niveau donné (génomique, transcriptome, protéome, cellules, tissu, organes, corps entier, environnement) ainsi qu'entre les niveaux correspondant à différentes échelles d'observation. La médecine des systèmes et des réseaux analyse ces interactions au moyen de graphes par la théorie du même nom. Elle bénéficie de l'avancée des méthodes numériques, dont les méthodes d'intelligence artificielle, pour caractériser des interactions nombreuses, complexes, non linéaires, redondantes et incluant des actions et des rétroactions, à partir de données hétérogènes de grandes dimensions.

Elle se développe de plus en plus en médecine, en général (p. ex., étude de l'axe cerveau-microbiote intestinal), mais aussi en cancérologie : de fait, la seule caractérisation moléculaire de la tumeur ne suffit pas à expliquer les mécanismes de réponse ou de résistance aux traitements, et il devient évident que la modulation de l'écosystème de la tumeur, notamment par l'immunothérapie, offre des perspectives prometteuses pour le contrôle de la maladie.

Le fondement de cette approche réside dans l'utilisation de bases de données intégrant des mesures de qualité à de multiples échelles (du génome au corps entier, voire à l'environnement) et en quantité suffisante. De telles bases sont pour l'instant rares, d'autant que l'intégration de données acquises à différentes échelles et de grandes dimensions, ainsi que l'étude des interactions entre observables, ne connaissent actuellement pas de solutions de référence, mais de multiples possibles approches. De plus, la validation des mécanismes suggérés au moyen des approches de médecine des systèmes et des réseaux est souvent complexe.

Le cas particulier de la modélisation en petite dimension

En épidémiologie, les données de toute nature, et correspondant à des échantillons de toute taille – d'une dizaine de sujets à plusieurs milliers voire millions – sont sources d'information pour répondre à des questions adaptées à l'échantillon dont on dispose. Les méthodes à employer pour l'analyse sont cependant différentes selon la nature des données et la taille des échantillons disponibles. Les données spatialisées de petite échelle, par définition acquises sur une population limitée, sont importantes pour étudier précisément l'impact de la localisation géographique des sujets, ou les effets associés aux spécificités d'un petit groupe de sujets. De plus, certaines données sont peu disponibles parce qu'elles ne sont pas fréquemment produites (p. ex., recherche de certaines mutations), pas recueillies systématiquement (p. ex., facteurs

Actions

- Soutenir les projets portant sur la médecine des systèmes et des réseaux en cancérologie, qui inclut l'étude des interactions entre observables mesurés à des échelles différentes
- Former le corps médical aux sciences numériques
- Favoriser l'interdisciplinarité et les mobilités entre spécialités

Actions

- Soutenir et coordonner les collections de données intelligentes de différentes disciplines, complémentaires des entrepôts de données constitués passivement, et les initiatives de liens entre ces collections pour permettre d'intégrer plus d'informations pour une même personne malade (contextualisation mutuelle)
- Soutenir les projets de modélisation en petite dimension en parallèle des projets de modélisation de grande dimension
- Promouvoir des projets d'intelligence artificielle consacrés aux données de petites dimensions et de nature hétérogène
- Favoriser l'interdisciplinarité (épidémiologie, sciences humaines, oncologie moléculaire, pathologie digitale, imagerie médicale)

environnementaux), ou qu'elles concernent des maladies rares (cancers rares). Cependant, il n'existe pas, aujourd'hui, de méthodes consensuelles pour la modélisation à partir d'échantillons de petites tailles (< 30 items).

LA DIFFICILE MAIS INDISPENSABLE STRUCTURATION DES DONNEES DE SANTE

La stratification fine des cancers et la personnalisation des traitements et des prises en charge reposent sur des informations de plus en plus nombreuses et précises. En corollaire, pour permettre l'évaluation des prises en charge et des parcours de soin en cancérologie, l'accès aux données de santé est un enjeu important. Ces données, qui sont produites lors des soins et stockées dans les entrepôts de données hospitaliers et le SNDS, constituent une mine d'informations en situation réelle qu'il est nécessaire d'exploiter de façon partagée et durable (critères FAIR). Toutefois, même avec des algorithmes puissants, ces données restent des données brutes, qui manquent beaucoup de précision. Il faut donc introduire de l'« intelligence » dans les données de population pour les structurer, les conformer à des référentiels internationaux, et en assurer la qualité et l'évolutivité, afin de prendre rapidement en compte l'innovation.

