



Vers la prévention et l'anticipation dans la pratique médicale : réflexions sur l'épistémologie des biomarqueurs dans le cas de la maladie d'Alzheimer

Jean Hache

► To cite this version:

Jean Hache. Vers la prévention et l'anticipation dans la pratique médicale : réflexions sur l'épistémologie des biomarqueurs dans le cas de la maladie d'Alzheimer. Philosophie. Université Panthéon-Sorbonne - Paris I, 2018. Français. $\langle$ NNT : 2018PA01H204 $\rangle$. $\langle$ tel-01982068 $\rangle$

HAL Id: tel-01982068

<https://theses.hal.science/tel-01982068v1>

Submitted on 15 Jan 2019

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

UNIVERSITÉ PARIS I PANTHÉON SORBONNE

UFR de Philosophie

Laboratoire de rattachement : Cetcopra

THÈSE

Pour l'obtention du titre de Docteur en Philosophie

Présentée et soutenue publiquement

le 19 janvier 2018 par

Jean HACHE

**Vers la prévention et l'anticipation
dans la pratique médicale**

**Réflexions sur l'épistémologie des biomarqueurs
dans le cas de la maladie d'Alzheimer**

Sous la co-direction de Bernadette BENSAUDE-VINCENT,

Professeur émérite de philosophie à l'Université de Paris 1 Panthéon-Sorbonne,

et de Xavier GUCHET, Professeur de philosophie des techniques à l'Université de
Technologie de Compiègne

Membres du Jury

- Bernadette Bensaude-Vincent, Professeur émérite, philosophie et histoire des sciences et des techniques, Université Paris 1 Panthéon-Sorbonne
- Catherine Bourgain, Chargée de recherche, génétique, Statistique et sciences sociales, CERMES 3, INSERM
- Bruno Dubois, Professeur, Neurologie, Université Pierre et Marie Curie et APHP
- Xavier Guchet, Professeur, philosophie des techniques, Université de Technologie de Compiègne
- Fabrice Gzil, Philosophe, Fondation Médéric-Alzheimer
- Maël Lemoine, Maître de conférences HDR, IHPST, Université François Rabelais (Tours)

*« Devant la complexité d'un organisme vivant,
notre entendement recule, déconcerté. »*

Henri Bergson

*« La connaissance est la seule chose qui
s'accroît lorsqu'on la partage »*

Sacha Boudjema

*« C'est quand la mer est calme qu'il faut
songer à la tempête »*

Catherine Rambert

*« Il y a une manière de contribuer au
changement, c'est de ne pas se résigner »*

Ernesto Sabato

REMERCIEMENTS

Un tel travail de thèse, mené après un parcours professionnel diversifié et passionnant, loin de la philosophie, n'aurait pas été possible sans l'accueil, l'ouverture et la bienveillance des nombreuses personnes que j'ai côtoyées ces dernières années et que je remercie sincèrement.

Je veux cependant remercier en premier lieu mes deux directeurs de thèse, Bernadette Bensaude-Vincent et Xavier Guchet, pour leur confiance et leur soutien tout au long de ce travail. Que souhaiter de mieux que ce duo pour qu'un scientifique se forme à un travail de recherche en sciences humaines.

Je remercie également les personnes qui ont accepté de faire partie de mon jury de thèse, Catherine Bourgain, Bruno Dubois, Fabrice Gzil et Maël Lemoine, et je ne manquerai pas de tirer parti de leurs remarques et suggestions, car la soutenance n'est pas une fin en soi, mais une étape dans un parcours à continuer.

Le choix de la maladie d'Alzheimer m'amène à remercier l'ensemble de L'Institut de la mémoire et de la maladie d'Alzheimer dirigé par le Professeur Bruno Dubois, et notamment Francis Nyassé qui n'a pas ménagé son temps lors de nos réunions de travail, ainsi que l'équipe du Professeur Philippe Amouyel à l'Institut Pasteur de Lille, en particulier Jean-Charles Lambert qui m'a tout spécialement éclairé sur l'évolution des approches génétiques.

Je voudrais également remercier Elodie Giroux et Valérie Gateau pour leur écoute et leurs conseils lors d'un comité de suivi à mi-parcours, ainsi que ceux qui n'ont pas hésité à lire et à discuter mes textes, notamment Philippe Bardy, Vincent Devictor et Jean-Paul Karsenty.

Merci également à tous mes « collègues », philosophes ou sociologues, qui m'ont accueilli comme un des leurs malgré nos différences de parcours et de situation. Je ne peux les citer tous, et je m'en voudrais de faire le moindre oubli, aussi je choisis de faire un remerciement collectif.

Enfin, je n'oublie pas ma famille, mon épouse Danielle pour son soutien inconditionnel, sa présence constante et son aptitude à faire toujours passer ses proches avant toute chose, mes enfants et leurs conjoints, mes petits-enfants dont je suis fier et qui maintiennent l'harmonie de la famille.

SYNOPSIS

L'objet du travail¹ est une réflexion de nature épistémologique autour de la notion de biomarqueurs et du défi de transférer des connaissances scientifiques du domaine des sciences du vivant à celui de la pratique médicale et clinique.

Un biomarqueur peut être simplement défini comme une caractéristique mesurée objectivement et évaluée comme indicateur de processus physiologiques ou pathologiques, voire comme indicateur de l'action de médicaments. Dans ce cadre, ce peut être un simple variant génétique comme une mutation ponctuelle, mais aussi le résultat d'une expression différentielle d'un gène, tout autant que la concentration considérée comme anormale d'un métabolite.

En fait, un biomarqueur, ce n'est pas qu'une molécule, mais plutôt le signe d'une altération, ayant une signification biologique et pouvant être associée à une pathologie. Cette notion doit par ailleurs être élargie en lui associant des composantes sociales, économiques et environnementales.

Pour appuyer la réflexion, le choix a été fait d'un terrain médical particulier, celui de la maladie d'Alzheimer, une maladie chronique présentant une phase asymptomatique longue et une évolution irrégulière. Ce choix a été complété par celui d'analyser les

¹ Tout au long de ce document, les citations sont directement celles tirées des ouvrages lorsqu'ils sont publiés en français, tandis que les citations écrites en une autre langue ont été traduites en français par moi-même.

techniques, et les outils d'analyses et d'interprétations associés, utilisables dans la pratique médicale, notamment les possibilités de rassembler des données en très grand nombre, les *Big Data*, sur un individu et son environnement dans toute sa complexité. Ces choix méthodologiques permettent de mieux appréhender le biomarqueur comme objet complexe, objet de connaissance, entouré d'incertitude.

L'accent est mis notamment sur la nécessité de contextualiser les questions épistémologiques, en tenant compte de la façon dont sont menées les recherches et leur transfert en clinique, avec l'importance que prennent les questions de standardisation et d'interopérabilité, d'un côté pour ne pas limiter les travaux à l'acquisition de données, de l'autre pour faciliter la question des modélisations. La transdisciplinarité, qui évite de ne s'appuyer que sur une discipline centrale, permet, par confrontation des disciplines, de dépasser les frontières fixées pour chacune d'elles.

Cela conduit à analyser la démarche explicative dans la connaissance médicale, et l'incertitude qui lui est associée. La question de la causalité d'un phénomène complexe n'apparaît pas obligatoirement pertinente, les corrélations révélant des tendances, qui ne disent pas précisément, ni même obligatoirement, le pourquoi de quelque chose, mais qui aident aux décisions d'action. Tout doit être repensé sous l'angle des interactions que tout être vivant entretient avec son milieu. La dimension temporelle de la médecine doit être prise en compte, notamment dans le cadre des maladies chroniques. Aussi, l'anticipation est au cœur de la problématique du biomarqueur, en invitant chacun - et en permettant à tous - à se surveiller, à se penser en fonction de ses facteurs de risque.

Enfin, en se focalisant sur le concept de biomarqueur, non pas molécule isolée, mais réseau d'interactions, voire algorithme, il s'agit de mieux comprendre comment se constituent une connaissance et une pratique, sans oublier pour autant leurs conséquences tant économiques ou socio-politiques qu'éthiques, les questions autour de la connaissance ne pouvant pas être correctement posées si l'on fait abstraction, d'une part des pratiques, d'autre part du rôle des institutions. Comprendre, à l'aide de biomarqueurs et des technologies associées, l'origine d'une maladie complexe, comme la maladie d'Alzheimer, que ce soit une « bifurcation » d'une situation de risques suite à des événements extérieurs, ou une évolution « normale » de facteurs de risque vers l'apparition de la maladie, avant même l'apparition de signes cliniques, amène la personne à devenir acteur de sa santé et à des retombées tant en clinique que dans les politiques publiques de la santé. Les biomarqueurs apparaissent ainsi comme un outil essentiel pour anticiper l'évolution de la condition d'un sujet et permettent de s'interroger sur les éléments transformant des facteurs de risque en une pathologie. En invitant chacun à se surveiller, à réfléchir à ses facteurs de risque, les biomarqueurs incitent à envisager un accompagnement dans une situation où les personnes n'ont encore aucun signe clinique d'une maladie évolutive.

SOMMAIRE

Remerciements	5
Synopsis	7
Introduction	15
Médecine personnalisée	21
Approche et méthodologie	33
Première partie : Revue de la recherche de marqueurs pour la maladie d'Alzheimer	
Chapitre I : Le contexte de la maladie d'Alzheimer	39
I-1 La maladie d'Alzheimer	44
I-2 Les lésions et les facteurs de risque	49
Chapitre II : Les biomarqueurs, une première approche	55
Chapitre III : Le panorama des techniques	65
III-1 Les technologies « -omics »	68
Le séquençage du génome	70
Le transcriptome	77
Le protéome	79
Le métabolome et le lipidome	80
Nanobiotechnologies	82
III-2 Les technologies physiques	83
Tomographie par émission de positrons	86
Électroencéphalographie	87

III-3 La quantification de soi	88
III-4 Les données environnementales	93
Chapitre IV : Face à l'avalanche des données, quelques outils	95
IV-1 La bioinformatique	95
IV-2 La biostatistique	100
IV-3 Les sources de données et les Big Data	108
Les Big Data, de quoi s'agit-il ?	110
Les données génétiques	112
Les outils d'analyse et l'émergence de nouvelles données	114
Chapitre V : Les biomarqueurs et la maladie d'Alzheimer	123
V-1 Les biomarqueurs de la maladie d'Alzheimer	124
Imagerie	124
Liquide céphalo-rachidien	125
Marqueurs sanguins	126
Marqueurs génétiques	131
V-2 Le diagnostic de la maladie d'Alzheimer et les controverses	139
V-3 La biologie des systèmes	147

Deuxième partie : Terrains et enjeux de recherche

Chapitre VI : Un terrain de recherche et un terrain clinique	155
VI-1 L'Institut Pasteur de Lille	156
VI-2 L'IM2A et le projet INSIGHT	159

Chapitre VII : Approche moléculaire versus approche multimodale	161
VII-1 Approche génétique	161
VII-2 Approche multimodale	165
VII-3 Complémentarités et tensions	169
Chapitre VIII : L'organisation des recherches	171
VIII-1 L'organisation des recherches	173
VIII-2 La recherche translationnelle	182
VIII-3 La question de la standardisation	186
VIII-4 La question de l'interopérabilité	188
Chapitre IX : La question de la démarche	191
IX-1 Data-driven versus hypothesis-driven	192
IX-2 La modélisation	196
Chapitre X : La mise en œuvre des techniques	203
X-1 Ruptures, continuités et limites des pratiques	206
X-2 Problème de l'hétérogénéité	210
X-3 La question de l'apprentissage	212
X-4 Conflit ou consensus entre les chercheurs et les cliniciens	213

Troisième partie : Trois questions épistémologiques

Chapitre XI : La question des recherches exploratoires dans la connaissance médicale	219
XI-1 Des données aux connaissances	222
XI-2 La question de l'incertitude	227
XI-3 La question de la causalité est-elle pertinente ?	232

XI-4 Causalité et corrélation	237
XI-5 Les mécanismes	246
XI-6 De la sollicitation à l'opportunité : dispositions et aptitudes	252
 Chapitre XII : La question de la représentation de la maladie d'Alzheimer	 267
XII-1 Vers une nouvelle compréhension de la maladie d'Alzheimer	272
XII-2 Vers une nouvelle nosologie	280
 Chapitre XIII : L'épistémologie, l'éthique et les enjeux économiques et sociaux	 287
XIII-1 L'anticipation	289
XIII-2 L'épistémologie et l'éthique	294
XIII-3 Les enjeux économiques et sociaux	301
 Conclusion et perspectives	 309
 Glossaire	 325
 Bibliographie	 331
Philosophie et sciences humaines	333
Sources scientifiques et médicales	343
 Résumés français et anglais	 354

Vers la prévention et l'anticipation dans la pratique médicale

Réflexions sur l'épistémologie des biomarqueurs dans le cas de la maladie d'Alzheimer

Introduction

L'objet de la recherche se situe dans le domaine de la médecine et de la santé et vise à une réflexion de nature épistémologique. Qu'un scientifique, au profil de chercheur-entrepreneur, s'intéresse à la médecine dans le cadre de réflexions de nature philosophique peut sembler inhabituel. En fait, c'est le résultat d'une conjonction entre plusieurs centres d'intérêt et plusieurs vécus :

- d'abord, un intérêt marqué pour tout ce qui touche à la vie, spécialement la vie humaine ;
- ensuite une formation scientifique de haut niveau, associant mathématiques, physique, chimie, biologie, et permettant une approche non seulement interdisciplinaire, mais également pléo-disciplinaire, par échanges réels, le progrès se faisant par usure, usure liée aux frottements entre des savoirs qui s'ignorent trop souvent ;
- enfin un parcours professionnel conduisant à associer sciences et développements techniques et problèmes de société, avec un souci constant de transfert des connaissances et des expériences.

La médecine est un champ vaste et complexe. Elle comporte, de fait, plusieurs volets. Elle est d'abord, et historiquement, curative, son objectif premier étant de soigner, en vue de guérir principalement, pour contrôler les douleurs à tout le moins. La médecine peut être régénérative, par greffes d'organes ou réparation de tissus par exemple, et la maîtrise future de la différenciation des cellules souches en différents types cellulaires ouvre de grandes perspectives. La médecine est aussi préventive, ainsi que l'illustrent les grands programmes de vaccinations et d'éducation sanitaire. S'y ajoute le profilage du risque, grâce notamment au séquençage du génome d'un patient. Elle peut enfin être prédictive, c'est-à-dire prédire à un individu sain qu'il est ou sera susceptible de développer, dans le futur, une pathologie particulière, alors qu'aucune manifestation pathologique n'existe au moment de cette prédiction.

Ces différents volets sont associés à l'évolution des connaissances dans les sciences du vivant qui ont bénéficié d'un certain nombre de grandes avancées. Ainsi cinq découvertes ou inventions, le microscope, la classification des êtres vivants, la théorie de l'évolution, la génétique et la structure de l'ADN ont largement contribué au développement des approches actuelles, mais c'est bien cette dernière découverte majeure qui est à la base du développement de la biologie moléculaire. Elle a conduit à s'intéresser à la structure des composés cruciaux dans les êtres vivants. Le séquençage du génome humain, et les travaux qui ont suivi sur l'expression et la traduction des gènes, ont ensuite ouvert de larges perspectives, en médecine notamment. La biologie de la fin du XX^e siècle a ainsi vu se succéder deux périodes. La première, la période génomique, dont le point d'orgue a sans doute été le séquençage du génome humain, repose sur le succès de la biologie moléculaire. Basée sur « l'invariant biologique fondamental »²

² Jacques Monod, *Le hasard et la nécessité : essai sur la philosophie naturelle de la biologie moderne*, Paris, Seuil, 1970, page 140.

qu'est l'ADN et le concept de programme génétique (et donc sur une génomique structurale), et sur la recherche des gènes, elle a montré l'impressionnante variabilité du génome, qui fait que chaque individu est unique, mais également que les gènes, finalement pas plus nombreux chez l'homme que chez d'autres organismes inférieurs, ne peuvent pas tout expliquer. Cela a ouvert la voie à la deuxième période, la période post-génomique, mettant en oeuvre d'autres méthodologies comme la génomique fonctionnelle, c'est-à-dire la compréhension du fonctionnement du génome, et la protéomique, c'est-à-dire l'étude du protéome ou ensemble des protéines présentes, à un moment donné, dans un environnement donné, dans un type cellulaire donné. Il en est ressorti, par exemple, qu'un gène peut avoir différentes fonctions selon son environnement cellulaire, ou qu'un même effet peut être dû à l'expression de différents gènes. À l'approche initiale de la biologie moléculaire succède une approche de complexité, la biologie des systèmes, avec l'émergence d'un nouveau concept, la « biologie dialectique »³, qui suggère un dialogue permanent entre les parties et le tout, abolissant ainsi les dichotomies entre « hérédité et environnement ou inné et acquis »⁴. Cet engouement pour des approches plus intégratives est d'abord le fait de biologistes et de chercheurs, mais se heurte dans la pratique aux technologies et instruments disponibles et, aujourd'hui encore, la modélisation des interactions entre les différents niveaux de complexité du vivant est encore un défi pour les bioinformaticiens. Tout processus de production des connaissances se situe le plus souvent dans un contexte institutionnel, même si des sociétés, comme *23andMe* par exemple, proposent sur le Web une production « privée » de savoir, et se trouve confronté aux choix des politiques de santé. Ainsi l'avènement des séquenceurs de nouvelle génération privilégie, ou du moins facilite, les analyses au niveau du génome dans sa

³ Richard C. Lewontin, S. Rose, KAMIN L.J. Kamin, *Nous ne sommes pas programmés : génétique, hérédité, idéologie*, trad. fr., Paris, La Découverte, 1985, page 29.

⁴ Ibid., page 29.

globalité et, pour reprendre la terminologie de Joëlle Vailly⁵, une « politique de la génétique » porte les promesses de grandes avancées médicales.

Le milieu et le génome sont à penser comme deux causes possibles d'un même phénomène observé, ce qui introduit la prise en compte de l'environnement dans l'étiologie des maladies multifactorielles. Milieu et génome sont dans une relation fonctionnelle, changeante au fur et à mesure du déroulement du phénomène, et ils ont une action réciproque l'un sur l'autre. L'analyse des génomes « a mis en évidence que tout n'était pas inscrit dans les séquences d'ADN, même aux niveaux moléculaire et cellulaire »⁶. Si on veut vraiment comprendre les processus complexes en jeu dans le vivant, il ne suffit pas de décrire les composants macromoléculaires élémentaires. Il faut comprendre comment ils s'organisent, comment ils fonctionnent. La compréhension d'un processus complexe ne se trouve pas dans la nature des composants élémentaires, mais dans la recombinaison et l'association de ces composants élémentaires.

Un système vivant, que ce soit une cellule ou un organisme, n'est jamais séparé de son environnement et cet environnement est actif sur une période de temps déterminée d'une part, évolutif dans le temps d'autre part. Par ailleurs, il est à la fois dépendant de son environnement et producteur de cet environnement.⁷

D'un point de vue conceptuel, la mise en évidence de l'insuffisance de la connaissance du génome pour prédire les fonctions précises des régions codantes, pour comprendre les processus complexes⁸, a confirmé la nécessité de passer à une vision plus systémique, s'appuyant sur des approches

⁵ Joëlle Vailly, *Naissance d'une politique de la génétique. Dépistage, biomédecine, enjeux sociaux*, Paris, PUF, 2011.

⁶ Henri Atlan, *Le vivant post-génomique, ou qu'est-ce que l'auto-organisation ?*, Paris, Odile Jacob, 2011, page 129.

⁷ Jean Hache, *Les enjeux des biotechnologies*, Colombelles, Editions EMS, 2005, page 104.

⁸ Evelyn Fox Keller, *Le siècle du gène*, Paris, Gallimard, 2003.

fonctionnelles et donc moins centrées sur le génome. L'avenir n'est plus uniquement dans le séquençage de l'ADN, mais dans la façon dont l'information portée par cette molécule est utilisée. Le gène est porteur d'une information, mais cette information, si elle existe, n'existe que dans un contexte, celui de la cellule, voire celui de tout l'organisme car un gène n'a pas de signification biologique indépendamment de tout ce contexte biologique⁹. On peut faire état, notamment, de l'importance des signaux de régulation reçus. Tout cela a ouvert la voie aux études dites épigénétiques.

L'épigénétique consiste en l'étude des mécanismes qui permettent le contrôle dans le temps et dans l'espace de l'activité de tous les gènes requis pour le développement d'un organisme complexe, du zygote à l'adulte.¹⁰

On peut donc dire que l'épigénétique prépare le terrain pour un contrôle organisationnel à un niveau supérieur à celui du génome. Comme l'a écrit Michel Morange, « l'être humain n'est pas plus prisonnier de ses gènes qu'un peintre ne l'est de sa boîte de couleurs ou un architecte des lois de la pesanteur »¹¹.

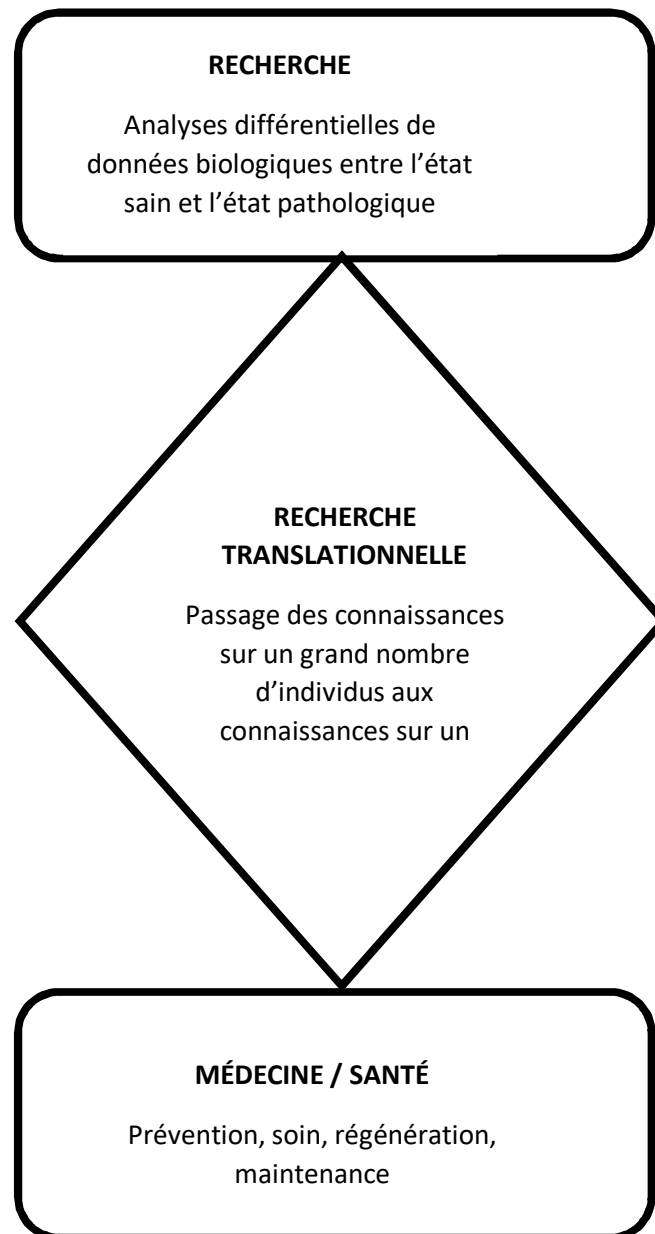
De ces évolutions dans les sciences du vivant, on peut retenir que la génomique et la post-génomique ont produit des quantités massives de données stockées dans des banques de données. La connaissance du génome humain dans sa totalité n'a pas permis de tout apprendre sur les secrets du vivant et sur les maladies, mais son association avec la post-génomique a permis l'évolution et la transformation d'un concept récent, celui de biomarqueurs que nous présenterons au chapitre II et que nous pourrions illustrer par un cas qui a fait l'objet de controverses, l'allèle 4 de l'apolipoprotéine E, présenté au chapitre V quand nous

⁹ P.E. Griffiths, « Genetic information : A metaphor in search of a theory », *Philosophy of Science*, 2001, 68(3) pp. 394-412.

¹⁰ R. Holliday, « DNA methylation and epigenetic inheritance », *Philos. Trans. R. Soc. Lond. B Bio. Sci.*, 326, 1990, pp. 329-338.

¹¹ Michel Morange, *La part des gènes*, Paris, Odile Jacob, 1998, page 189.

appliquerons le concept de biomarqueur au cas de la maladie d'Alzheimer. Il demeure néanmoins que le vrai défi est celui de transférer les connaissances scientifiques en biologie à la clinique, et donc de développer la « recherche translationnelle » à l'interface entre la recherche scientifique et la pratique médicale, comme le montre le schéma ci-après :



Avant d'aller plus loin dans la présentation de la problématique telle qu'elle va être abordée, il est nécessaire de s'interroger sur la notion de « médecine personnalisée ».

Médecine personnalisée

L'expression « médecine personnalisée » a plusieurs définitions selon le point de vue qui est privilégié. Longtemps, la médecine s'est définie comme un art et non comme une science. À mesure qu'elle devient une science, elle veut cependant rester un art, manifestant ainsi la volonté de prendre en compte chaque patient dans sa singularité. Ainsi, si la médecine est tout autant savoir-faire que savoir, connaissance pratique que connaissance théorique et si, de plus en plus, elle expose le corps humain à des pratiques expérimentales toujours nouvelles, elle reste cependant un rapport qui s'instaure entre deux personnes, un médecin/soignant et un patient/soigné. Ce rapport s'établit dans un dialogue singulier entre un médecin et un malade, mais il ne peut s'établir que si le médecin peut comparer, reconnaître, repérer, identifier. Cela passe par des techniques, c'est ainsi que le numérique, l'informatique et l'imagerie ont envahi la médecine.

Le colloque singulier patient-médecin est donc une caractéristique de base de la médecine. Aussi la médecine a toujours voulu être personnalisée. Le concept de médecine personnalisée n'est donc pas nouveau, toute bonne pratique médicale étant historiquement fondée sur le colloque singulier, mais l'introduction de nouveaux outils, les outils de la génétique notamment, amène à s'interroger sur le caractère flou et polysémique du concept, caractère polysémique qui cache un certain embarras. Elle peut signifier un regroupement à partir de certains marqueurs biologiques, ou ne considérer que l'individu et non la personne, ou encore elle peut simplement signifier une meilleure efficacité des diagnostics et des thérapies. Cette notion de médecine personnalisée est loin d'être claire et devient un concept à revisiter. La médecine personnalisée semble annoncer une médecine efficace, fiable, scientifique, mais amène à se poser la question de l'apport de données scientifiques en médecine.

En fait, l'expression « médecine personnalisée » a été utilisée à de nombreuses reprises dans des sens très différents. Dans les années 1970¹², elle était plutôt utilisée pour opposer le médecin de famille, qui suivait personnellement son patient, et le médecin spécialiste qui mettait en avant ses compétences particulières. Mais dès les années 1990, la médecine personnalisée a été d'abord confondue avec la pharmacogénétique, qui concerne la réponse d'un individu à un médicament, ou avec la pharmacogénomique, dont l'objet consiste davantage à prendre en compte l'information du génome dans sa totalité pour prédire l'action d'un médicament¹³. Au cœur de cette question réside l'idée que les individus diffèrent les uns des autres dans leurs réactions aux médicaments, selon la façon dont ils les métabolisent. C'est notamment associé aux polymorphismes de gènes codant pour les enzymes responsables de la détoxification des médicaments, les cytochromes P450 par exemple¹⁴.

Plus globalement, la médecine personnalisée est alors d'abord une médecine stratifiée rassemblant les individus dans des sous-populations différant quant à leurs réponses à un traitement particulier. Le profilage du risque reste cependant la forme la plus médiatisée de la médecine personnalisée, d'autant qu'il ouvre la voie à la prévention. Il s'agit de la caractérisation des prédispositions génétiques à certaines maladies, ce que permettent la mise en évidence des polymorphismes du génome, notamment les très nombreuses mutations ponctuelles, et les études d'association, et ce qui amène au concept de biomarqueurs dont nous reparlerons largement.

¹² W. M. Gibson, « Can Personalized Medicine Survive ? », *Canadian Family Physician*, 17 (8) : 29-88.

¹³ D. M. Roden et A. L. George, « The genetic basis of variability in drug responses », *Nature Reviews Drug Discovery*, 1 : 37-44.

¹⁴ La mise sur le marché par Roche en 2003 d'une puce de caractérisation des gènes CYP450 illustre l'implication des industriels dans cette acception de la médecine personnalisée. – Roche, 2003, « Roche Diagnostics launches the AmpliChip », press release, 25 juin.

Mais, dans notre travail, nous voulons dépasser le réductionnisme de la recherche biomédicale en lui associant des composantes sociales, économiques et environnementales, et donc en élargissant la notion de biomarqueurs, et en analysant ce que cet élargissement apporte à la connaissance. Plutôt que de parler de médecine personnalisée, il est plus justifié de parler de biomédecine, décrite par Jean-Paul Gaudillière¹⁵ comme la conjonction entre trois tendances : la molécularisation des sciences de la vie, l'essor des modèles animaux comme les souris transgéniques par exemple, et la transformation des pratiques cliniques en un champ expérimental. La biomédecine se réfère ainsi à un projet nouveau dans ses procédés, en utilisant des données en grand nombre, très diverses même si elles sont majoritairement moléculaires, pour déterminer la stratégie thérapeutique la meilleure pour un individu particulier.

Alors que la science aboutit à des résultats qui peuvent être probants à grande échelle, c'est-à-dire statistiquement probants, la médecine s'intéresse en effet à la situation d'un individu singulier. La médecine personnalisée, en ambitionnant de traiter le patient au lieu de la maladie, base son action sur une information complète, accessible et critique quant à l'interprétation qui peut être faite de données couplant probabilité et causalité, sans oublier de se poser la question de la qualité de l'information. Mais, les différents contextes de la médecine le montrent, la médecine personnalisée, médecine qui, en première approche, consiste à donner le bon traitement, à la bonne dose, au bon moment et au bon patient en utilisant les outils et les méthodes de la biologie, c'est-à-dire en prenant en compte les données scientifiques du patient, notamment son profil génétique, ne peut pas se baser uniquement sur la génomique. Il faut envisager d'aller plus loin en allant vers la

¹⁵ J. P. Gaudillière, *Inventer la biomédecine*, Paris, Editions la Découverte, 2002.

médecine des « 4P » telle que l'a présentée Leroy Hood¹⁶, à savoir une médecine personnalisée, préventive, prédictive, participative. Il s'agit d'une pratique médicale dont l'individu est le centre, qui utilise non seulement ses caractéristiques biologiques, dont les caractéristiques génétiques, mais également son mode de vie et ses habitudes alimentaires :

- *Personnalisée* est alors pris au sens du traitement plus précis de la maladie ;
- *Préventive* illustre que la prévention devient centrale dans la médecine individuelle, et pas seulement au sens de mesures de santé publique ;
- *Prédictive* fait apparaître le recours aux tests, leur utilisation étant pensée comme un des piliers de la personnalisation ;
- *Participative* affirme que l'individu devient davantage acteur de sa santé.

Cette médecine des 4P signifie qu'il est (ou sera) possible de décrire la susceptibilité d'un patient à telle maladie (*prédiction*) et à tel traitement (*personnalisation*), mais également de prendre des mesures correctives, notamment au niveau moléculaire, et même avant l'apparition de la pathologie (en ce sens *préemption* pourrait se substituer à *prévention* puisque la prévention vise à arrêter un processus en cours) et de conseiller un style de vie adapté (*participation*)¹⁷.

Xavier Guchet, dans son essai sur la médecine personnalisée¹⁸, relève que la médecine personnalisée cherche indiscutablement le bénéfice des patients par une application massive des connaissances et des technologies biomédicales,

¹⁶ www.p4mi.org

¹⁷ Elias Zerhouni, *Les grandes tendances de l'innovation médicale au XXIème siècle*, Leçon inaugurale au Collège de France, 18 mai 2011.

¹⁸ Xavier Guchet (2016) *La médecine personnalisée, un essai philosophique*, Paris, Les Belles Lettres.

mais « questionne la tension entre une médecine ultra-technologique qui confère une signification moléculaire à la personne et une médecine de soin centrée sur le patient ».

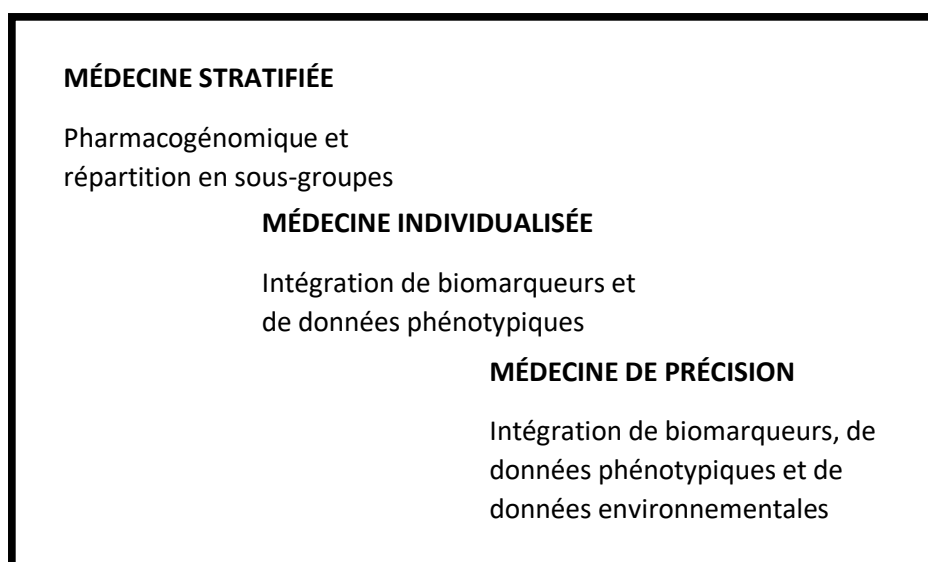
Tout mettre sous le vocable de médecine personnalisée peut n'être qu'un slogan qui, certes, suscite de l'espoir, mais qui, de plus, risque de créer des ambiguïtés ou de fausses promesses. Sans doute faut-il considérer que la médecine est dans une phase de transition, et que l'utilisation, dans des contextes variés, du qualificatif « personnalisée » amène à s'interroger sur l'importance du mot et de ce qu'il sous-entend.

Historiquement, ainsi que cela a déjà été noté, la médecine personnalisée, selon la définition communément admise et reprise par l'Agence Européenne du médicament (EMA)¹⁹, consiste à « donner au bon patient le bon traitement, chaque médicament étant donné à la bonne dose au bon moment ». Elle se présente ainsi comme une médecine « sur-mesure », alors qu'en réalité, comme nous l'avons relevé :

- plutôt qu'une médecine personnalisée, c'est souvent une médecine stratifiée qui subdivise la population en sous-groupes de patients à partir de marqueurs génétiques ou biologiques ;
- plutôt qu'une médecine personnalisée, c'est une médecine individualisée qui fait référence à l'individu biologique caractérisé par son génome (si on ne prend en compte que l'information génétique, on parle même de médecine génomique) mais aussi par son système immunitaire et son système nerveux.

¹⁹ EMA, « Personalised medicine towards the market and patients : the approval process. European Medicines Agency's perspective », Mai 2011.
http://ec.europa.eu/research/health/pdf/event06/13052011/marsa-papaluca-amati_en.pdf.

Cela conduit à introduire une autre formulation, en parlant de médecine de précision²⁰, voulant ainsi signifier une médecine qui vise à établir plus précisément la meilleure stratégie thérapeutique :



Cette formulation introduite, d'une part pour exprimer que toute bonne médecine est par définition personnalisée, d'autre part pour exprimer que les situations cliniques où il est réellement possible d'individualiser le choix du médicament sont extrêmement rares, ce qui peut apparaître comme deux raisons antinomiques, inclut la médecine 4 P, la médecine de précision pouvant intervenir :

- avant la maladie. C'est le domaine du diagnostic pré-symptomatique ou la prise en compte de facteurs de risque ;
- pendant la maladie, notamment pour adapter le bon traitement à la bonne dose, au bon moment ;
- après la maladie, pour son suivi, par l'utilisation de biomarqueurs par exemple.

²⁰ A. Katsnelson, « Momentum grows to make 'personalized' medicine more 'precise' », *Nature Med* 2013, 19(3), page 249.

Cela nous conduit à insister sur la question des temporalités car la combinaison de la génomique, de la biologie, de l'imagerie médicale, des sciences cognitives et de l'informatique, amène à prédire les maladies avant qu'elles n'apparaissent, avant que des symptômes ou des examens cliniques ne créent la moindre alerte. On retrouve la notion de convergence des technologies, aboutissant à une dissolution progressive de la frontière entre les sciences physiques et les sciences biologiques, mise en exergue dans ce qu'on appelle la convergence NBIC pour souligner l'interconnexion croissante entre l'infiniment petit (les nanotechnologies), la fabrication du vivant (les biotechnologies), les machines pensantes (l'informatique) et l'étude du cerveau humain (les sciences cognitives). Il faut ajouter la médecine anticipative qui conduit à la mise en place de moyens pour faire face le jour où le risque se réalise, non plus seulement dans un contexte individuel, mais également dans un contexte de santé publique. C'est dans ce contexte que nous avons choisi de centrer notre analyse sur les biomarqueurs dont les évolutions permettent la prise en compte des temporalités.

S'il est nécessaire d'introduire ce concept de médecine de précision et de médecine anticipative, il est aussi nécessaire de choisir un domaine médical pouvant servir de terrain de confrontation. Ce choix est celui de la maladie d'Alzheimer. Il n'est pas fortuit. En premier lieu, c'est une maladie chronique qui pose d'importants problèmes médicaux et sociaux. Pour reprendre l'expression de Fabrice Gzil²¹, c'est une maladie du temps, changeante et évolutive, qui cristallise peurs et angoisses, et qui présente une temporalité très particulière puisque son apparition est insidieuse, sa phase asymptomatique longue, son évolution irrégulière et accompagnée de phases d'aggravation soudaine et brutale, ses conséquences sur la vie quotidienne des personnes

²¹ Fabrice Gzil, *La maladie du temps. Sur la maladie d'Alzheimer*, Paris, PUF, 2014.

malades et de leurs proches importantes et difficiles à vivre, et c'est une maladie qui reste, à ce jour, inexorable tout en entraînant une perte progressive des capacités de discernement et de jugement. Il a également été estimé²² que si l'on pouvait retarder son apparition de cinq ans, les conséquences sur la qualité de vie des personnes seraient importantes et les dépenses induites seraient réduites de moitié, ce qui fait un lien avec la notion de médecine anticipative que peut permettre d'approcher la mise en œuvre de techniques d'aujourd'hui dans la pratique médicale. Ensuite, le choix de la maladie d'Alzheimer permet de se différencier des études, les plus nombreuses, menées sur les cancers qui ont mis en évidence la grande diversité des tumeurs et la possibilité de définir des thérapies ciblées, non seulement en fonction des profils génétiques des individus, mais également en fonction des caractéristiques propres à la tumeur. Contrairement aux cancers, le cas de la maladie d'Alzheimer ne se prête pas à l'analyse génomique de cellules spécifiques comme les cellules tumorales – la médecine personnalisée se distingue dans ce cas de la thérapie ciblée. Enfin, la maladie d'Alzheimer peut conduire à une certaine vision, une certaine anticipation²³ puisque, dans ce cas, il s'agit de retarder l'apparition de la maladie, voire de l'éviter, en comprenant quelle conjonction de facteurs enclenche la maladie chez des personnes présentant des facteurs de risque. En ce sens, médecine préventive et diagnostic précoce, à défaut d'être strictement reliés, vont dans le même sens dès l'instant où diagnostic précoce signifie diagnostic avant l'apparition des signes cliniques de la maladie, si bien qu'à ce stade on peut envisager deux évolutions vers la maladie, soit une « bifurcation » liée à un (ou des) événement(s) qui entraînerait une rupture dans l'évolution des facteurs de risque, soit une évolution « normale » mais avec des concentrations de facteurs de risque

²² Harald Hampel et al., « Development of biomarkers to chart all Alzheimer's disease stages : The royal road to cutting the therapeutic Gordian Knot », *Alzheimer's & Dementia*, 2012, 8, 312-336.

²³ A. Hedgecoe et P. Martin, « The Drugs don't work : Expectations and the Shaping of Pharmacogenetics », *Social Studies of Science*, vol. 33, n° 3, 2003, p. 327-364.

tels que les signes cliniques se manifesteront inévitablement. Nous reviendrons sur ces deux scénarios possibles quand nous aborderons la question de la nosologie de la maladie.

Dans le domaine de la santé, l'anticipation est une préoccupation réelle avec la multiplication des diagnostics et des outils de prédiction ou de prévention. Mais cette anticipation peut se décliner selon plusieurs conceptions²⁴ :

- L'anticipation prédictive s'appuie sur des probabilités pour construire des connaissances sur l'avenir.
- L'anticipation adaptative cherche à maximiser le potentiel évolutif et adaptatif des individus et des sociétés.
- L'anticipation projective, plus proche de la science-fiction, cherche à « penser l'impensable, pour débattre de l'impossible »²⁵.

Le diagnostic précoce revient à déterminer l'existence de la maladie avant l'apparition de signes cliniques clairs. Dans ce sens, il est plus du ressort de l'anticipation prédictive, avec une prise en compte des temporalités. Il s'agit d'identifier des personnes souffrant de quelques troubles de mémoire qui pourraient développer dans dix ou quinze ans la maladie d'Alzheimer. Nous verrons que, pour établir ces diagnostics, les médecins spécialistes sont encouragés à user de nouvelles techniques telles que la neuro-imagerie ou la ponction lombaire, destinée à détecter dans le liquide céphalo-rachidien la présence de certaines protéines. Mais leur utilisation ne fait pas l'unanimité. De nombreux commentateurs pensent que, « si les biomarqueurs sont prescrits dans le cadre d'un protocole de recherche, on peut penser que cela a un sens. Mais pratiqués à la

²⁴ Léo Coutellec et Paul-Loup Weill-Dubuc, « Les figures de l'anticipation, ou comment prendre soin du futur », *Revue Française d'Ethique Appliquée*, n° 2, 2016-1 : 14-18.

