

# DOSSIER DE VEILLE

UNITEC

MAI 2018

Les Dossiers de Veille ont pour objectif de vous offrir une analyse des grandes tendances numériques qui bouleversent le monde économique actuel

## Numérique et génomique s'allient contre le cancer

En 2003, après 13 années de recherche, le **Human Genome Project** aboutissait au **séquençage** presque parfait<sup>1</sup> de l'ADN humain. Cette prouesse a définitivement fait entrer la médecine dans une nouvelle ère, celle de la génomique. Selon les spécialistes, la médecine génomique<sup>2</sup> permettra de personnaliser totalement le traitement prescrit à un patient en fonction de l'origine génétique de la maladie. Décrypter le génome humain, c'est produire des millions de données<sup>3</sup> qu'il faut alors être en mesure de stocker et d'analyser en vue de comprendre le

fonctionnement génétique des maladies. Et ces nouvelles données s'ajoutent aux milliards de données déjà collectées sur les patients depuis des dizaines d'années. D'après Orange Healthcare, les données de santé devraient être **multipliées par 50 d'ici 2020**. Le recours aux outils informatiques modernes et aux nouvelles techniques de séquençage permet aujourd'hui d'envisager la lutte contre le cancer d'une manière différente : en axant l'approche autour de la donnée produite, le chercheur, le praticien et le patient peuvent profiter des avancées extraordinaires de la médecine moderne.



Photo from futuristech.info



Photo by rawpixel on Unsplash

<sup>1</sup> Il faudra attendre 2004 pour un séquençage total et parfait du génome humain.

<sup>2</sup> La **génomique** regroupe un ensemble d'analyses permettant d'étudier le génome d'un individu ou d'une tumeur, depuis la cartographie, l'identification et le séquençage des gènes d'une part, jusqu'à l'étude de leurs fonctions et du contrôle de leur expression d'autre part. Le développement des nouvelles...

technologies de séquençage offre des opportunités inédites de recherche et de soins dans le domaine des cancers, ce qui fait de la génomique un enjeu prioritaire.

<sup>3</sup> L'être humain possède environ 22.000 gènes soit une taille de 3400 Millions de paires de base (Mbp).

## 1- Vers un changement de paradigme ?

### a- Evolution dans la connaissance des cancers

Historiquement, les programmes de dépistage pour la détection et la prévention des cancers sont basés sur une approche par classification. Cette classification va permettre de définir des groupes thérapeutiques pour lesquels un protocole, prenant en compte toutes les possibilités thérapeutiques, sera établi. Cette classification utilise des outils simples comme les arbres de décisions, les cohortes, les groupes et les sous-groupes de patients afin d'établir le plus finement possible, à partir de caractéristiques, le type de cancer dont souffre un patient. La classification tumorale a pour objectif de proposer

le traitement le plus adapté au type de cancer identifié. Cette approche basée sur une collecte de données cliniques permet de classer les patients et les cancers dans des groupes et des sous-groupes d'appartenance auxquels on applique des protocoles de traitement. De cette classification a été tirée l'approche par stades d'évolutions des cancers, toujours utilisée aujourd'hui. A titre d'exemple, le cancer du sein était envisagé à travers des facteurs clinico-pathologiques, à savoir l'âge de la patiente, la taille de la tumeur, son grade, l'atteinte des ganglions lymphatiques et la sensibilité de la tumeur aux hormones pour évaluer le risque métastatique.



Photo by Jaron Nix on Unsplash

Mais les recherches depuis 20 ans montrent que le cancer est une maladie du génome. Il est le résultat d'une accumulation de modifications et de mutations de l'ADN perturbant le fonctionnement des cellules. Le cancer est une maladie polycausale puisqu'elle résulte d'une conjonction de facteurs dont certains sont d'ordre génétique tandis que d'autres, considérés comme aggravants, sont d'ordre environnemental. L'évolution de la recherche en oncologie et l'utilisation des techniques quantitatives permettent de repenser l'épidémiologie du cancer. Le séquençage à haut débit<sup>4</sup> (SHD) des tumeurs cancéreuses permet de mesurer l'expression génétique de chacune d'elle. « Cette technique donne la possibilité, à un coût relativement abordable désormais (environ 1 000 euros par patient) de décrypter tout le génome d'une personne », explique le Pr Le Caignec. L'objectif de cet examen consiste actuellement à « identifier de nouveaux gènes responsables de maladies rares pour lesquelles jusque-là, on ne connaissait pas le gène responsable ». « Une fois le gène identifié, cela permet chez d'autres patients qui présentent des maladies semblables, de vérifier qu'ils ont la même mutation ou non », poursuit-il. La société llumina est l'une des sociétés leader en matière de séquençage haut-débit. Sa machine HiSeq XTen génère quotidiennement plusieurs téraoctets de données génomiques, notamment à partir d'analyse de tumeurs cancéreuses.

Depuis juin 2016, la France s'est engagée de manière volontariste dans le SHD. A travers le plan France Médecine Génomique 2025 avec le déploiement prévu de 12 plateformes de séquençage très haut-débit sur l'ensemble du territoire à l'horizon 2020, dont deux sont déjà sorties de terre : la plateforme SeqOIA autour de l'AP-HP (Assistance Publique – Hôpitaux de Paris) et de l'Institut Curie et le projet Auragen, porté par les Hospices Civils de Lyon, le CHU de Grenoble, le CHU de Saint-Etienne, le CHU de Clermont-Ferrand, le Centre Léon Bérard, le Centre Jean Perrin et l'Institut de cancérologie de la Loire. Toujours en France, l'Institut national du cancer

coordonne et soutient la participation de la France au programme de l'International Cancer Genome Consortium (ICGC) à travers différents projets dont un autour du cancer du sein<sup>5</sup>. L'objectif est de contribuer à dresser le répertoire des altérations génétiques survenant dans ces cancers par le séquençage complet de certaines tumeurs. Les premiers résultats ont abouti au séquençage du génome entier de 560 tumeurs du sein puis à l'identification d'un catalogue de plus de 1 600 altérations suspectées d'être à l'origine du développement tumoral. Ces altérations portent sur 93 gènes différents et la quasi-totalité des tumeurs (95%) en présente au moins une. Cinq nouveaux gènes impliqués dans l'oncogénèse des cancers du sein sont identifiés ou confirmés. Cette cartographie tumorale est comparée à la cartographie génétique des patientes et permet, à l'instar du test MammaPrint mis au point par Agendia, de déterminer très précisément et génétiquement les risques d'occurrence d'une rechute si la patiente n'a pas recours à un traitement lourd comme une chimiothérapie. Grâce à l'analyse des données génomiques de la tumeur, les médecins peuvent déterminer a priori les traitements les mieux adaptés individuellement. Chaque tumeur étant unique, l'utilisation des techniques numériques permet d'établir une carte d'identité moléculaire permettant d'ajuster pour chaque patient le traitement le plus précis possible au regard des anomalies constatées.

<sup>4</sup> L'un des intérêts du séquençage haut débit, outre la rapidité avec laquelle le génome est décrypté, tient dans son coût qui s'élève désormais à moins de 100\$ par analyse. La société 23andMe par exemple propose un décryptage complet du génome à partir d'un échantillon de salive pour 99\$.