Les données médicoadministratives peuvent être appariées à des collections de données structurées, interopérables avec d'autres ressources (imagerie, omique, anatomopathologie, biologie, protocoles thérapeutiques, biobanques, etc.), pour créer des collections de données cliniques pérennes et standardisées, qui peuvent alors être mobilisées facilement pour contribuer à l'évaluation en vie réelle des prises en charge des maladies cancéreuses. Leur pertinence, pour répondre à des questions spécifiques sur la prise en charge, repose sur :

- une définition rigoureuse de la population couverte (population-cible des traitements, sans biais de sélection, sur des bases géographique et temporelle lisibles) ;
- la finesse des catégories diagnostiques (granularité similaire à celles des indications thérapeutiques) et pronostiques ;
- la description et la disponibilité des données concernant les traitements reçus ou les intentions de traiter ;
- les informations concernant les réponses aux traitements, les récurrences et le suivi à long terme.

Bien qu'attrayante, l'idée de créer ces collections se heurte à plusieurs écueils :

- la mise à niveau réglementaire adéquate ;
- le décloisonnement disciplinaire (épidémiologie, clinique, biologie, etc.) ;
- la structuration et l'intégration de bases de données multiples créées autour de projets cliniques souvent limités à une forme spécifique de cancer et à un type de protocole ;
- la mise en banque d'échantillons biologiques sur le long terme ;
- un financement de fond important sur la durée, qu'il est difficile de trouver malgré la rentabilité de la mutualisation de l'effort de collecte et de contrôle qualité.

Actions

- Encourager la transversalité, la mutualisation et l'interdisciplinarité pour la constitution de collections de données observationnelles « intelligentes », adaptées aux besoins des études en vie réelle
- Promouvoir l'utilisation des registres de cancer comme socles d'études observationnelles en vie réelle (appariés au SNDS, en réseaux européens et internationaux, données cliniques précises, sur échantillons représentatifs, selon des méthodologies standards internationales, interopérables)
- Encourager l'enrichissement des bases en fonction des besoins de la médecine de précision, et leur interconnexion avec d'autres collections de données comportant des annotations standardisées et validées

LE POINT CLE DU PARTAGE DES DONNEES ET DES CODES SOURCES

Bien qu'il soulève également des questions déontologiques et éthiques, le partage des données est un enjeu majeur de la recherche en santé de cette première moitié du 21^e siècle. Dans le domaine de la cancérologie, la quantité de données produites et le développement des méthodes d'analyse de ces données (bioinformatique, biostatistiques, apprentissage machine, etc.) devraient produire les prochaines avancées de rupture dans la prise en charge des cancers. C'est notamment le pari de la communauté européenne qui plaide pour la création de bases de données partagées. L'acronyme FAIR résume les caractéristiques que doivent présenter ces bases. Cependant, le partage et l'analyse des données massives engendrent de nouveaux modes de travail qui ne sont pas stabilisés et n'obéissent pas encore à une méthodologie bien établie. La mise en place des méthodes d'analyse est encore chaotique, les standards mal définis, et la manière de rapporter la méthodologie utilisée pour permettre une reproduction indépendante des résultats insuffisamment formalisée.

Les données sont de nature et d'échelles très variées : imagerie cellulaire, imagerie tissulaire, imagerie d'organe, imagerie du corps entier, multiples analyses moléculaires d'échantillons variés (omiques), etc. La création de telles bases de données issues de personnes malades, ou de sujets contrôles sains, obéit à une réglementation complexe, en particulier le RGPD (cf. infra). Le travail de constitution de bases de données de qualité, complètes, et scrupuleusement annotées et partagées, nécessite d'être bien identifié et reconnu (cf. infra). De plus, les résultats des analyses réalisées à partir de ces bases de données nécessitent d'être diffusés pour permettre la reproduction indépendante de résultats positifs et, inversement, éviter la réplication d'études produisant des résultats négatifs (même si les résultats négatifs restent difficiles à publier).