²⁵ *Ibid.*, page 16.

sauvage [...] ils n'ont aucun intérêt clinique pour les patients et ils ne doivent pas être utilisés »²⁶. Quant à la médecine préventive, elle est plus du ressort de l'anticipation adaptative dès l'instant où elle s'appuie sur l'hypothèse, dont nous reparlerons en discutant l'origine de la maladie, de l'apparition d'un événement provoquant une « bifurcation » vers la pathologie.

Ce choix d'un terrain médical est complété par un deuxième choix, celui d'analyser les techniques, et les outils associés, utilisables dans la pratique, notamment la possibilité de rassembler des données diversifiées, structurées ou non, mais accessibles en grandes quantités, les *Big Data*. Cela permet de prendre en compte non seulement les caractéristiques moléculaires et physiologiques, mais également les composantes individuelles, familiales, sociales, culturelles et professionnelles.

Que l'on parle de médecine personnalisée, de médecine de précision, de médecine d'anticipation, de maladie chronique, un concept revient systématiquement, celui de biomarqueur. Le biomarqueur est au cœur de la problématique étudiée puisque le travail de réflexion porte sur l'épistémologie des biomarqueurs, avec une attention particulière sur la complexité, en adoptant le point de vue de la technique d'une part, et en s'appuyant sur un terrain médical particulier, qui n'a pas fait encore fait l'objet d'analyses approfondies d'autre part.

En se focalisant sur le concept de biomarqueurs, il s'agit de mieux comprendre comment se constituent une connaissance et une pratique, sans oublier pour autant qu'elles ont des conséquences tant économiques ou socio-politiques qu'éthiques. Parler d'épistémologie – au sens strict l'étude de la connaissance – c'est admettre l'ambiguïté du mot « savoir ». En tant que nom, c'est

²⁶ Professeur Olivier Saint-Jean, responsable du service de gériatrie de l'hôpital européen Georges-Pompidou, *Le Monde Diplomatique*, N° 743, Février 2016

un fait, une connaissance, un produit, et c'est bien aussi le cas du biomarqueur. En tant que verbe, c'est un processus par lequel on acquiert des données et des connaissances, et c'est encore le cas du biomarqueur.

Mais retenir une approche épistémologique ne signifie pas ignorer les autres questions. À titre d'exemple, les questions éthiques peuvent se poser à plusieurs niveaux :

- Pour certains, l'éthique est avant tout une démarche prescriptive : elle doit donner des repères, rappeler les valeurs et les principes fondamentaux, permettre de dénoncer les mauvaises pratiques, et produire des recommandations.
- Pour d'autres, l'éthique est avant tout une démarche réflexive, interrogeant les valeurs qui doivent gouverner l'action, questionnant les principes qui commandent tant les choix individuels que les choix collectifs.

Les questions éthiques sont ainsi souvent centrées sur le patient. Elles se fondent sur des principes généraux²⁷ tels que l'autonomie, la bienfaisance, la non-malfaisance et la justice ou l'équité, principes initialement développés pour faire en sorte que les médecins se comportent dignement vis-à-vis de leurs patients. Cela aboutit à une « pratique fondée sur les valeurs »²⁸.

Cependant, les réflexions touchant à la médecine, aux technologies, aux *Big Data* rendent l'épistémologie et l'éthique interdépendantes. On ne peut donc pas séparer les questions

²⁷ Tom L. Beauchamp & James F. Childress, *Les principes de l'éthique médicale*, Paris, Les Belles Lettres, 2008.

²⁸ K. W. M. Fulford, « Facts/values : ten principles of value-based medicine », in J. Radden (Ed.) *Philosophy of Psychiatry : A Companion*, New York, Oxford University Press, 2004, 205-234.

K. W. M. Fulford, E. Peile, H. Carroll, *Essential Values Based Practice*, Cambridge, Cambridge University Press, 2012.

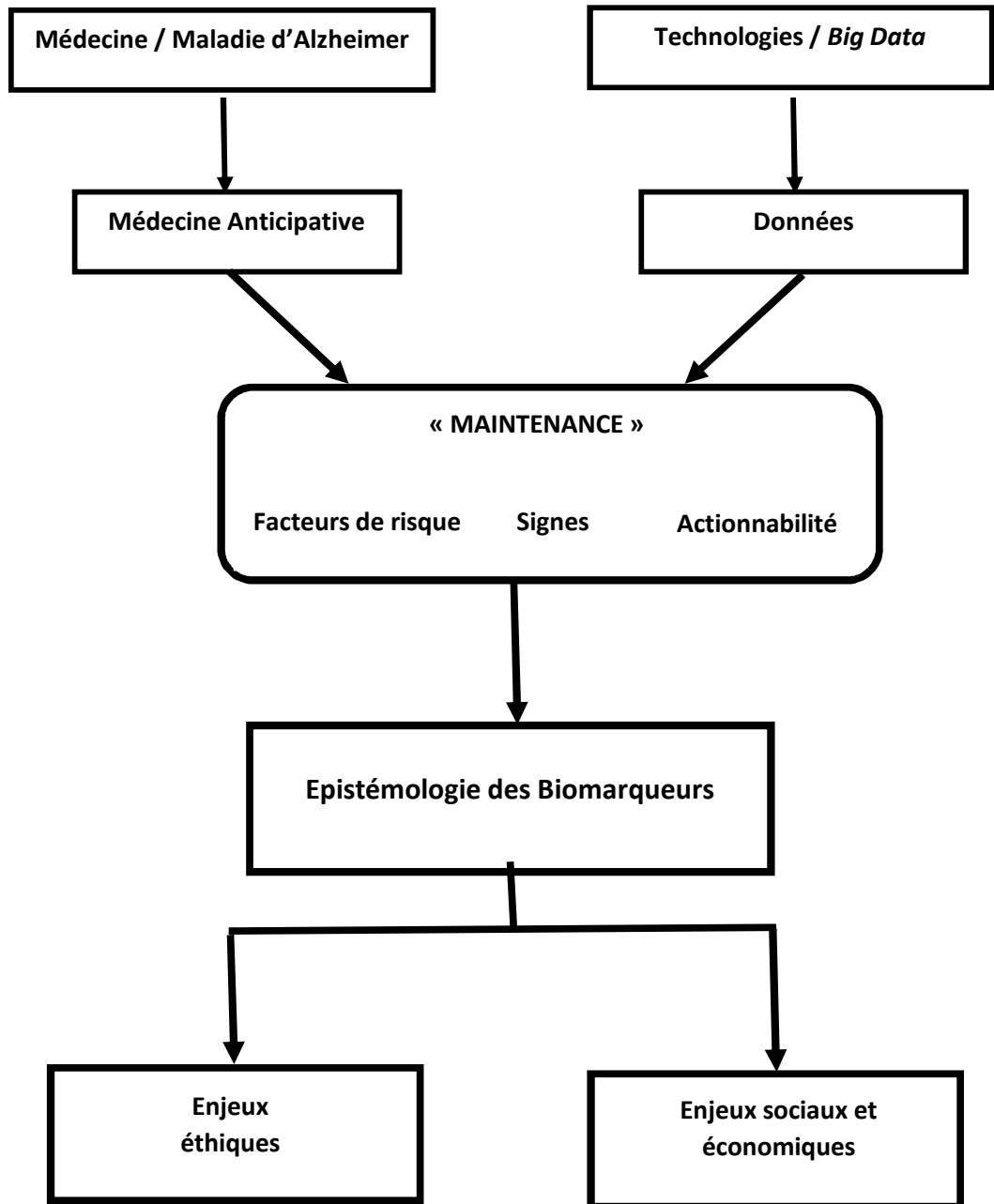
épistémologiques et les questions éthiques. La façon dont les savoirs appréhendent leur objet de connaissance est une question épistémologique tout autant qu'une question éthique. C'est pourquoi, dans le travail présenté, il a été retenu de ne pas envisager l'éthique, d'une part du point de vue du patient, mais plutôt en partant des acteurs que sont les biologistes, les médecins et les responsables de santé publique, d'autre part parce que les biomarqueurs, s'ils apportent des connaissances – et nous verrons lesquelles – posent aussi des questions éthiques, comme le degré de confiance qui peut être accordé à ces connaissances. Nous pouvons donc considérer en quoi les approches épistémologiques ont des conséquences éthiques et socio-politiques. En ce sens, les problèmes éthiques et socio-politiques sont abordés sous un autre angle. Il ne s'agit pas d'aborder exclusivement les biomarqueurs sous le regard de l'utilisation des nouvelles techniques biomédicales mais de considérer que, par les techniques, on retrouve des enjeux essentiels de la pratique médicale dans la mesure où l'arrivée de ces techniques est de nature à transformer et à reconfigurer en profondeur les pratiques de recherche et la clinique. Plus généralement, les questions, qu'elles soient épistémologiques, éthiques ou socio-politiques ne peuvent pas être bien posées si l'on fait abstraction, des pratiques d'une part, du rôle des institutions d'autre part. Ce n'est pas l'émergence des données en très grand nombre, ni les méthodes utilisées pour les obtenir, qui sont à l'origine des connaissances, mais plutôt la façon dont les données sont traitées et diffusées dans un contexte politique et économique, c'est-à-dire pour reprendre la terminologie de Sabina Leonelli, la façon dont les données « voyagent »²⁹.

²⁹ Sabina Leonelli, *Data-Centric Biology, a philosophical study*, Chicago, MI, The University of Chicago press, 2016

Approche et méthodologie

Si la problématique est de mieux approcher l'épistémologie des biomarqueurs, elle implique un travail de reformulation qui doit conjuguer plusieurs informations, parfois divergentes, et nécessite une méthodologie. Aussi, la première partie présente un panorama des techniques utilisables dans la pratique médicale et une revue de l'évolution de la recherche de marqueurs, notamment dans le contexte de la maladie Alzheimer. Puis, dans une deuxième partie sont présentés deux terrains complémentaires de confrontation, un terrain plus orienté recherche et un terrain à orientation médicale. Est ainsi abordée la contextualisation des questions épistémologiques. Des questions épistémologiques essentielles, comme la question de la causalité et du mécanisme de la maladie, ainsi que celle de son origine, constituent une troisième partie. Pour terminer, une approche vers une médecine anticipative et une médecine de maintenance est proposée, ainsi que les liens qui se font avec les enjeux éthiques et les enjeux économiques et sociaux.

La démarche peut être illustrée par le schéma présenté à la page suivante.



Ce schéma peut être complété par une métaphore, la métaphore du plongeur :

- On peut être un plongeur avec masque et tuba. On reste alors à la surface des choses, on voit les phénomènes de l'extérieur, on se protège sans réellement apprendre quelque chose.
- On peut être un plongeur en apnée. On sait alors peu de choses car la capacité de la cage thoracique amène à faire de constants allers et retours pour récupérer de l'air.
- On peut être plongeur avec des bouteilles. On est alors immergé dans le problème et sa complexité, on enrichit ses connaissances pas à pas et il y a beaucoup moins de limites car l'autonomie apportée par un appareillage extracorporel est plus importante que celle apportée par la retenue de la respiration.

Équipé des connaissances acquises par l'expérience, notamment des connaissances scientifiques et de leurs traductions en technologies et outils, nous espérons plonger avec des bouteilles, avec une attention aux plus petits détails.

Première partie

Revue de l'évolution de la recherche de marqueurs pour la maladie d'Alzheimer

Chapitre I

Le contexte de la maladie d'Alzheimer

Comme présenté dans l'introduction, le travail s'appuie sur un cas particulier, la maladie d'Alzheimer. Rappelons que ce cas a été retenu, d'une part parce que, contrairement aux cancers, il ne se prête pas à l'analyse génomique de cellules spécifiques comme les cellules tumorales – la médecine personnalisée se distingue dans ce cas de la thérapie ciblée -, d'autre part parce que la maladie d'Alzheimer est une maladie qui s'installe insidieusement sur plusieurs années après une longue période d'évolution silencieuse, et enfin parce qu'il a été estimé³⁰ que si l'on pouvait retarder son apparition de cinq ans, les conséquences sur la qualité de vie des personnes seraient importantes et les dépenses induites seraient réduites de moitié, ce qui fait un lien avec la notion de médecine anticipative que peut permettre d'approcher la mise en œuvre de techniques d'aujourd'hui dans la pratique médicale. Avant de la présenter, relevons dès maintenant qu'elle permet une approche globale schématisée par l'enchaînement suivant :

³⁰ Harald Hampel et al., « Development of biomarkers to chart all Alzheimer's disease stages : The royal road to cutting the therapeutic Gordian Knot », *Alzheimer's & Dementia*, 2012, 8, 312-336.

Exposition



Menaces



Stratégie d'adaptation



Evolution / Résultat

L'exposition renvoie aux prédispositions, aux facteurs de risques et fait référence aussi bien à des paramètres génotypiques ou phénotypiques, identifiables par des biomarqueurs, qu'à des éléments environnementaux et de modes de vie. Elle est donc associée à des techniques permettant d'acquérir ces paramètres.

Les menaces sont des événements susceptibles d'engager l'individu sur la voie d'une évolution défavorable, ce que nous pouvons appeler une bifurcation dans une évolution continue. On reviendra sur cette notion, mais on peut dire qu'une bifurcation intervient lorsqu'un petit changement d'un paramètre (un facteur de risque par exemple), ou l'apparition d'un nouveau facteur (suite à un événement par exemple) produit un changement majeur dans l'organisation d'un système (une évolution rapide vers des signes cliniques de la maladie d'Alzheimer par exemple). Elles sont aussi associées à des techniques dès l'instant où elles peuvent se concrétiser par des signaux d'alerte.

La stratégie d'adaptation comprend les ressources, aussi bien internes qu'extérieures, dont peut disposer la personne pour se

protéger. Les techniques permettent d'envisager des tests renforçant les atouts de la personne.

L'évolution/résultat traduit l'évaluation, tant par la personne que par la société, de la qualité de vie du malade Alzheimer. On peut associer cette phase à la présence au quotidien d'objets connectés par exemple.

Par ailleurs, avec l'augmentation de l'espérance de vie, la maladie d'Alzheimer représente un défi humain, social, médical et scientifique. Elle est un véritable enjeu de santé publique. Sa prévalence est imprécise, environ 870 000 cas en France, et croît d'une manière exponentielle avec l'âge. Son incidence est de 165 000 nouveaux cas par an et, à part quelques cas (moins de 1%) d'apparition très précoce, la maladie d'Alzheimer est une maladie de la vieillesse, une maladie liée à l'âge. C'est aussi une maladie multifactorielle pour laquelle on peut reconnaître au moins cinq types³¹.

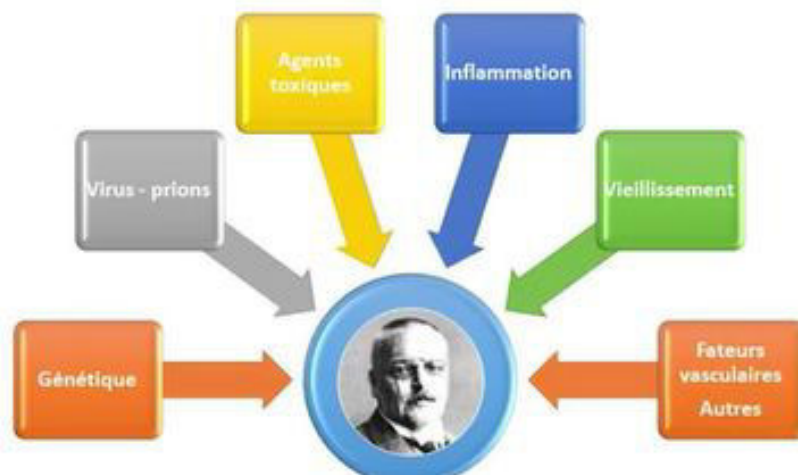
Entre le vieillissement qui peut conduire à une situation normale et le vieillissement qui conduit à une situation pathologique, il y a quelque chose qui n'est pas maîtrisé. Le souci du praticien est de diagnostiquer la maladie le plus précocement possible, avant même l'apparition de signes cliniques, de façon à retarder, voire maîtriser, son évolution. Nous reviendrons sur cette question du diagnostic car elle est sujette à discussion, la maladie d'Alzheimer n'étant pas, comme c'est souvent considéré en parlant d'une maladie de la vieillesse, une maladie du sujet âgé, mais une maladie de l'adulte mûr, une maladie biologique qui débute à 50-60 ans sans signes cliniques particuliers pour s'exprimer cliniquement à 70-80 ans.

³¹ Khalid Iqbal, Inge Grundke-Iqbal, « Alzheimer's disease, a multifactorial disorder seeking multitherapies », *Alzheimer's & Dementia*, Vol. 6, Issue 5, Sept. 2010, pp : 420-424.

Retenons également, dès maintenant, que, malgré d'intenses recherches, les causes de la maladie d'Alzheimer restent inconnues :

La ou les causes de la maladie d'Alzheimer

Malgré d'intenses recherches, la cause ou les causes de cette maladie restent inconnues.



La maladie d'Alzheimer est une maladie qui fait peur :

Alzheimer : une maladie qui fait peur

Elle menace des fonctions au centre des fondements de l'être humain



Elle peut être considérée comme un véritable tsunami puisqu'on estime qu'une personne sur 85 peut souffrir de la maladie d'Alzheimer en 2050³², et qu'elle atteint 5% de la population âgée de plus de 65 ans et 30% des personnes de plus de 85 ans. Avec cette maladie, on est face à une maladie troublante à plus d'un titre :

- elle pourrait être une « destinée » naturelle liée à l'âge, suite à un très long processus et donc sans être liée au vieillissement de la population ;
- il est difficile de définir son étiologie ;
- c'est une maladie du temps, présentant une grande complexité ;
- c'est une maladie qui efface progressivement la mémoire et avec laquelle il faut apprendre à vivre, tant pour le patient que pour son entourage ;
- c'est une maladie lente et progressive qui a des débuts insidieux.

La mémoire, première concernée par la maladie d'Alzheimer, est multiple. Elle peut être, notamment, implicite ou explicite, épisodique ou sémantique :

- La mémoire explicite garde les événements liés à l'apprentissage et regroupe l'histoire personnelle du sujet, son savoir, sa culture. La mémoire implicite permet de réaliser des tâches automatiques, par acquisition de savoir-faire ou par réaction à des émotions.
- La mémoire épisodique est la mémoire des événements et de leurs relations temporelles et spatiales. La mémoire sémantique est constituée d'éléments de mémoire se rapportant aux connaissances générales organisées en catégories (objets, faits, règles, concepts, propositions).

³² R. Brookmeyer, E. Johnson, K. Ziegler-Graham, H. M. Arright, « Forecasting the global burden of Alzheimer's disease », *Alzheimers & Dementia* 3, 2007, pp. 186-191.

Il en résulte qu'on n'évitera pas la maladie d'Alzheimer en faisant des sudokus, pour la bonne raison que nous avons plusieurs mémoires et que celles-ci nécessitent des apprentissages spécifiques. On peut ajouter que l'acte de mémorisation est sélectif et qu'il n'y a pas d'acte de mémoire sans oubli.

I-1 La maladie d'Alzheimer

La maladie d'Alzheimer est une affection neurodégénérative du système nerveux central qui se manifeste généralement à un âge avancé de la vie. Au plan clinique, elle se traduit par des troubles cognitifs (mémoire, orientation, langage, reconnaissance, programme gestuel) mais aussi par des troubles moteurs et psycho-comportementaux.

Grands domaines symptomatiques de la maladie d'Alzheimer



Les symptômes apparaissent d'abord de manière insidieuse, puis s'aggravent au fil des mois et des années, entraînant progressivement ce que les médecins appellent une perte d'autonomie³³. La première zone du cerveau atteinte est

³³

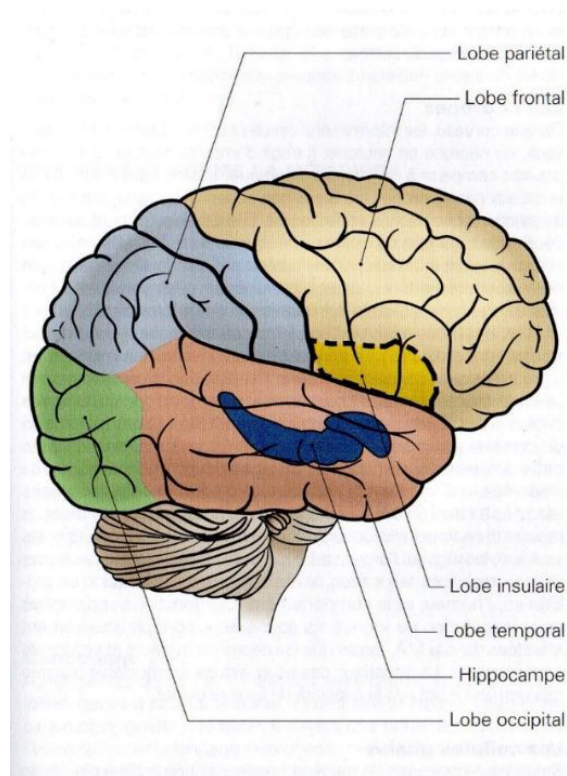
Fabrice Gzil, *La maladie d'Alzheimer : problèmes philosophiques*, Paris, Presses

l'hippocampe, en quelque sorte le péage de l'autoroute des souvenirs, si bien que la personne n'arrive plus à « fabriquer » de nouveaux souvenirs. Puis tout se propage dans le reste du cerveau et la maladie d'Alzheimer achève d'isoler le patient. Toutes les régions cérébrales n'ont pas la même vulnérabilité, mais les circuits les plus vulnérables sont ceux qui sous-tendent les fonctions mnésiques.

Le cerveau ne pèse que 1,3 kg à 1,4 kg, mais il comporte environ 100 milliards de neurones. Le cortex cérébral, la partie périphérique du cerveau, est le principal siège des fonctions mentales supérieures comme les fonctions cognitives ou la motricité. Il est subdivisé en cinq lobes cérébraux³⁴ :

- Le lobe frontal, situé à l'avant du cerveau, intervient dans le raisonnement, la planification, le langage et le mouvement.
- Le lobe pariétal, situé derrière le lobe frontal, est essentiellement consacré aux différentes perceptions sensorielles (goût, toucher, température, douleur).
- Le lobe occipital, situé à l'arrière du cerveau, permet de reconnaître et d'identifier les choses.
- Le lobe temporal, situé au-dessous des autres régions, est impliquée dans l'audition, le langage et la mémoire. C'est dans cette région qu'est l'hippocampe, une structure cérébrale essentielle au bon fonctionnement de la mémoire à long terme.
- Le lobe insulaire, situé derrière le lobe temporal, est impliqué dans la charge émotionnelle des souvenirs.

³⁴ Universitaires de France, 2009, page 2.
Fadi Massoud et Alain Robillard, *La maladie d'Alzheimer*, Annika Parance Editeur, Montréal, 2013, page 40.



Il est utile d'ajouter que la maladie d'Alzheimer affecte à la fois l'autonomie exécutionnelle (la capacité à effectuer par soi-même un certain nombre d'actes de la vie quotidienne) et l'autonomie décisionnelle (la capacité à se déterminer par soi-même en connaissance de cause).

En première analyse, quelques caractéristiques peuvent résumer la maladie d'Alzheimer :

- C'est, rappelons-le, une maladie dégénérative du cerveau dont la physiopathologie fait encore débat.
- C'est une maladie chronique dont l'évolution est fluctuante et très variable d'un individu à l'autre.
- C'est une maladie qui peut avoir des facteurs génétiques, mais pour l'évolution de laquelle la qualité de l'environnement et de l'accompagnement des personnes est déterminante.
- C'est une maladie dont le diagnostic n'est pas clair ainsi que nous l'analyserons.

La maladie d'Alzheimer s'installe donc et s'aggrave progressivement. Pendant la phase silencieuse de la maladie d'Alzheimer - rappelons qu'il peut se passer de dix à trente ans avant les premiers symptômes - la maladie progresse lentement et le cerveau utilise ses ressources et sa plasticité pour compenser la mort des neurones, utilisant en quelque sorte une « réserve cognitive » propre à chaque personne et retardant les premiers signes de la maladie. On peut néanmoins distinguer plusieurs phases³⁵ :

- un stade asymptotique ou pré-clinique pendant lequel les atteintes cérébrales seraient présentes sans traduction clinique ;
- un stade dit « prodromal » pendant lequel les symptômes commencent à apparaître, sans altérer significativement l'autonomie de la personne dans sa vie quotidienne ;
- un stade dit « démentiel »³⁶ où les troubles (légers, modérés puis sévères) retentissent de plus en plus gravement sur l'autonomie physique et psychique de la personne.

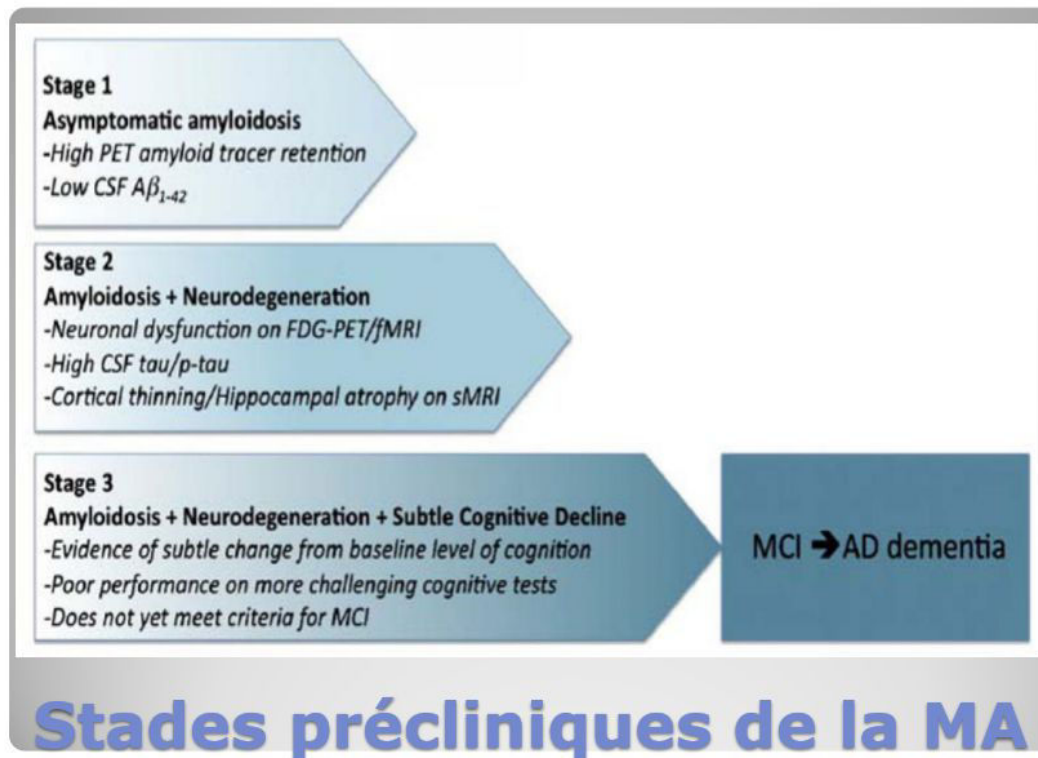
Le stade pré-clinique aboutit à ce qu'on appelle le *Mild Cognitive Impairment* (MCI). C'est un label utilisé pour caractériser la perte de mémoire d'une personne qui ne présente aucun signe de perte d'autonomie ou de démence. Il a été introduit par Ronald Petersen, un neurologue au Mayo Clinic in Rochester, Minnesota³⁷.

³⁵ Fabrice Gzil, Emmanuel Hirsch, *Alzheimer, éthique et société*, Paris Editions Eres, 2012, page 631.

³⁶ Il est à noter que la démence, ce n'est pas la folie, c'est la perte totale d'autonomie.

³⁷ Petersen R. C. and al., « Definition, Course, and Outcome of Mild Cognitive Impairment », *Aging, Neuropsychiatry and Cognition*, Vol. 3, 1996, pp : 141-147.

Le stade pré-clinique se répartit lui-même en plusieurs phases :



Ce pourra être un élément de diagnostic, mais doit-il être annoncé ? Les personnes MCI, certes, sont des personnes qui viennent consulter car elles ressentent des difficultés mnésiques et/ou cognitives, mais certaines vont évoluer vers une maladie d'Alzheimer, d'autres vont rester stables, d'autres encore vont récupérer et retrouver des performances cognitives dans les normes. Aujourd'hui il ne peut être annoncé à un patient que comme un facteur de risque nécessitant une surveillance particulière. Mais, et nous y reviendrons, cela montre l'importance de détecter le moment ou les conditions qui font que l'évolution « bifurquera » vers la maladie d'Alzheimer.

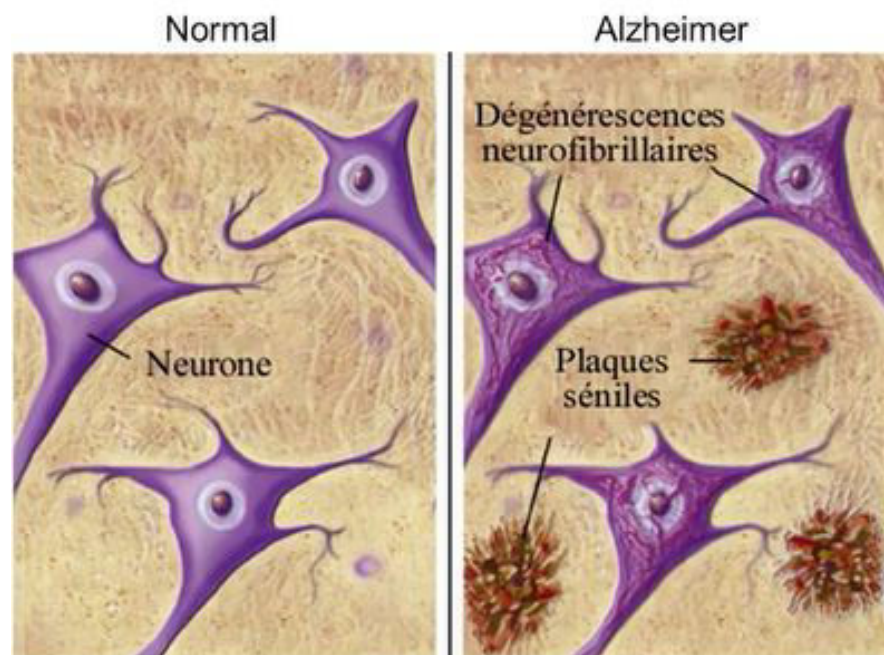
Il est en effet de plus en plus admis que les processus physiopathologiques de la maladie d'Alzheimer commencent plusieurs décennies avant l'apparition des signes cliniques et peuvent même exister sans que la maladie ne se manifeste par le moindre signe clinique ce qui posera la question de l'existence de

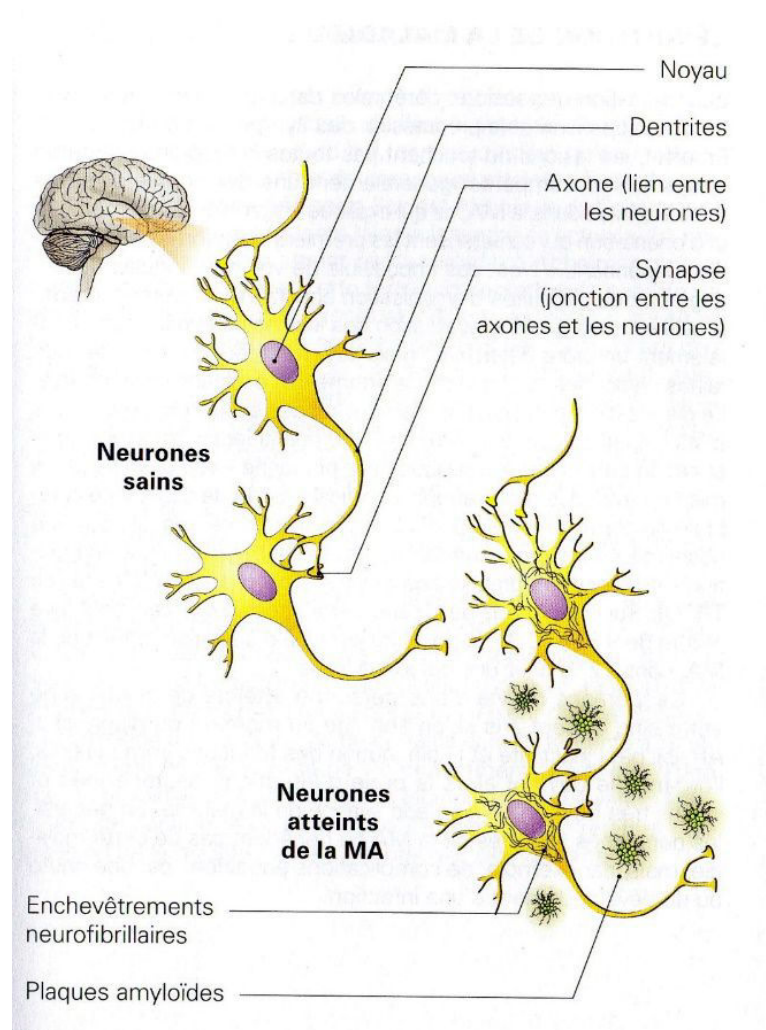
son commencement ou d'un événement déclencheur des signes cliniques.

I-2 Les lésions et les facteurs de risque

Les lésions de la maladie d'Alzheimer au plan cérébral sont connues. On reconnaît la maladie à deux lésions caractéristiques :

- les plaques séniles ou plaques amyloïdes (des dépôts extracellulaires, résultant de l'agrégation d'un peptide nommé bêta-amyloïde). Elles siègent entre les neurones ;
- les dégénérescences neurofibrillaires (un enchevêtrement des fibrilles intraneuronales résultant de l'hyperphosphorylation d'une protéine nommée Tau). Ces lésions conduisent à la perte neuronale et synaptique. Les dégénérescences neurofibrillaires affectent surtout les grands neurones pyramidaux de l'hippocampe et des couches corticales 3 et 5, à l'origine des connexions cortico-corticales et cortico-sous-corticales.

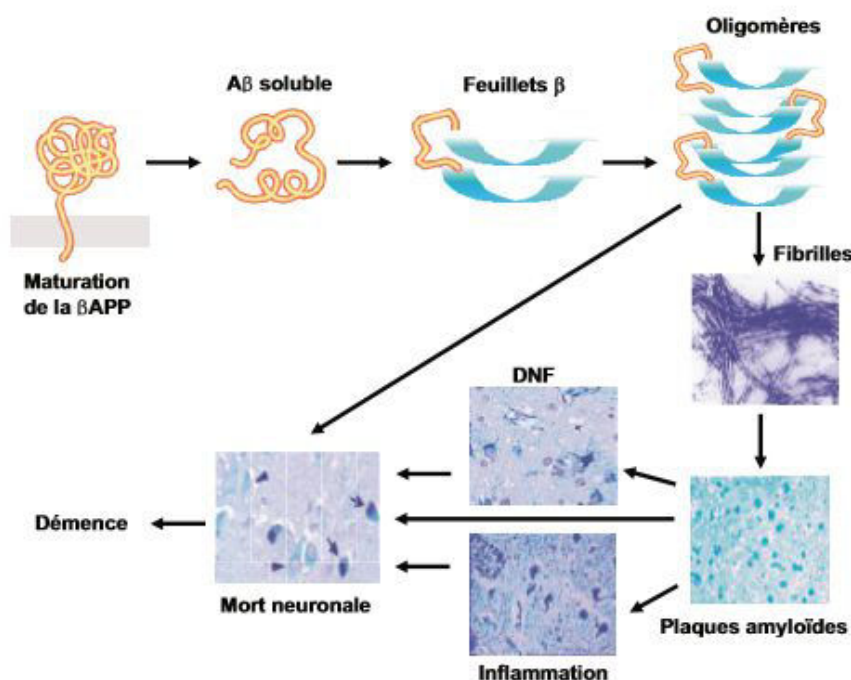




On considère ainsi que le mécanisme de la maladie d'Alzheimer découle de l'interaction délétère des deux protéines tau et amyloïde, mais la question de « la poule et de l'œuf » se pose :

- Pour certains, c'est la protéine amyloïde qui enclencherait le processus de dégénérescence neurofibrillaire. On parle alors de « cascade amyloïde ».
- Pour d'autres, la protéine tau serait le véritable responsable de la maladie clinique, la protéine tau anormale passant de neurone en neurone par l'intermédiaire des synapses.

L'explication dominante de la maladie d'Alzheimer reste, même si elle est contestée, l'hypothèse de la cascade amyloïde proposée en 1992 par John Hardy³⁸ :



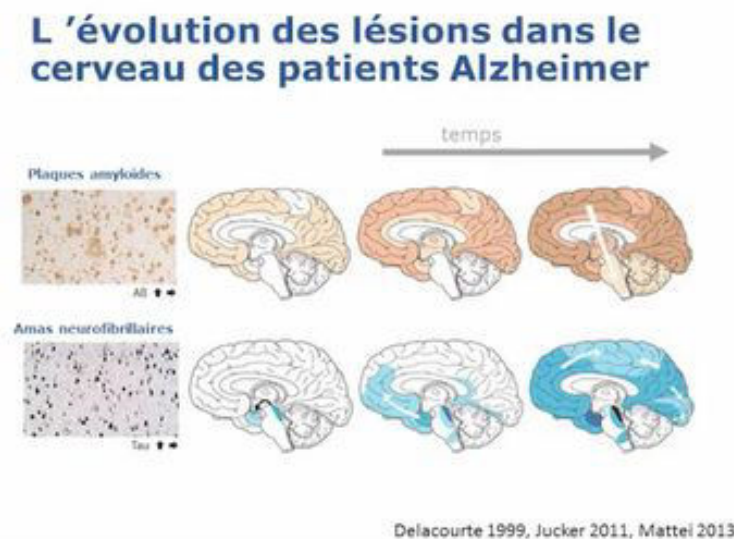
En fait, les deux fonctionnent sans doute en synergie et, pour reprendre la métaphore proposée par Bernard Croisile, « la pathologie tau met en place des tueurs au sein des neurones, mais l'activation de ces agents dormants est déclenchée par la pathologie amyloïde, qui joue le rôle du chef de gang »³⁹. Les plaques amyloïdes sont nécessaires au développement d'un Alzheimer, mais seules les dégénérescences neurofibrillaires, par leur topographie, leur étendue et leur progression sont corrélées à la sévérité de la démence.

Cependant, les lésions neuropathologiques de la maladie d'Alzheimer se développent bien avant le stade clinique de la

³⁸ John A. Hardy and Gerald A. Higgins, « Alzheimer's Disease : The Amyloid Cascade Hypothesis », *Science* 256, 1992, pp : 184-185.

³⁹ Bernard Croisile, *Alzheimer : Que savoir ? Que craindre ? Qu'espérer ?*, Paris, Editions Odile Jacob, 2014, page 68.

démence, quinze à vingt ans plus tôt⁴⁰ en moyenne et, dès l'âge de 47 ans, 50 % des personnes « normales » présentent déjà dans certaines régions de leur cerveau, des plaques amyloïdes⁴¹. A ce stade, la maladie, si on peut la juger comme telle, est parfaitement asymptomatique et peut parfaitement le rester. Aussi, la question de savoir si la plaque amyloïde est à l'origine du processus dégénératif reste ouverte même si la principale hypothèse étiopathologique reste souvent formulée autour de la cascade amyloïde, l'oligomérisation du peptide Abêta (bêta-amyloïde) étant un élément clé alors que la protéine Tau serait un élément nécessaire mais non suffisant.



Si on connaît assez bien les lésions, si de nombreux facteurs de risque⁴² sont reconnus, et si des mutations peuvent entraîner une augmentation de la quantité totale des différentes formes du peptide amyloïde Abêta, et en particulier une production préférentielle de la forme A β 42 reconnue comme responsable de la

⁴⁰ C. M. Hulette et al., « Neuropathological and neuropsychological changes in 'normal aging' : Evidence for preclinical Alzheimer disease in cognitively normal individuals », *J. Neuropathol. Exp. Neurol.*, 1998, 57, pp. 1168-1174.

⁴¹ Bruno Dubois, *Institut de la Mémoire et de la Maladie d'Alzheimer, Hôpital de la Pitié-Salpêtrière*, conversation privée.

⁴² Un facteur de risque n'est qu'un élément qui favorise la survenue d'une maladie sans en être directement responsable.

formation des dépôts amyloïdes⁴³, ce qui manque c'est la compréhension du déclenchement et deux approches se complètent, soit l'apparition de la maladie comme résultat de l'interaction d'éléments génétiques, soit l'apparition de la maladie comme résultat de facteurs liés à l'environnement, comme le mode vie.

En résumé, comme la maladie d'Alzheimer se développe au moins 15 ou 20 ans avant l'apparition des premiers signes visibles, tout l'enjeu est de pouvoir détecter les personnes malades – ou même à risque – avant les premiers symptômes. Pour cela la communauté scientifique et le monde médical placent beaucoup d'espoir dans la collecte de toutes les données pouvant être utiles à cette détection précoce, et portent tout particulièrement leur regard sur ce qu'on appelle le *Big Data*, ces milliards de données détenues par les géants de l'internet et des nouvelles technologies comme Google, mais également les données collectées dans les imageries médicales, par exemple. On peut espérer, en effet, qu'en combinant toutes les informations sur le cerveau des gens, avant les symptômes, on va pouvoir prédire s'ils sont sur la trajectoire de la maladie. Mais avant cela, il faut introduire une première approche des biomarqueurs.