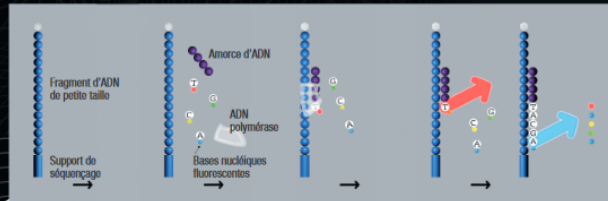
<sup>5</sup> Coordonné par le Pr Alain Viari (plateforme de bioinformatique Gilles Thomas, Fondation Synergie Lyon Cancer).

# La génomique

## PRINCIPE

Dans le cas d'un cancer, il s'agit de séquencer le génome des cellules cancéreuses pour dresser une carte d'identité de la tumeur et l'intégrer à la décision thérapeutique en vue d'un traitement personnalisé.

## Les étapes du séquençage



- 1 Préparation de l'ADN**  
Les molécules d'ADN extraites du prélèvement sont morcelées en fragments de petite taille (50-100 bases). Leurs extrémités sont modifiées pour les accrocher au support de séquençage.
- 2 Mélange**  
Les brins d'ADN immobilisés sont mis en contact avec les acteurs de la synthèse d'ADN : amorce d'ADN, ADN polymérase et bases nucléiques fluorescentes (A en bleu, T en rouge, C en jaune et G en vert).
- 3 Réplication**  
L'amorce d'ADN se fixe sur le brin matrice. L'ADN polymérase synthétise alors le brin complémentaire en ajoutant successivement des bases nucléiques qui forment les paires : A-T (ou TA) et C-G (ou GC).
- 4 Acquisition**  
Quand l'ADN polymérase ajoute une base nucléique, la synthèse du brin s'interrompt pour émettre un rayonnement de couleur spécifique à cette base incorporée. Après acquisition du signal lumineux, la polymérase peut ajouter la prochaine base nucléique. Cette opération est répétée jusqu'à obtention de lectures d'une longueur d'environ 200 bases.

**Prélèvements**  
Le patient se présente à l'hôpital pour deux types de prélèvements : cellules saines par prise de sang ; cellules cancéreuses récupérées directement au niveau de la tumeur.

**Séquençage à haut débit**  
A partir des prélèvements, des appareils automatisés (séquenceurs) effectuent le séquençage à haut débit des génomes du patient et de sa tumeur en quelques jours. Ils déterminent l'ordre des bases nucléiques (adénine-A, cytosine-C, guanine-G, thymine-T) pour reconstituer la séquence d'ADN (3 milliards de bases). C'est grâce à la lecture de plusieurs millions de séquences en parallèle que cette méthode a révolutionné les analyses en génomique.

**Analyse bio-informatique**  
A partir des données brutes issues du séquençage (100 giga octets environ pour un génome humain), des analyses vérifient la qualité des lectures, identifient, classent et trient les différentes séquences d'ADN à l'aide de logiciels performants. Après l'assemblage de milliard de lectures, la séquence du génome entier est ainsi reconstituée. La comparaison entre l'ADN « normal » et l'ADN tumoral permet ensuite de déceler les gènes et les mutations impliqués dans la pathologie. Ces analyses facilitent aussi la détection de variations du génome impliquées dans la réponse au traitement.

**Diagnostic et thérapie personnalisée**  
A partir du profil génomique de chaque tumeur, les médecins peuvent proposer à leurs patients des essais cliniques en vue de développer un traitement ciblé.

**Les mots du TOUT S'EXPLIQUE**

**Génome**  
Ensemble du matériel génétique (gènes) d'un individu.

**ADN polymérase**  
Complexe enzymatique intervenant dans la réplication, la réparation et la recombinaison de l'ADN au cours du cycle cellulaire.

Les défis du CEA-Ma 2016 N° 207  
Aurélie Loeuc et Fabrice Mathé (attribution), avec Diana Russo et Jean-François Delouis du CIA-IG

## b- Evolution des technologies permettant de mieux comprendre les cancers

Parallèlement à l'évolution de la connaissance des cancers, nous assistons depuis plus de 10 ans à une nouvelle approche dans l'appréhension des données et plus particulièrement des données médicales et de santé. L'augmentation de la puissance de calcul des ordinateurs associée à une nouvelle façon d'appréhender la quantité exponentielle de données produites a accompagné l'exploitation de données médicales de plus en plus nombreuses à mesure que

les techniques de mesures, de stockage et d'analyse des données évoluaient. De même, nous avons assisté à un changement de paradigme dans la recherche suite à l'émergence de ce que nous appelons le Big Data (ou données massives). Jusqu'à très récemment, toutes les études empiriques se basaient sur une approche hypothético-déductive, voire déterministe : une hypothèse était formulée, un jeu de données était collecté, un algorithme était programmé, et l'on obtenait un résultat qui validait ou invalidait l'hypothèse formulée. Dans ce cas, l'utilisation de l'algorithme relève de la prévision. Avec le Big Data, tout change : il n'y a plus de résultat unique correspondant à une hypothèse initiale : on entre un grand nombre de données



(pour la plupart hétérogènes car de sources différentes), l'algorithme les analyse et produit un ensemble de résultats non déterminés à l'avance. Dans ce second cas, l'absence d'hypothèse a priori fait de ce modèle un modèle prédictif. Les outils du big data, c'est-à-dire ceux qui sont capables de transformer une masse considérable de données en un output permettant de faire progresser la connaissance de la maladie sont de plusieurs ordres.

Lors du Challenge4Cancer (C4C), qui s'est déroulé entre le 5 novembre 2015 et le 5 mai 2016, une quinzaine d'équipes composées de participants de tous horizons (chercheurs, data-scientists, médecins, mais aussi designers et graphistes concevant de nouvelles formes de visualisation des données, ainsi que des professionnels des aspects juridiques et éthiques) se sont affrontés pour concevoir des projets ayant trait à l'épidémiologie du cancer. Trois d'entre eux ont été récompensés par le jury. Le premier, [Baseline](#), sert à modéliser l'incidence et la mortalité du cancer selon un grand nombre de facteurs. Le deuxième, [CancerViz](#), propose un outil de data visualisation pour accélérer la phase d'analyse exploratoire des données. Enfin, le troisième projet récompensé, [ELSE](#), se présente sous la forme d'un jeu éducatif de sensibilisation sur les facteurs de risques liés aux cancers.

Source : [Usbek & Rica](#)

Les premiers outils et ceux qui sont utilisés depuis longtemps par les scientifiques, sont les algorithmes. Ceux-ci ne datent pas d'hier<sup>6</sup> et peuvent être définis comme une suite finie et non ambiguë d'opérations ou d'instructions permettant de résoudre un problème ou d'obtenir un résultat, en somme il s'agit d'une *méthode générale pour résoudre un type de problèmes*.

En oncologie, les algorithmes sont de plus en plus utilisés comme outils prédictifs et d'aide au diagnostic. En 2011, une équipe de l'Université de Nottingham en Angleterre en collaboration avec la société [ClinRisk](#), a développé un [algorithme permettant de diagnostiquer plus tôt les personnes présentant un risque d'avoir le cancer](#) des poumons et ceux potentiellement atteints par le cancer de l'œsophage. Ces deux types de cancer sont difficilement décelables par un médecin parce que leurs symptômes sont communs et très peu spécifiques. L'algorithme permet d'associer d'autres facteurs à cette maladie : une toux difficile à soigner, une toux accompagnée de sang, une perte de poids, d'appétit et de l'anémie peuvent, lorsque combinés, signifier que ce type de cancer est présent dans les poumons. Une étude réalisée auprès de 375 médecins de famille en Angleterre, impliquant de 1 à 2 millions de patients âgés de 30 à 84 ans, a permis de démontrer que les patients jugés à risque par le logiciel de diagnostic (10% des patients) se sont révélés atteints du cancer du poumon ou de l'œsophage dans 77% des cas, dans les 2 années suivantes.