Les modèles numériques impliquent des paramètres qui sont identifiés à partir de données, typiquement appelées données d'apprentissage. Un défi actuellement posé par ces modèles concerne leur capacité à se généraliser (c.-à-d. à rester performants) et à continuer de fournir des observations ou prédictions valables lorsqu'ils sont appliqués à des données différentes de celles utilisées pour l'apprentissage. La capacité de généralisation des algorithmes d'IA peut se déterminer en évaluant ces algorithmes sur des données dites externes (c.-à-d. acquises de façon totalement indépendante et dans d'autres centres que les données d'entraînement). Pour assurer une indépendance parfaite, cette évaluation externe devrait être réalisée par des personnes autres que celles ayant développé l'algorithme. Elle nécessite d'avoir accès à l'algorithme et à son mode opératoire précis, soit sous forme de code directement exécutable, soit sous forme de code source à partir duquel on peut produire un exécutable adapté à son environnement de calcul.

Actuellement, les modèles d'IA sont parfois partagés dans des répertoires publics (GitHub, Sourceforge, etc.). Certains ne sont pas généralisables et nécessitent donc d'être réentraînés (*fine-tuning*) pour donner de bons résultats sur de nouvelles données. Dans ce cas, le partage du code devrait idéalement être accompagné des instructions nécessaires au réentraînement. Or l'entraînement repose souvent sur un grand nombre d'astuces techniques, laborieuses à expliquer dans un document d'aide. D'ailleurs, même lorsque les codes

Actions

- Mettre en place un registre des bases de données qualifiées (c.-à-d. caractérisées par des scores de qualité, cf. infra) pertinentes pour la cancérologie et un moteur de recherche associé, à l'image de *google dataset search*
- Défendre la notion selon laquelle, en recherche translationnelle comme en clinique, un résultat négatif, si l'étude est bien conçue et correctement réalisée, est une information importante et méritant d'être diffusée
- Établir un système de pré-enregistrement des études exploitant des bases de données, en incluant les hypothèses testées et le plan expérimental associé, à l'image de l'enregistrement des essais cliniques sur *clinicaltrials.gov*
- Associer à toute base de données une référence aux études qu'elle a générées (méthodologie et résultats), même si l'étude n'a pas été publiée

sont partagés, la documentation associée est souvent insuffisante pour mettre en œuvre le modèle de manière identique à l'auteur ou l'autrice du code. Lorsqu'ils sont fournis en données supplémentaires aux articles publiés, les codes ne sont bien souvent pas vérifiés, et ne sont donc pas toujours utilisables (données manquantes, impossibilité de savoir comment différents modules doivent être articulés, etc.). D'autres freins, plus classiques, entravent l'évaluation externe et indépendante des modèles : l'éventuelle propriété intellectuelle associée aux codes numériques (même si la propriété des modèles peut être assurée en les déposant auprès de l'agence de protection des programmes lorsque l'on souhaite protéger le code source), et le risque d'ingénierie inverse, qui dévoilerait le cœur de l'algorithme. Cependant, ne pas donner les clés des procédés expose au risque de voir les résultats non confirmés par des équipes externes, du fait d'un faible niveau de généralisation d'un code. Il est alors impossible de les valider sur des données externes, et donc de les certifier. Aujourd'hui, les mentalités évoluent : dans certaines disciplines (p. ex., la génomique), le partage des codes est quasiment obligatoire pour qu'une publication présentant le code puisse être acceptée.

L'INCONTOURNABLE QUALITE DES DONNEES

L'analyse de grands volumes de données, désormais permise par les méthodes statistiques adaptées aux données massives hétérogènes et évolutives et par les méthodes d'intelligence artificielle, est appelée à jouer un rôle majeur à plusieurs titres :

- mettre en évidence le rôle de certains facteurs jusqu'à présent négligés ;
- comprendre les mécanismes associés au développement des cancers, à leur évolution, à la résistance aux traitements, et à la survenue d'effets indésirables ou de séquelles ;
- optimiser la prise en charge de chaque malade en prenant en compte l'ensemble de ses spécificités.