⁴³ M. C. Chartier-Harlin, L. Araria-Goumidi, J.-C. Lambert, « Les formes tardives de la maladie d'Alzheimer : de la génétique à la biologie », *Médecine/Sciences*, 2002, , vol. 18, pp. 709-716.

Chapitre II

Les biomarqueurs, une première approche

Etant donné que les changements neurodégénératifs associés à la maladie d'Alzheimer apparaissent de nombreuses années avant l'apparition des signes cliniques, des biomarqueurs, dont nous devons analyser la signification, sont envisagés pour guider des interventions précoces⁴⁴. Le modèle dominant considère que la bêta amyloïde commence à s'agréger dans le cerveau de nombreuses années avant l'apparition de signes cliniques et que ces dépôts peuvent être mis en évidence par des techniques d'imagerie comme l'imagerie par émission de positrons ou des analyses du fluide céphalorachidien⁴⁵. Mais avant cela, il est nécessaire de dégager les différentes définitions de ce que sont les biomarqueurs.

De fait, il existe plusieurs définitions du biomarqueur⁴⁶. Si l'on reprend la définition proposée par les NIH (*National Institutes of Health*) américains, les biomarqueurs sont simplement définis comme une caractéristique mesurée objectivement et évaluée comme indicateur de processus physiologiques ou pathologiques, voire comme indicateur de l'action de médicaments.

⁴⁴ B. Dubois et al., « Revising the definition of Alzheimer's disease : a new lexicon », *Lancet Neurology* 9, 2010, p. 1118-1267.

⁴⁵ Le liquide céphalorachidien est le liquide biologique transparent dans lequel baignent le cerveau et la moelle épinière.

⁴⁶ K. Strimbu, J. A. Travel, « What are Biomarkers ? », *Curr. Opin. HIV AIDS*, 5(6), 2010, pp : 463-466.

Cela introduit une première distinction entre :

- Les biomarqueurs de diagnostic, notamment ceux qui permettent des diagnostics précoces des maladies comme la maladie d'Alzheimer qui présente une longue phase asymptomatique.
- Les biomarqueurs de pronostic qui permettent de prédire l'évolution d'une maladie chronique longue durée pour laquelle aucun traitement thérapeutique n'est disponible.
- Les biomarqueurs prédictifs pour prédire l'évolution d'une maladie suite à un traitement thérapeutique.
- Les biomarqueurs pharmacogénétiques associés aux réponses aux médicaments.

Ces biomarqueurs peuvent être de simples variants génétiques, mais aussi le résultat de la sur- (ou sous-) expression d'un gène, tout autant que la concentration de métabolites, considérée comme anormale par rapport à une concentration de référence.

Le développement de la science des gènes a fait passer de l'échelle du phénotype (et de la sélection) à l'échelle du génotype (et de la mutation). Ainsi les variants génétiques sont le plus souvent des mutations ponctuelles (*Single Nucleotide Polymorphisms* ou SNPs). Une mutation ponctuelle substitue un nucléotide (par exemple l'adénine) à un autre (par exemple la thymine), à un endroit déterminé du génome. Il y en a environ⁴⁷ trois millions, réparties sur l'ensemble du génome, dont des parties codantes, ce qui peut conduire, après les phases de transcription et de traduction, à des protéines différentes. Les mutations de

⁴⁷ Le nombre de SNPs est encore mal connu. Pour certains, on retrouve une mutation ponctuelle en moyenne toutes les 300 bases, ce qui conduit à 10 voire à 15 millions de SNPs au niveau populationnel. On peut ajouter qu'entre le génome de deux individus sans lien de parenté il y a environ 2 millions de SNPs (données fournies par Béatrice Desvergne, de l'Université de Lausanne, lors d'un séminaire au Centre Cavallès, à l'ENS, le 13 décembre 2016).

l'ADN constituent la source la plus connue de variation en biologie. Si ces mutations ne sont pas silencieuses, elles changent le fonctionnement de l'organisme considéré. Les SNPs sont souvent privilégiés dans la recherche de biomarqueurs⁴⁸. Ils sont généralement détectés dans des études d'association qui comparent une population de patients atteints d'une certaine maladie à une population de référence.

Si les biomarqueurs, ont un rôle privilégié⁴⁹, ils sont aussi très divers. Cette diversité peut se traduire par des données concernant l'ADN, mais aussi par un niveau d'expression d'un ensemble de gènes, la présence de diverses protéines, des concentrations de métabolites, ou une activité d'enzymes comme les kinases. Ce sont des molécules biologiques présentes dans le sang, dans les liquides corporels et les tissus organiques, ou dans les tumeurs elles-mêmes dans le cas des cancers, qui peuvent témoigner d'un processus normal ou anormal, ou de la présence d'un trouble ou d'une maladie.

A titre d'illustration, parallèlement à l'étude des mutations génétiques, les phénomènes d'épigénétique sont d'autres mécanismes sources de variations. Il est difficile de donner une définition unique de ce qu'est l'épigénétique⁵⁰ :

- Historiquement, ce terme a été introduit pour désigner tous les mécanismes mis en jeu pour passer du génotype au phénotype.
- Puis l'épigénétique est devenue l'étude des modulations de l'expression génétique à travers les modifications de la chromatine (l'ADN et son squelette protéique), notamment par méthylation des protéines histones.

Une des caractéristiques principales des modifications épigénétiques de la chromatine est qu'elles vont se transmettre de génération en génération.

⁴⁸ H.-J. Garchon, P. Dieudé, « Les biomarqueurs : en génomique, comment cela fonctionne-t-il ? », *Revue du Rhumatisme*, 78 (2011), 165-168.

⁴⁹ J. M. Meyer & G. S. Ginsburg, « The path of personalized medicine », *Current Opinion in Chemical Biology*, 2002, 6(4), pp. 434-438.

⁵⁰ E. Jablonka & M. J. Lamb, « The changing concept of Epigenetics », *Annals of the New York Academy of Sciences*, 981, 2002, p. 82-96.

L'épigenèse contribue à rebattre les cartes entre le rôle des gènes et celui de l'environnement. L'action des gènes se trouve sous la dépendance de facteurs qui leur sont extérieurs. L'épigenèse oblige à penser les liens entre inné et acquis, nature et environnement, sous forme d'une interaction complexe entre de multiples niveaux d'organisation. Le préfixe grec « épi » signifie « sur », l'épigénétique signifie donc qu'il y a quelque chose au-dessus de la génétique, elle se penche sur les facteurs, internes ou externes à l'individu, qui modulent l'expression des gènes. Ainsi, les modifications épigénétiques modifient la quantité d'expression d'un gène donné alors que les mutations de l'ADN modifient la structure des protéines et donc l'effet qualitatif résultat de l'expression du gène.

Nous en discuterons plus loin quand nous nous interrogerons sur la pertinence de la question de la causalité, mais le plus souvent, il n'y a pas de lien de causalité direct entre un biomarqueur et une maladie. Le biomarqueur est une altération, qui a une signification biologique, mais il ne renvoie pas à une situation préalable non-altérée. Nous avons déjà fait état de deux scénarios possibles. Dans le premier, les facteurs de risque ont toujours existé et c'est l'évolution de leurs concentrations qui aboutit aux signes cliniques sans qu'il soit possible d'en repérer l'origine. Dans le second, un événement extérieur ou l'apparition d'un nouveau facteur de risque crée une bifurcation se traduisant par une évolution plus rapide des facteurs de risque et donc dans l'apparition plus rapide des signes cliniques de la maladie d'Alzheimer. Cette altération décisive, si elle modifie le processus vers la maladie n'est cependant rien prise isolément. On peut d'ailleurs ajouter que certains biomarqueurs peuvent être représentatifs de plusieurs maladies – et à ce titre ils ne sont pas spécifiques – et qu'inversement une maladie ne peut être

diagnostiquée que si on détecte plusieurs biomarqueurs. Suivant Michel Foucault, on pourrait dire que, pour un biomarqueur, « Le propre du savoir n'est ni de voir ni de démontrer mais d'interpréter »⁵¹, en notant que la multiplication des biomarqueurs va amener des difficultés d'interprétation. Ajoutons néanmoins que leur présence, même en très faible quantité, peut permettre un diagnostic précoce d'une anomalie, simplement en les détectant, même si leur signification n'est pas clairement connue. On peut en effet admettre qu'un test basé sur un biomarqueur peut être utile sans qu'on ait besoin d'avoir une connaissance, ne serait-ce que partielle, de la fonction biologique associée à ce biomarqueur.

Quel qu'il soit, un biomarqueur doit faire l'objet d'un processus de développement comportant deux grandes phases :

- Une phase de validation permettant d'évaluer sa sensibilité, sa spécificité, sa reproductibilité, ce qui suppose, dans un premier temps d'identifier un biomarqueur candidat, puis de le caractériser et d'analyser ses liens avec la pathologie étudiée.

La sensibilité désigne la probabilité d'obtenir un test positif pour un individu porteur de la maladie, tandis que la spécificité désigne la probabilité d'obtenir un test négatif sur un individu non porteur de la maladie. Quant à la reproductibilité, elle vise à s'affranchir des effets aléatoires venant fausser les résultats.

- Une phase de qualification qui concerne le passage de la recherche médicale ou clinique à la routine clinique.

Pour cela, de nombreuses techniques peuvent être utilisées. Elles sont présentées dans le chapitre suivant, avant d'en analyser les limites et les controverses qui peuvent être associés à la non-prise en compte de facteurs comme les facteurs environnementaux et

⁵¹ Michel Foucault, *Les mots et les choses*, Paris, NRF Gallimard, 1966.

socioéconomiques. Nous pouvons cependant retenir dès maintenant que, selon le niveau d'analyse considéré (génome, transcriptome, protéome, métabolome, etc.), les biomarqueurs seront de types différents et ne différencieront pas les personnes les unes des autres (ainsi, si des variants génétiques singularisent une personne, des marqueurs sériques, détectés dans le sang, ne différencient pas deux personnes).

Le degré d'information diffère selon la nature du biomarqueur et les messages qu'il porte sont plus ou moins validés. Tout cela peut expliquer la difficulté du passage de la biologie à la médecine, la difficulté de leur application en médecine. La problématique du niveau de preuves de ces biomarqueurs est, en effet, essentielle. De fait, un biomarqueur est plus qu'une molécule. C'est plutôt un signal dont les variations présentent une corrélation statistique significative avec une pathologie et ces variations doivent pouvoir être détectées grâce un test validé et utilisable en clinique. Un biomarqueur a donc deux composantes, une composante biologique associée à ses variations et une composante pratique qui permet son utilisation en clinique. Cela implique la mesure simultanée d'un nombre important d'entités, suivant l'état de l'individu et une analyse faisant appel aux statistiques, tout en soulignant, dès maintenant, que les notions de moyenne et de variance, issues de statistiques sur une large population, n'ont pas de sens pour un individu particulier. Ce qui est important c'est d'intégrer⁵² toutes les informations, qu'elles soient quantitatives ou qualitatives. L'objectif final est de quantifier l'incertitude inhérente à chaque individu afin d'arriver à une plus grande certitude en médecine, en se basant sur des données scientifiques.

⁵² S. Huang, « The practical problems of post-genomic biology », *Nature Biotechnology*, 2000, 18, pp. 471-472.

Enfin, pour suivre Xavier Guchet⁵³, le biomarqueur peut avoir une double signification :

- Ce peut être un biomarqueur « actionnable » en ce sens qu'il permet le diagnostic, le pronostic et éventuellement l'intervention thérapeutique en identifiant des cibles moléculaires sur lesquelles agir.
- Ce peut être un biomarqueur « trace » qui intervient dans des cascades d'interactions moléculaires, suivies ainsi à la trace, qui résultent tant de phénomènes internes que de phénomènes externes liés à des événements naturels ou sociaux.

Il y a un « jeu » entre les caractéristiques génétiques, voire immunologiques, spécifiques de l'individu et l'environnement dans lequel il se trouve qui aboutit à un phénotype qui, du point de vue médical se traduit en paramètres physiopathologiques affectant l'individu. Nous retrouverons l'intérêt de s'intéresser à la biologie des systèmes, en relevant qu'il y a différents niveaux d'analyse, le réseau moléculaire et le réseau des organes par exemple.

La biologie des systèmes a pour objectif de comprendre les fonctions physiologiques d'un organisme à partir des mécanismes moléculaires, en parcourant les différents niveaux d'organisation de cet organisme. Elle suppose donc que le fonctionnement biologique implique plusieurs niveaux d'organisation, qu'il n'y a pas un niveau privilégié qui pourrait tout expliquer et que le niveau moléculaire, notamment le rôle des gènes, ne peut être compris qu'en prenant en compte les niveaux d'organisation supérieurs. Elle vise à réunir génomique, physiologie et connaissance clinique de l'individu et de son histoire.

⁵³ Xavier Guchet, *La médecine personnalisée, un essai philosophique*, Paris, Les Belles Lettres, 2016.

Cela revient à envisager les constituants moléculaires, non plus isolément, mais dans leur contexte, au sein d'un environnement, ce qui peut se résumer par une équation du type :

« Génome * Exposome = Phénomène »

Le concept d'exposome, construit par analogie au génome, illustre la nécessité de disposer de données sur l'environnement auquel peut être soumis une personne⁵⁴. L'exposome est ainsi l'ensemble des expositions auxquelles est soumis un individu, que ce soit dans son environnement ou dans son alimentation ou son mode de vie. Ce concept a été lui-même divisé en trois catégories⁵⁵ par Christopher Paul Wild qui, le premier, a introduit le concept d'exposome :

- l'environnement interne à l'organisme, notamment les concentrations en métabolites et en hormones ;
- l'environnement externe spécifique qui se réfère à des éléments comme l'exposition à des agents infectieux, à des rayonnements ou à différents polluants ;
- les facteurs externes généraux, notamment tout ce qui touche aux aspects socio-économiques et culturels.

Considérer le phénotype comme résultant d'une combinaison du génotype et de l'environnement⁵⁶ nécessite de développer des méthodes pour rassembler et traiter des informations sur l'environnement de la personne. Mais, tout cela conduit à des données. Elles peuvent se distinguer selon leur

⁵⁴ C. P. Wild, « Complementing the genome with an « exposome » : the outstanding challenge of environmental exposure measurement in molecular epidemiology », *Cancer Epidemiol. Biomarkers Prev* 2005 ; 14(8), pp.1847-1850, Epub 2005/08/17.

⁵⁵ C. P. Wild, « The exposome : from concept to utility », *Int. J. EPIDEMIOL.*, 2012 ; 41(1), pp. 24-32, Epub 2012/02/03.

⁵⁶ Dans l'équation, nous avons mis un « * » et non un « + » pour illustrer qu'il y a bien des interactions et non seulement des effets additifs. Nous y reviendrons.

caractère structuré ou non⁵⁷ et selon leur niveau hiérarchique, données élémentaires ou métadonnées. La santé est un secteur qui concentre d'énormes volumes de données très hétérogènes : imageries, ordonnances, comptes rendus d'analyse ou d'intervention, données administratives, publications scientifiques, sans oublier, désormais, les commentaires sur les réseaux sociaux. Il faut croiser toutes ces données et se demander ce que ces données ont en commun, notamment quelles propriétés les caractérisent le plus, des propriétés « catégoriques » ou des propriétés « dispositionnelles » par exemple.

Cela permet de donner dès maintenant une précision sur les *Big Data*. Dans le domaine de la santé, le « *Big* » ne signifie pas plus de patients ou de personnes, mais plus de variables par personne.

Retenons, à ce stade que, plutôt qu'une molécule, un biomarqueur est un ensemble de molécules analysées avec des algorithmes de bioinformaticiens. Nous approfondirons ce point mais, qu'en est-il des techniques susceptibles d'être utilisées ?

⁵⁷ Ainsi, une base de données orientée objet est un ensemble de données structurées, tandis que des documents, des images, des sons ne sont pas a priori structurés du point de vue d'un système d'information.

Chapitre III

Le panorama des techniques

Avant de faire un point plus précis sur les biomarqueurs et sur leur apport dans les connaissances sur la maladie d'Alzheimer et sur leurs applications cliniques, un panorama des techniques est nécessaire. Les technologies pénètrent l'ensemble de la société et modifient profondément le regard sur le monde et les comportements, qu'ils soient individuels ou collectifs. Les technologies et l'innovation technique portent, en effet, en elles-mêmes des conceptions différentes du monde.

C'est le cas dans le domaine de la santé et de la médecine où on observe ce que nous pourrions qualifier de course technologique au service d'une investigation médicale toujours plus approfondie. Ainsi, les techniques et technologies ont profondément modifié les pratiques médicales et la manière de mieux comprendre les maladies. La médecine constitue le « milieu associé » qui les rendent efficaces et pertinentes comme le souligne Gilbert Simondon⁵⁸. Selon lui, il est impossible de définir les objets techniques en tant qu'objets de connaissance, sans introduire la question des relations qu'ils entretiennent avec un milieu associé. Il faudra donc analyser les liens des techniques entre elles, ainsi que les réseaux, les infrastructures et les équipes

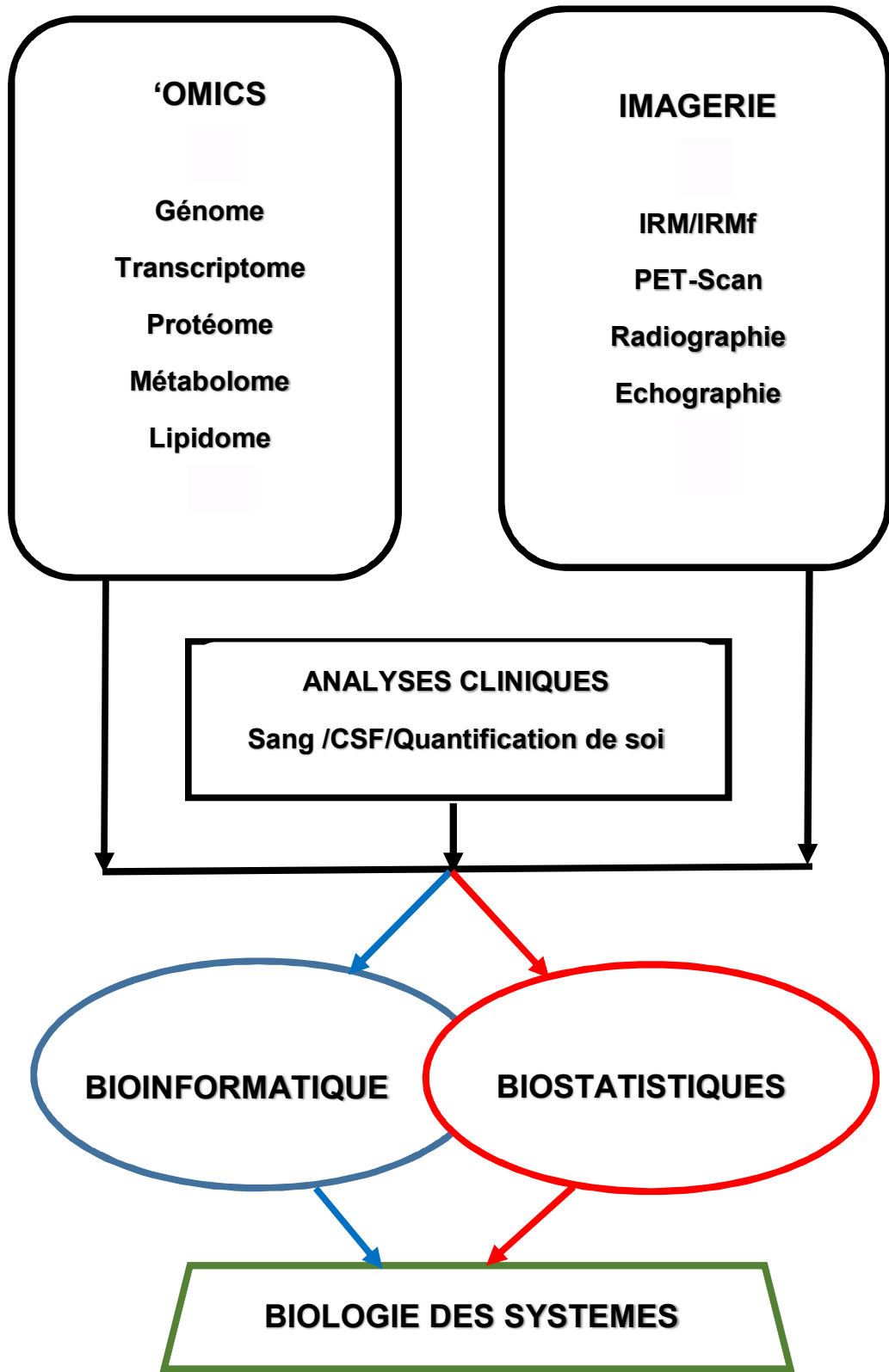
⁵⁸ Gilbert Simondon, *Du mode d'existence des objets techniques*, Paris, Aubier, 1989.

qui les utilisent. C'est rappeler que les questions, qu'elles soient épistémologiques ou éthiques, ne peuvent pas être bien posées si l'on fait abstraction, des pratiques d'une part, du rôle des institutions et des politiques d'autre part.

La médecine d'aujourd'hui est ainsi directement associée à des objets techniques car son développement passe par toutes les technologies d'acquisition des données, que ce soit des données biologiques acquises à très haut débit ou d'autres types de données, et par l'ensemble des outils informatiques qui permettent d'analyser et d'interpréter ces données. Une technique définit par elle-même ce que l'on peut obtenir ou appréhender ou regarder. Aussi, la complémentarité des méthodes et des techniques s'est imposée et il est nécessaire de comprendre comment les données sont produites et comment elles circulent et sont utilisées. Il convient donc de faire un panorama des objets techniques pour identifier quelles sont les sources de données et les outils d'analyse qui les accompagnent.

Les technologies pour la santé sont très diverses⁵⁹. Outre les technologies qui permettent d'observer et de comprendre les mécanismes du vivant, elles recouvrent l'ensemble des technologies utilisées sur le plan clinique, dont l'imagerie, la bio-ingénierie ou le traitement informatique des données. Elles sont aujourd'hui complétées par l'introduction de données en très grand nombre et l'utilisation des mathématiques, et face à la diversité et à la quantité des données, l'apport des statistiques et des probabilités a été considérable. Nous reviendrons sur ces points lors de l'analyse des questions épistémologiques. Nous nous limitons à ce stade à une présentation des principales techniques utilisées dans le domaine médical et de la santé, dont le panorama peut être schématisé de la façon suivante :

⁵⁹ <http://www.inserm.fr/thematiques/technologies-pour-la-sante/enjeux>



III-1 Les technologies « -omics »

Pour passer de caractéristiques biologiques moléculaires aux caractéristiques phénotypiques telles que l'apparition d'une pathologie, une cascade « -omics » s'enchaîne :

GENOME : 20 000 à 25 000 gènes



TRANSCRIPTOME : 100 000 transcrits



PROTEOME : 1 000 000 protéines

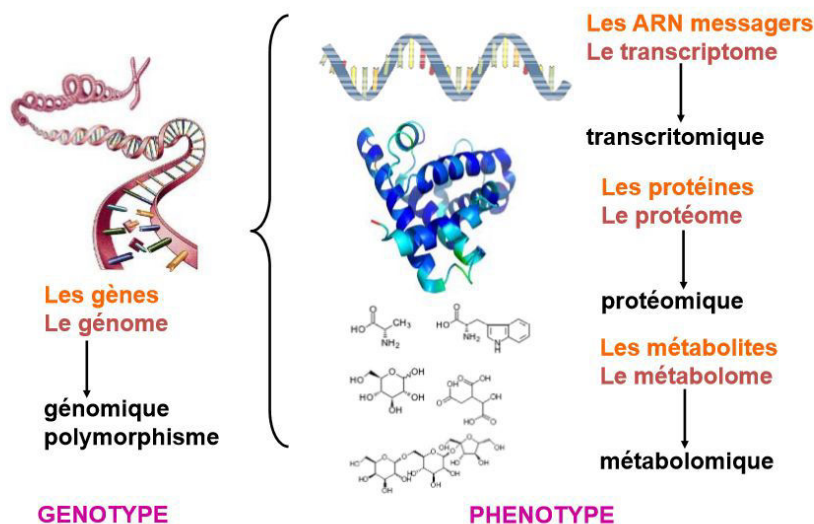


METABOLOME : 7 500 métabolites

Les techniques doivent permettre de rassembler des données au niveau des gènes contenus dans le noyau de chaque cellule (génome) mais aussi au niveau de l'expression des gènes (transcriptome) et de la traduction des gènes exprimés (protéome), pour ensuite modéliser les interactions (interactome et métabolome) et aboutir aux connaissances du vivant (biologie des systèmes, approche interdisciplinaire afin de comprendre des systèmes biologiques complexes).

Si la fonction d'un gène peut apparaître simple, la fonction des produits des gènes ne l'est pas et la réponse n'est pas unique. Seule « l'étude de ce qui gouverne l'expression d'un gène et sa traduction en une protéine ayant une certaine structure et une certaine fonction »⁶⁰ le permet. C'est ce qu'on appelle la post-génomique.

Si les approches massivement parallèles, les « -omics », permettent de caractériser l'ensemble des composants biologiques, l'ADN, les ARNs, les protéines, etc., il ne faut pas oublier que la complexité est de taille puisqu'un individu est constitué de près de 10^{14} cellules, que chaque cellule contient environ 22 000 gènes⁶¹, chaque gène codant en moyenne 6 ARNs messagers, ou ARNs codants, sans oublier des ARNs non codants faisant un vaste complexe d'interactions puisqu'on considère qu'entre deux gènes distincts plus de 360 millions d'interactions sont possibles. Quant aux protéines et peptides, on estime leur nombre entre 500 000 et 2 000 000.



⁶⁰ Jean Hache, *Les enjeux des biotechnologies*, Colombelles, Editions EMS, 2005, page 26.

⁶¹ Le nombre exact de gènes fait encore l'objet de discussions. On estime que, dans le génome humain, il y a entre 20 000 et 25 000 gènes codant pour des protéines.

Le séquençage du génome

L'acide désoxyribonucléique (ADN), est une macromolécule biologique, présente dans toutes les cellules, qui contient le génome, c'est-à-dire toute l'information génétique permettant le développement, le fonctionnement et la reproduction des êtres vivants. Les molécules d'ADN sont constituées d'un enchaînement de quatre nucléotides comportant une des quatre bases nucléiques, l'adénine (A), la cytosine (C), la guanine (G) ou la thymine (T). Séquencer le génome revient à déterminer l'ordre dans lequel se succèdent les nucléotides le long d'un brin d'ADN⁶².

Le séquençage de l'ADN par la méthode de synthèse enzymatique de Sanger a été utilisé pendant plus de 30 ans pour lire le code génétique des organismes vivants. La dernière génération des séquenceurs à capillaires, utilisant la technique Sanger, permet aujourd'hui de lire jusqu'à 2 millions de bases en une demi-journée.

Mais, dès l'année 2007, sont apparues sur le marché des machines dotées de débits de 50 à 1 000 fois supérieur. Ces séquenceurs dits de « nouvelle » génération ont permis de s'affranchir d'un certain nombre de biais de la méthode Sanger comme la nécessité de cloner l'ADN⁶³ à séquencer. La lecture de plusieurs millions de séquences en parallèle avec ces nouveaux séquenceurs à « haut débit » a considérablement accéléré les analyses en génomique⁶⁴.

⁶² Les molécules d'ADN sont formées de deux brins antiparallèles enroulés l'un autour de l'autre pour former une double hélice, la règle d'appariement étant un A face à un T et un C face à un G.

⁶³ Le clonage de l'ADN consiste à isoler un fragment d'ADN et à le multiplier à l'identique en l'insérant dans un vecteur, comme le bactériophage M13, permettant son amplification. C'est une opération longue et fastidieuse.

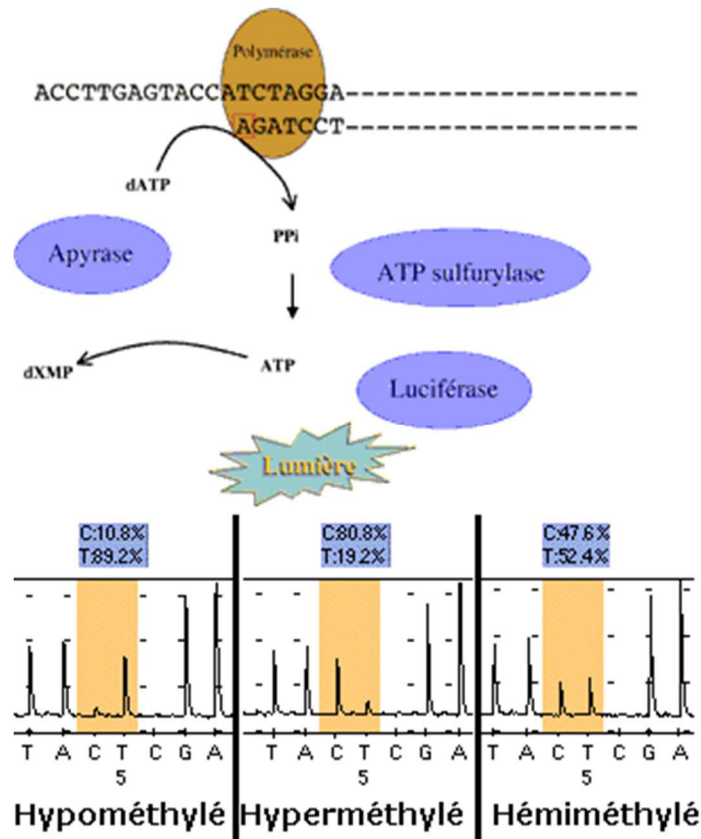
⁶⁴ M.L. Metzger, « Sequencing technologies - the next generation », *Nat Rev Genet*, 2010, 11, pp. 31-46.

Globalement les technologies de séquençage sont assez proches et fonctionnent en 3 étapes :

- la première consiste en la préparation et l'amplification des molécules d'ADN à analyser ;
- la seconde permet l'incorporation des bases complémentaires du brin à séquencer ;
- la dernière étape comprend la lecture de la séquence proprement dite.

Trois technologies sont apparues de façon quasi simultanée :

- le pyroséquençage, développé par Roche⁶⁵, assure la détection de la base incorporée à l'aide d'une réaction chemoluminescente, détectée par une caméra CCD. Cette technologie permet d'obtenir aujourd'hui jusqu'à 1 million de séquences pouvant atteindre jusqu'à 400 bases.



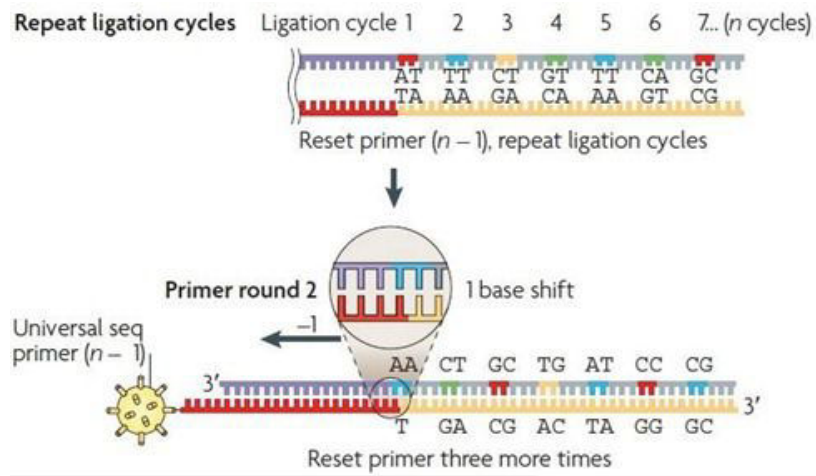
⁶⁵ <http://www.rochediagnostics.fr>

Les nucléotides ne sont pas ajoutés tous ensemble comme dans les réactions de séquence normales, mais l'un après l'autre. Si le nucléotide ajouté dans le milieu réactionnel correspond à celui attendu par la Polymérase, il est incorporé dans le brin en cours de synthèse en libérant un PyroPhosphate. Grâce à une ATP Sulfurylase, ce pyrophosphate est transformé en ATP qui est utilisé par une luciférase pour émettre un signal lumineux. C'est ce signal qui est capté par le séquenceur qui le reproduit sous la forme d'un pic sur le Pyrogramme. La hauteur de ce pic est fonction de l'intensité du signal lumineux, elle-même proportionnelle au nombre de nucléotides incorporés en même temps. On peut donc déduire la séquence de la taille des pics obtenus.

- le séquençage avec des terminateurs réversibles, développé par Illumina⁶⁶, réalise l'étape d'amplification sur un support solide sur lequel sont réalisées directement les réactions de séquence. L'extrémité des bases est protégée pour empêcher l'addition de bases supplémentaires à chaque cycle d'incorporation. Une lecture laser permet alors de détecter simultanément toutes les positions incorporées. Le clivage des fluorophores permet ensuite l'incorporation de la base suivante. La lecture est effectuée ainsi cycle après cycle. Cette méthode permet l'acquisition en parallèle de plus de 3 milliards de séquences de 100 bases de long.
- le séquençage par ligation, développé par Applied Biosystems⁶⁷, met en œuvre une réaction de séquençage par un système assez complexe de cycles de ligation et de clivage. Cette technique permet non seulement la lecture de la séquence mais inclut un système de correction des erreurs d'incorporations. Grâce à cette méthode, il est possible de lire jusqu'à un milliard et demi de séquences en parallèle de 75 bases de long.

⁶⁶ <http://illumina.com>

⁶⁷ <http://www.lifetechnologies.com>



Le processus consiste à ajouter par ligase et de manière répétitive, des blocs de 8 bases avec une paire de bases qui correspondront aux deux suivantes à synthétiser dans chaque brin d'ADN à séquencer (parmi 16 possibles, 4×4), précédée de trois bases et suivie de trois bases universelles et d'un marqueur au fluor qui seront éliminées. L'astuce consiste à ajouter pas à pas des séries de paires de bases détectées par fluorescence suivies de 3 bases universelles, d'enlever le tout et de recommencer l'opération quatre fois en la décalant d'une base, grâce à des amorces de taille différente (de 1 à 4 bases en moins que la première). Cela permet par recoupement d'identifier les bases de l'ADN et, qui plus est, de les mesurer chacune deux fois. Ce qui réduit le niveau des erreurs de lecture à 1 pour 1000, soit dix fois mieux que dans le pyroséquençage.

Le champ d'applications de ces nouvelles techniques est très vaste :

- l'analyse de nouveaux génomes, ou séquençage *de novo*, qui permet de travailler sur des génomes entiers ;
- le reséquençage, utilisé quand la séquence du génome de référence est déjà connue. Le séquençage à haut débit est alors employé pour connaître quelles sont les variations génomiques de l'échantillon qui est étudié en comparaison

avec celui pris comme référence. Ces approches sont certainement parmi les plus employées actuellement dans le domaine médical. A la frontière entre reséquençage et séquençage *de novo*, est aussi étudiée la métagénomique. Le but est de découvrir dans un mélange complexe l'ensemble des organismes qui le composent comme par exemple la flore intestinale⁶⁸. On peut ainsi imaginer qu'en séquençant le sérum de patients, il sera possible de déterminer quels sont les agents pathogènes qui y sont présents et qui sont à l'origine de la pathologie observée ;

- Le séquençage à haut débit est aussi utilisé pour les analyses fonctionnelles. Dans ce cas, l'important n'est pas de connaître la séquence d'ADN des échantillons, mais plutôt de quantifier le nombre et le type d'éléments biologiques présents. Il est par exemple possible de déterminer quelles quantités et quels types d'ARN s'expriment dans une cellule (RNA-Seq)⁶⁹, de détecter les phénomènes de transcription qui sont utilisés pour la régulation à travers les longs ARN non codants et toute la famille des petits ARN (Small RNA-Seq), de déterminer les modifications épigénétiques d'un génome en cartographiant ses sites de méthylation (MeDIP-Seq)⁷⁰.

Le séquençage à haut débit permet de s'affranchir des biais de l'hybridation, il dispose d'une gamme dynamique plus grande et la résolution des résultats obtenus est disponible à la base près. Ces outils sont donc intéressants pour déterminer les effets des molécules thérapeutiques lors de test *in vitro* sur des cultures de

⁶⁸ J. Qin, R. Li, J. Raes, M. Arumugam, K.S. Burgdorf, C. Manichanh, T. Nielsen, N. Pons, F. Levenez, T. Yamada et al., (2010) « A human gut microbial gene catalogue established by metagenomic sequencing », *Nature*, 2010, 464, pp.59-65.

⁶⁹ F.Ozsolak and P.M. Milos (2011) « RNA sequencing: advances, challenges and opportunities », *Nat Rev Genet*, 2011, 12, pp.87-98.

⁷⁰ P.W. Laird (2010) « Principles and challenges of genomewide DNA methylation analysis », *Nat Rev Genet*, 2010, 11 pp, 191-203.

cellules, *in vivo* sur des modèles animaux ou lors des essais cliniques. Le fait de pouvoir, soit être exhaustif sur l'ensemble de toutes les réponses génomiques possibles, soit cibler une région précise et passer un grand nombre d'échantillons en parallèle est un atout important pour la caractérisation et le criblage de biomarqueurs⁷¹.

Ces méthodes de séquençage à haut débit ont maintenant atteint la maturité technologique nécessaire à leur adoption large par la communauté scientifique et médicale. Elles ont conduit à la mise en place des plateformes de génomique très haut débit, présentées comme une mutualisation des moyens. En France, *France Génomique*⁷² rassemble et mutualise les ressources des principales plateformes françaises de génomique et de bioinformatique en donnant accès à un réseau de plateformes comme les plateformes nationales que sont le Génoscope et le Centre National de Génotypage regroupées au sein de l'Institut de Génomique du CEA à Evry. Toutes ces plateformes mettent en œuvre les nouvelles techniques de séquençage. À titre d'exemple, le Génoscope met en œuvre les trois méthodes présentées (séquenceurs *Roche*, *Illumina* et *Applied Biosystems*) et a participé, depuis sa création, à plus de 650 projets nécessitant le séquençage *de novo* de génomes et de métagénomes⁷³. En complément, des versions dédiées aux applications de diagnostic, comme le *GS Junior* de Roche, le *MiSeq* d'Illumina ou le *Ion Personal Genome Machine* d'Applied Biosystems sont maintenant sur le marché et de nouvelles générations d'appareils ambitionnent de lire directement la séquence de l'ADN sans avoir à l'amplifier et d'effectuer la lecture de l'ADN en temps réel.

⁷¹ H. Schwarzenbach, D.S. Hoon and K. Pantel(2011) « Cell-free nucleic acids as biomarkers in cancer patient », *Nat Rev Cancer*, 2011, 11, pp. 426-437.

⁷² <https://www.francegenomique.org>.

⁷³ <http://www.cns.fr>.

Le « Plan France Génomique 2025 »⁷⁴ cherche à répondre à quatre enjeux, un enjeu de santé publique (la médecine génomique révolutionne le parcours de soin), un enjeu scientifique et clinique (aller de l'exploration moléculaire d'une pathologie à un bénéfice thérapeutique pour le patient), un enjeu technologique (convergence des technologies de l'information avec les sciences de la vie et la santé), un enjeu économique (création d'une nouvelle filière industrielle).

Deux projets de plateformes, le projet SeQOIA en Ile de France et le projet AURAGEN en région Auvergne-Rhône Alpes, ont été retenus en juillet 2017. Elles devraient être opérationnelles dans moins d'un an et traiter 18 000 génomes par an d'ici 2019 sur chacune des deux plateformes. Plutôt centrées sur les cancers et les maladies rares, elles devraient servir au déploiement d'une douzaine d'autres plateformes d'ici cinq ans.

On peut également ajouter les séquenceurs basés sur des « nanopores »⁷⁵ qui permettent, à partir d'une analyse sur une molécule unique, de détecter un à un les nucléotides du brin d'ADN à séquencer en différenciant les signatures électriques spécifiques à chacune des quatre bases nucléiques.

Le génome personnel d'un individu représente approximativement 100 gigabytes de données, soit environ l'équivalent de 102 400 photographies, mais avec les technologies qui se sont développées, le coût d'obtention des données est inférieur aux coûts pour les stocker, les récupérer et les analyser. Obtenir l'information génétique intégrale d'un individu est devenu une réalité et son coût, le « génome à 1000 dollars »⁷⁶, rendra la technique applicable à grande échelle sur un marché de masse, même s'il faut relever que le coût du séquençage n'inclut pas le traitement des données.

Pour résumer, les nouvelles techniques de séquençage à très haut débit et à bas coût offrent de grandes perspectives dans

⁷⁴ « France Médecine Génomique 2025 », rapport de l'AVIESAN remis au Premier ministre le 22 juin 2016 (www.aviesan.fr).

⁷⁵ M. Wanunu, « Nanopores : A Journey Toward DNA Sequencing », *Physics of Life Review*, 9, 2012, 125-158.

⁷⁶ Kevin Davies, *The \$ 1000 genome, The Revolution in DNA Sequencing and the New Era of Personalized Medicine*, USA, Free Press 2010.

le domaine médical, mais il ne faut pas oublier qu'elles ne donnent des informations que sur le génome, l'ADN, et que le génome n'est qu'un niveau d'analyse. C'est pourquoi il faut s'intéresser à ce qui touche à l'expression des gènes, à leur traduction en protéines et à l'interaction des protéines entre elles dans des réseaux métaboliques qui mettent en œuvre d'autres technologies que les technologies de séquençage.

Le transcriptome

Le transcriptome est l'ensemble des ARN (messagers, ribosomiques, de transfert et autres espèces d'ARN) issus de la transcription du génome. La caractérisation et la quantification du transcriptome dans un tissu donné ou dans un type cellulaire, et dans des conditions données, permettent d'identifier les gènes actifs, de déterminer les mécanismes de régulation d'expression des gènes et de définir les réseaux d'expression des gènes. Les techniques couramment utilisées pour mesurer simultanément la concentration d'un grand nombre de types différents d'ARN messagers, sont généralement des puces à ADN (après reverse transcription) et, ainsi que présenté plus haut, le séquençage d'ARN à haut débit (RNA-seq). Les puces à ADN ont ouvert la voie à de nouvelles approches d'acquisition rapide de nombreuses informations biologiques. Par leur miniaturisation et leur automatisation, elles permettent d'analyser simultanément des milliers d'échantillons.