Plus récemment, des chercheurs de l'Université de Stanford ont utilisé un algorithme dit de *deep learning* qui, par apprentissage, est capable de [détecter les cancers de la peau](#) en se basant sur une photo de la lésion cutanée présumée. Grâce à un smartphone et à une application tous les dermatologues sont en mesure, en photographiant un grain de beauté d'évaluer le risque de tumeur maligne. Pour arriver à un tel résultat, les chercheurs ont alimenté l'algorithme avec plus de 130.000 photos de tumeurs cancérogènes afin que la machine soit en mesure d'évaluer avec une marge d'erreur très faible si

<sup>6</sup> Le mot algorithme vient du nom d'un mathématicien perse du IXe siècle, [Al-Khwārizmī](#). Source : Wikipédia

<sup>7</sup> C'est le même principe qui a été récemment utilisé par Google pour [prédire les risques de survenue d'une crise cardiaque](#) en observant la rétine et en confrontant cette observation avec 300.000 photos de patients.

une tumeur est maligne<sup>7</sup>. Un test a montré que le taux de recouvrement entre les prédictions de l'algorithme et les préconisations des dermatologues atteignait 91%. Une étude du [Harvard Medical School's Beth Israel Deaconess Medical Center](#) montre que le recours au Deep Learning<sup>8</sup> permet de faire chuter le taux d'erreur de diagnostic de près de 85%. L'idée n'est pas de remplacer l'homme par la machine mais de permettre aux praticiens d'améliorer leurs diagnostics grâce à l'apport de la machine. Ainsi, l'utilisation du Machine Learning<sup>9</sup> seul permet de faire passer le taux d'erreur de 3,5% (taux d'erreur observés chez les médecins) à 2,9%. Là où les résultats deviennent vraiment intéressants (notamment pour les patients) c'est lorsque l'algorithmie est associée au médecin, le taux d'erreur tombe à 0,5%<sup>10</sup>. Les chercheurs ont décidé de monter leur start-up PathAi afin de poursuivre le développement de leurs algorithmes et les mettre à disposition du plus grand nombre.



Photo by Chris Liverani on Unsplash

<sup>8</sup> Le Deep Learning est un ensemble de méthodes d'[apprentissage automatique](#) tentant de modéliser avec un haut niveau d'abstraction des données grâce à des architectures articulées de différentes transformations non linéaires ([Wikipédia](#))

<sup>9</sup> L'apprentissage automatique, ou *machine learning*, est une branche de l'intelligence artificielle qui utilise une variété de techniques statistiques, probabilistes et d'optimisation qui permettent aux ordinateurs d'« apprendre » des exemples passés et de détecter les modèles difficiles à distinguer des ensembles de données volumineux, bruyants ou complexes.

<sup>10</sup> On retrouve des résultats très proches en ce qui concerne les [cancers de la prostate](#).



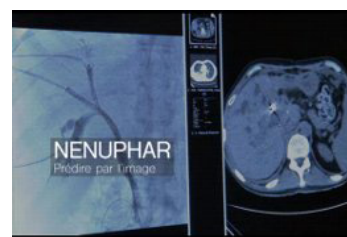
### Owkin : l'apprentissage fédéré et le transfert learning pour lutter contre le cancer

Cette [start-up française](#) fait beaucoup parler d'elle. En levant près de 11M auprès de [divers investisseurs](#), la jeune entreprise installée à Paris et à New York ambitionne de révolutionner le monde médical en proposant des algorithmes de Machine Learning aidant les praticiens dans leurs diagnostics. Contrairement aux gros acteurs américains positionnés sur le même marché, Owkin travaille sur des jeux de données qui ne sortent pas des systèmes d'informations des hôpitaux. Ainsi, les résultats obtenus respectent les contraintes réglementaires liées au respect de la vie privée des patients et à l'anonymisation des résultats. Pour ce faire, l'équipe utilise deux techniques éprouvées : l'apprentissage fédéré qui consiste à entraîner un algorithme sur la machine des utilisateurs de l'application (au sein d'un groupe hospitalier ou d'un laboratoire par exemple) et à partager les apprentissages réalisés sur la machine de chaque utilisateur. Le transfert learning vise à transférer des connaissances d'une ou plusieurs tâches sources vers une ou plusieurs tâches cibles. Pour cela, la start-up travaille avec des établissements comme l'Institut Curie et le Centre Léon Bérard.

Le recours à l'algorithmie dans le cadre du cancer permet également de faire évoluer et d'affiner les classifications tumorales. A titre d'exemple, une équipe de chercheurs allemands du [Deutsches Krebsforschungszentrum](#) a développé un [algorithme auto-apprenant](#) (machine Learning) avec pour objectif de distinguer les cancers du cerveau qui sont près de 100. Les oncologues ont testé leur algorithme sur 1 104 tumeurs du système nerveux central préalablement examinées manuellement et identifiées des erreurs de diagnostic dans 12% des cas. Formé avec des données de référence prélevées sur environ 2800 patients atteints de cancer, le programme a identifié 91 types de tumeurs à partir de ses profils de méthylation<sup>11</sup> de l'ADN et l'analyse de profil de signature génétique. Outre l'amélioration de la précision diagnostique, le programme permet d'identifier de nouvelles tumeurs rares en tant que telles - contrairement à la classification manuelle, où il existe une pression pour attribuer des tumeurs à des catégories connues, même dans des cas atypiques.

Le machine learning est également associé aux algorithmes en ce qui concerne l'univers de l'imagerie médicale<sup>12</sup>. L'utilisation de l'imagerie médicale en cancérologie a toujours plus mis l'accent sur la visualisation que sur le diagnostic. L'imagerie médicale recouvre la radiographie, la mammographie, la scannographie (scanner) et la tomodensitométrie (qui consiste en une série d'images radiographiques) qui utilisent un rayonnement (celui-ci peut être émis par radioactivité, par rayons ionisants, par ultrasons ou par rayons lumineux) pour produire des images de l'intérieur du corps humain, en vue de diagnostiquer une maladie ou une lésion, comme une fracture osseuse. L'intérêt de recourir aux outils big data dans le cadre de l'imagerie médicale est d'utiliser des millions de données issues de clichés afin de créer des patterns permettant à l'ordinateur, lorsqu'on lui présente une nouvelle image de l'analyser très rapidement, de la comparer aux millions d'images enregistrées et de déterminer avec une précision accrue (par rapport au corps médical) la présence ou non

d'indices relatifs à un cancer ([pré-diagnostic](#)). C'est ce qu'on appelle le [diagnostic assisté par ordinateur](#) (computer-aided diagnosis) qui peut aider le radiologue pour l'interprétation des mammographies par exemple ([et non le remplacer](#)). [Zebra](#) est une entreprise israélienne spécialisée dans l'utilisation du big data dans l'imagerie médicale. L'objectif de la société est de fournir en temps réel aux praticiens des résultats instantanés d'une très grande robustesse. L'utilisation des algorithmes de machine learning permet une amélioration constante des prédictions et peut même dans certains cas, réduire le nombre de biopsies. C'est l'objectif de [la collaboration](#) entre Median Technologies et le CHU de Nice qui ont lancé mi-mars un accord afin d'identifier des biomarqueurs tumoraux par imagerie médicale. Les données d'imagerie médicale de [l'étude AIR](#), collectées par le CHU de Nice, seront analysées, aux moyens d'algorithmes élaborés par [Median Technologies](#), pour identifier et caractériser les nodules pulmonaires visibles dans les scanners thoraciques des patients. L'entreprise [Sophia Genetics](#) qui développe, grâce à sa technologie SOPHIA, des tests génomiques vient de recruter le chercheur et mathématicien bordelais de l'INRIA Thierry Colin afin de bénéficier de 10 années de travaux scientifiques portant sur [la modélisation du cancer](#) à partir de l'analyse de l'imagerie médicale.