Or les méthodes d'analyse des données sont sensibles aux différences de nature technique affectant les données, ces différences agissant comme facteurs confondants lors de l'analyse. Dans ce contexte, la qualité des données conditionne la pertinence et la fiabilité de tous les résultats produits à partir de leur analyse. Les collections de données contrôlées quant à leur qualité sont donc précieuses et ont vocation à être réutilisées au-delà des soins ou de leur champ de recherche initial. Les critères FAIR ont été établis pour permettre leur maintien en qualité et leur partage sur la durée, et l'heure est à la multiplication des entrepôts de données, d'outils, de recommandations, initiés par de grands organismes ou sociétés savantes (NIH, NCI, etc.).

La prise en compte des écueils

Les données sont recueillies dans le cadre de la recherche et du parcours de soin au moyen de protocoles variables et d'instruments différents, et ne sont pas toujours annotées de façon identique. Ainsi, alors que l'exploitation des données massives produites par les soins (SNDS) devrait permettre d'aborder certaines questions dans d'excellentes conditions de puissance statistique, ni les grands nombres, ni la performance des algorithmes ne pourront surmonter l'obstacle d'une information manquante ou imprécise à la source (diagnostics codés de façon hétérogène et imprécise selon les

Actions

- Valoriser beaucoup plus fortement les contributions des équipes qui publient leurs modèles sous une forme utilisable et rendent ainsi possible des tests indépendants
- Exiger que les modèles soient rendus publics (éventuellement sous forme d'exécutables) dans les projets financés par des instances nationales
- Disposer d'un registre des modèles numériques et des études d'évaluation associées, avec préenregistrement de ces études d'évaluation, pour que soient ensuite colligés les résultats, positifs et négatifs, produits par chaque modèle dans des conditions correctement décrites

séjours hospitaliers, stades non codés, absence de mention des doses de traitement, etc.), ou de l'hétérogénéité dans le recueil même des données. Or si l'hétérogénéité des données est informative lorsqu'elle est due à de réelles variabilités liées aux diagnostics ou aux prises en charge et doit être prise en compte dans les modèles, l'hétérogénéité des données liée aux modalités de recueil et aux pratiques de codage biaise les analyses.

Des freins importants nuisent à l'homogénéité de ce recueil des données de santé, et limitent de ce fait leur interopérabilité : les champs lexicaux et les outils d'analyse sont parfois très différents entre spécialistes de différents domaines, certains types de données ne disposent pas de format de recueil standardisé, d'autres font référence à des cas particuliers mal documentés dans les référentiels existants, etc. De plus, ces données ne sont valables que dans un contexte donné, qui doit être explicité pour qu'elles puissent être correctement analysées et que les résultats des analyses puissent être interprétés. Il faut prendre en compte le fait qu'elles seront de plus en plus longitudinales. Par ailleurs, la pseudonymisation efface souvent des informations techniques importantes pour leur exploitation (p. ex., informations techniques de l'imagerie effacées lors de la pseudonymisation). Certains types de données ne disposent d'ailleurs pas de format de recueil standardisé, limitant leur interopérabilité. Les technologies évoluent rapidement ce qui impacte aussi la qualité et la nature des données. Enfin, les ressources (infrastructures, compétences et finances) nécessaires au nettoyage des données, à la constitution et au maintien des bases sont importantes et difficiles à combler.

Cela justifie donc la constitution de nouvelles bases complémentaires, évolutives et potentiellement plus riches que les bases existantes (telles que TCGA, ICCG, COSMIC, TCIA, etc.), munies de procédures d'actualisation/incrémentation et de canaux de communication entre les bases de données (stockage des liens qui permettent de remettre en correspondance des données, ou régénérer des informations). Ces bases de données 2.0 doivent être partagées, accessibles, standardisées, interopérables, réutilisables, pour que les scientifiques puissent travailler et comparer leurs résultats, et pour que les citoyens et citoyennes gardent confiance dans les sources d'informations exploitées par la science financée par l'État.