L'analyse de l'expression des gènes par les puces et les *microarrays* illustre les applications potentielles de ces technologies dans le domaine médical. Cela a conduit au développement de bases de données et de standards de présentation intégrant notamment les détails cliniques. Les *microarrays* permettent d'analyser simultanément l'activité de milliers de gènes. Les données obtenues avec des *microarrays* sont en effet,

généralement, des données d'expression relative et non absolue puisqu'elles résultent d'une analyse différentielle entre deux situations, une situation normale et une situation pathologique par exemple, tout en relevant qu'un individu normal n'existe pas, que ce n'est qu'un individu singulier et que donc il faut faire l'analyse différentielle entre un organe sain (des cellules saines) et un organe atteint (cellules tumorales) du même individu. Pour cela on peut, par exemple, séparer des cellules saines et des cellules tumorales d'un même échantillon de tissu, provenant d'un colon par exemple, à l'aide de techniques telles que la microdissection laser.

Pratiquement, de nombreux protocoles peuvent être utilisés pour obtenir le transcriptome. A titre d'illustration, peuvent être indiqués :

- l'extraction des ARN totaux par la technologie NucleoSpin dont une caractéristique importante est de prévenir la dégradation des ARNs en inhibant les enzymes de dégradation ;
- les bioanalyseurs d'ARN comme le RNA 6000 NanoChips qui permet de quantifier la qualité et l'intégrité des ARNs ;
- les analyseurs / lecteurs automatiques des hybridations comme l'Innoscan 900, associé au logiciel Mapix 5.0.0.

Ce qu'il faut retenir, c'est que le transcriptome, en permettant l'analyse de l'expression des gènes permet de passer du génome à l'épigénome⁷⁷. L'expression des gènes, étape intermédiaire essentielle pour passer du gène à la protéine, dépend non seulement de l'enchaînement des bases nucléiques du gène, mais aussi de l'environnement du gène, qui a de nombreuses composantes et dont l'étendue peut être variable. Pour illustrer, alors que le génome est le même dans toutes les cellules et qu'il ne dépend, sauf accidents, ni du temps (l'âge de l'individu), ni du lieu

⁷⁷ L'épigénétique (qui conduit à l'épigénome) désigne l'ensemble des mécanismes moléculaires qui affectent l'expression des gènes.

(l'organe), ni de l'état (sain ou pathologique) faisant en quelque sorte du génome une information statique, le transcriptome, au contraire, est dynamique puisque l'expression des gènes dépend de :

- l'âge (un nouveau-né, un adulte, un senior n'exprimeront pas les même gènes) ;
- l'organe (les gènes exprimés dans un neurone ne seront pas les mêmes que dans une cellule cardiaque) ;
- l'état de l'individu (les gènes s'exprimeront différemment dans une cellule saine et dans une cellule tumorale par exemple).

Le protéome

Quant à la protéomique, elle permet d'identifier et d'évaluer toutes les protéines, par type cellulaire ou par organe par exemple, et notamment leurs modifications post-traductionnelles qui interviennent après la traduction des ARNs messagers en chaînes polypeptidiques. Elle utilise des technologies matures, comme l'électrophorèse bidimensionnelle associée à la spectrométrie de masse, et des technologies encore en phase de développement comme les puces à protéines.

L'analyse du protéome comporte quatre étapes :

- l'extraction des protéines et leur solubilisation, à l'aide de différents tampons composés d'agents réducteurs, de détergents, voire de solvants organiques ;
- la séparation des protéines par électrophorèse bidimensionnelle, regroupant tout d'abord l'ensemble des protéines ayant le même point isoélectrique (pH auquel la protéine est électriquement neutre), puis séparant les protéines en fonction de leur masse moléculaire ;
- l'identification des protéines par spectrométrie de masse ;
- la comparaison des protéines identifiées aux protéines déjà disponibles dans les bases de données.

Rappelons que, si le code génétique existe, ce n'est qu'une correspondance statique entre une structure linéaire de bases nucléiques et une structure linéaire d'acides aminés conduisant à la protéine. Or une protéine n'est pas qu'une structure linéaire mais essentiellement une structure tridimensionnelle qui n'est pas codée directement par l'ADN et qui dépend éventuellement des modifications post-traductionnelles. Une même chaîne peut se replier de différentes façons et avoir différentes fonctions. C'est dire l'importance de la protéomique.

Le métabolome et le lipidome

La métabolomique étudie le rôle joué par les métabolites, notamment dans les relations entre statut physiologique et phénotype. Elle met en oeuvre toute une diversité de plates-formes depuis la chromatographie associée à la spectrométrie de masse jusqu'à la RMN, résonance magnétique nucléaire.

Parmi les métabolites, nombreux sont les métabolites lipidiques, et les lipides sont des régulateurs de la fonction du cerveau, impliqués dans les maladies neurodégénératives, dont la maladie d'Alzheimer⁷⁸. Les lipides sont effectivement importants, notamment parce que de nombreux processus moléculaires sont contrôlés par des protéines associées aux membranes et parce qu'ils sont impliqués dans beaucoup de voies métaboliques et de régulation de l'activité des protéines.

La lipidomique utilise, comme la protéomique, la spectrométrie de masse pour identifier les lipides, mais la phase de séparation n'utilise pas une électrophorèse bidimensionnelle mais une technique de chromatographie liquide haute performance⁷⁹.

⁷⁸ C. Czech, P. Berndt, K. Busch, and al., « Metabolite profiling of Alzheimer's disease cerebrospinal fluid », *PLoS ONE* 7 : e31501, 2012.

⁷⁹ Prifti et al., « FunNet : an integrative tool for exploring transcriptional interactions », *Bioinformatics*, 2008 Nov 15, 24 (22), pp. 2636-2638.

À tout cela, il faut ajouter toutes les interactions entre macromolécules dans une cellule ou un tissu : protéines-protéines⁸⁰, protéines-ADN, ADN-ARN, protéines-ARN. Ces interactions se situent à un niveau d'organisation supérieure à celui des composants moléculaires pris isolément.

De façon générale, le diagnostic moléculaire repose sur l'identification et la détection de marqueurs biologiques à l'échelle infra-cellulaire (protéines, acides nucléiques, ...). Nous verrons par la suite comment ce parc technologique est intégré dans la recherche de biomarqueurs potentiellement pertinents tant pour les connaissances des maladies que pour une utilisation clinique, débouchant ainsi sur des tests de diagnostic, présentant un certain nombre de qualités idéales sur lesquelles nous reviendrons également :

- un bon ratio sensibilité-spécificité, pour éviter les faux positifs et les faux négatifs ;
- une non invasivité
- une diffusabilité ;
- un coût abordable.

Les instruments d'analyse biologique à haut débit (biopuces ou *microarrays*, séquenceurs de nouvelle génération, procédés reposant sur la spectrométrie de masse) répondent à ces problématiques, d'une part par leur capacité à discriminer finement des molécules et des états moléculaires, d'autre part par la parallélisation qui permet de tester des milliers ou dizaines de milliers d'analytes simultanément.

⁸⁰ E. M. Phizicky, S. Fields, "Protein-protein interactions: methods for detection and analysis", *Microbiol. Rev.*, vol<; 59, n°1, mars 1995, pp. 94-123.

Nanobiotechnologies

Toutes ces évolutions ont été accompagnées du développement des nanotechnologies, et donc des nanobiotechnologies, qui, outre le parallélisme des mesures, s'est caractérisé par une miniaturisation de plus en plus importante qui s'est traduite par la possibilité d'utiliser du matériel biologique en quantité très réduite. Il est d'ailleurs apparu un nouveau terme, celui de nanomédecine, qui a rapidement été associé à celui de médecine personnalisée⁸¹.

Outre la miniaturisation des méthodes d'analyse existantes, les nanotechnologies, plus précisément les nanobiotechnologies⁸², combinaison de la biologie et de l'ingénierie nanométrique, ont ouvert de nouvelles approches, les séquenceurs nanopores dont nous avons déjà parlé, mais aussi des nanoparticules permettant d'augmenter la résolution et d'améliorer les contrastes des systèmes d'imagerie médicale comme l'imagerie par résonance magnétique (IRM). Tout en soulignant que cela ne rentre pas directement dans l'objet de notre travail, nous pouvons ajouter que les nanobiotechnologies permettent, de plus, de manipuler des organismes vivants ou de construire des matériaux inspirés de systèmes moléculaires biologiques et de vectoriser les médicaments de façon à libérer le principe actif là où on veut qu'il agisse. Elles rendent possible des avancées significatives, notamment dans le domaine de la précision.

⁸¹ K. K. Jain, « Nanobiotechnology and Personalized Medicine », *Prog. Mol. Biol. Transl. Sci.*, 2011, 104, pp. 325-354.

⁸² Marcel Lahmani, Patrick Boisseau, Philippe Houdy, *Les nanosciences – 3 – Nanobiotechnologies et nanobiologie*, Paris, Belin, 2007.

III-2 Les technologies physiques

La physique est souvent à la base des nouvelles technologies utilisées dans les sciences du vivant et la médecine post-génomique⁸³. En complément des systèmes d'imagerie, toujours plus résolutifs -photographie, radio, IRM, CT scanner, PET scan, échographie sont de plus en plus utilisés dans les applications médicales et répondent à des attentes - de nouvelles approches émergent régulièrement, comme les techniques SPR (*Surface Plasmon Resonance*)⁸⁴ par exemple.

L'application des connaissances de la physique fondamentale à la biologie n'est pas à l'origine des connaissances en physique (par exemple, il a fallu une trentaine d'années pour passer de la découverte de la RMN à son application dans l'IRM⁸⁵), Néanmoins, une fois utilisées en biologie, les techniques physiques trouvent leurs applications en médecine. Les techniques d'imageries médicales, en dehors de la radiographie traditionnelle, sont néanmoins encore récentes même si les principes physiques sur lesquels elles reposent sont connus depuis longtemps. En fait, leurs développements, outre des progrès d'ordre technique, n'ont été possibles que par les avancées de l'informatique, notamment l'accroissement exponentiel de la puissance des ordinateurs.

L'imagerie multiplie ses outils et apporte des informations de plus en plus sûres, ses modalités couvrant de larges domaines puisqu'aux rayons X se sont ajoutés les champs magnétiques, les rayonnements nucléaires, les ultrasons, les rayonnements

⁸³ Stephen F. Keevil, « Physics and medicine : a historical perspective », *The Lancet*, Vol. 379, April 21, 2012, pp. 1517-1524.

⁸⁴ T. Wilkop, A.S. Ramlogan, I.L. Alberts and al., « Surface plasmon resonance imaging for medical and biosensing », *Institute of Electrical and Electronic Engineers Sensors Conference*, Christchurch, New-Zealand, Oct 25-28, 22009, BSL-A6.

⁸⁵ RMN : Résonance magnétique nucléaire ; IRM : Imagerie par résonance moléculaire.

lumineux associés à des marqueurs fluorescents, etc. Elle donne des informations à plusieurs niveaux :

- au niveau structurel, en donnant accès à des informations sur la morphologie des organes, tissus ou cellules (par exemple, le volume ou la forme d'une éventuelle lésion) ;
- au niveau fonctionnel, en donnant accès à des informations concernant le fonctionnement et l'activité métabolique des organes, tissus ou cellules.

Pratiquement, on pourrait regrouper l'imagerie médicale en quatre catégories d'images :

- La radiographie traditionnelle, basée sur la propagation des rayons X.
- L'échographie qui utilise des ultrasons, utilisant les réflexions différentielles selon la nature des obstacles rencontrés par les ultrasons lors de leur propagation quand ce qu'on visualise n'est pas en mouvement, ou l'effet Doppler, caractérisé par un glissement en fréquence quand l'objectif cherché est en mouvement.
- La tomographie par émission de positrons, sur laquelle nous reviendrons, technique d'imagerie fonctionnelle utilisant des isotopes artificiels.
- L'imagerie par résonance magnétique (IRM) qui utilise la résonance qui intervient lorsqu'un système oscillant (le spin du noyau d'un atome d'hydrogène par exemple) est excité par un phénomène lui-même périodique (un champ magnétique en l'occurrence). Elle relève de la physique quantique.

L'IRM ne visualise pas directement, mais traite les données (la quantité des signaux est proportionnelle au débit sanguin) pour reconstituer des images, ce qui permet, par exemple, de suivre l'activation des zones du cerveau selon son activité. Cette

technique présente l'avantage d'être efficace et a-traumatique puisqu'elle ne nécessite pas l'ingestion d'une substance particulière. L'importance de l'imagerie ne fait par ailleurs que croître et de nouveaux systèmes spécifiques à la micro-imagerie se développent, tant en utilisant les ondes électromagnétiques ou les ondes acoustiques que des nanosystèmes. À titre d'exemple, des nanoparticules d'oxyde de fer améliorent les contrastes des images obtenues par IRM

Le développement de l'imagerie par résonance magnétique fonctionnelle (IRMf) rend perceptible l'activité cérébrale, avec ses 100 milliards de neurones et ses 100 billions de synapses. Néanmoins, et c'est vrai pour les autres techniques d'imagerie, si l'IRM permet de « voir » ce qui, sans elle, demeurerait caché au sein du corps humain, seul le médecin ou le clinicien est capable d'interpréter les images reconstituées et, nous y reviendrons quand nous aborderons les enjeux épistémologiques, il faut se demander si voir c'est savoir. L'imagerie médicale instaure, de fait, une relation entre le sujet et son corps vivant, ce qui donne toute sa validité à la proposition de Gilbert Simondon « la technique doit être comprise en tant que médiateur et non en tant qu'instrument »⁸⁶.

Le haut débit, vu à la fois comme un confort et une nécessité, est associé à l'image, mais chaque pathologie dispose de son type d'image. La compression des images, notamment pour réduire les temps de transfert, est une opération inévitable mais qui peut s'avérer redoutable, notamment pour les applications de diagnostic. Par ailleurs l'attrait de l'image peut conduire à un besoin d'accès (à distance éventuellement) potentiellement illimité. Il convient donc de s'interroger sur la logique du « juste nécessaire » et sur les besoins de traçabilité. Enfin, l'image est *a priori* complexe

⁸⁶ Gilbert Simondon, *Du mode d'existence des objets techniques*, Paris, Aubier, 1989, page 67.

et porteuse d'informations globalisées (métadonnées), si bien que le partage d'expertise autour de l'image doit sans doute être développé.

Tomographie par émission de positrons

La tomographie par émission de positrons (PET-Scan) a pris une place primordiale notamment dans l'étude des maladies neurodégénératives. Un positron, émis lors d'une désintégration radioactive, parcourt une petite distance et entre presque instantanément en collision avec un électron. Ils s'annihilent ainsi l'un l'autre et produisent dans ce processus deux photons qui partent dans deux directions opposées, formant l'un par rapport à l'autre un angle de 180°. C'est ce phénomène qui permet la reconstruction d'une image par l'identification du point d'émission de l'isotope le long de la ligne reliant les deux rayons gamma, et ainsi de produire une image du cerveau au travail. La TEP exige donc le recours à une molécule marquée à l'aide d'un isotope radioactif artificiel émettant des positrons au cours de sa désintégration. Par exemple, la florbetapir⁸⁷, utilisant un isotope du fluor (¹⁸F, période radio-active : 109,77 minutes), présente une grande affinité pour les plaques Aβ et donne une image durant 15 minutes commençant environ 50 minutes après l'injection avec un rayonnement gamma de 511 keV. On suit le devenir de la molécule marquée à l'intérieur du cerveau, une caméra détectant les effets de l'émission des positrons par les molécules marquées. Un traitement mathématique des données permet de reconstruire l'image de la position du radio-isotope au sein d'une tranche⁸⁸ du cerveau de quelques millimètres d'épaisseur. Ajoutons néanmoins que la tomographie par émission de positrons est une technique très coûteuse.

⁸⁷ Clark et al., « Use of florbetapir-PET for imaging beta-amyloid pathology », *JAMA*, 2011, vol. 305 (3), pp 275-283.

⁸⁸ D'où le nom de la technique, le grec τομή signifiant « section ».

Le développement des techniques non invasives, mais nécessitant l'injection de molécules marquées, pour détecter les plaques amyloïdes dans le cerveau humain par PET ont apporté des contributions remarquables dans la compréhension patho-physiologique de la maladie d'Alzheimer et a amélioré le diagnostic. Pratiquement, l'analyse est effectuée par interprétation visuelle entre des régions riches en plaques amyloïdes et des régions plus pauvres. Mais, pour des études longitudinales permettant de suivre l'évolution, il est nécessaire d'en faire une analyse quantitative. Par ailleurs, indépendamment des variabilités biologiques interindividuelles, il faut prendre en considération les variabilités techniques, notamment la résolution spatiale, et donc prendre en compte l'instrumentation utilisée pour acquérir les scans, les mouvements éventuels du patient lors de l'acquisition ainsi que les protocoles d'analyse d'image utilisés.

Electroencephalographie

L'électroencéphalographie (EEG) est une méthode d'exploration cérébrale qui mesure l'activité électrique du cerveau par des électrodes placées directement sur le cuir chevelu. L'EEG est un examen qui a l'avantage d'être non-invasif tout en renseignant sur l'activité neurophysiologique du cerveau au cours du temps. Il a montré tout son intérêt dans le pronostic de la maladie d'Alzheimer⁸⁹, notamment parce qu'il peut être utilisé soit quand la personne est au repos, soit quand la personne fait une tâche de mémorisation. Les signaux électriques, outre leur évolution dans le temps (intensité en fonction du temps), peuvent être analysés par transformée de Fourier (intensité en fonction de la fréquence) ou par ondelettes (temps en fonction de la fréquence).

⁸⁹ C. E. Jackson, P. J. Snyder, « Electroencephalography and event-related potentials as biomarkers of mild cognitive impairment and mild Alzheimer's disease », *Alzheimer's & Dementia*, 2008 Jan, 4 (1 Suppl 1) : S 137-143.

III-3 La quantification de soi

La quantification de soi consiste à capturer, stocker, analyser et partager ses données personnelles à l'aide d'objets du quotidien. La "quantification de soi" relève du domaine de la santé et pas de la maladie. La spécificité de la quantification de soi repose sur l'intégration poussée de capteurs avancés et de traitements de signaux dans des microsystèmes autonomes.

Les technologies individuelles en matière de santé bénéficient du développement considérable des réseaux, des solutions de communication et de la création de nombreux objets connectés qui permettent de récupérer des données, de les traiter et de les transmettre à distance. Le mobile est omniprésent aujourd'hui dans la vie quotidienne et il est devenu le principal mode de communication et d'interaction, tant à titre personnel que professionnel, pour échanger des informations ou des services.

Les technologies individuelles permettent de replacer l'individu au cœur de la prévention, voire des traitements, et nécessitent une participation active de ce dernier. Par ces technologies et ces mesures se manifeste une réappropriation du soi, intégrant une vision prospective avec la prise en compte des données passées et présentes, et la possibilité de se comparer aux autres. Au-delà des apparences individualistes de la démarche, la quantification de soi est potentiellement porteuse de bénéfices collectifs, notamment pour la santé ou l'économie.

Quels en sont les outils ? Tout d'abord les capteurs qui prennent les mesures et sont notamment logés dans des bracelets, des montres ou des balances. Non seulement les capteurs sont devenus plus petits et moins chers, mais ils sont également désormais tous sans fil et la durée de vie des batteries s'est radicalement améliorée. Les capteurs, ambiants ou embarqués, se

multiplient et permettent de recueillir un nombre croissant et diversifié de données, qu'elles soient biologiques, physiologiques, physiques, environnementales ou encore comportementales. C'est ce qu'apporte notamment l'usage des microsystèmes embarqués autonomes⁹⁰ et toutes les technologies d'e-santé⁹¹, par exemple la télémédecine avec le télédiagnostic, la télésurveillance ou l'imagerie diagnostique assistée par ordinateur. Par ailleurs, les smartphones, dont l'usage est désormais banalisé, permettent de recueillir, synthétiser et partager les données. La quantification de soi peut donc concerner tout un chacun.

Il y a tout intérêt à collecter des informations sur les paramètres physiologiques d'une personne dans son environnement naturel et sur le long terme. Par exemple, en analysant ses mouvements et ses activités quotidiennes il est possible de détecter des évolutions, voire des ruptures, d'un état de santé. C'est l'objet de l'actigraphie. Elle utilise des accéléromètres piézoélectriques pour enregistrer les mouvements, notamment les mouvements des bras, dans les trois dimensions, et mettre ainsi en évidence des comportements agités ou chaotiques.

Si l'intégration des différentes données *-omics* constitue un enjeu majeur pour reconstituer les états fonctionnels, il est nécessaire, cependant, pour expliquer les phénotypes observés et, par-là, essayer de comprendre les processus impliqués dans le développement d'une maladie, de combiner les informations moléculaires avec des données environnementales telles que l'alimentation, l'exposition à des agents infectieux ou à des agents chimiques, les médicaments. C'est l'objet des « *electronic health records* » (EHRs). Les EHRs ne relèvent pas à proprement parler de la quantification de soi. Ils relèvent des applications médicales

⁹⁰ <http://www.stic-sante.org>.

⁹¹ http://ec.europa.eu/information_society/activities/health/whatis_ehealth/index_en.htm.

à distance en permettant d'enregistrer, de conserver et de rendre accessible sur un réseau les dossiers des personnes. L'analyse de leur implication reste à faire. Ils peuvent illustrer en quoi la médecine personnalisée recouvre des approches et des pratiques très disparates, et permettre de poser les enjeux politiques, économiques et sociaux de la médecine personnalisée.

Les EHRs permettent, de fait, de piloter des services de santé pour une personne et de fournir des données pour la recherche⁹². Ces outils peuvent s'inscrire dans le maintien à domicile des personnes seules ou fragiles. Peu intrusifs, ils peuvent s'adapter automatiquement à l'hétérogénéité des conditions particulières de chacun et deviennent des éléments clés en recherche clinique⁹³. Ils nécessitent cependant des règles pour que les données puissent être partagées⁹⁴. Néanmoins, il faut souligner que les informations dans les EHRs se présentent souvent sous forme de textes, et donc sous forme narrative, ce qui ne facilite pas la mise en oeuvre de techniques de tris automatiques. En complément, les informations sont souvent parcellaires, incomplètes et leur exactitude discutable. Le problème de la fiabilité des données se pose donc. Il est généralement peu abordé quand les capteurs n'ont pour but que de connaître le bien-être d'une personne ou de permettre son maintien à domicile. Mais il doit l'être car il n'est pas envisageable de prendre des décisions fiables impliquant la santé sur la base de données incohérentes ou incomplètes.

⁹² La start-up californienne Syapse propose une plateforme qui agrège les informations du dossier médical personnel (les EHRs) avec des données plus complexes, comme les données génomiques, pour permettre aux médecins de faire des diagnostics plus poussés et d'adapter les traitements au patient.

⁹³ J. Powell, I. Buchan, « Electronic health records should support clinical research », *J. Med. Internet Res.*, Vol. 7, N° 1, Mar. 14, 2005.

⁹⁴ de Lusignan, N. Hague et al., « Routinely-collected general practice data are complex, but with systematic processing can be used for quality improvement and research », *Prim. Care Inform.*, Vol. 14, N° 1, 2006, pp. 59-66.

Ces systèmes rentrent dans le champ de la santé mobile qui recouvre, selon l'OMS, « les pratiques médicales et de santé publique reposant sur des dispositifs mobiles tels que téléphones portables, systèmes de surveillance des patients, assistants numériques personnels et autres appareils sans fil »⁹⁵. Les objets connectés permettent de récupérer des données, de les traiter et de les transmettre à distance. Ils offrent aussi, même si ce n'est pas directement notre objet, la perspective de l'*empowerment*⁹⁶.

L'*empowerment* est un processus par lequel une personne renforce sa capacité d'agir ainsi que sa capacité à identifier et satisfaire ses besoins, à résoudre ses problèmes et à mobiliser ses ressources de manière à contrôler sa propre vie.

Ce concept peut s'analyser comme la capacité qu'acquiert un individu à prendre des décisions lorsqu'il se trouve confronté à une situation ou à un problème spécifique, seul ou en interaction avec d'autres personnes dans le but de s'adapter à cette situation. Ces objets permettent en effet de l'auto-évaluation et du dépistage précoce. Soulignons que l'acquisition de ces données nécessite une approche « participative » qui suppose une implication forte de l'individu/patient mais retenons que valoriser l'*empowerment*, c'est penser le processus grâce auquel les fragiles – et les malades d'Alzheimer sont fragiles et vulnérables – parviennent à prendre une part active aux décisions qui les concernent. Le *quantified self* n'est introduit que comme un élément d'ouverture pour des études futures, mais il ne fait pas partie de notre étude parce qu'il n'est pas centré sur l'identification et la validation de biomarqueurs prédictifs.

Tous ces nouveaux instruments offrent des opportunités non négligeables. On peut considérer que le progrès technique transforme les conditions de la connaissance en permettant

⁹⁵ Organisation mondiale de la santé, *mHealth – New horizons for health through mobile technologies*, *Global Observatory for eHealth series – Volume 3*, p. 6.

⁹⁶ M. H. Bacqué, C. Biewener, *L'empowerment, une pratique émancipatrice*, Paris, La Découverte, 2013.

l'acquisition rapide d'un grand nombre de données dans les situations les plus diverses, et en intégrant les méthodes statistiques pour les interpréter. Il faut compléter cette présentation par une réflexion, à venir, de portée politique sur le type de « sujet », ou de « soi », qui se trouve redéfini dans la mise en œuvre de ces technologies.

Retenons pour le moment que toutes ces technologies sont supposées faciliter le maintien à domicile des personnes atteintes de la maladie d'Alzheimer, notamment dans les premières phases de son évolution, en améliorant leur vie quotidienne tout en permettant leur suivi à distance par les professionnels de santé. Des enquêtes, qui restent à analyser, permettront de confronter le discours à la réalité mais cela sort du périmètre de notre étude, qui porte sur les biomarqueurs.

un certain nombre d'exemples d'objets connectés:

- Le service complet d'assistance et de surveillance du domicile et du malade développé chez *Samsung* après le rachat de *SmartThings*. Il comporte notamment une application destinée à stimuler la mémoire des patients afin qu'ils puissent reconnaître leurs proches.
- *Eyezheime*, une invention française d'une lunette qui permet la reconnaissance faciale des proches enregistrés dans une base de données de l'utilisateur.
- Le bracelet de géolocalisation *Bluegard* de la société *Bluelinea* équipé d'une puce RFID (identification par radiofréquence) qui permet de repérer l'endroit où se trouve la personne et d'éviter toute errance.
- La montre *Vivago* composée de capteurs de mouvements et d'activités, qui collecte également en permanence des mesures physiologiques sur l'utilisateur.
- Le système de video-vigilance *Edao*, développé par *Link Care Service*¹ qui permet aussi le suivi des traitements thérapeutiques.
- L'application *Moves* qui permet de garder les traces de ses activités journalières tout en détectant les dérives par rapport aux habitudes de la personne pour alerter les professionnels de santé en cas de détection d'une situation anormale.

Ces objets connectés doivent permettre le développement de pratiques de santé en liaison avec du personnel médical, mais également hors d'un contexte médical proprement dit. Ils permettent de suivre en continu les variations dans le temps des données d'un individu. Outre le maintien plus longtemps à domicile des patients et l'aide apportée aux aidants ces objets et techniques apportent un plus dans la connaissance de la maladie et de son évolution car les données sont acquises en situation de vie, dans l'environnement même du malade, et non pas dans un centre spécialisé de tests et d'analyses. Par ailleurs, les applications s'améliorent par des mises à jour permanentes, les données récoltées faisant l'objet d'une analyse régulière des médecins en interface avec les ingénieurs systèmes

III-4 Les données environnementales

Il reste à prendre en compte les données de l'environnement qu'a introduit le concept d'« exposome ». Ce concept veut signifier et mettre en exergue l'importance de connaître l'environnement auquel est exposé l'individu. A priori très variable, cet environnement, ainsi que cela a été rappelé, peut être divisé en trois catégories⁹⁷, l'environnement interne, l'environnement externe spécifique et l'environnement externe général.

Les techniques « -omics » et les EHRs, par exemple, donnent accès à des données pour les deux premières catégories. Par contre, pour ce qui concerne l'environnement socio-économique ou le contexte culturel, l'identification de biomarqueurs associés à des expositions environnementales n'est pas facile, même si la tendance à vouloir « moléculariser » l'environnement existe⁹⁸. En dehors de l'utilisation d'objets connectés qui peuvent

⁹⁷ WILD C. P., " The exposome: from concept to utility", *Int. J. Epidemiol.*, 2012, 41(1), pp. 24-32.

⁹⁸ Le projet « Exposomics » en est une illustration même s'il se limite particulièrement à

donner des indications, il faut faire appel à d'autres approches, notamment des approches par questionnaires. On peut ajouter que les applications mobiles, les appareils connectés, les services de télémédecine ouvrent des perspectives, en particulier dans le suivi des maladies chroniques et des maladies à évolution lente.

l'exposition à des polluants dans l'air ou à des contaminants dans l'eau.

Chapitre IV

Face à l'avalanche de données, quelques outils

Ainsi qu'il a déjà été soulevé, la prise en compte des données biologiques, notamment les données génétiques, et la nécessité de comprendre comment les facteurs génétiques interagissent avec d'autres facteurs comme les facteurs socio-économiques et psychosociaux, nécessitent, d'une part d'utiliser la bioinformatique et la biologie statistique, d'autre part des outils permettant d'analyser l'avalanche de données susceptibles d'être rassemblées.

Aussi, avant de faire état des biomarqueurs dans le contexte de la maladie d'Alzheimer, et des controverses qui leur sont associées, nous proposons un panorama de la bioinformatique et de la statistique ainsi que des sources de données et des outils d'analyse de ces données qui constituent des *Big Data*.

IV-1 La bioinformatique

La bioinformatique, dans son acception la plus large, se réfère au traitement informatique de toute information biologique⁹⁹, mais dans la pratique, se réfère aux séquences génomiques, protéomiques, voire aux autres '*omics*'. En fait, elle regroupe l'ensemble des concepts et des techniques nécessaires à l'interprétation informatique de l'information biologique, à savoir :

- La bioinformatique des séquences, qui traite de l'analyse de données issues de l'information génétique contenue dans la séquence de l'ADN ou dans celle des protéines qu'il code.

⁹⁹ J.M. Claverie & C. Notredame, *Bioinformatics for Dummies*, Hoboken, NJ, Wiley, 2003.

- La bioinformatique structurale, qui traite de la reconstruction, de la prédiction ou de l'analyse de la structure 3D ou du repliement des macromolécules biologiques (protéines, acides nucléiques).
- La bioinformatique des réseaux, qui s'intéresse aux interactions entre gènes, protéines, cellules, organismes, en essayant d'analyser et de modéliser les comportements collectifs d'ensembles de briques élémentaires du Vivant.

La bioinformatique est passée d'une approche traditionnelle de la biologie, ne s'intéressant qu'à un seul gène, à une approche combinant *Data Mining* et analyse intégrative qui facilite les déterminations fonctionnelles du génome¹⁰⁰. Envisager la recherche de réseaux d'interactions par une analyse globale du génome et de l'épigénome ouvre la voie à de nouvelles connaissances. Le développement des séquenceurs de nouvelle génération, déjà évoqué, aboutit à la connaissance de la totalité du génome d'un individu à un coût raisonnable et dans un délai court, ce qui permet d'envisager l'utilisation de la connaissance du génome entier en clinique¹⁰¹. Le problème est de transformer les séquences lues en applications cliniques, ce qui nécessite une approche multi-étapes.

Il faut d'abord assembler les données obtenues. S'il s'agit d'un re-séquençage (séquençage d'une personne dont on a déjà la séquence), on utilise des algorithmes de hachage ou la transformation de Burrows-Wheeler

¹⁰⁰ P. Bertone, M. Gerstein, « Integrative data mining : the new direction in bioinformatics », *IEEE Eng. Med. Biol. Mag.*, 20 (4), 2001, pp. 33-40.

¹⁰¹ M. N. Bainbridge, W. Wisniewski, D. R. Murdoch, J. Friedman, C. Gonzaga-Jauregui, I. Newsham, « Whole-genome sequencing for optimized patient management », *Sci. Transl. Med.* 2011, 3 (87)re3.

La transformée de Burrows-Wheeler réalise une permutation des lettres du mot d'entrée de telle sorte que les lettres identiques vont être regroupées formant ainsi des longues séquences identiques facilement compressibles. Elle est très utilisée en génomique par exemple pour les problèmes d'alignement de lectures courtes issues des nouvelles techniques de séquençage d'ADN ou d'ARN et pour des problèmes de comptage de mots (détection de répétitions).

tandis que dans le cas d'un nouveau séquençage, on utilise la méthode des graphes de de Bruijn¹⁰².

Le graphe de de Bruijn est un graphe orienté qui permet de représenter les chevauchements de longueur $n-1$ entre tous les mots de longueur n sur un alphabet donné, un alphabet de quatre éléments (A, T, C, G) dans le cas du génome :



le graphe dirigé contient les sous-séquences de longueur 21 de la séquence, et les sous-séquences adjacentes dans la séquence sont liées par une flèche. En combinant la représentation en graphe de toutes les séquences, on obtient un graphe plus gros qui contient un chemin représentant la séquence génomique de laquelle proviennent les séquences

¹⁰² D. R. Zerbino, E. Bimney, « Velvet algorithms for de novo short read assembly using de Bruijn graphs », *Genome Research*, Vol. 18 , N° 5, 2008, pp : 821-829.

Il faut ensuite identifier les variants génomiques dans le génome assemblé. C'est une étape importante dans la recherche de biomarqueurs. Plusieurs algorithmes sont disponibles comme le *Genome Analysis Tool Kit (GATK)*¹⁰³.

Il faut enfin associer les variants identifiés à des effets. Pour cela, on regarde si le variant était connu ou est nouveau, et s'il se trouve dans une partie codante ou une partie non-codante du génome. S'il est dans une partie codante, on regarde s'il produit un changement dans la séquence d'une protéine, s'il est dans une partie non-codante on regarde s'il peut entraîner des modifications d'épissage (passage à la transduction). Dans les deux cas, ces variants peuvent provoquer des effets cliniques.

De toute évidence, l'informatique est devenue un apport fondamental à la biologie moléculaire. Les moyens informatiques sont naturellement utilisés pour le stockage ou la gestion des données mais également pour l'interprétation de ces données. En absence d'expériences, le traitement informatique des séquences peut par exemple déceler la fonction biologique potentielle d'un gène par la recherche de critères spécifiques (signaux, structures secondaires ou tertiaires...) ou par la recherche de similitudes entre séquences. Cela ne signifie pas que la science et la connaissance ne soient gouvernées que par les données, ces données ne pouvant être analysées et interprétées que si l'on tient compte du contexte et des hypothèses qui ont conduit à leur acquisition. Par ailleurs, la bioinformatique doit encore passer de la biologie à la médecine. Une première approche consiste à passer des 'omics aux études d'association. Ces études visent à caractériser les fréquences alléliques de nombreux gènes afin de mettre en évidence les allèles qui diffèrent entre deux groupes, un groupe de

¹⁰³ A. McKenna, M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis, A. Kernytsky, « The Genome Analysis Toolkit : a MapReduce framework for analyzing next-generation DNA sequencing data », *Genome Research*, 2010, 20 (9), pp ; 1297-1303.

patients atteints d'une pathologie et un groupe témoin de contrôle. Si on fait cette étude au niveau du génome, on parle de GWAS (*Genome-Wide Association Studies*) et on cherche généralement à mettre en évidence des mutations ponctuelles (SNPs) comme marqueurs génétiques. Mais on peut aussi faire des études d'association à d'autres niveaux, comme celui de l'expression des gènes ou de l'épigénétique par exemple.

Les études d'association peuvent se faire au niveau des gènes, au niveau de l'expression des gènes, au niveau de l'épigénome, au niveau de l'exposome ou au niveau du phénotype :

- les études d'association au niveau du génome (*Gene-Wide Association Studies* – GWAS) visent à associer la susceptibilité à certaines maladies à des variations génétiques dans le génome. Elles sont basées sur l'analyse des mutations ponctuelles qui existent tout au long du génome. Pour analyser la relation entre ces mutations et des caractéristiques phénotypiques, la susceptibilité à une maladie par exemple, on peut utiliser des méthodes de régression logistique qui ont l'inconvénient d'être linéaires, ou des méthodes bayésiennes qui permettent d'introduire des interactions plus complexes et de prendre mieux en compte les connaissances biologiques déjà acquises.
- Au niveau de l'expression des gènes (*Gene Expression-Based Gene Wide Association Studies* – eGWAS), on utilise les données disponibles en grande quantité dans les bases de données sur l'expression des gènes (données obtenues généralement suite à des expérimentations mettant en oeuvre des puces ADN après une étape de reverse-transcription) et on les combine pour calculer la probabilité de trouver les mêmes gènes exprimés différemment dans un grand nombre d'échantillons provenant de malades atteints de la même maladie.
- L'épigénomique étudie les modifications chimiques de l'ADN qui jouent un rôle important dans l'activité et l'expression des gènes. C'est le résultat d'un processus dynamique entre le génome et des facteurs environnementaux tels que l'alimentation ou l'exposition à certains polluants par exemple. Les études d'association au niveau épigénomique (*Epigenome Wide Association Studies* – EpWAS) visent à étudier l'influence de modifications épigénétiques sur des caractères phénotypiques.
- Les données des EHRs permettent de renverser l'approche des études d'association. Au lieu de faire le lien d'un caractère phénotypique unique avec des mutations dans le génome (GWAS), on peut identifier les caractères phénotypiques associés à un génotype particulier. C'est l'objet des *Phenome-Wide Association Studies* – PheWAS.

On peut ainsi s'ouvrir à d'autres champs que ceux habituellement retenus par la biologie, à savoir la génomique et la post-génomique en intégrant des données plus qualitatives.

Ainsi, si, en complément de la connaissance du génome, l'analyse du transcriptome et celle du protéome sont indispensables, il faut cependant considérer qu'elles ne permettent pas de dégager les informations utiles si elles sont considérées individuellement. Seule l'intégration des technologies de l'information permet d'en dégager les informations pertinentes. La bio-informatique cherche à faire une analyse globale des données, notamment des données dynamiques issues du transcriptome et du protéome. « La plupart des processus cellulaires sont, en effet, enracinés dans les interactions dynamiques entre un grand nombre de biomolécules et c'est un réseau d'interactions qui détermine la fonction biologique »¹⁰⁴.

IV-2 La biostatistique

La statistique a une terminologie souvent trompeuse, les termes comme significatif, erreur ou hypothèse ayant un sens différent de celui de l'usage courant. Un résultat statistiquement significatif ne signifie pas qu'il est scientifiquement important. Par exemple, avec de grandes tailles d'échantillons, de petites différences peuvent être statistiquement significatives alors qu'elles ne sont pas importantes. Quand on cherche à extrapoler à un ensemble complet, une population par exemple, les propriétés mises en évidence sur un échantillon (qu'on supposera représentatif), on parle de statistique inférentielle. L'inférence

¹⁰⁴ Jean Hache, *Les enjeux des biotechnologies*, Colombelles, Editions EMS, 2005, page 29.

statistique consiste ainsi à induire les caractéristiques inconnues d'une population à partir d'un échantillon issu de cette population.

Les bio-statistiques ne sont pas nouvelles. On peut rappeler qu'en 1760, le mathématicien Daniel Bernouilli a proposé de déterminer l'efficacité de l'inoculation de la petite vérole en calculant le gain moyen d'espérance de vie des personnes inoculées et a montré que, pour le même âge, les risques de mort étaient de 1/64 pour la variole contre 1/473 pour l'inoculation, en concluant l'intérêt de l'inoculation effectuée avant l'âge de cinq ans¹⁰⁵.

La mathématisation est vite devenue une aide à la rationalisation progressive de l'expérimentation, une aide face à la variabilité biologique, la moyenne statistique renvoyant au collectif et la variance réintroduisant l'individuel. L'introduction des tests statistiques au XX^e siècle avec le développement du test du χ^2 (106) en biologie dans les années 1930 et les travaux de R. A. Fisher, d'abord en recherche agronomique, aboutiront à une méthode formelle de randomisation¹⁰⁷, le premier essai clinique véritablement randomisé étant l'évaluation de l'effet de la streptomycine sur la tuberculose en 1948¹⁰⁸.

L'histoire de l'émergence des bio-statistiques montre l'importance de la modélisation probabiliste dans les sciences de la vie. Le terme de probabilité, que l'on peut faire remonter à Pascal, renvoie à la volonté de mathématiser les jeux de hasard, tandis que le terme statistique est polysémique puisqu'il recouvre à la fois une

¹⁰⁵ Anne Fagot-Largeault, *Les causes de la mort : histoire naturelle et facteurs de risques*, Paris, Vrin, 1989.

¹⁰⁶ Le test du χ^2 , prononcé « khi-deux » ou « khi carré », est un test statistique permettant de tester l'adéquation d'une série de données à une famille de lois de probabilités ou de tester l'indépendance entre deux variables aléatoires.

¹⁰⁷ H. M. Marks, « L'irruption de la preuve statistique en médecine », *La Recherche*, 1999, N° 316, PP : 76-81.

¹⁰⁸ R. P. d'Arcy Hart, "A change in scientific approach : from alternation to randomised allocation in clinical trials in the 1940's", *British medical Journal*, 1999, N° 319, PP : 572-573.

méthodologie mathématique de recueil de données et une méthodologie d'analyse des données recueillies, l'inférence statistique.

Par ailleurs, la statistique fait appel à des concepts abstraits et est à l'interface entre les mathématiques et les sciences, ce qui nécessite de l'aborder sous les deux angles. Cela peut être illustré par la notion de p-valeur, qui est une probabilité.