<sup>11</sup> La méthylation est une modification chimique consistant en l'ajout d'un groupe méthyle (CH<sub>3</sub>) sur un substrat. La méthylation se retrouve couramment en biologie, où elle est effectuée par des enzymes. L'ADN peut être méthylé au niveau des cytosines, mais aussi au niveau des protéines qui l'entourent, les histones.

<sup>12</sup> Selon le Larousse il s'agit d'une « spécialité médicale consistant à produire des images du corps humain vivant et à les interpréter à des fins diagnostiques, thérapeutiques (imagerie interventionnelle) ou de surveillance de l'évolution des pathologies. »

## Le Big Data aide les oncologues à détecter les cancers

Pour la première fois, des médecins japonais ont pu détecter [les premiers signes d'une leucémie grâce à la puissance de Watson](#). Une patiente âgée de 60 ans avait d'abord été hospitalisée pour une leucémie myéloïde chronique. Devant l'apparente inefficacité du traitement prodigué à la patiente, les médecins ont décidé d'employer une autre technique. C'est alors qu'ils ont fait appel à des chercheurs de l'Institut de Médecine de l'Université de Tokyo pour comparer l'information génétique de la patiente aux 20 millions d'études oncologiques auxquelles ils ont pu avoir accès, et ce grâce à Watson. Très vite, le système comprend qu'il s'agit d'un autre type de leucémie et recommande un nouveau traitement, qui s'est révélé beaucoup plus efficace. Bien d'autres logiciels d'aide au diagnostic à partir de l'exploitation des données existent. En France, la start-up parisienne [Khresterion](#) a mis au point un logiciel de ce type. Il aide le personnel soignant des patients atteints de diabète et de cancers à prendre des décisions grâce à l'intelligence artificielle.

d'une base d'apprentissage suffisamment qualifiée. Grâce à un [processus d'apprentissage supervisé](#) à partir de données collectées massivement, l'algorithme identifie automatiquement les tumeurs hépatiques, délimite les contours de la prostate pour la radiothérapie ou permet de compter le nombre de cellules au niveau microscopique avec une performance similaire à celle d'un œil humain expert.

La fédération de centres de lutte contre le cancer (CLCC) Unicancer et le groupe biopharmaceutique Oncodesign ont lancé un projet nommé [OncoSnipe](#) dont l'objectif est le développement et la mise en œuvre d'approches d'intelligence artificielle contre le cancer et plus particulièrement les problématiques de résistance aux traitements anticancéreux. « L'objectif est d'améliorer la modélisation des problématiques de résistances aux traitements anticancéreux par une contextualisation des données cliniques, omiques et d'imagerie par l'enrichissement de métadonnées pertinentes issues des dossiers patients, » indique Philippe Genne, Fondateur et PDG d'Oncodesign. « A l'issue de ce projet, Oncodesign sera capable d'identifier de nouvelles cibles pour traiter la résistance et l'insensibilité aux traitements, c'est un catalyseur puissant pour nos programmes de découverte, en partenariat ou en licensing. Cette technologie est clé dans le contexte de la médecine de précision, au sein duquel nous aspirons à nous développer fortement dans les années à venir. »

La détection des cancers est également l'un des objectifs de [l'intelligence artificielle](#). Cette dernière peut être définie comme un algorithme capable de reproduire artificiellement les capacités cognitives de l'être humain. En faisant appel à la neurobiologie computationnelle, c'est-à-dire aux [réseaux neuronaux artificiels](#), des chercheurs ont mis au point un logiciel capable d'apprendre par lui-même. L'intérêt des réseaux de neurones artificiels est basé sur le principe de l'induction à savoir que l'algorithme est suffisamment puissant pour apprendre et évoluer de manière indépendante de l'humain à partir





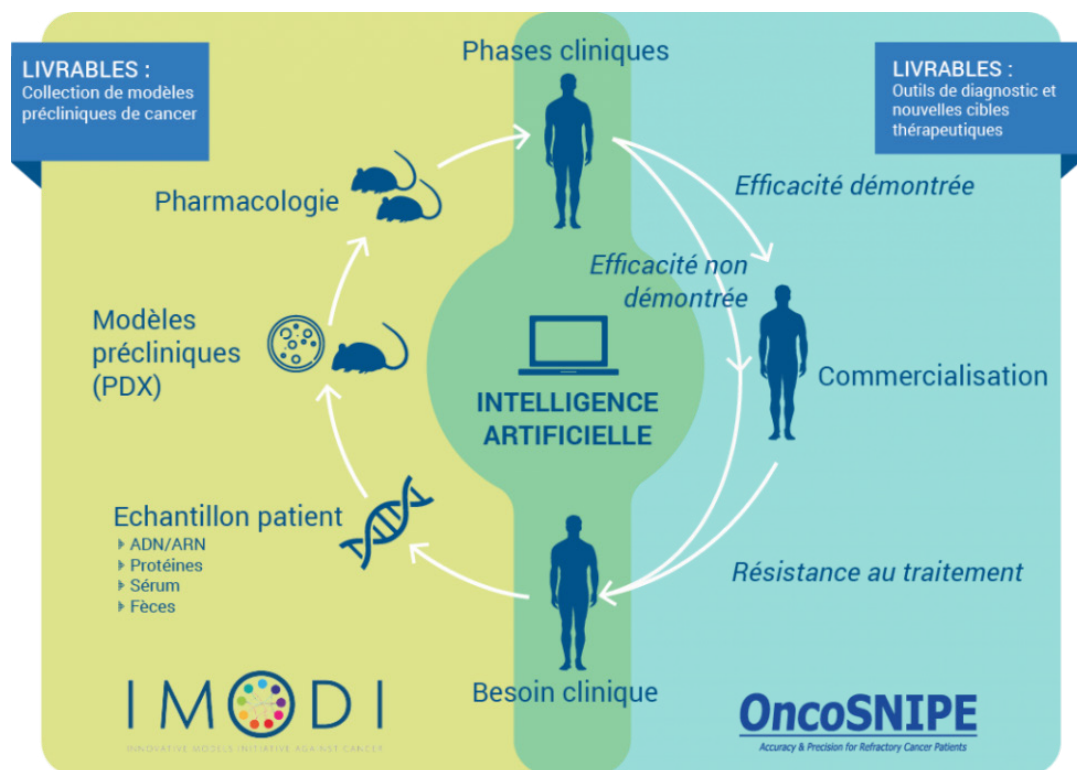
### c- Des perspectives nouvelles dans la lutte contre les cancers

L'analyse des données massives permettent également d'améliorer l'épidémiologie c'est-à-dire la connaissance des facteurs influant sur la santé et les maladies de populations. De nombreuses initiatives nationales émergent un peu partout dans le monde. Aux Etats-Unis, la Precision Medicine Initiative a été lancée en 2015 par le Président Obama avec pour objectif de coordonner l'ensemble des acteurs du monde médical autour de la médecine de précision. 215 millions de dollars étaient alors mobilisés pour constituer une base de données génomiques sur laquelle l'ensemble des praticiens pourraient s'appuyer afin d'offrir à chacun le traitement contre la maladie le plus personnalisé qui soit.