La reconnaissance de la constitution d'une base de données de qualité

L'évaluation est au cœur de la vie professionnelle des scientifiques. Elle s'exerce continuellement et conditionne la progression de carrière et les ressources auxquelles ils ou elles accèdent pour conduire leur travail. Dans le domaine de la biologie des cancers, elle s'appuie essentiellement sur la production de publications scientifiques après évaluation par des pairs. Malgré l'adhésion, par un nombre croissant d'institutions – dont l'HCERES, le CNRS, l'Inserm et la plupart des universités – à la déclaration de San Francisco sur l'évaluation de la recherche (DORA, 2012), le critère de référence de l'évaluation est encore très fréquemment le facteur d'impact du journal dans lequel un article est publié. Le nombre de citations de chaque article est beaucoup plus rarement utilisé, et encore moins la qualité du contenu de cet article, puisque la qualité est supposée être reflétée par celle du journal qui a accepté la publication.

Actions

- Promouvoir, en concertation avec les sociétés savantes, la définition de formats standard de données interopérables pour chaque type de données
- Établir des consensus pluridisciplinaires et des guides de bonnes pratiques pour le recueil de données, à diffuser par les sociétés savantes et les organismes internationaux (prévoir éventuellement deux types de format : minimum et élaboré)
- Former les personnels au contrôle qualité et à l'usage des classifications/recommandations existantes
- Alerter contre une pseudonymisation excessive des données, qui freine leur réutilisation optimale dans différents contextes

Ces indicateurs quantitatifs obéissent à une certaine logique et ont un réel intérêt, mais ils ont aussi des limites et des effets pervers bien identifiés par DORA. L'une des limites de l'indicateur « publications » est une mauvaise reconnaissance du travail considérable que représente la constitution (collecte, vérification, structuration, standardisation, interopérabilité, mise à disposition) d'un corpus de données fiable et bien structuré, alors même que la qualité des bases de données est appelée à jouer un rôle essentiel en médecine, notamment en cancérologie. Il devient donc nécessaire de reconnaître la constitution d'une base de données de qualité au même titre qu'un travail expérimental conduisant à un article dans une revue scientifique de bon niveau. Cela aurait, en outre, la vertu d'encourager une recherche plus ouverte et plus coopérative, mettant en valeur le partage de données.

La remise en cause de pratiques propres à la biologie et à la médecine, consolidées par plusieurs décennies d'expérience au niveau international et soutenues par le lobbying des éditeurs qui ont su tirer avantage de cette organisation, sera longue et difficile. Néanmoins, les instances nationales et internationales devraient réviser progressivement leurs critères d'évaluation pour prendre en compte l'activité des scientifiques qui contribuent efficacement aux progrès de la recherche par des pratiques distinctes de celles aujourd'hui prises en compte.

Actions

- Soutenir la reconnaissance de l'implication de la communauté scientifique dans la constitution et la mise à disposition de bases de données de grande qualité en promouvant la mise en place de scores de qualité et de complétude des données
- Items à promouvoir :
 - ❖ auteurs ou autrices de la base avec mention des règles de propriété intellectuelle ;
 - ❖ contenu : définition du contexte de constitution, transparence des sources et des auteurs ou autrices des données, localisation des données, complétude des informations ;
 - ❖ usage : nombre de publications qui y font référence, nombre de bases de données dérivées ou liées ;
 - ❖ impact : questions auxquelles la base a permis de répondre ;
 - ❖ actualisation : améliorations apportées au cours du temps pour améliorer la qualité de l'information.
- Identifier une entité, nationale, européenne ou internationale, susceptible de promouvoir un tel système afin qu'il devienne un standard

LE DEFI DE LA PROTECTION DES DONNEES PERSONNELLES ET DE L'HARMONISATION DES PRATIQUES

Les prochaines innovations de rupture en matière de prévention et de traitement des cancers reposent dorénavant sur notre capacité à produire et analyser des données massives partagées *via* des banques de données. Cependant, dès lors que ces bases comportent des données personnelles, la confidentialité de celles-ci doit être protégée, qu'elles soient collectées à partir de dossiers médicaux, issues de bases de données médicoadministratives, produites par l'analyse de matériel biologique ou provenant de source déclarative.