Une hypothèse étant testée, « l'hypothèse nulle », la p-valeur est la probabilité d'obtenir une différence aussi grande, ou plus grande, que celle observée si l'hypothèse nulle est vraiment correcte. Ainsi une p-valeur de 0,03 signifie que, même si deux populations de personnes (ou patients) avaient les mêmes moyennes, 3% des études semblables à celle qui a été menée rapporteraient une différence au moins aussi grande que celle qui a été observée. Une mauvaise interprétation consisterait à dire « s'il y a seulement 3% de chance que la différence ait été causée par le hasard, alors il devrait y avoir 97% de chance qu'elle résulte d'une véritable différence ». En fait, la seule chose que l'on puisse dire c'est que, si l'hypothèse nulle était vraie, 97% des études aboutiraient à une différence plus petite que celle qui a été observée et 3% des études aboutiraient à une différence aussi grande ou plus grande.

Retenons simplement que plus la p-valeur est faible, plus la présomption contre l'hypothèse nulle est forte.

En fait, au lieu de considérer la p-valeur, il faut mieux retenir le « *odds ratio* ». L'*odds* de la survenue de quelque chose, par exemple la présence d'un biomarqueur dans une cohorte de patients atteints de la même maladie, est le rapport entre la probabilité que l'événement (la présence du biomarqueur) arrive et la probabilité qu'il n'arrive pas (l'absence du biomarqueur). Ainsi, si dans une cohorte de patients, la probabilité de mettre en évidence un variant génétique est de 25 %, l'*odds* de l'événement « le variant génétique est présent » est de $25/75 = 0,33$. Quand on parle d'*odds ratio*, on fait le rapport du même événement dans deux échantillons ou cohortes différents. C'est une mesure de l'effet de la taille de

l'échantillon décrivant la force d'une association ou la non-indépendance entre deux données.

L'*odds ratio* (OR), également appelé rapport des chances, rapport des cotes ou risque relatif rapproché, est une mesure statistique, souvent utilisée en épidémiologie, exprimant le degré de dépendance entre des variables aléatoires qualitatives¹⁰⁹. Il est utilisé en inférence bayésienne et en régression logistique, et permet de mesurer l'effet d'un facteur. Il se définit comme le rapport de la cote¹¹⁰ d'un événement arrivant à un groupe A d'individus, par exemple une maladie, avec celle du même événement arrivant à un groupe B d'individus. Si la probabilité qu'un événement arrive dans le groupe A est p , et q dans le groupe B, le rapport des cotes est :

$$\frac{p/(1-p)}{q/(1-q)} = \frac{p(1-q)}{q(1-p)}.$$

L'*odds ratio* est toujours supérieur ou égal à zéro. Si on garde l'exemple d'un biomarqueur, s'il est :

- proche de 1, le biomarqueur est indépendant du groupe ;
- supérieur à 1, le biomarqueur est plus fréquent dans le groupe A que dans le groupe B ;
- bien supérieur à 1, le biomarqueur est beaucoup plus fréquent dans le groupe A que dans le groupe B ;
- inférieur à 1, le biomarqueur est moins fréquent dans le groupe A que dans le groupe B ;
- proche de zéro, le biomarqueur est beaucoup moins fréquent dans le groupe A que dans le groupe B.

Cet indicateur pourra être envisagé pour caractériser un biomarqueur comme un biomarqueur prédictif d'une maladie, mais

¹⁰⁹ A. Jammal, G. Loslier, R. Allaurd(1988) Dictionnaire d'épidémiologie. Edisem/Maloine, St-Hyacinthe/Paris, p. 124-125

¹¹⁰ La cote est à entendre comme celle d'un cheval de course dans un grand prix. Un cheval à 3 contre 1 a une chance sur 4 de gagner. De façon générale, un cheval à a contre b a une probabilité p de gagner telle que $(a/b + 1) = 1/p$

nous verrons que pour cela l'*odds ratio* entre une cohorte de patients et une cohorte témoin doit être significativement bien supérieur à 1.

L'utilisation des statistiques permet d'autres applications. Elles permettent tout d'abord possible de réduire le nombre de données :

- en éliminant des données considérées comme aberrantes, encore faut-il définir ce que sont des données aberrantes ;
- en limitant la recherche des informations quand les données sont fortement corrélées entre elles, en faisant une analyse en composantes principales par exemple.

L'Analyse en composantes principales (ACP) est une méthode de statistique multivariée consistant à transformer des variables liées entre elles, ou "corrélées", en nouvelles variables décorréelées les unes des autres. Ces nouvelles variables sont les "composantes principales". L'ACP est une approche à la fois géométrique (les variables sont représentées dans un nouvel espace, selon des directions d'inertie maximale) et statistique (la recherche porte sur des axes indépendants expliquant au mieux la variance¹¹¹ des données). Le toilettage des données se fait sans perdre le caractère pertinent de données qui pourraient sembler aberrantes. Pour cela, si les méthodes d'analyse en composantes principales sont utilisées, il ne faut pas oublier que cela reste une analyse statistique et il faut rester prudent quant à la signification biologique du résultat, d'autant que, par nature, la variabilité biologique existe, et quant à la signification clinique de ces résultats qui peuvent n'être d'aucune utilité pour les praticiens.

Parmi les méthodes statistiques, les méthodes de régression sont largement utilisées. La plus simple est la régression linéaire, mais la plus courante dans le domaine biomédical est sans doute la régression logistique¹¹². Elle s'utilise généralement lorsque la

¹¹¹ La **variance** est une mesure servant à caractériser la dispersion d'un échantillon ou d'une distribution. Elle indique de quelle manière la série statistique se disperse autour de sa moyenne. Une petite variance est signe que les valeurs sont proches les unes des autres alors qu'une variance élevée est signe que celles-ci sont très écartées. La variance est la moyenne des écarts au carré à la moyenne. Sa racine carrée est l'écart-type.

¹¹² D. W. Hosmer, S. Lemeshow, *Applied Logistic Regression*, Second edition, Wiley, 2000.

variable à expliquer est qualitative, le plus souvent binaire (l'apparition de la maladie d'Alzheimer par exemple). Les variables explicatives peuvent être par contre soit qualitatives, soit quantitatives. La variable dépendante est habituellement la survenue ou non d'un événement (maladie ou autre) et les variables indépendantes sont celles susceptibles d'influencer la survenue de cet événement, c'est-à-dire les variables mesurant l'exposition à un facteur de risque ou à un facteur protecteur par exemple. Elle reste une simplification mathématique¹¹³ de phénomènes complexes, et repose théoriquement sur des conditions (les individus doivent être sélectionnés de façon aléatoire, indépendamment les uns des autres), dont le respect est trop peu souvent vérifié¹¹⁴.

Une autre approche est le modèle de Cox¹¹⁵. C'est un modèle de régression largement utilisé dans le cadre de l'analyse des données de survie, il prend en compte le temps et des données censurées (le point final, la maladie d'Alzheimer par exemple ne s'est pas encore produit). Ce modèle¹¹⁶ a gagné en popularité dans différents domaines dès lors que la variable dépendante étudiée représente la durée jusqu'à un événement terminal, et que l'étude est limitée dans le temps.

A ces techniques de régression, il faut ajouter l'analyse ROC¹¹⁷ (*Receiver-Operating Characteristic*) pour évaluer et comparer des modèles. La courbe ROC est construite à partir des

¹¹³ Pour une seule variable X par exemple, le modèle logistique détermine $\text{Logit } P = \alpha + \beta X$ avec $\text{Logit } P = \ln P - \ln(1-P)$ où P est la probabilité de Y sachant X.

¹¹⁴ Pour la régression logistique, la variable dépendante suit généralement une loi de Bernoulli de paramètre p (p la probabilité moyenne pour que l'événement se produise).

¹¹⁵ COX David, « Modèles de régression et tables de survie », *Journal de la Société Statistique Royale*, 1972, Séries B 34 (2) : pages 187 à 220.

¹¹⁶ Le modèle à risques proportionnels de Cox est de la forme suivante : $h(t|x) = h_0(t) \exp(xb)$ où $h_0(t)$ représente le risque de référence et $b = (b_1, \dots, b_p)'$ représente le vecteur des coefficients de régression.

¹¹⁷ T. Perneger, A. Perrier, « Analyse d'un test diagnostique, Courbe ROC ou *receiver-operating characteristic* », *Revue des maladies respiratoires*, Vol 21, N° 2, avril 2004, pp. 398-401.

mesures telles que la sensibilité et la spécificité [sensibilité fonction de (1 - spécificité)] et l'aire sous la courbe permet de caractériser la précision d'un modèle. L'aire sous la courbe (*Area Under the Curve* – AUC) correspond à la probabilité pour qu'un événement positif soit classé comme positif par le test sur l'étendue des valeurs seuil possibles.

Pour un modèle idéal, on a AUC=1, pour un modèle aléatoire, on a AUC=0.5. On considère habituellement que le modèle est bon dès lors que la valeur de l'AUC est supérieure à 0.7. Un modèle bien discriminant doit avoir une AUC entre 0.87 et 0.9. Un modèle ayant une AUC supérieure à 0.9 est excellent.

Enfin, il faut envisager que les variables entrant dans un modèle ne sont pas figées et que de nouveaux paramètres peuvent être ajoutés. Cela permet d'introduire « *l'integrated discrimination improvement* (IDI) »¹¹⁸. L'IDI est défini par $IDI = (IS_{new} - IS_{old}) - (IP_{new} - IP_{old})$. Dans cette équation, IS est l'intégrale de la sensibilité, IP est l'intégrale correspondante de "1 - spécificité." tandis que "new" se réfère au modèle incluant un nouveau paramètre (un biomarqueur supplémentaire par exemple) et "old" à l'ancien modèle.

Ces quelques exemples montrent la richesse des outils de statistique, mais aussi la nécessité de dépasser ces outils et de les appréhender dans ce qu'ils apportent. Il faut surtout s'interroger sur la signification biologique et médicale des analyses statistiques, et également assurer à ces analyses statistiques une certaine confiance.

La confiance ne se décrète pas. Elle est le résultat de pratiques, notamment des articulations de la science et de la médecine qui aboutissent à de nouvelles pratiques médicales. Ainsi qu'il a déjà été évoqué, la science et la médecine interagissent

¹¹⁸ M. J. Pencina et al., « *Evaluating the added predictive ability of a new marker: from area under the ROC curve to reclassification and beyond* ». *Stat Med* 2008;27(2), pp.157-172.

désormais de manière continue, et nous pouvons considérer que ce sont ces interactions qui ont donné naissance, ainsi que l'a analysé Jean-Paul Gaudillière, à la biomédecine¹¹⁹. Ces interactions conduisent à l'utilisation de technologies qui modifient tant les diagnostics que les pronostics et les thérapies et ont abouti à une transformation des pratiques cliniques en un champ expérimental par le biais de l'augmentation des essais cliniques et des études épidémiologiques prospectives¹²⁰.

L'articulation recherche biologique / pratiques cliniques produit de nombreuses données de plusieurs origines. Cela conduira à s'interroger sur l'usage des outils statistiques dans le contexte d'une avalanche de données par individu mais d'un nombre d'individus limité au regard de cette masse de données.

La détermination du nombre d'individus pour lesquels acquérir des données de façon à avoir un certain degré de confiance peut s'appuyer sur les travaux de David Simel¹²¹. A partir des notions de sensibilité (probabilité que le test soit positif si l'individu est malade) et de spécificité (probabilité que le test soit négatif si l'individu n'est pas malade), ce dernier définit :

- une probabilité positive (positive likelihood ratio, LR+) définie par le rapport sensibilité / (1-spécificité) ;
- une probabilité négative (negative likelihood ratio, LR-) définie par le rapport (1- sensibilité) / spécificité ;

et propose une méthode de calcul du nombre d'individus à caractériser. Par exemple, dans le cas de la maladie d'Alzheimer, si on considère qu'un individu sur cinq va convertir les facteurs de risque en maladie, que la sensibilité et la spécificité sont de l'ordre de 90% et que l'on retient un intervalle de confiance de 95%, il faut

¹¹⁹ Nous ferons état des travaux de Jean-Paul Gaudillière sur l'histoire de la médecine.

¹²⁰ Jean-Paul Gaudillière, *Inventer la biomédecine*, Paris, Editions La découverte, 2002.

¹²¹ D. L. Simel, G. P. Samba, D. B. Matchar, " Likelihood ratios with confidence : sample size estimation for diagnostic test studies", *J. Clin. Epidemiol.*, 1991, Vol. 44, N° 8, pp. 763-770.

rassembler au moins 300 individus si les critères retenus sont un $LR + \geq 2.0$ et un $LR - \leq 0.5$.

Pratiquement, il faut considérer disposer de deux catégories de données :

- Les données en grande masse, comme les données -*omics*, qui peuvent faire l'objet d'une approche de fouille de données, sans hypothèse préalable.
- Les données sans flux massif, comme les réponses à un questionnaire administré par un expert à un patient par exemple, pour lesquelles les méthodes d'analyse statistique, reposant sur des modèles, trouveront toute leur utilité.

Il faut donc envisager d'articuler deux approches complémentaires, une approche ayant une composante globale ou collective et une approche plus individualisante.

Relevons enfin que, qui dit données en grande masse, dit traitement probabiliste, la probabilité pouvant être considérée comme un concept mathématique permettant de mesurer l'incertitude, et qui dit incertitude, dit risque et renvoie à un processus complexe d'analyse et de quantification¹²². Cela nécessite la prise de conscience d'une omniprésence des incertitudes.

IV- 3 Les sources de données et les *Big Data*

Si les outils existent et se développent il reste néanmoins nécessaire d'agréger des données très nombreuses et très hétérogènes, et de prendre en compte les processus dynamiques.

¹²² André Beauchamp, « Risque (évaluation et gestion) », in Gilbert Hottois et Jean-Noël Missa, *Nouvelle encyclopédie de bioéthique. Médecine, Environnement, Biotechnologie*, Bruxelles, De Boeck Université, 2001, page 710.

Ces données, des *Big Data*, constituent des savoirs et contribuent à l'émergence de nouvelles pratiques.

Le phénomène *Big Data*, la production, le recueil et stockage de données en grand nombre, est ainsi investi de grandes ambitions dans le champ biomédical : mettre à jour de nouvelles vérités scientifiques, comprendre l'étiologie des maladies, identifier de nouvelles réponses thérapeutiques ou préventives, les *Big Data* amenant trois grands changements¹²³ :

- Le fait que nous soyons capables d'analyser de grands ensembles de données.
- La prise en compte du désordre bien réel des données, sans privilégier leur exactitude.
- La prise en compte des corrélations sans faire une quête perpétuelle d'une insaisissable causalité.

Cela s'accompagne de la sophistication des technologies développées pour produire, stocker et disséminer les données utilisées en médecine, les données génomiques par exemple, mais pas uniquement, sans oublier, comme le souligne Sabina Leonelli, que les circonstances dans lesquelles les données sont collectées, les différents niveaux d'organisation, des gènes et cellules aux organismes, populations et même écosystèmes, auxquels elles se réfèrent, affectent profondément les voies à travers lesquelles elles sont interprétées¹²⁴ et que les institutions jouent un rôle de premier plan dans les pratiques de traitement et de diffusion des données¹²⁵. Grâce aux technologies, notamment les technologies

¹²³ Viktor Mayer-Schönberger et Kenneth Cukier, *Big Data, La révolution des données est en marche*, trad. Hayet Dhifallah, Paris, Editions Robert Laffont, 2014, page 29.

¹²⁴ Sabina Leonelli, « Introduction : Making sense of data-driven research in the biological and biomedical sciences », *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43 (2012), pp : 1-3.

¹²⁵ Sabina Leonelli, *Data-Centric Biology, a philosophical study*, Chicago, MI, The University of Chicago Press, 2016.

d'analyse moléculaire à très haut débit, la médecine intensifie sa démarche d'objectivation et de quantification mais crée de nouveaux cadres épistémiques. Comprendre les mécanismes d'une maladie, identifier des cibles moléculaires sur lesquelles agir, prévenir ou retarder l'apparition d'une maladie – et c'est important dans le cas de la maladie d'Alzheimer pour laquelle il n'y a pas aujourd'hui de traitement satisfaisant ni de possibilité de guérison –, faire évoluer les pratiques de la recherche souvent focalisées sur un modèle biomédical qui, trop souvent, en privilégiant une hypothèse – la cascade amyloïde par exemple dans le cas de la maladie d'Alzheimer - réduit la complexité d'une question, ... tels sont les enjeux épistémiques qu'il faut analyser sous le regard des technologies et des outils utilisés par les médecins et cliniciens.

Les Big Data, de quoi s'agit-il ?

Les *Big Data* sont associées aux technologies de l'information et de la communication et présentent quelques caractéristiques importantes :

- les données sont produites en permanence ;
- les données sont stockées indéfiniment, du moins en théorie, car il existe des limites au stockage, peu comprises aujourd'hui, qui nécessitent des sauvegardes régulières, sans oublier que les biobanques reposent sur la volonté des politiques scientifiques d'en assurer le financement ;
- les données sont traitées massivement, et ce sont ces traitements qui peuvent aboutir à des connaissances¹²⁶ ;
- les données produisent elles-mêmes de nouvelles données.

¹²⁶ Ibid., page 45.

Les données sont construites et le processus de fabrication est aussi important que la donnée elle-même. Quand on parle de *Big Data*, on peut avoir en tête deux types de données, les données « larges » obtenues sur une grande population, et les données « profondes » obtenues en acquérant de nombreuses données biologiques et cliniques pour chaque individu. Dans le domaine médical, quand on parle de *Big Data*, ce n'est pas que la quantité qui est déterminante, mais surtout l'association, dans une même analyse, de données variées et non structurées, afin d'en déduire des informations qu'il aurait été impossible de mettre en évidence avec les analyses classiques de données structurées. On n'a pas affaire à une population, mais à des individus, aussi la situation est différente. Il y a effectivement un nombre très important de paramètres différents et hétérogènes, mais ces paramètres ne concernent qu'un nombre limité de personnes, des patients ou un groupe de référence, et généralement pas une population importante¹²⁷. Quoi qu'il en soit, avec les *Big Data*, on parle de plusieurs caractéristiques, parfois illustrées par les « 4V » canoniques :

- le volume de données ;
- la vitesse avec laquelle on a accès aux données ;
- la variété/diversité des données et des sources de données, et donc leur hétérogénéité ;
- la variabilité des données, c'est-à-dire leur caractéristique dynamique.

On devrait y ajouter « 2V », la valeur, car les données sont porteuses de sens et de connaissance et ont une valeur économique, et la véracité des données. Outre le volume, ce qui

¹²⁷ Ce n'est évidemment pas le cas des études épidémiologiques qui, elles, doivent avoir une signification au niveau d'une population.

caractérise également les données, cela a déjà été noté, c'est leur hétérogénéité :

- ce peut être une hétérogénéité liée aux phénomènes biologiques, notamment au niveau d'analyse (molécule, cellule, tissu, ...) ;
- ce peut être une hétérogénéité liée aux bases de données, au stockage et à la sélection des données ;
- ce peut être une hétérogénéité liée aux traitements des données et à la modélisation.

Les données sont des briques élémentaires d'information et plusieurs, voire de nombreuses, données sont nécessaires pour donner du sens et produire de l'information. Pour étudier, traiter et si possible prévenir les maladies qui sont le résultat d'interactions complexes entre la génétique et les facteurs environnementaux, on a besoin d'avoir accès à une multitude de données, dont des données biomédicales et des données environnementales.

Les données génétiques

Indépendamment des données phénotypiques et des données environnementales, les données génétiques sont celles qui profitent le plus des avancées technologiques. Le développement de la génomique, de la transcriptomique, de la protéomique, de la métabolomique, bref de toutes les approches « -omics », donne naissance à un nombre de données sans précédent et les nouvelles générations de séquenceurs ne font qu'amplifier toujours plus le phénomène. De nombreuses bases de données moléculaires sont disponibles, par exemple, pour ne citer que les plus connues, Genbank¹²⁸, ENA¹²⁹ ou DDBJ¹³⁰. Mais les infrastructures comme Genbank ne répondent qu'à un type de

¹²⁸ <http://www.ncbi.nlm.nih.gov/genbank/>

¹²⁹ <http://www.ebi.ac.uk/ena/>

¹³⁰ <http://www.ddbj.nig.ac.jp/>

données, alors que les nouvelles bases de données sont destinées à stocker une variété de données, produites automatiquement et obtenues expérimentalement, et à les rendre accessibles à des groupes de chercheurs ou de cliniciens pouvant avoir des cultures épistémiques très différentes.

Une base de données est un ensemble structuré et organisé permettant le stockage de grandes quantités d'informations afin d'en faciliter l'utilisation (ajout, mise à jour, recherche et éventuellement analyse). Elles sont toutes organisées en fonction d'un modèle de données qui peut être de différents types, le modèle de base de données relationnelle étant l'un des plus utilisés. Il repose sur un certain nombre de règles, notamment :

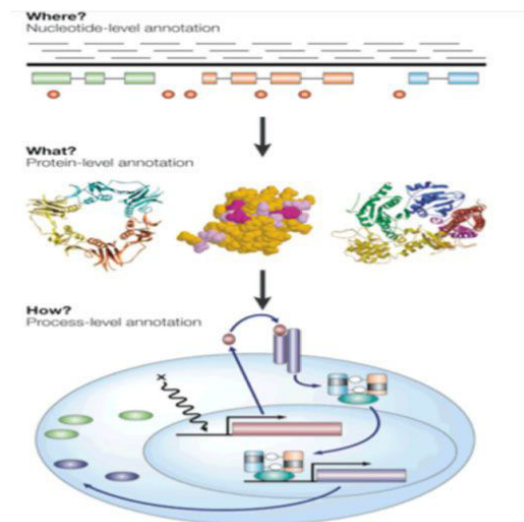
- Une règle d'unicité, toute information étant représentée d'une et d'une seule manière ;
- Une règle de garantie d'accès, toutes les données devant être accessibles sans ambiguïté ;
- Une règle de traitement des valeurs nulles, afin de pouvoir utiliser des données dont certains champs d'informations sont incomplets ;
- Une règle d'insertion, de mise à jour et d'effacement.

Les bases de données permettent notamment d'annoter les génomes, notamment le génome humain à différents niveaux. C'est un processus complexe qui peut être subdivisé en trois catégories :

- l'annotation syntaxique, qui permet d'identifier les séquences présentant une pertinence biologique (gènes, signaux, répétitions, ...) ;
- l'annotation fonctionnelle, qui permet de prédire les fonctions et produits potentiels des gènes préalablement identifiés (similitudes de séquences, motifs, structures, ...) et de

collecter d'éventuelles informations expérimentales (littérature, jeux de données à grande échelle, ...)

- l'annotation relationnelle, qui permet de déterminer les interactions que les objets biologiques préalablement identifiés sont susceptibles d'entretenir (familles de gènes, réseaux de régulation, réseaux métaboliques, ...).



Where? = annotation syntaxique

What? = annotation fonctionnelle

How? = annotation relationnelle

(d'après Lincoln Stein, *Nature Reviews Genetics*, 2, 493-503)

Les outils d'analyse et l'émergence de nouvelles données

Si on peut disposer de données pour les phénotypes, pour l'environnement et pour la génétique, il s'agit alors de voir comment intégrer ces données pour résoudre l'équation :

$$\textbf{Génome * Exposome = Phénotype}$$

Rappelons que cette équation, déjà présentée, manifeste le « jeu » entre les caractéristiques génétiques spécifiques de l'individu et l'environnement dans lequel il se trouve, aboutissant à un phénotype qui, du point de vue médical, se traduit en paramètres physiopathologiques affectant l'individu.

De nouvelles technologies et de nouveaux outils ont été nécessaires pour profiter de l'opportunité des *Big Data*, et seules des approches multidisciplinaires permettent d'en déduire des informations utiles. Il faut en effet :

- s'appuyer sur des outils mathématiques, notamment les statistiques et les méthodes d'optimisation et de réduction des données ;
- développer des techniques d'analyse des données, tels que le *Data Mining* (fouille de données), les méthodes d'apprentissage et les réseaux de neurones.

Le *Data Mining* est l'opération qui permet d'extraire une connaissance à partir des données. Il est parfois assimilé à ce que Fayyad et al.¹³¹ ont appelé le KDD (*Knowledge Discovery in Databases*). Il permet de trouver les données pertinentes dans une grande masse de données¹³² en utilisant soit un modèle statistique, comme une distribution par exemple, soit une approche d'apprentissage, soit une approche algorithmique. Son objet est de mettre en évidence des faits cachés dans une masse de données. Qu'il y ait des variables cachées dans une très grande masse de données est plutôt réaliste. Mais du point de vue épistémologique, c'est en quelque sorte un intermédiaire entre le rationalisme qui suppose que toute connaissance est fondée sur le raisonnement, et l'empirisme pour qui l'expérience est la source de toute

¹³¹ U. M. Fayyad, G. Piatetsky-Shapiro, P. Smyth, "From Data Mining to Knowledge Discovery in Databases", *AI Magazine* Vol. 17 N° 3, AAAI 1996, pp. 37-54.

¹³² Amos Bairoch, *Comment rassembler l'information : présent et futur*, Genève, Dept. Instruction Publique, 2006.

connaissance. Avec le *Data Mining*, on peut considérer que la connaissance, fondée sur des échanges entre des situations, est en construction perpétuelle et que cette construction est fondée sur l'emploi de techniques toujours renouvelées et sur l'accès à des données toujours plus nombreuses.

Pour ce faire, on peut utiliser le *clustering* ou regroupement. C'est le procédé qui consiste à examiner un ensemble de données, à les rassembler dans des groupes et à calculer une « distance » entre les groupes. Deux points d'un même groupe sont peu distants, tandis que deux points dans des groupes différents sont beaucoup plus distants.

Il faut donc définir ce qu'est un point, ce qu'est un espace et ce qu'est une distance. Un point est une donnée qui appartient à un certain espace. On peut considérer qu'un espace n'est qu'un ensemble de points, mais on peut aussi associer à une donnée des coordonnées dans un espace euclidien à n dimensions et donc associer à une donnée un vecteur dont on connaît les composantes dans l'espace. On peut alors très classiquement mesurer la distance entre deux points, distance qui n'est jamais négative (le zéro étant la distance d'un point à lui-même), qui est symétrique (la distance de A à B est la même que la distance de B à A), et qui respecte l'inégalité triangulaire (la distance de A à B et à C n'est jamais inférieure à la distance pour aller de A à C directement).

Les algorithmes de regroupement peuvent suivre deux stratégies fondamentalement différentes :

- La stratégie hiérarchique ou agglomérative qui remplit les groupes au fur et à mesure en définissant préalablement la distance maximale qui peut séparer deux données et qui peut également définir *a priori* le nombre de groupes. Cette stratégie est limitée à un ensemble de données peu volumineux ;
- La stratégie qui consiste à trouver le groupe le plus représentatif d'une donnée et à assigner cette donnée à ce groupe.

Les algorithmes les plus courants, qui assignent un point à un groupe, sont les algorithmes « *K-means* ». Ils supposent que l'espace est euclidien et que le nombre k de *clusters* est connu à l'avance (même si, par itérations successives on peut déterminer k). On choisit le nombre k de *clusters* et on assigne à chaque *cluster* un point, qui est le centre de gravité initial du *cluster*. Puis, pour chaque donnée, on mesure la distance aux k points ainsi définis et on assigne la donnée au *cluster* pour lequel la distance est la plus courte. L'algorithme BFR¹³³ (algorithme de Bradley, Fayyad et Reina) est une variante adaptée aux espaces euclidiens de grande dimension.

L'algorithme des nuées dynamiques est une généralisation de l'algorithme des *K-means*. Le grand intérêt de cette méthode de classification est sa simplicité, notamment à la mise en œuvre. En outre, il existe une variante des nuées dynamiques permettant de travailler à partir d'un tableau de distances. Le défaut de cette méthode est qu'il faut obligatoirement fixer le nombre de classes a priori.

Les grandes étapes d'un tel algorithme sont les suivantes :

1. Déterminer une partition initiale, c'est-à-dire regrouper les individus de façon arbitraire en K classes ;
2. Calculer les représentants de chacune des classes, c'est-à-dire les individus (existants ou virtuels) optimisant un critère (exemple, l'inertie interclasse) ;
3. Affecter chaque individu à la classe qui lui est la plus proche ;
4. Retourner à l'étape 2 jusqu'à ce qu'aucun des individus ne change de classe.

Les techniques d'exploration de données (*Data Mining*) se développent pour analyser la masse d'informations récupérées et stockées. Elles conduisent à la production des « index

¹³³ P. S. Bradley, U. M. Fayyad, C. Reina, « Scaling clustering algorithms to large databases », *Proc. Knowledge Discovery and Data Mining*, pp. 9-15, 1998.

composites » permettant d'acquérir des connaissances ou de prendre des décisions sans avoir à explorer toutes les données dans le détail. Dans le cas où les paramètres sont beaucoup plus nombreux que les échantillons disponibles, l'approche *Data Mining*¹³⁴ de fouille de données doit être associée à des méthodes d'apprentissage. Partant de données en très grand nombre, il s'agit de trouver des relations qui peuvent exister entre elles.

L'apprentissage peut être un apprentissage supervisé au cours duquel on se focalise entièrement sur la caractérisation d'une variable cible. C'est le cas par exemple quand on l'utilise comme aide au diagnostic en médecine. Ce peut être un apprentissage non supervisé qui peut aboutir à de nouvelles segmentations, et donc à la création d'une typologie de patients, et à la définition des politiques adaptées à chaque segment ainsi découvert. Cet apprentissage non supervisé utilise des règles d'association pour découvrir l'ensemble des relations probabilistes directes présentes au sein des données. La segmentation de variables qui en résulte permet de découvrir de nouveaux concepts et de synthétiser des variables cachées. Parmi les méthodes d'apprentissage, les méthodes de graphe permettent d'interconnecter les variables pertinentes¹³⁵.

En complément du *Data Mining*, des technologies cherchent à se rapprocher du cerveau humain qui peut capturer et analyser une myriade de données sensibles à chaque instant, et cela de façon efficace et robuste, le système de calcul du cerveau étant différent de celui d'un ordinateur puisqu'il partitionne et stocke individuellement, si bien que les calculs sont effectués de façon distribuée et en parallèle. On peut ainsi, quand on analyse des données en très grand nombre, les classer dans des groupes

¹³⁴ S. Tuffery, *Data Mining et Statistique décisionnelle : l'intelligence des données*, Paris, Technip, 2007.

¹³⁵ C. Bishop, *Pattern Recognition and Machine Learning*, Springer, 2006.

(*clusters*) avec une plus ou moins grande résolution¹³⁶. C'est là l'origine de nouvelles technologies de traitement de données comme *Hadoop*¹³⁷. *Hadoop* se présente comme un framework *open source* mettant à la disposition des développeurs et des administrateurs un certain nombre de briques essentielles :

- Les éléments d'infrastructure permettant la mise en place de *clusters* de stockage et de traitement des données. Le stockage des données s'appuie le *Hadoop Distributed File System* (HDFS), un système de fichiers permettant d'héberger les données à traiter sous forme de fichiers distribués ;
- *Map/Reduce* : Un *framework* Java de traitement distribué qui permet de développer des tâches de traitement sur les données du cluster.

Hadoop peut être considéré comme un système de traitement de données évolutif pour le stockage et le traitement par lot de très grandes quantités de données.

*Map/Reduce*¹³⁸ partage les données en sous-systèmes, dispersés dans des milliers d'ordinateurs jusqu'à ce qu'ils puissent être traités directement, avant de recombinaison les différentes solutions pour apporter une solution unique au problème d'origine (dans *Map/Reduce*, *Map* est la partie qui localise les données et résultats de chaque machine, tandis que *Reduce* agrège l'ensemble des résultats à la fin). Ce système tire son efficacité de la puissance du parallélisme tout en évitant les difficultés liées aux connexions entre des milliers de machines.

¹³⁶ On peut illustrer cela en prenant des images provenant d'un satellite. Avec des images de faible résolution, on verra les structures nuageuses associées à des phénomènes atmosphériques, tandis qu'avec des images de grande résolution, seules de petites cibles comme une ville, voire une rue apparaîtront.

¹³⁷ <http://hadoop.apache.org>.

¹³⁸ J. Deam, S. Ghemawat, « MapReduce : simplified data processing on large clusters », *ACM* 51 (1), 2008, pp. 107-113.

Une application de *Map/Reduce* dans le domaine des biotechnologies a abouti au GATK (*Genome Analysis Tool Kit*)¹³⁹ et au développement de *Crossbow*¹⁴⁰, outil qui permet de génotyper un individu en environ 3 heures en détectant 3,7 millions de mutations ponctuelles avec une précision supérieure à 99% pour un coût de l'ordre de 100 dollars.

Tous ces outils apparaissent très puissants mais ils ne peuvent être développés que par de grands professionnels de l'informatique et, jusqu'à maintenant, bien que leur potentiel soit largement reconnu dans les environnements académiques, leur utilisation à une grande échelle et en routine n'est pas encore développée.

On peut cependant relever la collaboration de l'*Institute for Genomics and Multiscale Biology* de l'Ecole de Médecine du Mont Sinaï avec la société *Cloudera*, un leader de la technologie *Hadoop*, pour appliquer les technologies *Big Data* pour « diagnostiquer, comprendre et traiter les maladies »¹⁴¹. Cette collaboration qui a démarré il y a deux ans n'a pas encore abouti à des publications. On peut cependant relever que le Docteur Eric Schadt, directeur de l'institut, a été fortement influencé par l'ouvrage de Thomas Kuhn¹⁴² et qu'il se veut visionnaire dans l'utilisation de la biologie computationnelle. Il considère que les approches classiques de la biologie moléculaire ne sont pas assez complexes et qu'il faut introduire des ruptures en utilisant les technologies permettant d'exploiter les *Big Data*. Pour cela il cherche à exploiter des données provenant tant des techniques de séquençage que des données d'expression des gènes et de traduction des gènes exprimés en protéines, pour aboutir à des diagnostics personnalisés et à la définition de traitements.

¹³⁹ A. McKenna et al., « The genome analysis toolkit : a MapReduce framework for analyzing next-generation DNA sequencing data », *Genome Research*, 2010, 20 (july), pp. 1297-1303.

¹⁴⁰ B. Langmead, M. C. Schatz, J. Lin, M. Pop, S. L. Salzberg, « Searching for SNPs with cloud computing », *Genome Biol.*, 2009, 10, R134.

¹⁴¹ Jeff Hammerbacher, directeur scientifique de Cloudera, <http://www.marketwire.com/press-release/Cloudera-Chief-Scientist-Jeff-Hammerbacher-Teams-With-Mount-Sinai-School-Medicine-1676135.htm>.

¹⁴² Thomas Kuhn, *La structure des révolutions scientifiques*, Paris, Editions Flammarion, Collection Champs sciences, 2008.

Ainsi, les technologies sont encore dans l'enfance et sont loin d'être suffisamment matures pour être utilisées de façon routinière, d'autant que programmer *Hadoop* suppose un très haut niveau d'expertise en java d'une part, que la sortie des résultats doit être visualisée de façon adaptée à une utilisation en dehors de la recherche d'autre part. Par ailleurs, si ces approches permettent de traiter de nombreuses données en parallèle, elles le font d'un point de vue statique, sans prendre en compte l'aspect dynamique des problèmes. Enfin, un algorithme ne peut être neutre dès l'instant où on lui demande de faire des choix pour trier, filtrer ou ordonner les informations.

Chapitre V

Les biomarqueurs et la maladie d'Alzheimer

La maladie d'Alzheimer se développe au moins 15 ou 20 ans avant l'apparition des premiers signes visibles. Tout l'enjeu est de pouvoir détecter les personnes malades – ou même à risque – avant les premiers symptômes. Etant donné que les changements neurodégénératifs associés à la maladie d'Alzheimer apparaissent de nombreuses années avant l'apparition des signes cliniques, des biomarqueurs, dont nous discuterons l'intérêt, sont envisagés pour guider des interventions précoces¹⁴³. Le modèle dominant considère que la bêta amyloïde commence à s'agréger dans le cerveau de nombreuses années avant l'apparition de signes cliniques et que ces dépôts peuvent être mis en évidence par des techniques d'imagerie comme l'imagerie par émission de positrons ou des analyses du fluide céphalorachidien¹⁴⁴, ou encore que des marqueurs sanguins ou des marqueurs génétiques peuvent permettre de détecter des personnes à risque en phase asymptomatique. On peut espérer qu'en combinant toutes les informations sur le cerveau des gens, avant les symptômes, on va pouvoir prédire s'ils sont sur la trajectoire de la maladie. Dans une première partie, nous présentons l'état de la situation, notamment les marqueurs candidats potentiels. Puis, dans une deuxième partie, nous abordons la question des controverses associées au diagnostic de la maladie d'Alzheimer.

¹⁴³ B. Dubois et al., « Revising the definition of Alzheimer's disease : a new lexicon », *Lancet Neurology* 9, 2010, p. 1118-1267.

¹⁴⁴ Le liquide céphalorachidien est le liquide biologique transparent dans lequel baignent le cerveau et la moelle épinière.

V-1 Les biomarqueurs de la maladie d'Alzheimer

Dans cette partie, nous présentons un état de l'art concernant les différents marqueurs envisagés pour la maladie d'Alzheimer, en analysant successivement les marqueurs issus des techniques d'imagerie, les marqueurs présents dans le liquide céphalo-rachidien, puis les marqueurs sanguins, dont les marqueurs génétiques.

Imagerie

Historiquement, la maladie d'Alzheimer a été reconnue à deux lésions caractéristiques, apparaissant bien avant les signes cliniques de la maladie :

- Les plaques séniles ou plaques amyloïdes, constituées de dépôts extracellulaires résultant de l'agrégation d'un peptide nommé bêta-amyloïde ($A\beta$), et siégeant entre les neurones.
- Les dégénérescences neurofibrillaires résultant de l'hyperphosphorylation d'une protéine, la protéine Tau.

Aussi, les technologies qui permettent une imagerie directe de l'accumulation de $A\beta$ dans le cerveau humain apparaissent particulièrement prometteuses :

- La technologie de tomographie par émission de positrons (PET) utilisant un radiotraceur comme ^{11}C -labeled Pittsburgh compound B (PiB) confirme l'importance des dépôts de $A\beta$ chez les personnes manifestant des signes de déficience cognitive légère¹⁴⁵.
- L'imagerie par résonance magnétique fonctionnelle (fMRI) montrent que les dépôts de $A\beta$ commencent principalement dans la région de l'hippocampe¹⁴⁶.

¹⁴⁵ W. E. Klunk et al., « Imaging brain amyloid in Alzheimer's disease with Pittsburgh Compound B », *Ann Neurol* 2004, 55, 306-319.

¹⁴⁶ R. L. Buckner et al., « Cortical hubs revealed by intrinsic functional connectivity : mapping, assessment of stability, and relation to Alzheimer's disease », *J Neurosci* 2009, 29, 1860-

Cela a conduit à recommander un modèle, combinant plusieurs techniques d'imagerie pour évaluer l'évolution de la maladie d'Alzheimer¹⁴⁷.

Liquide céphalo-rachidien

Le liquide céphalo-rachidien, issu d'une ponction lombaire, est considéré comme la meilleure source de biomarqueurs pour la maladie d'Alzheimer car il interagit directement avec l'espace extracellulaire du cerveau et peut ainsi être le bon reflet des altérations biochimiques qui s'y produisent.

A priori, personne n'envisage qu'un seul biomarqueur suffise, mais certains défendent qu'une combinaison de biomarqueurs issus du liquide céphalo-rachidien permet de diagnostiquer une maladie d'Alzheimer, même à un stade précoce. La concentration en Aβ42 diminue bien avant l'apparition des signes cliniques tandis que l'augmentation de la concentration en protéine *Tau* semble beaucoup plus tardive. Cependant, selon Flora Duits¹⁴⁸, un rapport « protéine *Tau* totale / Abêta42 > 0,52 » représente un profil caractéristique de la maladie d'Alzheimer, et ce d'autant plus que l'évolution de ces marqueurs semble capable de différencier les personnes qui évoluent du stade préclinique au stade prodromal de celles qui ne suivent pas cette évolution¹⁴⁹, le profil biologique en faveur de la maladie d'Alzheimer combinant une diminution de la protéine amyloïde et une augmentation de la protéine *Tau* totale et phosphorylée. Il est cependant relevé, et ce point doit être

1873.

¹⁴⁷ Jack CR Jr, et al., « Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade », *Lancet Neurol*, 2010, 9, 119-128.

¹⁴⁸ F. H. Duits & al., « The cerebrospinal fluid « Alzheimer profile » : Easily said, but what does it mean ? », *Alzheimer's & Dementia*, 2014, pp. 1-11.

¹⁴⁹ N. Mattsson, K. Blennow, H. Zetterberg, « CSF biomarkers : pinpointing Alzheimer pathogenesis », *Ann. NY Acad. Sci.*, 2009, 1180, pp. 28-35.

approfondi, que la spécificité de ces biomarqueurs diminue avec l'âge. Il faut ajouter que la ponction lombaire présente un caractère invasif, parfois mal accepté.

Marqueurs sanguins

D'autres travaux se développent pour identifier des biomarqueurs, associés à la maladie d'Alzheimer, détectables directement dans le flux sanguin, ce qui permettrait une approche non-invasive, moins coûteuse et facilement généralisable, tout en pouvant, c'est du moins l'objectif, être adaptée à la détection de signes significatifs avant la survenue des phases symptomatiques.

Cependant, le sang est un milieu complexe et l'approche par voie sanguine présente plusieurs difficultés.

La première est due au fait que la maladie d'Alzheimer est plutôt une maladie de la vieillesse, si bien que les personnes susceptibles d'être concernées souffrent généralement d'altérations diverses. Leur profil sanguin représente l'accumulation des altérations de tous les tissus, pas uniquement du cerveau. Dans le protéome sanguin, hautement complexe, il faut développer des méthodes pour identifier ce qui ressort de la maladie d'Alzheimer de ce qui ressort des autres altérations de l'organisme.