Toujours aux Etats-Unis, en 2016, le National Cancer Institute a lancé le Cancer MoonShot<sup>13</sup> dont l'une

des 10 recommandations est de construire un écosystème de données sur le cancer utilisables par tous les acteurs concernés, des chercheurs aux patients en passant par les médecins. L'American Society for Clinical Oncology (ASCO) a lancé l'initiative CancerLinQ dont l'objectif est de collecter et de rendre accessible l'ensemble des données médicales concernant les patients américains traités pour un cancer afin d'identifier des modèles prédictifs puissants. On peut également citer l'initiative Cancer Genome émanant d'une collaboration entre le National Cancer Institute (NCI) et le National Human Genome Research Institute (NHGRI) aboutissant au Cancer Genome Atlas (TCGA) une cartographie génomique de plus de 33 types de cancers générant plus de 2,5 petaoctets de données portant sur 11.000 patients. Cette cartographie et la base de données associée étant accessibles publiquement aux chercheurs et aux praticiens américains.

<sup>13</sup> Cette initiative a été financée à hauteur de 1,8 milliards de dollars sur 7 ans.



En France, dès 2013, le [programme AcSé](#) (Accès sécurisé à des thérapies ciblées innovantes), porté depuis par [l'INCa](#) et [l'ANSM](#), s'est donné pour objectif de permettre l'accès équitable à l'innovation thérapeutique en France. En ce sens, les patients pourront potentiellement bénéficier d'une thérapie ciblant une anomalie spécifique présente dans leur tumeur. Cette anomalie aura été identifiée au préalable par l'une des [28 plateformes régionales de génétique moléculaire coordonnées par l'INCa](#). Cette initiative a permis de créer en 2016, le [Groupe Immuno-Oncologie](#) portant sur les immunothérapies anticancéreuses.

En 2015, à l'initiative du laboratoire Roche et du laboratoire ouvert la [Paillasse](#), était lancé le projet [Epidemium](#), projet transdisciplinaire (faisant appel à des chercheurs en santé, des sociologues, des data scientists, des designers, des graphistes...) ouvert et collaboratif dont l'objectif est d'observer et traiter les données afin de reconnaître les situations à risques. [Le projet](#) met à disposition, sous la forme d'une plateforme collaborative, l'ensemble des bases de données participantes ainsi que les outils méthodologiques permettant de les traiter. L'idée est de permettre aux chercheurs de lancer des nouvelles recherches, d'approfondir certains résultats mais aussi de tester de nouvelles idées. Le tout grâce aux outils Big Data et aux millions de données collectées.

Finalement, l'articulation entre l'avancée technologique liée à l'exploitation de données toujours plus massives et les progrès liés à la recherche autour cancer permet de dessiner les nouveaux contours d'une médecine qui entre réellement dans une ère de la médecine prédictive, c'est-à-dire, où les médecins et les chercheurs seront capables de prévoir un risque de développer un cancer chez certains patients à partir de l'analyse de leur génome. En analysant plus finement les données et en les comparant avec les données du patient, il sera possible de mieux cibler la prévention sur les causes et leurs facteurs aggravants.

Ces progrès permettent également de s'orienter vers une médecine toujours plus personnalisée. Toutes les données collectées grâce à la génétique vont permettre d'adapter les traitements à certains types de patients, en fonction de leur profil génétique. C'est le cas notamment pour certaines affections comme les cancers du sein et du côlon. Le partage des données recueillies au niveau national et mondial, entre différents établissements de santé, conduira petit à petit à une adaptation des diagnostics et surtout à des traitements aux cas difficiles et rares. « Il conviendra néanmoins de prendre garde à toujours qualifier, contextualiser et relativiser correctement les données. Contrairement au marketing, on n'est pas dans le même niveau de responsabilité pour la santé » prévient toutefois [Manuel Gea](#), PDG cofondateur de [BMSsystems](#), spécialisé dans la recherche de nouveaux traitements et solutions de prévention.

## 2- Les défis du Big Data dans la lutte contre le cancer

Si les progrès dans la recherche contre le cancer s'accroissent de jour en jour, il n'en reste pas moins que nous sommes encore à l'aube d'une véritable révolution. Pour que les outils informatiques soient parfaitement efficaces à chaque étape de la lutte contre la maladie (depuis les études génomiques jusqu'à l'accompagnement du patient dans son traitement), de nombreux défis restent à relever.

### a- Les Défis techniques

**I. Les données sont très hétérogènes : par nature, par leur format, par leur dispersion.**

L'un des principaux défis pour la science est le défi technique. Si les technologies Big Data sont aujourd'hui mûres pour avaler et traiter des quantités faramineuses de données, d'autres challenges se présentent aux acteurs.

Les données collectées étant produites par d'innombrables contributeurs, elles sont très hétérogènes. Premièrement, elles sont très dispersées. Si de nombreuses initiatives ont été lancées pour collecter et centraliser les données (comme nous l'avons vu précédemment avec les projets américains ainsi que le projet Epidemium entre autres), il n'en reste pas moins que de nombreuses données sont encore à « aller chercher à la main » là où elles se trouvent. Elles sont produites par des acteurs aussi différents que les médecins généralistes ou spécialistes, les laboratoires d'analyses médicales, les groupes hospitaliers, les laboratoires de recherches, les bases de données publiques, les bases de données privées et les patients eux-mêmes (ce que l'on nomme le *quantified self*). Chaque acteur produit des données de natures très différentes. Les données des médecins (généralistes ou spécialistes) sont principalement des données textuelles (ordonnances et suivi médical) ou physiologiques. Les données des laboratoires sont des données biologiques c'est-à-dire des informations biologiques obtenues par la réalisation d'analyses sur des échantillons biologiques issus de prélèvements<sup>14</sup>, le plus souvent sous forme numérique. La recherche médicale produit quant à elle des données génomiques portant à la fois sur les patients et les tumeurs. Les radiologues produisent des données d'imagerie médicale (images en 2D et en 3D). Enfin, les données épidémiologiques ou encore les données scientifiques viennent également alimenter les bases de données sur le cancer. En oncologie, des dizaines de paramètres cliniques, biologiques, d'imagerie et de génétique sont systématiquement recueillis.