Des contraintes multiples

Pour résoudre ce paradoxe et protéger la confidentialité des données personnelles, l'Europe s'est dotée en 2016 du GDPR (*General Data Protection Regulation*), dont chaque État membre a sa propre lecture. En France, le RGPD complète la Loi Informatique et Libertés de 1978, et chaque institution applique sa propre grille d'interprétation de la réglementation, en général sous la forme de contraintes additionnelles. De plus, des exigences de protection supplémentaires à la conformité au RGPD peuvent s'imposer, comme l'hébergement agréé « données de santé » des bases de données de santé identifiantes, ou de conformité au référentiel des « entrepôts de données de santé », ou encore de compatibilité SNDS.

Par ailleurs, les bases de données sont protégées par la propriété intellectuelle qui reconnaît (et valorise) à leur établissement propriétaire la plus-value de son investissement dans la structuration, l'harmonisation et la qualité des données. Les partenariats industriels imposent en outre des circuits juridiques différents des circuits académiques pour l'accès aux données.

Les bases de données doivent également progressivement obéir aux règles FAIR et à différents standards et référentiels permettant d'assurer l'intégrité des données partagées, leur utilisation optimale et leur pérennité.

Des freins multiples

Les exigences de mise en conformité des bases de données, ainsi que des pratiques de collecte et de partage, avec ces différentes réglementations (notamment le RGPD) sont presque impossibles à satisfaire sans un accompagnement juridique, informatique et financier adéquats. La saturation des circuits juridiques a induit des retards de plusieurs mois à plusieurs années pour les recherches planifiées sur les bases de données de cancer, creusant le déficit d'attractivité de jeunes scientifiques, de production scientifique et de compétitivité internationale.

De surcroît, la constitution des bases et le partage des données sont ralentis et limités par des confusions :

- sur la notion de propriété intellectuelle des bases de données, et non de propriété des données (p. ex., le médecin qui produit la donnée brute lors des soins revendique un droit sur les bases de données et leur exploitation ; une agence sanitaire revendique le droit de disposer d'une cession de la base de données au titre de sa mission) ;
- sur la notion de droit, pour les personnes malades, de s'opposer à l'inclusion de leurs données dans une base de données, qui nécessite une information précise et standardisée, à tort interprétée par certains médecins comme l'exigence d'un consentement.

La nécessité d'une harmonisation

Les réglementations sur la protection des données sont conçues pour permettre l'utilisation large et dans les meilleures conditions de données qui auparavant n'étaient pas accessibles aux scientifiques. Cependant, la complexité des réglementations, leurs multiples interprétations et le déficit en capacité de support juridique et technique des organismes sont actuellement un des obstacles majeurs à la constitution de bases de données de recherche compétitives sur le cancer.

Si la France et l'Europe se sont dotées d'une philosophie et d'une réglementation, l'utilisation de serveurs de données hors UE, en particulier pour les infrastructures créées en dehors des États membres, soulève aussi des questions de souveraineté digitale.

Actions

- Mettre en place une coordination nationale des démarches de mise en conformité réglementaire des bases de données, de gouvernance de l'accès aux données, de conventionnement d'accès aux données et de mise en œuvre des règles FAIR
- Soutenir la création de ressources informatiques et juridiques mutualisées pour garantir et mettre en œuvre les compatibilités RGPD, HDS (Hébergement de données de santé), EDS (Entrepôt de données de santé) et SNDS
- Investir dans des supports informatiques adéquats environnés de personnels garantissant la stabilité et la qualité des bases de données
- Prendre appui sur des initiatives existantes :
 - ❖ STANDS4PM (*European standardization framework for data integration and data-driven in silico models for personalised medicine*): guider les équipes de recherche dans le partage des données en recherche sur le cancer en médecine personnalisée
 - ❖ GAIA-X : initiative franco-allemande pour définir des règles communes d'utilisation d'espaces digitaux (mais, avec plus de 300 membres, dont beaucoup hors Europe, l'objectif de souveraineté digitale a été écarté)