La seconde est associée à l'existence de la barrière hémato-encéphalique (BHE)¹⁵⁰ qui contrôle et restreint les échanges entre le sang et le cerveau. Certes, elle sert à réguler le milieu dans le cerveau en le séparant du sang et à le protéger des agents pathogènes, des toxines et des hormones circulant dans le sang,

¹⁵⁰ La BHE est constituée d'une paroi vasculaire continue faite de cellules endothéliales cérébrales jointives, entourée d'autres cellules dont les astrocytes qui forment une enveloppe supplémentaire. Cette double paroi exerce une fonction de sélection et de tri des substances par un mécanisme dynamique pour apporter des substances nutritives vers le cerveau ou au contraire réguler l'élimination des déchets.

mais elle constitue un filtre extrêmement sélectif dans les deux sens de l'échange et restreint donc le mouvement des grosses protéines du cerveau vers le sang. Avec l'âge, cette barrière peut être altérée et cette altération débute dans l'hippocampe, région du cerveau impliquée dans la mémorisation, et il se peut que cette altération empêche la barrière d'éliminer le peptide A-bêta, ce qui peut déclencher la maladie d'Alzheimer.

Néanmoins, par analyse sanguine, on peut mettre en évidence la baisse de concentration de l'APP (*amyloid precursor protein*), des marqueurs de l'oxydation (réduction de l'activité d'antioxydants), des marqueurs associés au système immunitaire ou à des changements micro-vasculaires, mais ces marqueurs sont loin d'être validés et de faire consensus.

En fait, la véritable avancée dans les marqueurs sanguins réside dans l'utilisation des méthodes protéomiques. Les techniques protéomiques, la possibilité de rassembler des cohortes plus importantes et de coordonner des développements multicentriques ont permis d'identifier des ensembles de protéines susceptibles de servir de biomarqueurs pour la maladie d'Alzheimer. Ainsi, Ray¹⁵¹ et ses collègues ont identifié un ensemble de 18 protéines, présentes dans le plasma sanguin, notamment des protéines endocrines, et corrélées avec des niveaux de Abêta et de protéines Tau dans le liquide céphalo-rachidien. La difficulté, cependant, est de valider ces profils protéiques de façon reproductible. Plus récemment, Abdul Hye¹⁵² et ses collègues ont analysé 26 protéines sur des cohortes de 1148 personnes, dont 476 personnes diagnostiquées comme ayant la maladie d'Alzheimer, 220 personnes ayant des atteintes cognitives sans signes cliniques,

¹⁵¹ S. Ray & al., « Classification and prediction of clinical Alzheimer's diagnosis based on plasma signaling proteins », *Nat. Med.*, 2007, 13, pp. 1359-1362.

¹⁵² Abdul Hye & al., « Plasma proteins predict conversion to dementia from prodromal disease », *Alzheimer's & Dementia*, 2014, article in press.

au stade MCI (*Mild Cognitive Impairment* ou déficience cognitive légère) et 452 personnes âgées ne présentant aucun trouble. Ce qui est intéressant, c'est qu'ils en ont tiré 10 protéines (Clusterine, RANTES, Transthyréline, Cystatine C, A1AcidG, ICAM1, PEDF, A1AT, ApoC3, CC4), sans oublier l'ApoE, pouvant être potentiellement des marqueurs de la conversion du stade préclinique (MCI) au stade clinique prodromal puis au stade démentiel.

Tout ceci est cependant sujet à controverse car chaque équipe utilise ses propres méthodes et ses propres protocoles. Même si ces résultats suscitent beaucoup d'espoir, le manque de procédures normalisées, ne serait-ce qu'au niveau des réactifs utilisés pour les analyses, notamment les anticorps, font qu'il est difficile de valider et de généraliser les résultats, les différences obtenues entre différentes approches pouvant refléter plus des différences dans les techniques et les méthodes que dans les processus biologiques pathologiques. Il faudrait systématiquement utiliser des techniques physiques comme la spectrométrie de masse pour rendre comparables les résultats obtenus par différentes équipes sur de larges cohortes de personnes.

Il faut aussi relever qu'aucune des protéines identifiées par les différentes équipes ne présentent une véritable spécificité. C'est le cas, par exemple, de celles faisant partie du profil proposé par Abdul Hye et ses collègues :

- La clusterine, une apolipoprotéine qui peut former des complexes avec l'Abéta et pour lesquelles des études¹⁵³ ont montré que son taux dans le sang était plus élevé chez les personnes atteintes de la maladie d'Alzheimer, présente sans doute une certaine spécificité, mais son taux au stade pré-symptomatique est-il suffisamment discriminant ?

¹⁵³ E. M. Schrijvers et al., « Plasma clusterin and the risk of Alzheimer disease », *JAMA*, 305 (13), pp. 1322-1326.

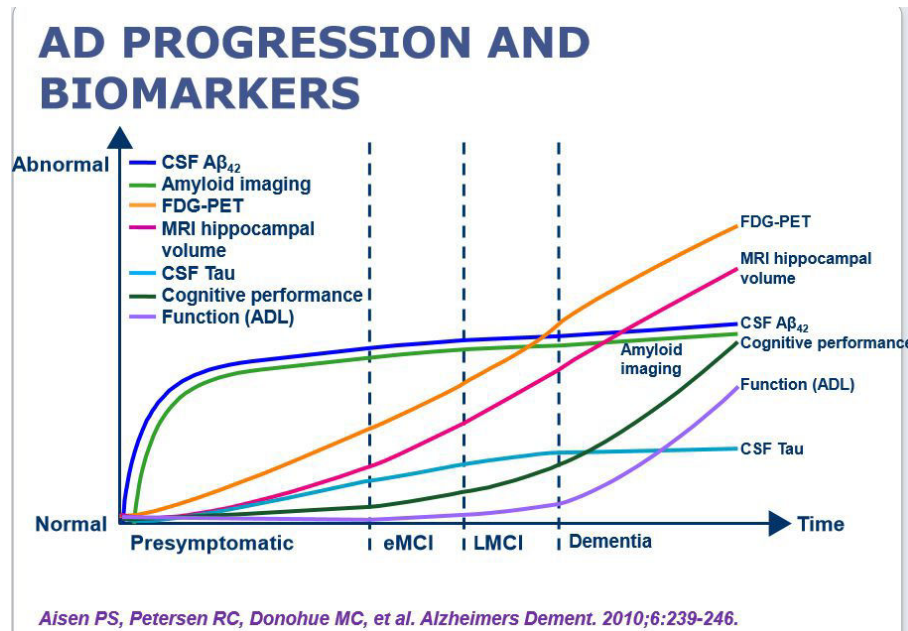
- La RANTES (*Regulated on activation, normal T cell expressed and secreted*) est une cytokine ayant un pouvoir chimiotactique¹⁵⁴. Elle est exprimée dans plus de cent maladies humaines.
- La transthyréline (TTR), très présente dans le plasma et le liquide cérébro-spinal, est plutôt un marqueur précoce de la dénutrition.
- La cystatine C est un marqueur de la fonction rénale.
- L'ICAM1 est une molécule d'adhésion intermoléculaire.
- La PEDF (*Pigment epithelium-derived factor*) présente une multitude de fonctions, dont des fonctions antitumorales.
- L'A1AT, alpha-1 antitrypsine, est connue essentiellement pour protéger les tissus contre des enzymes produites par des cellules inflammatoires.
- L'ApoC3 est plutôt utilisée comme facteur de risque de maladies coronariennes.

On peut ajouter également, que dans les études protéomiques, on a rarement accès à toutes les protéines et que celles qu'on analyse sont souvent celles pour lesquelles on dispose de méthodes analytiques. Le développement de l'électrophorèse bidimensionnelle, avec des gradients de pH suffisamment larges et précis, et donc résolutifs, couplée à la spectrométrie de masse, ouvrirait sans doute d'autres perspectives et permettrait d'aboutir à des profils de protéines toujours plus spécifiques.

Par ailleurs, demeure une question essentielle : les biomarqueurs sanguins détectés ne reflètent pas nécessairement la cause de la maladie d'Alzheimer et n'ont qu'une valeur probabiliste qui nécessite de les associer à d'autres informations, comme l'IRM par exemple.

¹⁵⁴ La chimiotaxie est le phénomène par lequel des cellules se dirigent ou dirigent leurs mouvements en fonction de certaines espèces chimiques présentes dans leur environnement.

Il reste les biomarqueurs génétiques, que nous allons aborder. A ce stade, on peut cependant relever que les biomarqueurs de la maladie d'Alzheimer sont donc nombreux :

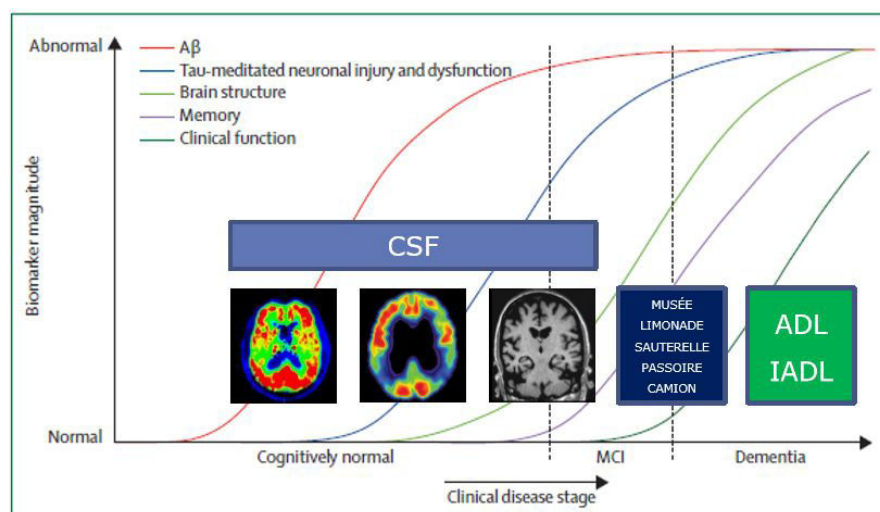


Ils peuvent également être plus ou moins associés aux différentes phases de la maladie :

Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade

Lancet Neurol 2010; 9: 119-28

Clifford R Jack Jr, David S Knopman, William J Jagust, Leslie M Shaw, Paul S Aisen, Michael W Weiner, Ronald C Petersen, John Q Trojanowski



Enfin, et nous y reviendrons, sans doute faut-il considérer qu'il n'y a pas un niveau de causalité privilégié et qu'il faut envisager plusieurs niveaux d'organisation du fonctionnement biologique.

Marqueurs génétiques

Outre des molécules biologiques présentes dans le sang ou dans les liquides corporels, des mutations du génome peuvent témoigner d'un processus normal ou anormal, ou de la présence d'un trouble ou d'une maladie. C'est le cas pour la maladie d'Alzheimer.

Nous ne nous intéressons pas aux formes héréditaires de la maladie d'Alzheimer. Elles sont exceptionnelles, mais trois gènes sont impliqués, codant soit pour le précurseur APP (*Amyloid Precursor Protein*) du peptide *Abêta* qui présente plus de 30 mutations dont au moins 25 peuvent causer la maladie d'Alzheimer¹⁵⁵, soit pour les présénilines 1 et 2 (PS1 et PS2), gènes fortement influencés par des facteurs environnementaux car des études sur des jumeaux ont montré qu'il pouvait y avoir jusqu'à 10 ans d'écart dans la survenue de la maladie chez des vrais jumeaux¹⁵⁶. Des mutations pathogènes au sein de ces gènes provoquent des modifications du métabolisme du précurseur du peptide amyloïde pouvant entraîner une augmentation de la quantité totale des différentes formes du peptide amyloïde *Abêta*, et en particulier une production préférentielle de la forme *Aβ42* reconnue comme responsable de la formation des dépôts amyloïdes¹⁵⁷.

¹⁵⁵ A. Goate and al., "segregation of a Missense Mutation in the Amyloid Precursor Protein Gene with Familial Alzheimer's Disease », *Nature* 349 (1991) : 704-6.

¹⁵⁶ L. Tilley and al., « Genetic Risk Factors in Alzheimer's Disease », *Molecular Pathology* 51, n°6 (1998) : 293-304.

¹⁵⁷ M. C. Chartier-Harlin, L. Araria-Goumidi, J-C. Lambert, « Les formes tardives de la maladie d'Alzheimer : de la génétique à la biologie », *Médecine/Sciences*, 2002, , vol. 18, pp. 709-716.

Nous nous intéressons plutôt à l'identification de marqueurs génétiques pouvant déboucher sur des tests génétiques de la maladie d'Alzheimer non-héréditaire, avec le souci d'aller vers une stratégie de prévention de cette maladie. Pour cela, sont largement utilisées les études d'association au niveau du génome, les GWAS (*Gene-Wide Association Studies*), déjà introduites. Ces GWAS ont largement le potentiel d'identifier des biomarqueurs et leur utilisation pour suivre le développement et la progression de la maladie¹⁵⁸. Les GWAS sont utilisées pour détecter les variants génétiques, notamment les mutations ponctuelles ou SNPs, supposés avoir un lien avec une maladie, donc augmenter le risque de la maladie, sans pour autant avoir un lien de cause à effet. L'intérêt des GWAS est de considérer le génome dans sa totalité et de détecter les SNPs qui apparaissent associés à des traits de la maladie d'Alzheimer tels que la perte de mémoire ou le dépôt de plaques amyloïdes.

Toutes les GWAS sur la maladie d'Alzheimer confirment l'importance du gène de l'apolipoprotéine E (ApoE).

Le gène APOE, situé sur le chromosome 19, est essentiel pour le métabolisme, le transport et la transformation des lipides et du cholestérol LDL. Il présente trois isoformes différentes chez l'homme, les allèles APOEε2, APOEε3 et APOEε4. Chaque variant produit une protéine différente et c'est le variant ε3 qui est le plus répandu dans la population humaine. En 1993, le généticien Allen Roses fut le premier à associer explicitement l'allèle APOEε4 et l'augmentation du risque de la maladie d'Alzheimer non familiale¹⁵⁹. Cette association fut confirmée par plus de cent laboratoires dans le monde, en deux à trois ans¹⁶⁰.

¹⁵⁸ L. Bertram, H. Hampel, « The role of genetics for biomarker development in neurodegeneration », *Prog. Neurobiol.*, 2011, 95, 501-504.

¹⁵⁹ E. H.Corder, A. D. Roses and al., "Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer's Disease in Late Onset Families", *Science* 261 (1993) : 921-23.

¹⁶⁰ D. Roses Allen, « Apolipoprotein E and Alzheimer's Disease : A Rapidly Expanding Field with Medical and Epidemiological Consequences », *Annals of the New York Academy of Sciences* 802, n° 1 (1996) : 50-57.

La présence de l'allèle e4 du gène de l'ApoE est considérée comme un facteur de risque important de la maladie d'Alzheimer. On estime que le gène de l'ApoE pourrait être impliqué dans au moins 20 % des cas de maladie d'Alzheimer¹⁶¹. Le génotype ApoEe4e4 (homozygote e4 pour ApoE) est présent chez environ 2 % de la population française, conférant un risque cumulé à 75 ans d'être atteint de cette maladie d'environ 30 %, et de plus de 50 % à 85 ans¹⁶². Mais, si, clairement, l'allèle APOEe4 est un gène de susceptibilité à la maladie d'Alzheimer, il n'est néanmoins ni nécessaire ni suffisant pour provoquer cette maladie¹⁶³, ce qui peut signifier que plusieurs voies métaboliques sont impliquées dans la maladie et que toutes ne concernent pas l'apolipoprotéine E.

Le gène APOE4, un statut épistémique évolutif

Dès 1993, Allen Roses et son équipe à la *Duke University* publièrent un article¹⁶⁴ montrant explicitement l'association entre l'APOEe4 (APOE4) et une augmentation du risque de la maladie d'Alzheimer. Cette association fut confirmée en deux à trois ans par plus de 100 laboratoires dans le monde¹⁶⁵. Cette découverte a suscité beaucoup d'espoir car, pour de nombreux chercheurs, elle devait permettre de traiter la maladie d'Alzheimer.

En fait, il est vite apparu que l'APOE4 n'est qu'un gène de susceptibilité qui n'est ni nécessaire ni suffisant pour causer la maladie d'Alzheimer¹⁶⁶. Si le gène intervient bien dans une voie métabolique pouvant aboutir à la formation des plaques et des enchevêtrements fibrillaires, d'autres voies

¹⁶¹ V. S. Rao, A. Cupples, C. M. Van Duijn, et al., « Evidence for major gene inheritance of Alzheimer disease in families of patients with and without apolipoprotein E epsilon 4 », *Am. J. Hum. Genet.*, 1996 ; 59, pp. 664-675.

¹⁶² F. Gzil, E. Hirsch, *Alzheimer, éthique et société*, Paris, Editions Eres, 2012, page 607.

¹⁶³ La moitié des individus homozygotes pour l'allèle APOEe4 n'ont jamais la maladie d'Alzheimer, et entre 30% et 60% de ceux qui développent la maladie d'Alzheimer ne portent pas l'allèle APOEe4.

¹⁶⁴ A. D. Roses et al., « Gene Dose of Apolipoprotein E Type 4 Allele and the Risk of Alzheimer's Disease in Late Onset Families », *Science* 261 (1993) : 921-23.

¹⁶⁵ A. D. Roses, et al., « Apolipoprotein E and Alzheimer's Disease : A Rapidly Expanding Field with Medical and Epidemiological Consequences », *Annals of the New York Academy of Sciences* 802 n°1 (1996) : 50-57.

¹⁶⁶ La moitié des personnes homozygotes pour APOE4 n'ont jamais la maladie d'Alzheimer et 30 à 60 % des gens qui développent cette maladie ne portent pas l'allèle 4 de l'apolipoprotéine E.

métaboliques existent et, même si l'association de l'APOE4 avec la maladie d'Alzheimer a été parfaitement établie, il n'y a pas de consensus sur la façon dont cette association se traduit sur le plan physiopathologique¹⁶⁷. Par ailleurs, des anthropo-biologistes comme Alan Templeton ont fait remarquer que les génomes sont généralement organisés en « clusters » de gènes reliés fonctionnellement et qu'il est dangereux de ne considérer qu'un seul gène¹⁶⁸.

Dès 1996, il a été proposé de ne pas utiliser l'APOE pour des tests prédictifs de la maladie d'Alzheimer mais simplement pour confirmer un diagnostic de la maladie d'Alzheimer probable d'individus présentant des symptômes de démence¹⁶⁹ mais des industriels s'étaient déjà saisis de l'APOE4. En 1994, la société Genica Pharmaceuticals de Montréal lançait un premier test basé sur l'APOE4 pour la clinique et en 1996, la société Athena Neurosciences introduisait le « ADmark ApoE Genetic Test ». Ces deux sociétés voulaient commercialiser des tests prédictifs de la maladie d'Alzheimer, mais la résistance des cliniciens, qui voulaient un diagnostic précoce de la maladie d'Alzheimer et non la mise en évidence d'un facteur de risque, et les problèmes éthiques qui ont accompagné cette résistance, ont conduit les industriels à ne présenter ces tests que comme des tests de confirmation de la maladie d'Alzheimer pour des patients présentant des signes de démence, détectés par ailleurs¹⁷⁰ : « *APOE genotyping is useful only when applied to people already diagnosed with probable AD by clinical evaluation* »¹⁷¹.

A la même période, le laboratoire pharmaceutique Warner-Lambert se préparait à lancer le premier médicament contre la maladie d'Alzheimer, la Tacrine, un inhibiteur de l'acetyl cholinesterase qui devait ralentir le développement des symptômes de démence, approuvé en 1993 malgré les doutes émis par la Food and Drug Administration concernant la méthodologie et la validité des essais préalablement au lancement¹⁷². Le gène APOE4 est apparu au cours d'essais.

¹⁶⁷ L. Bertram et R. E. Tanzi, « Alzheimer's Disease : One Disorder, Too Many Genes », *Human Molecular Genetics* 13, n° 90001 (January 13, 2004) : R137.

¹⁶⁸ A. R. Templeton, « The Complexity of Genotype-Phenotype Relationship and the Limitations of Using Genetic 'Markers' at the Individual Level », *Science in Context* 11, n° 3 (1998) : 373-89.

¹⁶⁹ A. D. Roses, et al., « Apolipoprotein E and Alzheimer's Disease : A Rapidly Expanding Field with Medical and Epidemiological Consequences », *Annals of the New York Academy of Sciences* 802 n°1 (1996) : 50-57.

¹⁷⁰ Adam Hedgecoe, *The Politics of Personalised Medicine, Pharmacogenetics in the Clinic*, Cambridge UK, Cambridge University Press, 2004

¹⁷¹ Ibid., page 50.

¹⁷² FDA 1991, « Tacrine as a treatment for Alzheimer's dementia : an interim report from the FDA », *New England Journal of Medicine* 324(5), 1991 : 349-52.

Ainsi, en 1995, une étude menée par Judes Poirier¹⁷³ montrait que les patients porteurs de l'allèle APOE4 répondait moins bien au traitement par la Tacrine que ceux porteurs d'autres allèles de l'apolipoprotéine E, ouvrant ainsi la porte à l'entrée de la pharmacogénétique dans la clinique.

Cela a conduit à des tensions entre les chercheurs et les cliniciens dont les besoins sont différents. Alors que les biologistes souhaitent faire avancer les recherches sur la maladie d'Alzheimer, notamment les recherches génétiques, les cliniciens souhaitent disposer d'un diagnostic précoce et fiable pour la maladie d'Alzheimer. Pour les cliniciens, d'une part l'APOE4 n'apporte qu'une connaissance a posteriori sur quelque chose qu'ils savent déjà, d'autre part un test pharmacogénétique basé sur l'APOE4 crée trop de problèmes éthiques et sociaux dans leur pratique clinique, la relation entre l'APOE4 et les chances de développer la maladie d'Alzheimer étant encore trop complexe pour indiquer un risque à un patient.

Aujourd'hui, l'APOEε4 est toujours considéré comme un facteur de risque majeur, puissant statistiquement, mais qui ne permet en rien de pronostiquer l'avenir ou mettre en place une thérapeutique. Dans la mesure où il est appréhendé comme facteur de risque, il n'est pas recommandé de proposer un diagnostic moléculaire de l'APOEε4, quelles que soient les conditions. Donc, *a priori*, un sujet participant à une recherche sur la maladie d'Alzheimer n'a théoriquement pas à être informé du résultat qui le concerne sur son allèle de l'APOE puisqu'il n'a pas d'implication clinique pertinente à ce jour (si ce n'est pour la recherche). Il y a cependant une exception si le sujet est homozygote APOEε4 puisque ses enfants seront nécessairement porteurs de l'allèle 4. On considère alors que l'information doit être divulguée car elle ne concerne pas que le sujet mais aussi sa descendance. On peut d'ailleurs ajouter que le gène APOE4 est un puissant agrégateur familial. Dans les familles où a été identifiée une agrégation du facteur APOE4, la demande de participer à la recherche est forte

¹⁷³ J. Poirier et al., « Apolipoprotein E allele as a predictor of cholinergic deficits and treatment outcome in Alzheimer disease », Proceedings of the National Academy of Sciences 92 (1995) : 12260-4.

car la crainte d'être un jour malade comme tel ou tel membre de la famille est réelle et forte.

Etant donné que le gène APOEε4 n'est ni nécessaire ni suffisant pour provoquer la maladie d'Alzheimer, l'hypothèse peut être faite que d'autres gènes peuvent être impliqués et les études GWAS doivent pouvoir les mettre en évidence. Dans le cas de la maladie d'Alzheimer, il est vite apparu qu'il n'y avait pas quelques variants qui pouvaient expliquer la maladie, mais que de nombreux gènes, avec peu d'effets, apparaissaient comme pouvant être importants. Si toutes les GWAS sur la maladie d'Alzheimer confirment bien que le gène APOEε4, notamment dans sa forme homozygote, est un facteur de risque, elles montrent également que plus de 30 variants, sur différentes voies métaboliques, peuvent être associés à la maladie d'Alzheimer¹⁷⁴. En particulier, en plus de l'APOE, trois gènes, le CLU (clusterine), le CR1 et le PICALM semblent fortement associés à la maladie d'Alzheimer¹⁷⁵ :

- La clusterine, produite par le gène CLU, est associée à l'élimination des débris cellulaires et à l'apoptose ;
- Le gène CR1 (*complement receptor 1 gene*) est associé au fonctionnement du système immunitaire, et sa sur-expression entraîne des dommages dans les tissus, notamment ceux du système nerveux central ;
- Le PICALM intervient dans le transport des molécules dans les cellules nerveuses et joue un rôle dans la formation de la mémoire.

¹⁷⁴ Cela a conduit à la constitution d'une « AlzGene » database. Voir BERTAM Lars et TANZI Rudolph E., « Genome-wide Association Studies in Alzheimer's Disease », *Human Molecular Genetics* 18, n° R2 (2009) : 270-81.

¹⁷⁵ Gabrielle Strobel, « Paper Alert : GWAS Hits Clusterin, CR1, PICALM Formally Published », *Alzheimer Research Forum*, September 7, 2009, <http://www.alzforum.org/new/detail.asp?id=2233>.

Il faut s'interroger sur l'interprétation que l'on donne aux GWAS. Tient-on compte de l'impact biologique des gènes ou donne-t-on trop d'importance aux valeurs statistiques, notamment à la p-valeur ? Par exemple, si 1000 individus entrent dans l'étude et qu'on obtient une p-valeur de 0,003, en testant 10 000 individus, on obtiendra une p-valeur supérieure, ce qui sera analysé comme plus significatif, alors que l'information biologique reste exactement la même. En fait, comme cela a été dit, au lieu de considérer la p-valeur, il est préférable de retenir le « *odds ratio* ». C'est une mesure de l'effet de la taille de l'échantillon décrivant la force d'une association ou la non-indépendance entre deux données. La difficulté est que dans les études GWAS, les *odds ratios* ne sont pas grands.

Par ailleurs, dans l'enthousiasme qui a suivi le séquençage du génome humain, de nombreux scientifiques, en s'appuyant sur quelques constatations isolées, ont supposé que certaines variations dans le génome étaient plus souvent associées à certaines maladies. Cette hypothèse, dite des variations fréquentes, n'a pas vraiment donné de résultats, car « la majorité des variations fréquentes n'ont pas permis d'établir de correspondance biologique avec une maladie et n'ont pas d'utilité clinique en ce qui concerne le pronostic ou le traitement »¹⁷⁶. Cela a amené certains à dire, soit qu'il ne fallait s'intéresser qu'aux variations génétiques rares¹⁷⁷, soit que l'importance de l'environnement des gènes a une telle influence sur l'expression des gènes qu'il doit être pris prioritairement en compte. La difficulté additionnelle, c'est que ces variants rares, tout en pouvant avoir un effet important ne sont pas détectés par la majorité des GWAS car ces études ne se focalisent généralement que sur des variants qui apparaissent dans au moins 5% de la population étudiée.

¹⁷⁶ Stephen Hall, « Génome et médecine : la révolution retardée », Pour la Science, N° 399, 2011, pp. 36-43.

¹⁷⁷ Une variation génétique est considérée comme rare si la mutation est présente chez moins de deux pour cent de la population.

D'un autre côté, les fonctions associées au génotype APOE, par exemple, sont insaisissables car les effets sont modifiés par la présence d'autres gènes et des variables environnementales. Il y a donc de profondes difficultés pour estimer le risque d'avoir la maladie d'Alzheimer sur la base d'un génotypage APOE. Tout cela relève de la complexité du vivant. Les organismes vivants sont hétérogènes et composés de plusieurs parties en interaction dynamique les unes avec les autres et à différentes échelles, depuis l'échelle moléculaire jusqu'à l'organisme entier en passant par l'échelle cellulaire et l'échelle des organes. Une approche trop géo-centrée doit être remplacée par une approche plus centrée sur l'épigénétique, c'est-à-dire non seulement sur la base du génome de la personne, mais également sur son environnement moléculaire interne et sur l'ensemble de ses conditions de vie et du contexte dans lequel elles se situent.

Toutes les études s'expliquent par le souci d'aller vers une stratégie de prévention de la maladie d'Alzheimer, et on peut rappeler que plus de 1200 études GWAS ont été menées sur cette maladie. Elles ont pour objet de comprendre comment les gènes et leurs produits interviennent tout au long de la vie des individus, c'est-à-dire de documenter en détail les voies métaboliques impliquées au cours du temps. Mais il faut noter que les données de départ doivent être contextualisées si on veut espérer que les gènes ainsi mis en évidence soient réellement associés à la maladie d'Alzheimer, dont la complexité n'est plus à démontrer. Il y a plusieurs processus impliqués et les gènes identifiés codent pour des molécules qui ont plusieurs fonctions. C'est ainsi que certains gènes peuvent suggérer que le système immunitaire et une inflammation peuvent avoir une implication causale dans la maladie d'Alzheimer, ce qui pourrait laisser supposer que l'inflammation peut précéder le dépôt amyloïde et remettre en cause le modèle dominant. La prise en compte de l'APOE a d'ailleurs été combattu

initialement car le modèle présenté¹⁷⁸ suggérait que les allèles APOEε2 et APOEε3 étaient des allèles protecteurs, protégeant la protéine *Tau* des dommages pouvant conduire à la mort neuronale, et ainsi à la maladie d'Alzheimer, tandis que l'allèle APOEε4 supprimait ce rôle protecteur. Un tel modèle n'était pas conforme au modèle dominant pour expliquer la maladie d'Alzheimer, modèle basé sur la bêta-amyloïde et la cascade amyloïde. En rejoignant Ludwig Fleck¹⁷⁹, on retrouve la notion de « collectif de pensée », le collectif de pensée pouvant être défini comme « la communauté de personnes qui échangent des idées et interagissent intellectuellement »¹⁸⁰. Cela traduit que c'est par la rencontre que de nouvelles idées émergent, mais également que la production de connaissances se fait dans un contexte social, psychologique, idéologique, matériel et technique donné. Fleck insiste notamment sur le fait que ce style de pensée est le fait de scientifiques, et non de cliniciens, même s'il est consolidé ou conforté par le contexte social.

V-2 Le diagnostic de la maladie d'Alzheimer et les controverses

Les images de l'imagerie par résonance magnétique ou tomographie par émission de positrons, les données cliniques, génétiques, protéomiques, biologiques obtenues sur le sang ou le fluide cérébrospinal, les tests de mémoire sont autant de données, pas nécessairement « *big* » au sens où elles s'accumulent à grande vitesse et se présentent sous des formes très diverses, mais plutôt « *big* » en terme de volume. En ce sens, la communauté scientifique place beaucoup d'espoir dans la collecte de toutes les données pouvant être utiles à cette détection précoce, et porte tout

¹⁷⁸ A. D. Roses, « Apolipoprotein E alleles as risk factors in Alzheimer's disease », *Annual Review of Medicine* 47 (1996) : 387-400.

¹⁷⁹ Ludwig Fleck, *Genèse et développement d'un fait scientifique*, Trad. Française, Paris, Les Belles lettres, 2005.

¹⁸⁰ Ibid., page 74.

particulièrement son regard sur ces *Big Data*. Les données collectées dans les imageries médicales présentent, parmi tant d'autres, un intérêt majeur.

Mais, si les travaux sur les biomarqueurs associés à la maladie d'Alzheimer sont nombreux, le vrai défi reste de passer de corrélations fortes à des relations de cause à effet entre un marqueur diagnostic et un phénomène biologique. La thèse de Russo / Williamson¹⁸¹, dont nous reparlerons au sujet du lien causalité / corrélation, propose ainsi de combiner approches corrélationnistes et approches mécanistiques. Que les biomarqueurs soient particulièrement utiles en recherche ou pour le développement de médicaments ou de nouvelles voies thérapeutiques, c'est largement admis, mais que ces biomarqueurs soient utilisés en médecine, que ce soit à des fins de diagnostic ou de personnalisation de traitements, cela ne va pas de soi car le biomarqueur n'est pas plus un symptôme que le signe d'un état pathologique. Cela revient à dire que, selon le contexte d'utilisation, un biomarqueur a plusieurs modes d'existence, ce qui doit être analysé.

Certes, dans certaines situations, la présence d'une mutation a un effet certain. C'est évidemment le cas pour les maladies mono-géniques pour lesquelles l'altération qui existe dès la conception dans l'ADN de l'individu se concrétisera par la maladie. C'est aussi le cas, par exemple, pour l'herceptine qui a été à l'origine de grands espoirs pour la médecine personnalisée. En introduisant cette molécule pour traiter le cancer du sein dans les années 1990, les laboratoires Roche ont démontré qu'il était possible d'anticiper le résultat du traitement grâce à un simple test immunohistochimique sur les patientes, le traitement n'agissant que chez les femmes sur-exprimant la protéine transmembranaire HER-2, causée par l'amplification du gène codant pour cette protéine. Ce n'est pas la situation la plus courante et l'incertitude augmente considérablement quand on est face à de nombreux gènes de susceptibilité.

¹⁸¹ F. Russo et J. Williamson, « Interpreting Causality in Health Sciences », *International Studies in the Philosophy of Science*, 2007, 21 (2), pp : 157-170.

Il faut ainsi s'interroger sur l'importance d'un gène comme le gène de l'APOE comme biomarqueur génétique (associé à d'autres biomarqueurs comme la bêta-amyloïde et la protéine *Tau*) comme prédicteur ou diagnostiqueur de la maladie d'Alzheimer en phase pré-symptomatique et se demander si le biomarqueur est une molécule ou plutôt un ensemble de molécules analysées avec des algorithmes des bio-informaticiens.

Le diagnostic clinique de la maladie est, ou semble, décorrélé de ces biomarqueurs. C'est un point important car cela amène à se demander si les biomarqueurs peuvent avoir une signification pour les cliniciens. Indépendamment de la maladie d'Alzheimer, on peut illustrer ces difficultés par le cas de l'utilisation de l'herceptine pour traiter le cancer du sein, utilisation qui nécessite l'accompagnement d'un test de surexpression d'une protéine, la protéine HER2.

La sensibilisation des cliniciens britanniques à l'utilisation de l'herceptine

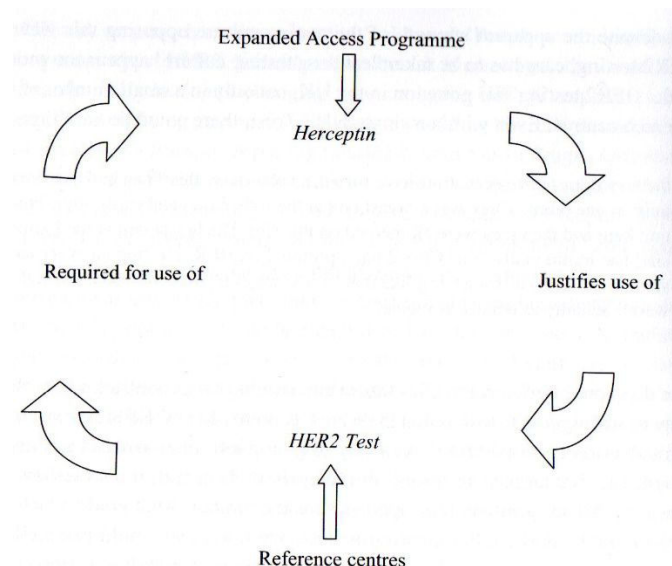
Pour décider si une patiente atteinte d'un cancer du sein peut recevoir l'herceptine comme traitement, le clinicien doit évaluer le niveau d'une protéine, la HER2. Pour cela un test immunohistochimique, proposé par Roche qui commercialise l'herceptine en Europe, est disponible et est proposé comme test d'accompagnement de la décision thérapeutique.

Cela suppose que les cliniciens acceptent ce test. Or Roche s'est trouvé confronté à des différences de « culture du test ». En Europe, environ 40 % des cliniciens l'acceptent sauf au Royaume Uni où ce taux n'est que de 6 %¹⁸². Pour les cliniciens britanniques, un test n'est réalisé que s'il y a une vraie raison de le réaliser. Or, selon eux, HER2 est un facteur pronostic trop limité et il n'apporte rien de

¹⁸² Adam Hedgecoe, *The Politics of Personalised Medicine, Pharmacogenetics in the Clinic*, Cambridge UK, Cambridge University Press, 2004, page 107.

plus probant quant au choix du traitement thérapeutique contre le cancer du sein.

Pour surpasser ces réticences, Roche a soutenu et financé, au début des années 2000 et pendant 30 mois, trois centres de référence à Nottingham, Glasgow et Londres¹⁸³. Ces centres avaient pour vocation de réaliser les tests HER2, en utilisant le test développé par le laboratoire pharmaceutique, sur des échantillons envoyés de tous les coins du pays, Roche assurant étroitement le contrôle qualité des tests effectués. Parallèlement, et alors que l'herceptine suivait les procédures britanniques pour l'autorisation d'utilisation, Roche donna l'accès à ce médicament aux cliniciens de façon à réaliser le cycle suivant¹⁸⁴ :



Aujourd'hui, la culture du test a évolué au Royaume Uni. Mais ce rappel illustre l'implication des industriels dans la sensibilisation et l'acceptation de leurs produits et tests par les cliniciens qui ont besoin « d'incitations » pour dépasser leurs réticences.

Si l'on revient au diagnostic clinique de la maladie d'Alzheimer, il s'appuie sur la détection d'une atteinte cognitive

¹⁸³ Ibid., page 110.

¹⁸⁴ Ibid., page 110.

dominée par des troubles mnésiques auxquels peut s'associer un syndrome aphaso-apraxo-agnosique¹⁸⁵. Le « a » privatif dans aphasie, apraxie, agnosie, voire apathie, dérange. On le retrouve sous le terme général d'anosognosie issu du grec *nosos* (signifiant maladie) et du grec *gnosis* (qui porte la signification de connaissance), le « a » privatif donnant une signification d'absence à la maladie. L'atteinte de la mémoire est ainsi, le plus souvent, le premier signe de la maladie. Elle se concrétise par des troubles de la mémoire des faits récents, des oublis répétés ou des difficultés de recevoir des informations nouvelles par exemple. Dans ce contexte, l'anosognosie est simplement une altération de la capacité à reconnaître des déficits dans le fonctionnement sensoriel, perceptif, moteur, affectif ou cognitif¹⁸⁶. Cependant, l'anosognosie n'accompagne pas nécessairement la maladie d'Alzheimer. Des malades Alzheimer peuvent être conscients de leurs troubles, même à des stades avancés de la maladie, et donc être nosognosiques.

C'est pourquoi des critères complémentaires ont été proposés, notamment les critères du NINCDS-ADRDA¹⁸⁷. Ils proposent :

- des critères pour le diagnostic clinique de la maladie d'Alzheimer probable basés notamment sur une démence diagnostiquée sur des données cliniques documentées par MMS (*Mini-Mental State* ou test de Folstein)¹⁸⁸ ou tout test équivalent, sur des altérations d'au moins deux fonctions

¹⁸⁵ L'aphasie est un trouble du langage, l'apraxie un trouble de l'habileté dans les gestes, l'agnosie un trouble de la perception.

¹⁸⁶ S. Kotler-Cope, C. J. Camp, « Anosognosia in Alzheimer disease », *Alzheimer Disease and Associated Disorders*, 9 (1), 1995, pp. 52-56.

¹⁸⁷ NINCDS : *National Institute of Neurological and Communication Disorders and Stroke*. ADRDA : *Alzheimer's Disease and Related Disorders Association*.

¹⁸⁸ M. F. Folstein and al., « 'Mini-mental State' : A Practical Method for Grading the Cognitive State of Patients for the Clinician », *Journal of Psychiatry Research*, Vol. 12, N° 3, 1975, pp : 189-198. Le seuil pathologique de la maladie d'Alzheimer varie entre un MMS de 22 à 26/30.

cognitives, sur une diminution progressive de la mémoire, sur l'absence de trouble de la conscience et sur un début tardif des signes cliniques, pas avant 40 ans ;

- un diagnostic de maladie d'Alzheimer probable basé sur la présence d'une aphasie, d'une apraxie, d'une agnosie, sur une diminution des activités quotidiennes et des troubles du comportement et sur des analyses biologiques normales du liquide céphalo-rachidien ;
- des critères pour le diagnostic de maladie d'Alzheimer certaine, à savoir les critères pour le diagnostic de maladie d'Alzheimer probable, complétés par la mise en évidence d'altérations histopathologiques caractéristiques obtenues par biopsie ou autopsie.

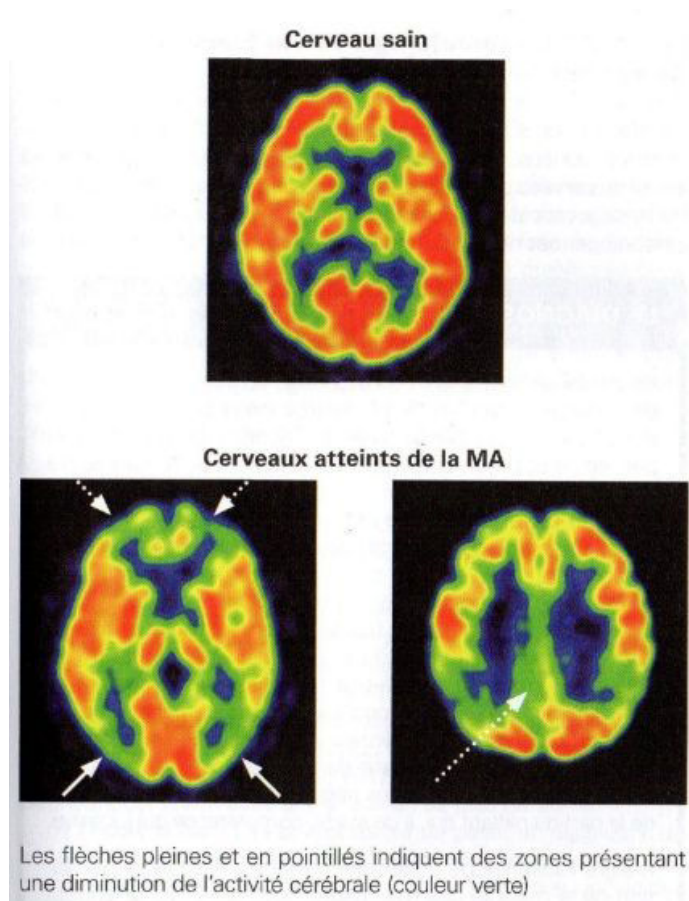
On le voit, ces critères n'apportent pas grand-chose au niveau de la certitude car attendre une autopsie n'est pas satisfaisant. Faire un diagnostic post-mortem après examen des lésions cérébrales apporte, certes, des connaissances, mais ne pouvoir évoquer qu'une probabilité de pathologie, et le plus souvent à un stade tardif de la maladie, n'est pas acceptable. On peut simplement relever un élément important incompatible avec la maladie d'Alzheimer, à savoir le début soudain de troubles.