## II. Des données à organiser

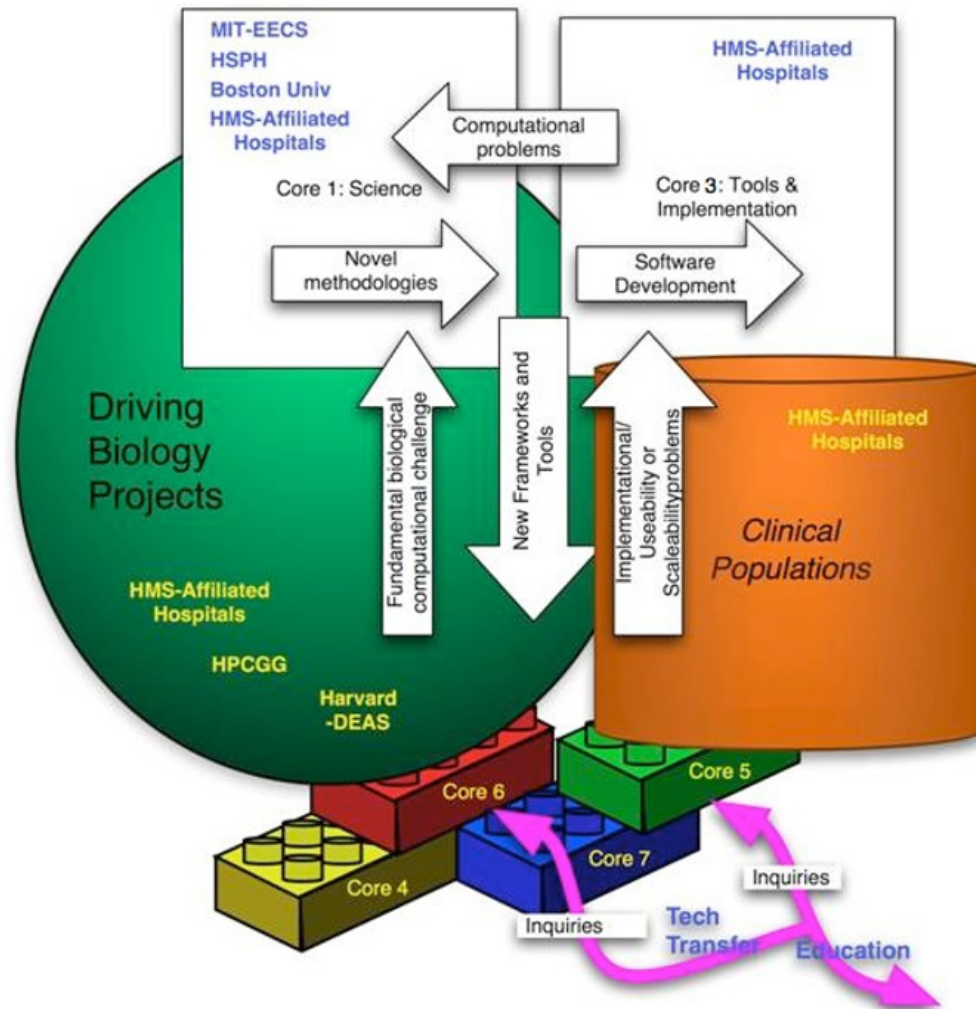
Toutes ces données, pour être valorisées dans le cadre de la recherche contre le cancer, doivent être organisées et structurées. De plus en plus, les laboratoires de recherche mutualisent des entrepôts de données afin d'augmenter la puissance de calcul des ordinateurs et de baisser les coûts de maintenance. Les organismes de recherche français

disposent tous de serveurs de stockage et de supercalculateurs. C'est par exemple le cas du [Mésocentre de calcul intensif aquitain \(MCIA, Bordeaux\)](#), partagé par les universités de Bordeaux et les laboratoires CNRS, Inra, Inria et Inserm de la région. Créé en 2011, il a pour objectif de mettre à disposition des laboratoires de recherche et des entreprises d'Aquitaine un plateau technique de qualité et un lieu d'échange d'expériences et de compétences dans le domaine du calcul intensif. A Toulouse, universités, centres de recherches, entreprises et pouvoir publics ont créé un Groupement d'Intérêt Scientifique (GIS) afin d'offrir aux acteurs de la recherche des outils et des moyens dont ils ne peuvent disposer seuls. Le GIS [GénoToul](#) coordonne les plateformes toulousaines autour des sciences du vivant, couvrant ainsi les problématiques allant de la cellule jusqu'aux populations entières. 12 plateformes sont à disposition des professionnels. Pour ce qui concerne la recherche contre le cancer, Génotoul gère le [Centre de Ressources Biologiques humaines \(CRBh\)](#) qui regroupe les 3 CRB mis en œuvre par le CHU de Toulouse : Germethèque, CRB-Cancer et Toulouse Bio-Ressources. Plus particulièrement, la mission des CRB est la préparation, le conditionnement, le stockage et la mise à disposition des ressources biologiques humaines, organisées en grandes collections thématiques. Les ressources biologiques peuvent provenir de deux origines : 1- les protocoles de recherche biomédicale, et 2- la filière de soins, où les prélèvements destinés à l'analyse médicale peuvent être secondairement requalifiés pour la recherche.



Mésocentre de Calcul Intensif Aquitain (MCIA)

<sup>14</sup> Le [projet BIOBANQUES](#) coordonné par l'Inserm, en partenariat avec l'Institut national du cancer, l'Institut Pasteur, l'Établissement français du sang (EFS) et plusieurs centres hospitaliers universitaires et universités (CPU), mobilise les acteurs du réseau français des biobanques, des tumorothèques et des centres de ressources microbiologique de microorganismes (mBRCs) afin d'organiser l'accès aux ressources biologiques et données nécessaires aux projets scientifiques d'excellence.



Organisation de l'information médicale dans le cadre du projet i2b2

Ces plateformes intègrent donc des données hétérogènes (par leur nature, leur forme, etc.) et complexes. Afin de les traiter avec la plus grande efficacité, leur structuration est nécessaire. C'est l'un des objectifs du [i2b2 tranSMART](#) de Boston aux Etats-Unis qui a créé une communauté ouverte de professionnels et de scientifiques internationaux autour de laquelle est organisée le partage, l'intégration, la standardisation et l'analyse de données hétérogènes produites par la recherche et le monde de la santé. L'ensemble de la démarche est axé autour de l'open source et l'open innovation. Les CHU de Rennes, de Bordeaux ou encore l'Hôpital européen Georges Pompidou (Paris) sont partenaires de l'initiative et collaborent avec le i2b2

(Informatics for Integrating Biology and the Bedside).

Pour finir, ces initiatives de mutualisation de l'information et de la donnée autour du cancer permettent aussi d'organiser l'ensemble de la chaîne de partage de l'information entre des acteurs éloignés géographiquement et dont les métiers sont différents. En regroupant en un même lieu tous les acteurs de la lutte contre le cancer, en leur proposant des outils puissants afin de stocker, traiter et analyser des données hétérogènes elles permettent un meilleur partage des connaissances, un transfert des savoirs du milieu académiques vers les acteurs économiques plus efficace et une prise en compte plus sereine des patients et de leurs maladies.

## b- Défis juridiques et éthiques

Les nouvelles approches de la lutte contre le cancer, en permettant d'individualiser les traitements grâce à l'ensemble des données fines collectées sur le patient, son environnement, l'ensemble de son historique médical, permettent de collecter des dizaines de données personnelles sur chacun des patients, leur état de santé et leurs maladies.

La question du respect de la vie privée et des conséquences de la mise à disposition à une multitude d'acteurs, tant sur le sol national français qu'au niveau international se pose de plus en plus. A titre d'exemple, nous avons vu qu'une société privée américaine comme Illumina génère quotidiennement des téraoctets de données sur le génome humain grâce à des données fournies quotidiennement par ses clients. Ces données sont soumises aux lois américaines et l'entreprise stipule dans ses conditions liées à [sa politique de confidentialité](#) que ces données et informations peuvent être utilisées, traitées et transférées aux États-Unis et dans d'autres pays ou territoires qui n'offrent pas forcément le même niveau de protection que le pays où réside le patient.

La question se pose également lorsque l'on pense aux politiques volontaristes des GAFAM<sup>15</sup> d'investir le monde de la santé. Les géants américains d'internet ont tous une volonté de pénétrer le marché de la santé qui représentera d'ici quelques années plusieurs milliards de dollars avec un taux de croissance de 25% par an dans les cinq prochaines années. L'entrée des GAFAM sur ce marché bouleverse les équilibres, redessine la carte mondiale de la santé et de la recherche et inquiète jusqu'aux pouvoirs publics. Par le rachat par IBM d'entreprises liées à la santé comme [Merge Healthcare](#), ou le rapprochement entre Google et Sanofi débouchant sur la création d'une entreprise, [Onduo](#), spécialisée dans l'analyse Big Data et le diabète. Qualcomm, de son côté, s'est associé avec Novartis avec pour objectif de connecter des inhalateurs « intelligents » de Novartis à la plateforme [Life's2net](#), afin que les données

stockées sur un cloud permettent de suivre l'observance thérapeutique du patient. IBM et Google proposent même des plateformes cloud et des services de traitement de la donnée massive à l'usage des chercheurs, des médecins, des développeurs ou des data analysts. Avec la plateforme [Genomics](#), Google met à disposition des professionnels des outils d'une puissance incomparable bénéficiant du savoir-faire de l'entreprise américaine. La question est de savoir ce que Google fait des données que tous les utilisateurs injectent dans l'outil, comment il traite la confidentialité de ces données. La même question se pose pour IBM et son outil d'intelligence artificielle [Watson for Oncology](#). Cette question peut également se poser pour Apple et son AppleWatch qui collecte des données personnelles en continu et pour l'ensemble des objets connectés de santé, qu'ils appartiennent à des start-up ou à des grands groupes.

La protection de la vie privée et, dans le cadre de la santé, de celle des malades est une question essentielle en France. Le [Ministère de la Santé](#) rappelle, à ce titre, en avril 2018, la loi, modifiée en 2016, L.1111-8 du code de la santé publique, qui stipule que « toute personne physique ou morale qui héberge des données de santé à caractère personnel recueillies à l'occasion d'activités de prévention, de diagnostic, de soins ou de suivi médico-social pour le compte de personnes physiques ou morales à l'origine de la production ou du recueil de ces données ou pour le compte du patient lui-même, doit être agréée ou certifiée à cet effet ». Il propose une [nouvelle procédure de certification de l'hébergement des données de santé](#). Depuis le 10 avril 2017, le [Système National des Données de Santé](#) est entré en service. Il s'agit d'une base de données centralisant toutes les informations de santé des français et regroupant toutes les données provenant des feuilles de soins, des consultations, de l'hospitalisation et des achats de médicaments.

<sup>15</sup> Google, Apple, Facebook, Amazon, Microsoft.

Ce nouveau système de stockage et de partage de l'information confidentielle des patients français soulève quelques inquiétudes. A ce sujet, la **FMF**, le syndicat professionnel de médecins libéraux français, alerte sur « *le risque élevé de perte de confidentialité des données personnelles des patients* », mais aussi sur les risques de discriminations à l'embauche et à l'assurance, liés aux critères de santé.

De son côté, la CNIL travaille également à la protection de la vie privée des patients et des malades. Pour cela, **elle a fait évoluer le chapitre IX de la Loi « Informatique et Libertés »** en juin 2017 pour redéfinir ce qui est applicable en matière de recherche, d'étude ou d'évaluation dans le domaine de la santé. Le chapitre IX de la loi distingue deux grandes catégories de recherches : d'une part, les recherches impliquant la personne humaine et, d'autre part, les recherches, études et évaluations n'impliquant pas la personne humaine. Dans le premier cas, il s'agit des recherches organisées et pratiquées sur l'être humain en vue du développement des connaissances biologiques ou médicales. Il en existe trois types, en fonction du niveau d'implication du patient :

- Des recherches interventionnelles qui comportent une intervention sur la personne non justifiée par sa prise en charge habituelle ;
- Des recherches interventionnelles qui ne comportent que des risques et des contraintes minimales dont la liste est fixée par arrêté du ministre chargé de la santé, après avis du directeur général de l'Agence nationale de sécurité du médicament et des produits de santé ;
- Des recherches non interventionnelles qui ne comportent aucun risque ni contrainte dans lesquelles tous les actes sont pratiqués et les produits utilisés de manière habituelle.

Pour les secondes, il s'agit des recherches dans le domaine de la santé qui n'appartiennent pas aux recherches impliquant la personne humaine. Sont en particulier visées les recherches nécessitant exclusivement la réutilisation de données de santé à caractère personnel (par exemple celles issues de dossiers médicaux, de cohortes existantes ou du SNDS).



**CNIL.**  
COMMISSION NATIONALE  
INFORMATIQUE & LIBERTÉS

*Données personnelles :  
vous avez des droits !*

La CNIL s'inscrit dans la lignée du **RGPD**<sup>16</sup> (Règlement général sur la protection des données) qui entrera en vigueur le 25 mai 2018 et qui redéfinit le cadre européen relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel. En ce qui concerne la donnée de santé, le RGPD offre une définition plus large de cette dernière : « *les données à caractère personnel relatives à la santé physique ou mentale d'une personne physique, y compris la prestation de services de soins de santé, qui révèlent des informations sur l'état de santé de cette personne* ». Selon le **Cabinet Plasse-raud**, « *le considérant 35 vient préciser cette définition générale, ajoutant que les données de santé peuvent aussi bien se rapporter à l'état de santé passé, présent ou futur d'une personne. Elles comprennent également les informations sur une personne physique collectées lors de son inscription en vue de bénéficier de services de soins de santé (numéro ou symbole spécifique attribué à une personne pour l'identifier de manière unique), les informations obtenues lors de la prestation de soins elle-même (résultats de tests ou d'examens médicaux) ou encore les données permettant d'identifier une maladie, un état physiologique ou biomédical.* »

Les données produites en oncologie n'échapperont pas à ce nouveau règlement. Ce dernier...