Si tout ne survient pas d'un coup, c'est qu'il existe une phase préliminaire qui précède la phase clinique. Ainsi que cela a déjà été indiqué, elle a été identifiée sous le vocable de « *mild cognitive impairment* » (MCI) par le neurologue américain Ronald Petersen¹⁸⁹. Le MCI, en français « déficience cognitive légère », définit un déclin cognitif plus déficitaire que celui attendu pour l'âge et le niveau socioculturel du sujet, sans être pour autant sévère au point d'être au stade de démence. Le MCI n'est pas une maladie,

¹⁸⁹ R. C. Petersen et al., « Mild cognitive impairment : Clinical characterization and outcome », *Arch. Neurol.*, 1999, 53 (3), pp : 303-308.

mais un syndrome¹⁹⁰, c'est-à-dire un ensemble de symptômes cliniques relevant de plusieurs causes

Puisque la maladie d'Alzheimer évolue sur un large spectre allant de l'individu cognitivement normal à la forme démentielle, en passant par une phase de déficits cognitifs légers, et puisque des biomarqueurs et des techniques d'imagerie sont aujourd'hui disponibles, de nouveaux critères diagnostiques ont été définis en 2011¹⁹¹. Ils intègrent, aux signes cliniques, les nouvelles données de biologie, d'imagerie et de neuropsychologie comme l'illustre le schéma qui suit :



¹⁹⁰ Bernard Croisile, *Alzheimer : que savoir ? que craindre ? Qu'espérer ?*, Paris, Editions Odile Jacob, 2014, page 147.

¹⁹¹ G. M. Mc Kahnnet al., « The diagnosis of dementia due to Alzheimer's disease : Recommendations from the National Institute on Aging and the Alzheimer's Association workgroup on diagnostic guidelines for Alzheimer's disease », *Alzheimer's Dement.*, 2011, 7 (3),pp : 263-269.

Cela a élargi la définition de la maladie d'Alzheimer si bien qu'on peut vivre pendant de nombreuses années avec une maladie d'Alzheimer non démentielle sans qu'elle ait d'impact majeur sur le quotidien. On est de fait passé d'une définition clinico-pathologique (clinique + autopsie) à une entité clinico-biologique (clinique + biomarqueurs)¹⁹².

Le diagnostic biologique avec des biomarqueurs permet de détecter l'installation des lésions, mais les résultats liés aux biomarqueurs posent cependant un certain nombre de questions, notamment :

- le problème des seuils de détection ;
- l'appréciation du risque associé à cette détection, les études étant encore peu nombreuses et sur de petits échantillons ;
- la question de la temporalité.

Des biomarqueurs peuvent être positifs une quinzaine d'années avant l'apparition des symptômes. Il ne sert à rien d'annoncer de mauvaises nouvelles très en avance, d'autant que des erreurs de diagnostic ou des approximations liées aux incertitudes associées à ce diagnostic peuvent également entraîner des problèmes juridiques. C'est pourquoi, pour le moment, en l'absence de signes cliniques, les biomarqueurs sont essentiellement des outils de recherche. Cette situation est probablement transitoire, l'approche de la santé évoluant avec le suivi de toujours plus de biomarqueurs.

Idéalement, on pourrait souhaiter un diagnostic dès le déclenchement du processus dégénératif, mais il faut être vigilant pour bien distinguer entre maladie potentielle et maladie effectivement exprimée.

¹⁹² B. Dubois et al., « Advancing research diagnostic criteria for Alzheimer's disease : The TWG-2 criteria », *The Lancet Neurology*, 2014, 13 (6), pp : 614-629.

Le diagnostic précoce de la maladie d'Alzheimer, voire l'anticipation prédictive, pose des enjeux éthiques et épistémologiques, d'autant plus qu'il n'existe pas réellement d'accompagnement thérapeutique. On peut en effet considérer que l'anticipation du diagnostic et de l'annonce de la maladie pose des questions graves, surtout dans un contexte de faiblesse des réponses médicales et de fragilité des structures de suivi et d'hébergement, ce qui impose sans doute de définir des règles, surtout dans la façon de prendre en compte les facteurs de risque. À la prédiction est associée l'incertitude, dès lors notamment que la maladie est difficile à saisir. De plus, annoncer un diagnostic de maladie d'Alzheimer, c'est certes annoncer un Alzheimer, mais c'est d'abord et avant tout dire à la personne qu'elle est malade, que les troubles dont elle souffre sont pathologiques alors qu'un grand nombre de patients ne se sentent pas malades. Cela amène à rappeler que la précocité d'un diagnostic est souvent accompagnée d'incertitudes qui peuvent être considérables. Nous reviendrons sur l'incertitude, mais la multiplication des biomarqueurs à considérer¹⁹³ va encore accroître l'impression d'incertitude.

V-3 La biologie des systèmes

L'afflux massif de données issues du séquençage de génomes entiers, et des différentes technologies -omiques, ne fait que confirmer la complexité des phénomènes biologiques et la nécessité de parcourir différents niveaux d'organisation pour mieux comprendre le fonctionnement biologique d'un système. Cela revient à dire qu'il n'y a pas un niveau de causalité privilégié et que le rôle des gènes ne peut être compris qu'en dépassant le niveau moléculaire pour prendre en compte les niveaux d'organisation

¹⁹³ La génétique, la biologie moléculaire, l'imagerie, etc, vont amener davantage de complexité.

supérieurs¹⁹⁴. Cela revient à envisager les constituants moléculaires du vivant, non plus isolément, mais dans leur contexte, comme le contexte social et économique et dans leur environnement. Pour suivre la proposition de John Dewey¹⁹⁵, on pourrait remplacer « contexte/environnement » par « situation », terminologie qui met en évidence le caractère évolutif des conditions (facteurs conceptuels, matériels, sociaux, institutionnels) impliquées dans le développement des connaissances. Ce qui compte alors, ce sont les réseaux de relations entre les différents constituants, avec, comme conséquence, non pas une causalité définie mais une causalité distribuée. Dans les maladies complexes, le rôle des gènes doit être repensé car leur expression dépend de l'expression d'autres gènes, mais aussi de facteurs épigénétiques, métaboliques et environnementaux. Sans doute faut-il rejeter tout jugement de valeur sur le génome d'un individu et parler plutôt de « vulnérabilité génétique » à tel ou tel environnement, telle ou telle activité, tel ou tel stress, etc., que de prédisposition génétique. Thomas Heams l'énonce clairement : « Même si l'on empêche constitutivement le fonctionnement d'un gène crucial pour la survie des bactéries, celles-ci sont capables, en exprimant de très nombreux gènes préalablement « éteints », de trouver des solutions inattendues pour survivre dans un milieu supposé hostile quand ce gène est invalidé. [...] Cela incite à penser le génome d'un organisme non pas tant comme le support d'un programme, mais comme une boîte à outils dans laquelle la cellule pioche tant qu'elle peut dans le cadre d'un comportement exploratoire, jusqu'à ce que, si le milieu lui en laisse le temps, elle trouve une « solution ». Le génome est donc une mémoire de solutions évolutives archivées et mobilisables »¹⁹⁶.

¹⁹⁴ D.. Noble, « Claude Bernard, the first systems biologist, and the future of physiology », *Experimental Physiology*, 2008, vol. 93, pp. 16-26.

¹⁹⁵ John Dewey, *Logic : The Theory of Inquiry*, New York, Holt, Rinehart and Winston, 1938.

¹⁹⁶ Thomas Heams, « Expression stochastique des gènes et différenciation cellulaire », in J.J. Kupiec, O. Gandrillon, M. Morange, M. Silberstein (dir.), *Le hasard au coeur de la cellule. Probabilités, déterminisme, génétique*, Paris, Syllepse, 2009, page 36.

Cela pose la question fondamentale de la robustesse des systèmes biologiques, c'est-à-dire de leur étonnante capacité à continuer à fonctionner malgré des capacités internes et externes¹⁹⁷, dit d'une autre façon leur capacité à maintenir leurs caractéristiques phénotypiques fondamentales à travers les variations internes (génétiques) ou externes (environnementales). Cette robustesse peut être produite par de la redondance¹⁹⁸, mais dans la plupart des cas, la robustesse d'un système tient à son mode d'organisation. D'un point de vue conceptuel, la mise en évidence de l'insuffisance de la connaissance du génome pour prédire les fonctions précises des régions codantes, pour comprendre les processus complexes¹⁹⁹, a confirmé la nécessité de passer à une vision plus systémique, s'appuyant sur des approches fonctionnelles et donc moins centrées sur le génome.

La biologie des systèmes²⁰⁰ peut y aider en permettant de mieux comprendre la régulation des gènes et leur pleiotropie, c'est-à-dire les divers effets qu'un gène ou un ensemble de gènes peut avoir sur diverses fonctions. Cela introduit l'épigénétique et la prise en compte de tous les facteurs autres que les facteurs génétiques, comme les facteurs environnementaux ou les facteurs associés au mode de vie²⁰¹, qui peuvent intervenir sur l'expression et la régulation des gènes, ainsi que sur la traduction (et même les

¹⁹⁷ A. Wagner, *Robustness and Evolvability in Living Systems*, Princeton, NJ, Princeton University Press, 2005.

¹⁹⁸ Ainsi on a deux reins et un seul suffit pour vivre.

¹⁹⁹ Evelyn Fox Keller, *Le siècle du gène*, Paris, Gallimard, 2003.

²⁰⁰ Harald Hampelet al., « Development of Biomarkers to Chart All Alzheimer's Disease Stages : The Royal Road to Cutting the Therapeutic Gordian Knot », *Alzheimer & Dementia* 8, n°4 (2012) : 312-36.

²⁰¹ On peut se demander s'il faut continuer à parler en termes de facteurs et ne pas adopter l'approche dialectique de Richard Lewontin (*La Triple Hélice : les gènes, l'organisme, l'environnement*, Paris, Seuil, 2003) pour qui les phénomènes observables à tout instant font partie de processus, qui ont une histoire et un futur, dont les voies ne sont pas uniquement déterminées par leurs unités constitutives. Ils sont composés d'unités dont on peut décrire les propriétés, mais l'interaction de ces unités, lors de leur constitution engendre des complexités qui font que les produits obtenus sont qualitativement différents des parties constitutives.

modifications post-traductionnelles, telles que la méthylation) des gènes exprimés.

Progressivement, la biologie s'est mise à s'intéresser autant aux transferts d'information qu'aux transferts d'énergie et de matière, et la génétique s'est ainsi imposée. S'il a été vite clair que l'organisme était co-produit par son génome et par son environnement, l'information (l'épigénétique) contenue dans l'ensemble de la machinerie qui « lit » le message génétique, que Lewontin a appelé « la triple hélice », est devenue une préoccupation croissante, traduisant le fait qu'aucun texte n'a de sens s'il n'est traduit par un lecteur.

Il faut donc intégrer des données génomiques et phénotypiques ainsi que des informations sur l'environnement du sujet ou de l'organisme. On retrouve l'équation déjà présentée :

$$\text{« Génome * Exposome = Phéno me »}$$

qui illustre la nécessité de disposer de données sur l'environnement auquel peut être soumise une personne²⁰².

La biologie des systèmes est l'étude intégrative, par modélisation, des gènes, des protéines, des réactions biochimiques et d'autres processus physico-chimiques considérés en tant que systèmes dynamiques d'éléments en interaction. C'est une approche de choix pour s'intéresser aux réseaux, qu'ils soient métaboliques²⁰³, génétiques ou de signalisation, et pour prendre en compte les évolutions en fonction du temps. Elle conduit à des modèles et ces modèles peuvent aider à comprendre les relations multi-causales et les maladies complexes, tout en pouvant donner un sens aux résultats des expérimentations - omiques.

²⁰² C. P. Wild, « Complementing the genome with an « exposome » : the outstanding challenge of environmental exposure measurement in molecular epidemiology », *Cancer Epidemiol. Biomarkers Prev* 2005 ; 14(8), pp.1847-1850, Epub 2005/08/17.

²⁰³ Il est intéressant de noter une évolution du vocabulaire. On est passé de la voie métabolique au réseau métabolique.

La biologie des systèmes est associée aux modèles mathématiques²⁰⁴, ce qui permet notamment de prendre en compte les aspects dynamiques de phénomènes car ces modèles utilisent des équations différentielles (temps) ou des équations aux dérivées partielles (temps et espace). Cet aspect dynamique est important pour comprendre les mécanismes de rétroaction qui sont « la base de toutes les formes de régulation, de contrôle et de coordination à tous les niveaux dans les organismes vivants »²⁰⁵

Un modèle décrit un phénomène, mais ne l'explique pas²⁰⁶. Le mieux est sans doute de construire des modèles extrêmement simplifiés des réseaux dans lesquels sont insérés les différents constituants, en ignorant la nature et les détails des interactions moléculaires. Un modèle prédictif n'est pas nécessairement un modèle explicatif. Les corrélations peuvent être utilisées pour une prédiction probabiliste, mais n'expriment pas une causalité. C'est une limite fondamentale pour les applications médicales car, pour celles-ci, il est nécessaire de disposer de modèles explicatifs. La question de la contribution respective des méthodes inductives et des travaux théoriques se pose avec le développement de la biologie des systèmes.

La place des biomarqueurs est à interroger, notamment en les replaçant dans le processus de diagnostic multimodal rassemblant un faisceau d'arguments conduisant vers un diagnostic de probabilité.

²⁰⁴ O. Wolkenhauer & A. Muir, « The complexity of cell-biological systems », in C. Hooker (Ed.), *Philosophy of complex systems*, 2011, Amsterdam, Elsevier, pp. 355-385.

²⁰⁵ O. Wolkenhauer et al., « A plea for more theory in molecular biology », in P. Bringmann, E. C. Butcher, G. Parry & B. Weiss (Eds), *Systems Biology – Applications and Perspectives*, Berlin, 2007.

²⁰⁶ D.M. Kaplan et C. F. Craver, « The explanatory force of dynamical and mathematical models in neuroscience : A mechanistic perspective », *Philosophy of Science*, 2011, 78, 601-627.

Deuxième partie

Terrains et enjeux de recherche

Chapitre VI

Un terrain de recherche et un terrain clinique

Après avoir présenté le contexte de la maladie d'Alzheimer et fait un panorama des technologies associées à l'évolution des connaissances dans les sciences du vivant, et abordé le problème des *Big Data* et des outils d'analyse, la confrontation au terrain permet de mieux analyser comment, dans la pratique, sont abordées les approches et sont utilisés les outils.

Deux terrains complémentaires ont été retenus :

- Le premier est plus dédié à la recherche. Le site qui a été retenu est celui de l'Institut Pasteur de Lille et de l'équipe du professeur Philippe Amouyel, directeur de l'UMR 744 Inserm et du laboratoire d'excellence DISTALZ
- Le deuxième a une orientation médicale. Le site est l'Institut de la Mémoire et de la Maladie d'Alzheimer dirigé par le professeur Bruno Dubois à l'Hôpital de la Pitié Salpêtrière.- Charles Foix

Nous présentons rapidement ces deux terrains d'analyse, puis nous présenterons, dans le chapitre suivant, comment ces deux ensembles mettent en œuvre les technologies et les outils actuels.

VI-1 L'Institut Pasteur de Lille

L'objectif de recherche de l'unité dirigée par le professeur Philippe Amouyel (unité de recherche *Facteurs de risque et déterminants moléculaires des maladies liées au vieillissement*) consiste en l'acquisition de connaissances sur l'impact des facteurs environnementaux et de susceptibilité génétique, ainsi que de leurs interactions, dans la survenue, l'évolution et la prise en charge des maladies liées au vieillissement chez l'Homme, notamment les maladies neurodégénératives.

Ces recherches sont abordées grâce à l'intégration, au sein même de l'unité, d'un savoir-faire en épidémiologie et d'une maîtrise d'outils de biologie moléculaire adaptés à l'approche épidémiologique, intégrant les avancées récentes de la génomique à haut-débit, de la transcriptomique, de la protéomique et de la bioinformatique.

Cette unité fait partie du Labex (laboratoires d'excellence) DISTALZ, Développement de stratégies innovantes pour une approche transdisciplinaire de la maladie d'Alzheimer. Son objectif est d'explorer les processus biologiques impliqués dans la maladie d'Alzheimer, en particulier à la lumière des découvertes récentes de la génomique, et de développer de nouveaux biomarqueurs de la maladie et de nouvelles cibles thérapeutiques.

Ainsi qu'il a déjà été souligné, Les lésions provoquées dans le cerveau par la maladie d'Alzheimer restent longtemps silencieuses, puis entraînent des manifestations perceptibles au fur et à mesure qu'elles s'étendent et touchent des zones importantes pour le fonctionnement cérébral. Si bien que lorsque les signes cliniques apparaissent, la maladie a déjà fait des dégâts dans le

cerveau depuis souvent plus d'une dizaine d'années. Pour traiter la maladie avant que les lésions deviennent irréversibles, il faut trouver des marqueurs qui permettront d'identifier la maladie le plus tôt possible, avant les premiers signes, perceptibles cliniquement que 5 à 10 ans plus tard. Les recherches actuelles s'orientent sur trois types de marqueurs : des marqueurs dans le cerveau que l'on peut détecter à l'imagerie ou biologiquement, des marqueurs sanguins et des marqueurs génétiques. C'est sur ce dernier aspect que se concentre l'unité lilloise du Pr Philippe Amouyel. Elle a contribué à la découverte de la presque totalité des gènes connus de susceptibilité à la maladie, notamment par la méthode GWAS (*genome wide association study*) autrement dit l'analyse complète du génome d'un très grand nombre d'individus. C'est le travail, dans cette unité, de l'équipe dirigée par le Docteur Jean-Charles Lambert qui étudie les déterminants moléculaires de la maladie d'Alzheimer et des troubles cognitifs. L'objectif est d'identifier les déterminants environnementaux et génétiques conduisant à l'altération des fonctions cognitives. L'équipe développe une analyse comparative du transcriptome des gènes localisés dans les régions chromosomiques identifiées dans la maladie d'Alzheimer à début tardif. Des tissus cérébraux de patients atteints et non atteints ainsi que des modèles cellulaires spécifiques sont analysés sur puces ADN pangénomiques *Affymetrix*. Cela permet de mettre en évidence un certain nombre de gènes candidats dans les régions chromosomiques d'intérêt et de converger plus rapidement sur de nouveaux gènes candidats. Par ailleurs, des analyses bioinformatiques permettent de caractériser des événements d'épissages alternatifs de ces gènes pouvant expliquer une part de la susceptibilité individuelle.

Jean-Charles Lambert est aujourd'hui à la tête de la plus grosse équipe de génétique Alzheimer en France, avec 15 membres, chercheurs, techniciens, ingénieurs. De gros projets de séquençage des parties codantes du génome portant sur 5 000

malades et 5 000 témoins et du génome complet sur 800 malades et 200 témoins sont en cours, ce qui nécessite des capacités de stockage des données énormes, sans qu'on puisse dire à ce jour ce qui va en sortir. Les données obtenues, des *Big Data*, ne sont plus analysables simplement, ce qui donne un rôle important aux bio-informaticiens et nécessite que les biologistes leur fassent confiance. Sabina Leonelli²⁰⁷ va plus loin en introduisant ce qu'elle appelle le « *Data Journey* », la circulation des données des producteurs aux utilisateurs, c'est-à-dire l'ensemble des circonstances matérielles, sociales et institutionnelles qui permettent le « *packaging* » et le transport des données. Les procédures de packaging comprennent la sélection, le formatage, la standardisation et la classification des données, ainsi que le développement de méthodes de récupération, d'analyse, de visualisation et de contrôle qualité. Quant au transport, il passe par une décontextualisation / recontextualisation, la décontextualisation étant un moyen pour que les données perdent la signification qui leur est attribuée dans leur contexte de recherche original, la recontextualisation permettant aux utilisateurs d'évaluer la signification potentielle des données en évaluant leur provenance par la consultation des données précisant la façon dont elles ont été acquises.

Si Jean-Charles Lambert a contribué à la découverte de 21 des 25 gènes identifiés comme ayant un rôle dans la maladie d'Alzheimer, ce qui constitue déjà une carte génétique bien fournie de la maladie, il considère que quelques gènes restent sûrement à découvrir mais qu'il faut plutôt se concentrer aujourd'hui sur les actions de ces gènes à l'intérieur des cellules. Il faut en effet donner du sens à ce qu'on trouve statistiquement en revenant à la biologie.

²⁰⁷ Sabina Leonelli, *Data-Centric Biology, a philosophical study*, Chicago, The University Chicago Press, 2016.

Il faudra revenir sur ce point. Pour revenir à la biologie, on va vers la biologie des systèmes, qui dépend de ce qui a déjà été défini et connu, si bien que, d'une certaine façon, on force la connaissance vers ce qui est déjà défini, ce qui peut aboutir à ce que la biologie des systèmes devienne vite tautologique. C'est un véritable problème épistémologique. Si l'on peut considérer que les données de séquence sont un exemple classique de données biologiques produites en l'absence d'une question de recherche spécifique, il y a cependant toujours un cadre de référence théorique préalable auquel on rapporte des données. En généralisant ce que dit Paul Edwards²⁰⁸ au sujet des études sur le climat et le réchauffement climatique, on peut souligner que les données sont chargées de théorie, et que données et modélisation sont des activités interdépendantes, les modèles étant eux-mêmes chargés de données. On verra que la modélisation est un des principaux moyens d'interprétation des données.

VI-2 L'IM2A et le projet INSIGHT

L'Institut de la mémoire et de la maladie d'Alzheimer (IM2A), dirigé par le professeur Bruno Dubois est le premier centre français de recherche clinique totalement dédié à la prévention, au diagnostic précoce, à la recherche et au traitement de la maladie d'Alzheimer et des démences, au sein d'un bâtiment de 3 000 m², au cœur de l'hôpital de la Pitié-Salpêtrière à Paris. Aujourd'hui, l'IM2A travaille à percer les secrets de la maladie d'Alzheimer en créant l'interface clinique/recherche indispensable pour développer de nouveaux algorithmes diagnostiques et définir des biomarqueurs pertinents. Demain, l'IM2A, grâce à cette connaissance, se veut être le Centre Expert international pour le développement de nouvelles thérapies et pour la formation de chercheurs dans le domaine

²⁰⁸ Paul N. Edwards, *A Vast Machine : Computer Models, Climate Data, and the Politics of Global Warming*, Cambridge, MA, MIT Press, 2010.

de la recherche translationnelle. Le patient est au cœur de ses travaux.

"Pour répondre à l'enjeu médical et scientifique que représentent les maladies de la mémoire, le patient doit être au cœur de toute action : car il détient un secret qu'il nous faut comprendre. A l'IM2A, le patient va contribuer à notre connaissance de ces maladies en participant activement aux études de recherche clinique. La recherche sur les maladies de la mémoire ne peut en effet avancer qu'avec les patients".

Professeur Bruno Dubois, Directeur de l'IM2A

Le leitmotiv de la recherche précompétitive est le diagnostic précoce : il s'agit d'identifier des personnes atteintes de quelques troubles de mémoire qui pourraient développer dans 10 ou 15 ans la maladie d'Alzheimer. Ce changement d'approche repose sur une nouvelle définition de la maladie, élaborée par une équipe de chercheurs internationaux dirigée par le Pr Bruno Dubois²⁰⁹. On n'attend plus le stade de la démence pour diagnostiquer la maladie, mais sont proposés des critères de diagnostic qui incluent tous les stades de la maladie, notamment celui qui existe avant l'apparition des symptômes, le stade prodromal.

La problématique de l'équipe de la Pitié-Salpêtrière est ainsi, si l'on peut dire, simple : sachant que la prévalence de la maladie d'Alzheimer diminuerait de moitié si l'on retarde le début de la maladie de 5 ans, la question est de savoir si tous les sujets à risque vont évoluer vers la maladie, et si oui, quand, et peut-on identifier des indicateurs de conversion ? C'est l'objet du projet INSIGHT, présenté plus loin.

²⁰⁹ B. Dubois et al., « Revising the definition of Alzheimer's disease : a new lexicon », *Lancet Neurology* 2010, 9 (11), pp : 1118-1127.

Chapitre VII

Approche génétique versus approche multimodale

Les deux terrains se complètent, l'un étant plus orienté vers la recherche, l'autre vers la pratique clinique, mais différent sensiblement par leurs approches, l'une se focalisant essentiellement sur les aspects génétiques, l'autre prenant en compte beaucoup plus de variables.

VII-1 Approche génétique

À l'Institut Pasteur de Lille, Jean-Charles Lambert, déjà présenté, est un biologiste, spécialiste de la génétique de la maladie d'Alzheimer. Au sein de l'unité du professeur Philippe Amouyel, on peut rappeler qu'il a collaboré à la découverte de 21 des 25 gènes²¹⁰ actuellement identifiés dans cette pathologie.

Avec l'arrivée de la bio-informatique, Jean-Charles Lambert a connu une véritable révolution technologique dans une évolution des approches qu'il présente de la façon suivante :

- Dans les années 80, on ne pouvait guère étudier que 15 variants génétiques quand on cherchait à mettre en évidence des facteurs de risque pour une maladie. Par ailleurs, chaque chercheur travaillait de façon isolée, et les résultats présentaient de nombreux faux positifs.
- En 1993, par l'approche gène-candidat a été mis en évidence le gène ApoE et, notamment, son allèle e4 qui, par son importance, cachait le reste. Durant cette période, il faut faire face à une difficulté et une nécessité. La difficulté avec

²¹⁰ J-C. Lambert et al., « Meta-analysis of 74,046 individuals identifies 11 new susceptibility loci for Alzheimer's disease », *Nature Genetics*, 2013, Dec 45(12), pp : 1452-1458.

la maladie d'Alzheimer est de disposer de patients (malgré les progrès des tests neuropsychologiques et du diagnostic des cliniciens, et malgré le développement des techniques d'imagerie et des techniques moléculaires). La nécessité est d'augmenter la population témoin pour être sûre qu'elle soit représentative, malgré l'existence éventuelle de faux témoins en raison de la temporalité de la maladie.

- On passe ensuite à une population de taille importante et mieux caractérisée. Du point de vue méthodologique, le projet international HapMap²¹¹ a permis de se contenter 500 à 600 000 variants (au lieu des 10 millions présents dans le génome) après avoir montré que si on analyse un variant on obtient en fait des informations sur plusieurs variants.
- Puis vint le développement des puces, de plus en plus denses (de 20 000 à 5 millions de « spots »), avec la standardisation et l'automatisation et la mise en place d'une structuration clinique facilitant l'accès aux échantillons d'ADN, en 2009 pour la maladie d'Alzheimer. Il n'en est résulté que 3 nouveaux déterminants génétiques suite à des cohortes comportant 2 000 malades et 6 000 témoins, et en analysant des variations supérieures à 1%.
- Enfin se sont constitués des consortiums, comme IGAP, permettant de mettre en commun les efforts et d'augmenter considérablement le nombre de cas. Des variants rares (variabilité < 1%) ont été recherchés et cela a conduit aux 25 déterminants génétiques connus.

L'International Genomics of Alzheimer's Project (IGAP), lancé en 2011, est une collaboration mise en place pour découvrir et cartographier l'ensemble des gènes de susceptibilité à la maladie d'Alzheimer. La création de l'IGAP permet donc de rassembler une base de données partagée par l'ensemble de ces chercheurs, qui inclut les données génétiques de plus de 40.000 personnes.

²¹¹ HapMap est une base de données génétiques à très grande échelle dont l'objectif était de déterminer la fréquence d'occurrence de millions de SNPs, et d'étudier les corrélations entre les types d'associations de ces mutations ponctuelles et des maladies.
Voir : <http://hapmap.ncbi.nlm.nih.gov>

Il n'y a pas à proprement parler de ruptures, mais plutôt une accélération et une amplification des données car, sur un plan technique, les premières étapes d'un travail de recherche consistent à transférer, stocker, sécuriser et analyser des données. Pour avoir des ordres de grandeur, nous pouvons noter que :

- Un échantillon de 10 000 individus pour 600 000 variables (SNPs) représente 50Go (giga-octets) de données brutes, correspondant, après filtrage de l'information, à environ 2Go de données traitées.
- Le séquençage des parties codantes (exome) de 1000 individus représente 30 To (tera-octets) de données brutes pour des données effectives traitées de 10 Go.
- Quant au séquençage du génome complet de 1 000 individus, il nécessite un espace de travail de 500 To pour un volume de données effectivement traitées de 200 Go.

Les données brutes sont contrôlées en qualité, nettoyées en quelque sorte, et les mutations à analyser sont sélectionnées par des outils statistiques. Par ailleurs, au terme du traitement des données, les données générées sont sécurisées, c'est-à-dire qu'elles sont systématiquement dupliquées et affectées à deux unités physiquement distinctes. Avec les *Big Data*, on est obligé de passer d'une recherche organisée en laboratoires autonomes à une recherche plus distribuée collaborativement, et on peut analyser si l'analogie avec les recherches en physique ne permettrait pas de progresser. Ajoutons, d'une part que dans le domaine des *Big Data*, il faut savoir déléguer car il est impossible de tout maîtriser (ainsi, un statisticien n'est pas un biologiste et vice-versa), d'autre part que

la quantité d'informations à manipuler dans un laboratoire tel que celui de l'Institut Pasteur de Lille nécessite de l'externalisation²¹².

Les études d'association, basées sur une approche dite gène candidat, sont privilégiées. L'utilisation des études d'association s'avère utile pour mettre en évidence des interactions entre les facteurs étudiés (génétiques ou environnementaux) ou des effets relativement faibles sur le risque de développer l'affection. Cependant, les limites de cette approche résident dans la sélection des gènes étudiés, dans la qualité et la taille des populations utilisées. En effet, les protéines impliquées dans la physiopathologie connue de la maladie d'Alzheimer sont nombreuses et les gènes candidats ainsi définis multiples. Cette multiplication des analyses tend à favoriser la mise en évidence d'associations liées seulement au hasard.

Ces gènes, ce sont des pistes dans le processus pathologique mais, on ne sait pas pour le moment quel est leur ordre d'intervention ni quel est leur rôle. Il reste encore sûrement quelques gènes à découvrir mais on dispose d'une carte génétique déjà bien fournie de la maladie et on se concentre aujourd'hui sur les fonctions de ces gènes à l'intérieur des cellules. Il faut en effet donner du sens à ce qu'on trouve statistiquement en revenant à la biologie. Cela peut conduire à une nouvelle dynamique dans les stratégies de recherche avec le développement de nouveaux cadres théoriques visant à donner du sens à la multiplication en très grand nombre des données empiriques issues des technologies utilisées.

L'approche qui consiste à déterminer, à partir des mesures d'expressions différentielles de gènes, quels gènes pourraient être

²¹² L'Institut Pasteur de Lille s'appuie sur le Centre de ressources informatiques de Lille (CRI).

impliqués dans les processus pathologiques comme celui de la maladie d'Alzheimer avant l'apparition des symptômes cliniques, nécessite de construire des modèles de réseaux d'interactions entre tous les constituants moléculaires d'une cellule, notamment des réseaux de régulation entre gènes ou des réseaux d'interactions entre protéines. La question qui se pose est de savoir si l'on peut inférer la structure et les mécanismes de régulation des systèmes biologiques à partir de données portant sur leur comportement, auquel cas l'accumulation de données sur l'expression des gènes, les interactions entre protéines, etc., suffirait pour construire des modèles représentant les mécanismes moléculaires sous-tendant le fonctionnement des cellules et des organismes. Cela semble difficile vu la complexité du vivant. Pour choisir entre différents modèles possibles et déterminer les paramètres du modèle, il faudrait une quantité de données beaucoup plus grande que ce que l'on produit actuellement dans un contexte *Big Data*. La solution le plus souvent adoptée est en fait de construire des modèles extrêmement simplifiés de ces réseaux sans prendre en compte la nature et les détails des interactions moléculaires.

Il ne faut pas oublier qu'une cellule est un système complexe caractérisé par des processus, souvent non linéaires, qui, de plus, se produisent à différentes échelles spatio-temporelles. L'aspect dynamique est important pour comprendre l'émergence des fonctions biologiques et la théorie des systèmes dynamiques peut être utile pour cela, d'autant que de nombreux mécanismes de rétroaction sont souvent à la base des régulations dans les organismes vivants.

VII-2 Approche multimodale

L'objet du projet INSIGHT coordonné par l'IM2A est de travailler sur des algorithmes de prédiction en mettant en évidence

une dynamique du processus de conversion des facteurs de risque de la maladie, avant l'apparition des premiers signes cliniques. Sachant qu'on a tous les mêmes risques d'avoir la maladie, tout en ne l'exprimant pas tous de la même façon, il s'agit d'identifier des facteurs de conversion comme l'âge, ou des facteurs génétiques, des allèles particuliers comme l'ApoEε4 par exemple. Pour cela, l'étude retient des personnes ne présentant aucun symptôme clinique de la maladie d'Alzheimer, mais ayant consulté pour des problèmes de mémoire et présentant des taux de Abeta42 ou de protéines Tau anormaux.

Le projet INSIGHT repose sur une cohorte de personnes (plus de 400 sujets âgés entre 70 et 85 ans) se plaignant de troubles de la mémoire (*SMC Subjective Memory Complainer*), sans aucun autre symptôme, qui présentent l'avantage d'être motivés pour participer à une recherche clinique.

Avant de prendre sa décision, chaque participant reçoit un formulaire d'information et de consentement qui précise :

- Le cadre général et les objectifs de la recherche biomédicale.
- Les critères d'inclusion, notamment les examens préalables à effectuer.
- Le déroulement de l'étude (une durée de cinq ans, avec des visites (et des tests ou analyses) semestrielles).
- Les bénéfices attendus.
- Les risques prévisibles.
- Les droits des participants, notamment en ce qui concerne les données personnelles et l'information sur les résultats de la recherche.
- Le cadre législatif de la recherche médicale.

Si, après un délai de réflexion, la personne accepte de participer à la recherche, elle signe alors le document donnant son consentement pour participer au projet de recherche.

L'étude est une étude monocentrique pour s'affranchir des biais que peuvent apporter des procédures d'acquisition et de traitement dans une approche multicentrique.

L'approche est multimodale. Elle s'appuie, d'une part sur l'extraordinaire potentiel de recherche de l'UPMC, notamment dans le domaine des mathématiques appliquées et des biostatistiques, d'autre part sur des ressources du site de la Salpêtrière et notamment de l'Institut du cerveau et de la moelle épinière (ICM)²¹³ dans les domaines particuliers de la neuroimagerie, de la génétique, des neurosciences cognitives, de la biologie et de l'électrophysiologie. Cette approche multimodale, associée à des modèles biophysiques pour faire des analyses de corrélation, est mise en œuvre sur la cohorte pour étudier les interactions entre les différents biomarqueurs susceptibles d'être utilisés pour la prédiction de la survenue de la maladie d'Alzheimer. Parallèlement sont suivis les profils d'expression des gènes et des analyses de biomarqueurs associés à la nutrition.

Sont ainsi suivis, sur une période de 5 ans, et sur la base d'une périodicité semestrielle :

- La charge amyloïde dans le cerveau par imagerie PET et analyse de la concentration en Abeta 42 dans le liquide céphalo-rachidien
- Les fonctions du cerveau par IRM, EEG, actigraphie²¹⁴ ainsi que des tests de mémoire (MCT).
- Le métabolisme du cerveau par PET scan
- Les marqueurs transcriptomiques, lipidomiques et nutritionnels.

²¹³ L'ICM (Alexis Genin et Alexandra Auffret) construit, de façon structurée, la base de données qui agrège des données très hétérogènes, et dont les informations sont incomplètes, ce qui nécessite d'avoir une approche bayésienne.

²¹⁴ L'actigraphie permet de suivre l'activité motrice d'un individu dans son environnement habituel, à l'aide d'accéléromètres piézoélectriques portés au poignet ou à la cheville.

L'objectif de l'étude est de déterminer la meilleure combinaison de biomarqueurs ou le meilleur algorithme pour prédire la survenue de la maladie d'Alzheimer, soit par l'évolution continue de certains paramètres, soit par l'apparition d'un événement se traduisant par de petits changements de certains paramètres provoquant ce que nous appelons une bifurcation vers la maladie d'Alzheimer au stade prodromal. Pour cela, on peut, à titre d'exemples, :

- définir les différences dans la charge amyloïde déterminée par imagerie PET entre les sujets qui bifurquent vers la maladie d'Alzheimer et ceux qui ne bifurquent pas.
- identifier la dynamique du processus de déclin cognitif par IRM fonctionnelle et EEG.
- évaluer l'intérêt de l'actigraphie.
- suivre les évolutions de l'expression des gènes.

L'objectif est aussi de décrire rétrospectivement les variables qui caractérisent la phase préclinique de la maladie d'Alzheimer et les facteurs qui contribuent à la progression du stade préclinique au stade symptomatique.

L'information à traiter est hétérogène, d'une part en raison des composants concernés (ADN, ARN, protéines), d'autre part en raison de la caractéristique génétique propre à chaque individu, sans compter que les techniques, comme les procédés de séquençage par exemple, ne sont nullement uniformes. La biologie est une science constamment aux prises avec l'hétérogénéité. Deux ou trois laboratoires travaillant en parallèle sur un même questionnement produisent fréquemment des résultats non superposables. Par exemple, le dosage de marqueurs dans le liquide céphalo-rachidien, effectué en France dans une vingtaine de centres, conduit à des données difficilement superposables. Qu'en

penser alors qu'il existe des dizaines de milliers de biomarqueurs dans les sciences biomédicales ? On ne peut agréger que ce qui est normalisé, la normalisation constitue donc un enjeu de première importance.

VII-3 Complémentarités et tensions

Les deux terrains, l'institut Pasteur de Lille et l'IM2A sont complémentaires, même si l'on peut percevoir un lien conflictuel entre la biologie et la clinique. L'explication des pathologies s'appuie de plus en plus sur les contributions de la biologie cellulaire et de la biologie moléculaire, et c'est bien l'approche des équipes de l'Institut Pasteur de Lille, mais on peut se demander si l'innovation biologique permet de résoudre le problème médical de la maladie d'Alzheimer.

Les données massives et hétérogènes constituent un phénomène nouveau, et pratiquement incontournable aujourd'hui, la recherche visant à faire émerger des connaissances à partir d'un volume considérable de données. Mais si du côté de la recherche on cherche à profiter de l'ensemble des données disponibles en provenance de toutes les équipes, mettant ainsi à profit le « voyage des données » et le traitement de décontextualisation / recontextualisation nécessaire à leur utilisation, la recherche clinique a une approche pragmatique, pas nécessairement explicative, et se concentre pratiquement uniquement sur les données acquises par leurs équipes.

Il faut ainsi trouver un point d'équilibre entre la recherche scientifique et la recherche clinique sur la maladie d'Alzheimer. Les deux approches se complètent mais se heurtent à des tensions, essentiellement temporelles, car la recherche clinique cherche à

stopper la progression de la maladie à un stade où la récupération est encore possible, sans forcément avoir la certitude du mécanisme d'action sur le processus physiopathologique, tandis que la recherche scientifique a la compréhension de la maladie et de ses mécanismes comme finalité. Il faut réussir à harmoniser les deux approches de façon à ce que des actions basées sur des approches statistiques ne perdent pas de vue la finalité de la compréhension de la maladie. On est ainsi amené à s'intéresser à une zone intermédiaire, la zone entre la recherche qui cherche à comprendre et à expliciter les mécanismes de la maladie, et le soin pour lequel aucune voie thérapeutique n'est encore efficace et disponible.

Chapitre VIII

L'organisation des recherches

Les technologies et les *Big Data* associées risquent de décontextualiser les pratiques médicales, alors que, selon Michel Foucault²¹⁵, il n'y a pas de vivant hors de son contexte. Il est donc important d'analyser dans quel contexte se posent les questions épistémologiques et donc d'analyser comment s'organisent les recherches.

Selon Michel Foucault dans *Les mots et les choses*²¹⁶, chaque époque est définissable par son *épistémè*, c'est-à-dire, non seulement par un ensemble de problématiques, d'hypothèses et de méthodes de recherche qui constituent un invariant de cette époque, mais aussi par un certain découpage du réel en objets identifiables, par des énoncés qu'il est possible de produire à propos de ces objets. Avec l'importance que prennent les données, on peut considérer qu'une nouvelle *épistémè*, axée sur l'information, est en voie de constitution. Mais sans doute faut-il s'interroger sur la définition même d'information, définition qui ne peut pas être que technologique – les notions de programme et de code sont abandonnées aujourd'hui dans les sciences du vivant -, mais plutôt vue dans un contexte de processus complexes. Ce qui a été dit sur l'évolution de la génétique l'illustre bien. Au moment du

²¹⁵ Michel Foucault, *Naissance de la clinique*, Paris, PUF/Quadrige, 2012.

²¹⁶ Michel Foucault, *Les mots et les choses*, Paris, Gallimard, 1966.