<sup>16</sup> Considérant 35 et art. 4 du règlement (UE) 2016/679 du Parlement Européen et du Conseil du 27 avril 2016 relatif à la protection des personnes physiques à l'égard du traitement des données à caractère personnel et à la libre circulation de ces données, et abrogeant la directive 95/46/CE (règlement général sur la protection des données)

dernier s'appliquera à tous les acteurs du monde de la recherche et de la santé travaillant avec des données issues de citoyens européens. La question est alors de savoir dans quelle mesure, la mise en place du RGPD aura un impact sur l'utilisation des données de santé collectées en Europe au niveau de la recherche mondiale contre le cancer. Cette question renvoie alors plus généralement au débat entre l'intérêt médical (la lutte contre les cancers nécessitant de grandes masses de données) vs les problématiques éthiques (le patient doit pouvoir exercer son consentement à l'utilisation de ses données personnelles de manière éclairée). « *La révélation d'une pathologie pourrait avoir des conséquences catastrophiques pour l'individu concerné, vis-à-vis d'un employeur ou d'un organisme de crédit par exemple* », confirme Hélène Guimiot-Bréaud, chef du service de la santé à la Cnil. « *Une vigilance absolue s'impose concernant le traitement de ces données.* » C'est en outre pour cette raison que [la CNIL vient d'épingler la CNAM](#) (Caisse Nationale d'Assurance Maladie) pour sa mauvaise gestion de la sécurité du Système national d'information inter-régimes de l'assurance maladie<sup>17</sup> (Sniiram).



Photo by Drew Hays on Unsplash

<sup>17</sup> Le Sniiram recense les identités et les caractéristiques de tous les patients, toutes les prestations remboursées dans le cadre des soins réalisés en médecine de ville, ainsi que la consommation de soins en établissement»

<sup>18</sup> Du nom de cette actrice américaine qui a décidé de subir une double mastectomie suivie d'une ablation des ovaires suite à la découverte chez elle de la présence d'une [mutation de son gène BRCA1](#) augmentant significativement ses chances de contracter un cancer.

<sup>19</sup> Avis n°124, page 45.

Mais l'utilisation des données, notamment des données génomiques, à des fins de prévention pose également des questions éthiques. Aux Etats-Unis, la FDA (Food and Drug Administration) vient d'autoriser la mise en vente d'un test prédictif des cancers du sein, de l'ovaire et de la prostate vendu par la société [23andme.com](#) basé sur l'analyse génomique big data. A partir d'une analyse ADN de la salive, la société est capable d'identifier les mutations des gènes BRCA1 et BRCA2 susceptibles d'augmenter le risque de cancer du sein et de l'ovaire chez la femme, le cancer du sein et de la prostate chez l'homme. Les dangers d'une démocratisation et d'un déploiement des tests génétiques à grande échelle est de constater une augmentation du recours à la chirurgie préventive, autrement appelé le Syndrome Angelina Jolie<sup>18</sup>, basé sur la seule foi d'une corrélation et non d'une causalité. La FDA de rappeler qu'une augmentation du risque ne signifie pas un cancer à coup sûr. La survenue d'un cancer s'accompagne également de facteurs environnants : le tabagisme, l'obésité, une hygiène de vie inadaptée, etc. Dans [son avis n°124](#) « *Réflexion éthique sur l'évolution des tests génétiques liée au séquençage de l'ADN humain à très haut débit* », le [Comité Consultatif National d'Éthique pour les sciences de la vie et de la santé](#) expose également les conséquences d'un déploiement massif de tests basés sur la génomique. Parmi celles-ci, l'impact psychologique est indéniable. Selon les auteurs « *l'annonce du résultat du test génétique, on pense notamment à celui d'un test de paternité biologique, ou à l'annonce possible d'un risque si le test est effectué dans le cadre médical, conduit le sujet à réaliser un travail de deuil anticipé d'un avenir sur lequel il se - ou s'est construit*<sup>19</sup>. ». La recherche – du moins anglo-saxonne<sup>20</sup> – ne semble pas faire fi de cette remarque. Une équipe de l'Université de Californie Los Angeles (UCLA) a mis au point [une nouvelle méthode d'analyse de l'ARN](#) (*survival analysis of mRNA isoform variation*) permettant d'estimer beaucoup plus précisément le temps de survie des patients et l'influence des traitements sur ce délai.

<sup>20</sup> En France, les [Etats Généraux de la Bioéthique](#) ont été lancés à l'initiative du CCNE afin d'orienter les grands axes d'une nouvelle Loi de Bioéthique après une consultation des spécialistes comme des citoyens sur les questions relatives à la santé et à l'éthique.

## Conclusion

Les avancées technologiques associées aux progrès fulgurants de la recherche contre le cancer ouvrent une nouvelle voie à la lutte contre cette maladie qui fait chaque année près de 9 millions de victimes dans le monde. Le recours aux outils Big Data devient une nécessité dès lors que l'ensemble du corps médical produit quotidiennement plusieurs millions de données qu'il est indispensable de faire parler. Les outils à disposition sont toujours plus pointus et plus efficaces<sup>21</sup> menant, de ce fait, très rapidement vers une médecine préventive, personnalisée, de précision, participative, prédictive et « patient-orientée », dite la médecine 6P. La conjonction de ces deux mouvements rend optimistes les chercheurs et les professionnels de la santé sur les capacités des traitements de plus en plus personnalisés et de plus en plus efficaces à comprendre, analyser, diagnostiquer et traiter les diverses formes de cancer. Les outils numériques, aussi évolués soient-ils, ne remplaceront jamais les praticiens. Tout au plus, ils les aideront et les accompagneront dans la pratique de leurs métiers, même si certains se montrent plus pessimistes, à l'image du Docteur Guy Vallancien, pour qui la machine et l'intelligence artificielle remplaceront le médecin qui verra son rôle se transformer vers celui de conseiller. La prolifération des technologies dites wearables<sup>22</sup> pourrait accélérer ce basculement. Ainsi, comme l'avance Eric Schadt, directeur de l'Institut Icahn, « *bientôt, vous verrez que les capacités allouées à ces types de capteurs, en parallèle de leurs développements, seront à même de fournir des profils [médicaux] toujours plus précis à leurs usagers. J'estime que, d'ici cinq à dix ans, les informations les plus précises sur votre santé seront davantage présentes à l'extérieur de votre système de santé (comprendre ici dans le cloud) qu'à l'intérieur de celui-ci.* »

## Auteur

Alexandre BERTIN

Responsable Veille et Prospective chez Unitec

Nous remercions Thierry Colin, chercheur à l'INRIA et porteur du projet Nénuphar pour ses précieux conseils lors de la rédaction de ce dossier.

### Pour aller plus loin...

[5 propositions pour la médecine du futur - Un enjeu majeur pour la France - Rapport du Pr André Syrota et d'Olivier Charmeil](#)

[Vers un consentement éclairé dynamique - Henri-Corto Stoeklé, Jean-François Deleuze, Guillaume Vogt, Christian Hervé](#)

[Big Data and Prevention : from Prediction to Demonstration - HEALTHCARE DATA INSTITUTE](#)

[Commission Innovation 2030 : Un principe et sept ambitions pour l'innovation, 5ème ambition : La médecine individualisée, 2013 - Commission sous la présidence d'Anne Lauvergeon](#)

[From Bench to Bedside: Deep Learning's Journey in Healthcare - Cynthia Burghard, Research Director with IDC Health Insights](#)

<sup>21</sup> On parle ici des 5V du Big Data : volume (capacité de traiter des grandes bases de données), véricité (qui doivent être de très bonne qualité), vitesse (de manière de plus en plus rapide), variété (à partir de données hétérogènes), valeur (qui créent une vraie valeur pour le producteur).

<sup>22</sup> Les technologies portables sont tous les vêtements ou accessoires extra ou intra corporels permettant d'acquérir, en temps réel, des données sur la personne qui porte ces technologies ainsi que sur son environnement.

DOSSIER DE  
VEILLE

by

UNITEC