Projet Génome Humain, la génétique était entièrement dominée par le dogme central qui considérait que l'information était le secret de la vie et qu'elle se logeait exclusivement au sein de l'ADN sous la forme mixte d'une mémoire et d'un programme, tout en obéissant aux règles écrites dans la double hélice de la molécule. Avec la post-génomique, notamment la protéomique et la biologie des systèmes, tout n'est plus dans l'ADN et l'information est plutôt considérée comme une propriété émergente.

Par émergence, on entend que, au fur et à mesure que les systèmes acquièrent des degrés de plus en plus élevés de complexité organisationnelle, ils présentent de nouvelles propriétés dont l'existence ne peut être prédite à partir des lois gouvernant les systèmes les plus simples.

On peut, à ce stade, faire référence à Gilbert Simondon²¹⁷ qui proposait de remplacer la notion d'information par la notion de potentiels, au sens de la physique. Pour lui, « l'information n'est pas une chose, mais l'opération d'une chose arrivant dans un système et y produisant une transformation »²¹⁸. Par ailleurs, il est intéressant de relever ce qu'écrit Henri Atlan dans *Le Vivant post-génomique*. En critiquant la notion d'information, il dénonce « d'une part son caractère seulement probabiliste qui semble ignorer toute question de signification et, d'autre part, l'impossibilité de création d'information »²¹⁹. Enfin, La simple production de données n'équivaut pas à l'élaboration d'une information scientifiquement rigoureuse, pertinente et cliniquement utile. La quantité d'informations pertinentes est loin d'avoir augmenté dans les mêmes proportions que l'information tout court. Ce qui est en cause n'est autre que la relation entre les savoirs et les actions à mettre

²¹⁷ Gilbert Simondon, « Forme, information et potentiels », conférence de 1960.

²¹⁸ Gilbert Simondon, « L'amplification dans les processus d'information (1962) », dans *Communication et information : cours et conférences*, Chatou, La Transparence, 2010, page 159.

²¹⁹ Henri Atlan, *Le Vivant post-génomique, ou qu'est-ce que l'auto-organisation ?*, Paris, Odile Jacob, 2011, page 33.

en place, même si la formule « connaître pour agir » peut s'avérer inopérante faute de traitements adaptés contre la maladie. La bonne information nécessite de faire une première distinction entre le fait de communiquer sur une potentialité par le biais de processus biologiques et de communiquer sur la maladie elle-même.

En adoptant le point de vue des techniques, c'est-à-dire en considérant que les techniques influent sur les pratiques médicales, nous mettons un premier accent sur l'importance de la pluralité disciplinaire qui ne doit pas se « réduire à une simple juxtaposition de disciplines chacune ayant sa spécificité »²²⁰.

VIII-1 L'organisation des recherches

En partant des *Big Data*, en cherchant à s'appuyer sur la production massive de données, on cherche à mettre au jour de nouvelles connaissances, qu'elles soient scientifiques ou médicales, en considérant que la puissance des techniques informatiques utilisées tant pour collecter les données, issues de disciplines et de terrains très différents, que pour les traiter permettront de mettre en évidence des corrélations, voire des causalités, jusqu'alors insoupçonnées.

Retenons tout d'abord que les différentes techniques, que ce soient les techniques *'omics* ou les techniques d'imagerie, exigent pour leur conception et leur mise en œuvre une collaboration de compétences entre divers spécialistes. Le médecin ou le clinicien n'est qu'un des acteurs et il ne faut pas réduire le personnel technique, voué au fonctionnement des appareils, à une fonction subalterne. L'analyse médicale effectuée par le médecin est directement associée à la qualité des données et des images qui lui

²²⁰ Leo Coutellec, *La science au pluriel. Essai d'épistémologie pour les sciences impliquées*, Versailles, Editions Quae, 2015, page 14.

sont fournies. En se référant à Ludwik Fleck, on peut mesurer la difficulté des collaborations entre chercheurs de disciplines différentes et médecins ou cliniciens. L'époque où, selon Edgar Morin, « certains scientifiques croient naïvement que ce que leurs outils ne peuvent pas appréhender n'existe pas »²²¹, n'est sans doute plus d'actualité, mais chacune des actions menées a une certaine direction, sa propre direction. Cette direction peut, ne serait-ce que momentanément, ne pas être correcte, mais ce qui compte, c'est la direction générale. Cela est bien illustré par l'analogie que fait Fleck avec les fleuves se jetant dans la mer : « il suffit que suffisamment d'eau [le travail de l'ensemble du collectif] coule dans les fleuves et qu'un champ de gravité [l'ambiance dominante] existe, pour que les fleuves se jettent finalement dans la mer. [...] Le résultat ne provient pas de la direction momentanée de chaque goutte mais de la direction générale de la gravitation »²²². Les données ne parlent qu'en fonction des questionnements et des intérêts de ceux qui les interrogent et le chercheur les interroge souvent à partir des connaissances dont il dispose déjà.

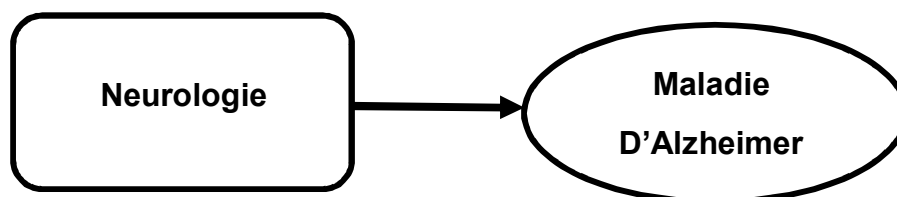
Cela change la façon de faire la recherche biomédicale. Dans tous les cas, il faut établir un climat de confiance. Il s'agit de faire travailler conjointement des acteurs dont les expériences, les compétences et les méthodes sont ou peuvent être très différentes. Les acteurs de ces recherches doivent former des réseaux et utiliser des espaces de stockages partagés, ce qui suppose des investissements pour des infrastructures technologiques autour des *Big Data*. Il est nécessaire d'organiser les collaborations de recherche pour éviter la duplication des efforts d'une part, pour augmenter les informations acquises d'autre part. C'est l'objet des plateformes et des consortiums de recherche. Le chercheur ou le

²²¹ Edgar Morin, *Science avec conscience*, Paris, Fayard, 1892, page 67.

²²² Ludwik Fleck, *Genèse et développement d'un fait scientifique*, traduit par Nathalie Jas, Paris, Les Belles Lettres, 2005, page 140.

médecin doit ainsi accepter de ne pas tout maîtriser, ce qui suppose de faire confiance à l'autre, et ce qui nécessite de former des réseaux qui se concrétisent par des espaces de stockage partagés entre les différents acteurs. La mise à disposition des données en ligne ne les rend pas automatiquement utilisables. Notamment il faut intégrer différents types de données, produites par différents instruments et différentes techniques, mais il faut aussi que soit documentée la provenance des données, ce qu'on appelle les métadonnées, à savoir la nature et la provenance des matériels biologiques testés, les auteurs de l'expérimentation ayant abouti aux données et leur laboratoire d'attachement, l'époque à laquelle les données ont été acquises. Ainsi, un portail Web comme *BioSharing* est consacré à l'assemblage de « standards communautaires développés dans les sciences de la vie pour rendre les données ainsi que les détails expérimentaux disponibles de manière standardisée »²²³. D'autres outils comme AraCyc ou MetaCyc²²⁴ permettent aux chercheurs et cliniciens de combiner et de visualiser des données génomiques, transcriptomiques et métaboliques de façon unifiée. De fait, il faut revoir la place des disciplines dans la compréhension d'un « objet complexe ».

Classiquement, tout objet était associé à une discipline :



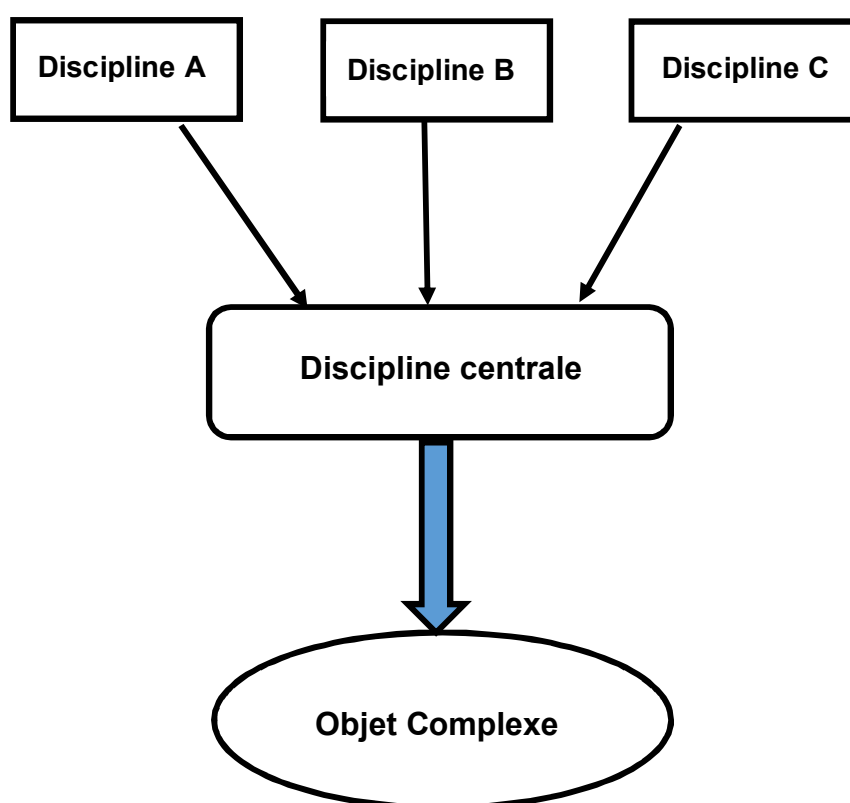
Il s'agit d'un véritable enfermement qui ne peut pas aboutir à une compréhension correcte d'un objet complexe car un tel objet ne peut pas dépendre que d'une seule discipline particulière. Le

²²³ <http://www.biosharing.org>.

²²⁴ Sabina Leonelli, *Data-Centric Biology, philosophical study*, Chicago, MI, The University of Chicago, Press, 2016, page 147.

monde de la recherche se caractérisant par une grande fragmentation des savoirs, il a alors fallu faire appel à une approche interdisciplinaire.

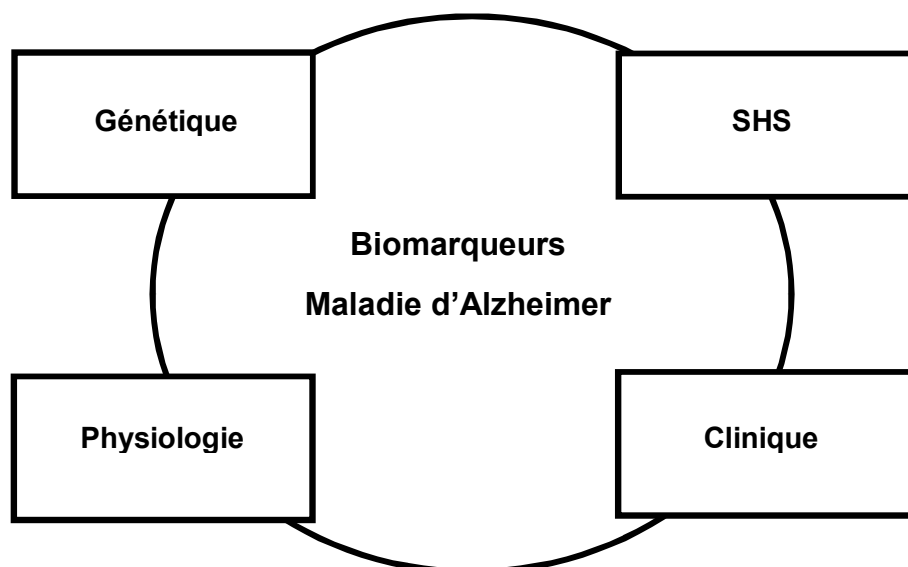
Néanmoins, cette prise en compte d'une approche interdisciplinaire s'est faite, dans un premier temps avec une discipline centrale qui « capturait » d'autres disciplines pour améliorer la compréhension de cet objet :



Cette interdisciplinarité est en grande partie illusoire, et ne permet pas une véritable vision transversale. Avec des objets complexes comme la maladie d'Alzheimer ou les biomarqueurs, on ne peut plus raisonner en disciplines et la convergence de disciplines elle-même n'apparaît plus également comme une approche suffisante. Il faut mettre en relation de la science et de l'humain, tenir compte des hétérogénéités et ne pas privilégier une discipline particulière. Il semble plus utile de ne pas considérer une discipline centrale mais de décentrer toutes les disciplines et de mettre au centre l'objet complexe que l'on cherche à comprendre.

Il faut même plutôt considérer que chaque discipline ne peut pas tout expliquer et amener chaque discipline à dire ce qu'elle ne peut pas expliquer au vu de l'état de ses connaissances actuelles. Ainsi – et on rejoint là l'approche insuffisante de la médecine génomique –, amener la biologie moléculaire et la génétique à dire qu'elles ne peuvent pas expliquer la maladie d'Alzheimer et ses évolutions peut amener d'autres disciplines, par exemple les sciences humaines et sociales, à transférer leurs connaissances ou propositions. Comme « objet de connaissance », la maladie d'Alzheimer, et les biomarqueurs associés, convoquent plusieurs disciplines, chacune ayant son style de pensée :

- La génétique est convoquée pour l'identification des gènes impliqués dans l'apparition, puis l'évolution de la maladie ;
- La physiopathologie conduit à des modèles de différentes voies physiologiques possibles, comme celui de la cascade amyloïde ;
- L'imagerie biomédicale est directement couplée aux pratiques expérimentales et aux équipements ;
- Les essais cliniques, y compris leur composante psychologique et les tests de mémoire, sont associés à l'analyse statistique et au calcul des probabilités.



On peut parler de transdisciplinarité, par confrontation des disciplines. Le résultat n'est pas la somme des apports individuels de chaque discipline, il y a un véritable dépassement des frontières fixées pour chaque discipline.

De fait, la complexité des phénomènes nécessite de transcender les différences culturelles, pas uniquement en engageant une démarche multidisciplinaire ou interdisciplinaire. Il s'agit d'aller dans le sens d'une démarche *pléo-disciplinaire* en rendant possible un travail conjoint, simultané ou parallèle, appuyé sur une interaction organisée dans une structure commune conçue à cet effet. C'est dire que les différentes disciplines doivent se rassembler, non pas par de la communication, mais par des échanges réels, les progrès se faisant par usure, usure liée aux frottements entre des savoirs qui trop souvent s'ignorent.

Outre leur intervention dans les essais cliniques, les SHS doivent être associées au système de descriptions textuelles des connaissances impliquées dans les expériences biologiques et cliniques et à la façon dont on peut traiter l'attention portée aux objets et aux processus afin que le processus d'interaction entre l'expérimentation conduisant aux données et l'individu se prêtant à cette expérimentation soit prise en compte par des techniques d'observation, des dispositifs d'enregistrement ou des enquêtes par exemple. En ce sens, les SHS pourraient intervenir dans la validation ou le rejet de certains biomarqueurs.

Ainsi, nous insistons, aucune approche, prise isolément, ne permet une compréhension correcte de l'objet que sont les biomarqueurs ou la maladie d'Alzheimer, une discipline isolée ne permet d'en faire qu'une description partielle. Il faut au contraire faire appel à plusieurs disciplines, mais pas chaque discipline dans son intégralité, plutôt certaines composantes de la discipline, de façon à éviter le pouvoir hégémonique de l'une d'elle.

Cela a conduit à considérer une nouvelle façon d'acquérir des connaissances, une science basée sur les données, ce qui suppose de comprendre comment ces données sont disséminées auprès des différents acteurs et comment elles sont utilisées pour aboutir à une nouvelle connaissance. On retrouve l'importance du packaging des données pour établir la confiance et du « voyage » des données de leur site de production vers de nombreux autres sites de recherche biologique ou clinique. Ces deux caractéristiques sont déterminantes pour l'épistémologie centrée sur les données. Les plateformes ont été créées dans ce but, d'une part pour mutualiser les investissements, ensuite pour les concentrer dans des pôles d'excellence, enfin pour faciliter le recrutement des personnels compétents pour faire fonctionner les équipements de pointe. Elles permettent l'émergence de nouvelles formes de collaboration entre le secteur public et le secteur privé et la mise en commun de ressources et de technologies. En France, la création d'un Réseau national de génopoles en février 1999 en est un exemple.

Un autre exemple est celui de la SMIP (*Stratified Medicine Innovation Platform*)²²⁵ créée en Grande-Bretagne pour développer la médecine personnalisée, en associant des laboratoires académiques, des laboratoires privés et des organismes de financement, et en définissant des procédures précises d'information des patients se prêtant aux recherches. Managée par le *Technology Strategy Board*²²⁶, la plate-forme gère l'ensemble des activités depuis le financement des projets de

²²⁵ *Stratified Medicine in the UK : Vision and Roadmap*. Technology Strategy Board, Driving Innovation Technology Strategy Board, in : <http://www.innovateuk.org>.

²²⁶ Le *Technology Strategy Board*, l'agence de l'innovation du Royaume-Uni, est un peu l'équivalent d'OSEO Innovation (qui a été fusionnée avec CDC Entreprises et les fonds stratégiques d'investissement pour créer la Banque Publique d'Investissement) en France, avec cependant une différence essentielle, c'est que l'agence ne se limite pas à financer des projets, mais y participe concrètement.

recherche jusqu'au développement de nouveaux produits et est très active dans l'évolution des réglementations. L'agence joue un rôle de coordination entre les acteurs du monde industriel, du monde académique et les Pouvoirs Publics. Sept organisations de premier plan, comme le *Technology Strategy Board*, le *Medical Research Council* ou le *National Institute of Health and Clinical Excellence* par exemple, collaborent avec des groupes pharmaceutiques, comme Astra Zeneca ou GlaxoSmithKline et des sociétés de diagnostic, dans le domaine de la cancérologie, de l'immunologie et des maladies inflammatoires. Les projets sont organisés de façon à ce que différents aspects d'un problème soient menés en parallèle, avec mise en commun du matériel biologique et des résultats, et suivi mensuel par des comités de pilotage.

Quant aux données, elles sont rassemblées dans des bases de données. Ce sont moins les données que le sens qui est donné par la présentation de ces données qui est importante. Parmi les données, il y a de signaux, qui contiennent réellement une information, et des traces, qui reflètent plutôt des comportements et l'objet est de « se servir des traces pour trouver la meilleure relation entre les signaux »²²⁷.

Rappelons que les bases de données, ensembles structurés et organisés permettant le stockage de grandes quantités d'informations afin d'en faciliter l'utilisation, sont toutes organisées en fonction d'un modèle de données qui peut être de différents types : modèle hiérarchique, modèle en réseau, modèle relationnel, modèle orienté objet, modèle semi structuré, modèle associatif, modèle EAV (*Entity-Attribute-Value data model*) ou encore modèle contextuel, le modèle de base de données relationnelle étant l'un des plus utilisés.

²²⁷ Dominique Cardon, *A quoi rêvent les algorithmes, nos vie à l'heure des Big Data*, Paris Seuil, , 2015, p. 63.

Ces bases de données sont à la disposition des chercheurs, des développeurs et des médecins et, aujourd'hui, toute information de séquence qui fait l'objet d'une publication, doit être déposée dans une banque de données.

Un autre moyen est la mise en place de consortiums. Un consortium est constitué d'entités institutionnelles gardant leurs spécificités, mais trouvant un lieu de pouvoir où elles discutent ensemble. Si les plates-formes permettent la mise en commun de ressources technologiques, les consortiums, de leur côté, permettent la mise en commun de ressources sociales²²⁸. Ils sont importants, d'autant plus qu'il faut mener conjointement le développement de tests diagnostiques et le développement de nouvelles voies thérapeutiques.

De nombreux consortiums se sont mis en place. On peut citer le *Biomarkers Consortium*²²⁹ mis en place aux Etats-Unis par le NIH (*National Institutes of Health*) pour identifier et développer les biomarqueurs permettant de faciliter le développement de nouveaux médicaments, en facilitant la collaboration entre les équipes de R&D du secteur public et du secteur privé. Ce consortium se focalise sur les cancers, les maladies inflammatoires et les maladies du système nerveux. De plus, il établit des règles et recommandations pour faciliter le passage à la clinique.

En Europe, le consortium PREDICT, mis en place en 2009 dans le cadre du 7ème programme cadre, est un autre exemple d'un large réseau structuré en médecine personnalisée. Il se focalise sur l'identification des biomarqueurs en cancérologie, en associant des laboratoires publics, des centres hospitaliers et des partenaires privés, grands groupes ou PME. Son budget est de 5,8

²²⁶ R. R. Nelson, « What enables rapid economic progress : what are the needed institutions », *Research Policy*, 2008, 37, pp. 1-11.

²²⁹ <http://www.biomarkersconsortium.org>

millions d'euros. On peut citer également EPEMED (*European Personalized Medicine*), une association européenne à but non lucratif qui accompagne le développement de la médecine personnalisée et qui propose une plate-forme d'harmonisation, la médecine personnalisée imposant le développement d'un environnement réglementaire et la maîtrise des informations pour les patients et les cliniciens.

VIII-2 La recherche translationnelle

Les biologistes cherchent à comprendre les mécanismes de la vie et le rôle des différents niveaux d'analyse tandis que les médecins et cliniciens cherchent à préserver la santé et le bien-être des êtres humains. Le vrai défi est de transférer les connaissances scientifiques en biologie à la clinique, et donc de développer la « recherche translationnelle » qui permet un accroissement et un partage des connaissances tout au long du processus allant de la recherche à la clinique. Elle s'est progressivement imposée pour réduire le temps entre la recherche de base et le développement clinique, d'une part parce que les allers/retours entre les observations cliniques et la recherche sont permanents, d'autre part parce que la pression pour satisfaire des besoins médicaux non satisfaits et les impératifs économiques et industriels des laboratoires pharmaceutiques amènent à tester des produits avant d'en comprendre tous les mécanismes d'action.

La recherche translationnelle est une démarche pour passer de la recherche à la pratique médicale. Son objet premier est de mettre en évidence les déterminants principaux d'une maladie et d'appliquer cette connaissance pour améliorer les pratiques cliniques et la santé des populations. Elle vise à une application des résultats de la recherche dans la pratique clinique. C'est en fait un transfert durant lequel certaines méthodes d'intégration et certains

développements d'un programme de recherche sont utilisés en recherche clinique. Il ne faut pas simplement la considérer comme une connaissance qui va d'un contexte A à un contexte B mais plutôt comme l'adaptation d'une méthode d'enquête²³⁰. La traduction (*translation*) peut être comprise comme la phase ultime de l'intégration de données, car elle aborde la question médicale de façon beaucoup plus complexe que ce que peut modéliser la science.

Les relations entre la recherche scientifique pure et la recherche clinique, voire la recherche au chevet du malade, ne sont plus des relations de phases successives, mais des relations d'interactions, ce qui nécessite un travail progressif et équilibré de disciplines très diverses. Il est ainsi essentiel de combiner tout un ensemble de méthodes et de développer des interprétations à plusieurs niveaux. Ainsi que le relèvent Alberto Cambrosio et ses collègues²³¹, les essais cliniques, part essentielle de la recherche translationnelle, en associant les biomarqueurs pour prédire les effets d'une thérapie, ne sont plus simplement des tests de validation, mais de véritables expériences cliniques pouvant aboutir, pour reprendre le mot de Cooper²³², à une « surprise expérimentale ». Cela résulte de plusieurs tensions :

- Raisonner avec un ensemble de données cliniques homogènes ou prendre en compte les différences entre les patients intégrés dans les essais, les échantillons utilisés et les techniques mises en œuvre.

²³⁰ Maureen A. O'Malley and al., « The role of integration in molecular systems biology », *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43 (2012), 58-68.

²³¹ Nicole C. Nelson, Peter Keating, Alberto Cambrosio, Adriana Aguilar-Mahecha, Mark Basik, « Testing devices or experimental systems ? Cancer clinical trials take the genomic turn », *Social Science & Medicine*, 111 (2014) 74-83.

²³² M. Cooper, « The pharmacology of distributed experiment – user-generated drug innovation », *Body & Society* 18 (2012) 18-43.

- Appliquer rigidelement le protocole prévu pour les essais ou faire évoluer les protocoles au cours des essais, pour tenir compte notamment de l'évolution des techniques de biologie moléculaire et des nouvelles connaissances.
- Considérer les participants aux essais comme représentatifs de la population ou comme le modèle d'un système.

Ces tensions illustrent les conflits éventuels entre test (le mécanisme d'action est supposé connu, on cherche à le valider) et découverte (on cherche toujours de nouvelles connaissances), entre recherche et application, entre stabilité/inflexibilité et continuité/flexibilité.

Le principal obstacle de la médecine translationnelle reste la question de l'individualisation, non pas au sens de processus de construction de l'individu mais au sens où il faut déduire ce qui est relatif à un individu particulier à partir de résultats obtenus sur un ensemble d'individus. D'une part, on n'est jamais dans la situation où on sera capable de donner une réponse personnelle, individuelle. D'autre part, le cadre économique et social ne permet pas de donner aux malades toutes les technologies identifiées comme nécessaires dans la recherche translationnelle, illustrant le conflit qui peut exister entre la recherche clinique et la routine clinique.

4

De ce point de vue, l'IM2A a adopté une approche novatrice qui place le patient au cœur de la recherche. Considérant qu'un diagnostic précoce et sûr est un pré-requis indispensable pour la recherche et le développement de nouveaux traitements contre la maladie d'Alzheimer, l'IM2A met à la disposition des patients atteints de troubles de la mémoire un plateau technique exceptionnel regroupant les ressources dans les domaines de la neuro-imagerie, de la génétique, des neurosciences cognitives, de

la biologie, de l'électrophysiologie notamment. Il propose une chaîne ininterrompue du patient au chercheur comportant :

- Une consultation médicale spécialisée dans la maladie d'Alzheimer.
- Un hôpital de jour pour faciliter les investigations cliniques.
- Un centre mémoire de ressources et de recherches (CMRR) pour coordonner la prise en charge des malades.
- Un centre d'essais thérapeutiques.
- Un centre de coordination d'études de cohortes
- Des espaces de travail facilitant le travail interdisciplinaire et l'interaction entre cliniciens et chercheurs pour et autour des patients.

Il se situe ainsi à mi-chemin entre l'hôpital et le laboratoire, avec une approche systématique de recueil des données.

Enfin, il ne faut pas occulter le conflit qui peut exister entre l'acquisition de nouvelles connaissances par les chercheurs et la résistance à leur utilisation par les praticiens. Il peut résulter de différentes raisons depuis l'incertitude technique ou les contraintes financières jusqu'à leur utilité ou les risques de nature éthique qui peuvent accompagner certains tests. C'est ainsi que le rôle du gène APOEε4, associé au départ à un traitement contre la maladie d'Alzheimer, traitement utilisant la Tacrine, a longtemps été rejeté par la communauté clinique de la maladie d'Alzheimer²³³. Indépendamment des recherches sur la causalité, le débat porte sur l'utilisation des tests génétiques en clinique, sans parler de leur utilisation en phase asymptomatique, et la signification des risques génétiques. A titre d'illustration, la société américaine *Athena Neurosciences* a introduit « *the ADmark ApoE Genetic Test* », d'abord comme test prédictif, puis, sous la pression des cliniciens,

²³³ Adam Hedgecoe, *The Politics of Personalised Medicine, Pharmacogenetics in the Clinic*, Cambridge UK, Cambridge University Press, 2004.

uniquement comme test de diagnostic de la maladie d'Alzheimer²³⁴, l'introduction du test ayant été vue comme une opération uniquement commerciale. Cela conduit à se poser la question de l'utilité clinique et de la praticabilité du diagnostic APOE différentiel et à se demander s'il faut donner le résultat du test à un patient – cela a déjà été évoqué -, alors qu'il n'a pas de valeur prédictive, et ce d'autant plus si la personne n'a pas d'antécédents familiaux qui pourraient le justifier, auquel s'ajoute le fait qu'il n'y a pas de traitement efficace contre la maladie d'Alzheimer. Dans un contexte de faiblesse des réponses médicales et de fragilité des structures de suivi et d'hébergement, l'anticipation du diagnostic et l'annonce d'une maladie comme la maladie d'Alzheimer pose des questions graves, ce qui impose sans doute de définir des règles, surtout dans la façon de prendre en compte les facteurs de risque. A la prédiction est associée l'incertitude, dès lors notamment que la maladie est difficile à saisir. De plus, annoncer un diagnostic de maladie d'Alzheimer, c'est, rappelons-le, certes annoncer un Alzheimer, mais c'est d'abord et avant tout dire à la personne qu'elle est malade, que les troubles dont elle souffre sont pathologiques alors qu'un grand nombre de patients ne se sentent pas malades. Sans doute faut-il s'interroger sur le fait de savoir si les chercheurs et les cliniciens doivent avoir des points de vue indépendants ou des points de vue qui convergent vers un consensus.

VIII-3 La question de la standardisation

L'explosion des données moléculaires a nécessité de développer des méthodes pour pouvoir les partager et donc de définir des standards pour cela. Ces standards portent autant sur la description de l'expérimentation ayant permis l'obtention des données que sur l'exécution proprement dite de l'expérience²³⁵.

²³⁴ Adam Hedgecoe, Ibid. page 37.

²³⁵ S. A. Chervitz, E. W. Deutsch, D. Field, H. Parkinson, J. Quackenbusch, P. Rocca-Serra, « Data standards for Omics data : the basis of data sharing and reuse », *Methods Mol.*

Ces différents standards :

- couvrent le minimum d'information nécessaire pour décrire une expérience comme l'illustre le MIBBI (*Minimal Information about Biological and Biomedical investigation*)²³⁶,
- facilitent l'interopérabilité et l'échange des données²³⁷,
- permettent l'interaction entre les différents « -omics »²³⁸.

À titre d'illustration, l'*European Medicines Agency* (EMA) établit ainsi des *guidelines* pour la médecine personnalisée et les biomarqueurs mais ne délivre des autorisations de mise sur le marché que pour les médicaments, et non pas pour les tests, dont la précision du diagnostic est pourtant extrêmement importante.

Le problème de la fiabilité des données a déjà été rapidement posé. Elles ne doivent être ni corrompues, ni incohérentes et, dans la mesure du possible, ni incomplètes. C'est une condition importante pour que les techniques d'exploration de données (*Data Mining*) permettent d'aboutir à des « index composites » sans avoir à exploiter les données dans le détail.

Quelles que soient les approches utilisées pour obtenir des données, il faut rappeler que tous les détails concernant les conditions expérimentales spécifiques et les procédures utilisées sont nécessaires pour être capable d'interpréter correctement les résultats. Il est donc important que ces données soient enregistrées avec les données d'observation ou les résultats, d'autant qu'il faut considérer que les *Big Data*, en raison de leurs différents modes d'obtention, sont souvent bruitées, dynamiques, hétérogènes, corrélées et peu fiables. Pour résumer, les données sont de plus en

Biol., Clifton, NJ, 2011, 719, pp. 31-69.

²³⁶ <http://mibbi.sourceforge.net/>.

²³⁷ <http://www.cdisc.org/>.

²³⁸ <http://www.ebi.ac.uk/ontology-lookup/>.

plus nombreuses, mais différent énormément :

- dans leur format et leur accessibilité,
- dans la façon dont elles ont été obtenues, notamment les technologies et expériences qui les ont produites,
- par le lieu, le laboratoire qui les ont obtenues.

Pour qu'elles aboutissent à une nouvelle connaissance, il faut intégrer cette diversité²³⁹. Quant aux systèmes d'information médicaux, ils sont de nature composite, associant de multiples composants hétérogènes.

Retenons cependant que, dès lors que les conditions expérimentales deviennent suffisamment stabilisées et standardisées, l'obtention de résultats toujours insatisfaisants signifie que l'hétérogénéité mesurée n'est pas un artefact expérimental, mais une propriété de ce que l'on observe.

VIII-4 La question de l'interopérabilité

Un des problèmes majeurs de l'utilisation des bases de données concerne l'interopérabilité, tant au niveau du contenu et des différents objets qui s'y trouvent stockés qu'au niveau des différents outils informatiques qu'on va mettre en face de ces bases de données. Cette interopérabilité ne peut être garantie qu'à condition de définir et d'utiliser des standards²⁴⁰.

L'interopérabilité se définit comme l'échange de données de nature différente, au sein de domaines qui peuvent être différents. L'interopérabilité est ainsi la capacité que possède un produit ou un

²³⁹ Sabina Leonelli, « Integrating data to acquire new knowledge : three modes of integration in plant science », *Studies in History and Philosophy of Biological and Biomedical Sciences*, 44, 2013, pp. 503-514.

²³⁸ Ce peut être illustré par le processus d'ECBD (Extraction des Connaissances à partir de Bases de Données ou *Knowledge Discovery in Databases*), Voir Frawley, Piattzsky-Shapiro and Smyth in « knowledge Discovery in Databases », 1991, AAAI/MIT Press.

système, dont les interfaces sont intégralement connues, à fonctionner avec d'autres produits ou systèmes existants ou futurs et ce sans restriction d'accès ou de mise en œuvre. C'est une nécessité dans le domaine médical. Dès l'instant où de nombreux acteurs et de nombreuses sources de données sont nécessaires, il est impératif, cela a déjà été dit, que des standards existent.

Le manque de standards pour coder des signaux, la diversité des plates-formes informatiques créent des difficultés lorsque l'on veut partager des données, voire même des modèles. Utiliser des données scientifiques ou expérimentales, des données cliniques, des données spécifiques à tel ou tel individu, nécessite un langage commun qui permette de former une unité explorable informatiquement. C'est la préoccupation des biologistes et des techniciens des différentes plates-formes. C'est aussi l'objet des ontologies²⁴¹ de référence²⁴². Il existe ainsi les cGP ou cartes Genotype-Phenotype (*causally cohesive Genotype-Phenotype mappings*)²⁴³ qui partent de l'idée que, dans un modèle dynamique d'un trait physiologique complexe, une part importante de la variation génétique sous-jacente se manifeste à travers les paramètres du modèle, paramètres qui sont eux-mêmes des traits phénotypiques. On peut ajouter que la commission européenne, dans le rapport *Semantic Health*²⁴⁴, a précisé ce que devait être une ontologie dans les systèmes d'information et les identificateurs permettant le référencement informatique des entités sémantiques dont le libellé permet la communication avec les utilisateurs.

²⁴¹ L'ontologie constitue en soi un modèle de données représentatif d'un ensemble de concepts dans un domaine, ainsi que des relations entre ces concepts. Elle caractérise le vocabulaire et les mots clés du domaine concerné tout en distinguant les faux amis. L'ontologie est notamment utilisée dans les méthodes d'apprentissage automatique pour la reconnaissance des modèles.

²⁴² « OBO foundry », *Open Biological and Biomedical Ontologies*, <http://www.obofoundry.org>

²⁴³ A. B. Gjuvsland, J. O. Vik, D. A. Beard, P. J. Hunter, S. W. Omholt, « Bridging the genotype-phenotype gap : what does it take ? », *The Journal of Physiology*, 591 (Pt8), 2013, pp. 2055-2066.

²⁴⁴ « Semantic Interoperability for better health and safer Healthcare. Research and deployment roadmap for europe », *SemanticHealth* report January 2009.

De la même façon, l'IHE-Europe²⁴⁵ est une association consacrée à l'interopérabilité en technologie de l'information pour la santé. Elle propose des guides d'implémentations pour les analyses de biologie médicale, les systèmes d'imagerie ou les systèmes de radiologie par exemple. Elle s'appuie notamment sur la norme HL7 (*Health level seven*)²⁴⁶, une norme internationale qui concerne l'échange de données sous forme de message et qui définit des spécifications techniques pour les échanges informatisés de données, notamment de données cliniques.

²⁴⁵ www.ihe-europe.net.

²⁴⁶ www.hl7.org.

Chapitre IX

La question de la démarche

La pratique médicale donne de plus en plus un rôle dominant au diagnostic, le diagnostic ayant une double implication. C'est d'une part une évaluation individuelle de la maladie, d'autre part une donnée statistique épidémiologique, ce qui entraîne une corrélation du diagnostic et des pratiques de prévention. Mais, de fait, les stratégies diagnostiques sont le plus souvent sous-déterminées par les connaissances biologiques. Comme le relate Ronan Le Roux²⁴⁷, l'espoir mis dans les technologies, notamment dans les instruments très haut débit, vient compenser le manque de connaissances sur les causes des maladies chroniques²⁴⁸ et, plus généralement, des maladies multifactorielles. On en vient alors à s'en remettre aux instruments sans en passer par la compréhension des mécanismes en jeu dans ces maladies. Cela amène à des choix sur les niveaux d'analyse. Par exemple, les partisans d'une approche purement génomique mettront en avant les technologies de séquençage, tandis que les partisans d'une approche basée sur les protéines et leurs interactions privilégieront la protéomique. Ainsi, comme l'écrit Emmanuel Petricoin, « la génomique a mené à la notion de 'médecine personnalisée', mais je crois que la route menant à la médecine personnalisée sera pavée de protéines [...] On ne peut continuer à espérer qu'un test génétique, séparé jusqu'à six degrés des aspects fonctionnels de la maladie [cancer], mènera aux thérapies plus précises dont on a besoin »²⁴⁹. En outre, il ne

²⁴⁷ Ronan Le Roux, " Nanotechnologies for Cancer Detection, Instruments, Paradigms and the Two Dogmas of Molecular Diagnostics", article à paraître.

²⁴⁸ On parle maintenant plutôt de maladies complexes que de maladies chroniques.

²⁴⁹ Emmanuel Petricoin, in Barbara Culliton, « Nanomedicine – the power of proteins : a

faut pas sous-estimer la relation d'influence réciproque entre technologies et paradigmes, ni l'effet de mode qui peut modifier la popularité relative d'une approche (ainsi les nouvelles techniques de séquençage ont refocalisé les approches sur la génomique, au détriment des puces qui donnent plus accès aux différences d'expression). De fait, il y a compétition entre les approches et cette compétition est renforcée par le fait que les équipes de chercheurs ou de médecins ne disposent que de certaines technologies (qu'ils privilégient alors) et que les coûts doivent être maîtrisés (ce qui amène à faire des choix), alors qu'il ne devrait pas y avoir compétition entre les technologies, mais plutôt complémentarité. On y reviendra en analysant les conséquences des *Big Data* sur la façon de faire la recherche, qu'elle soit scientifique ou clinique.

IX-1 Data-driven versus Hypothesis-driven

Relevons d'abord que statistique et méthodologie sont deux domaines indissociables. D'une part, la façon d'acquérir des données ne peut être médiocre si ce qu'on attend de la statistique doit permettre d'aboutir à des conclusions claires. D'autre part, les données n'ont de sens, ou d'intérêt, que si elles peuvent être analysées et interprétées. Cette remarque préliminaire n'a pour objet que de rappeler que l'opposition « *hypothesis-driven* » / « *data-driven* » n'est pas dichotomique et qu'il existe une véritable dualité statistique / méthodologie d'acquisition des données. Depuis Bacon (*Novum Organon*) jusqu'à Popper²⁵⁰ ou même Lakatos²⁵¹, le rôle de l'expérimentation était de tester des hypothèses. Mais

conversation with Lance Liotta and Emanuel Petricoin », *Health Affairs* 27(4), pp. 310-314.

²⁵⁰ K. R. Popper, *The logic of scientific discovery*, Harper and Row, New York 1968.

²⁵¹ I. Lakatos, *The methodology of scientific research programmes : Philosophical papers volume 1*, Cambridge University Press, Cambridge, 1978.

l'expérimentation est devenue une approche épistémique en soi par l'accumulation de données ²⁵².

Les instruments deviennent de véritables moteurs de recherche. Cela ne signifie pas qu'un déterminisme technologique se substitue à un autre déterminisme comme le déterminisme génétique. Cela signifie qu'une logique instrumentale oriente le cours des recherches en se substituant à la logique des problèmes à résoudre. Les instruments et techniques servent, de fait, à accumuler des données. Comme l'écrivent Bernadette Bensaude-Vincent et Dorothee Benoit-Browaeys, « la recherche à haut débit procède d'un choix méthodologique qui valorise l'accumulation de données en vue de permettre des comparaisons et des combinaisons. [...] Cette méthode transforme la théorie en calcul combinatoire »²⁵³. Certes, avoir de nouvelles possibilités ne signifie pas nécessairement savoir mieux, mais tous ces nouveaux instruments offrent des opportunités non négligeables en permettant des analyses à différents niveaux, un niveau moléculaire et un niveau plus global par imagerie notamment. On peut considérer que le progrès technique correspond à un accroissement de la connaissance. D'un côté, il est, en effet, le résultat d'un processus d'accumulation, de diffusion et d'utilisation croissante de la connaissance, que cette connaissance résulte de nouvelles lois, de nouvelles régularités, ou qu'elle résulte de nouvelles combinaisons de moyens techniques. De l'autre, il transforme les conditions de la connaissance, en permettant l'acquisition rapide d'un grand nombre de données dans les situations les plus diverses, et en intégrant les méthodes statistiques pour les interpréter.

²⁵² S. Leonelli, "Making sense of data-driven research in the biological and biomedical sciences", *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43, 2012, p ; 1-3.

²⁵³ Bernadette Bensaude-Vincent, Dorothee Benoit-Browaeys, *Fabriquer la vie, où va la biologie de synthèse ?*, Paris, Seuil, 2011, page 58.

L'utilisation expérimentale des instruments à haut débit, dans un cadre revendiqué a-théorique (recherche de corrélations sans hypothèses), continue ainsi de susciter des espoirs, bien qu'elle permette rarement d'identifier des biomarqueurs moléculaires cliniquement pertinents²⁵⁴. En comparant des cellules saines et des cellules malades et en cherchant ce qui les différencie à l'échelle moléculaire, on fait une observation a-théorique, sans hypothèses *a priori*, on cherche simplement à identifier des corrélations sans proposer de mécanisme explicatif. Dans un monde de données en petites quantités, la recherche des causes et l'analyse des corrélations commencent par la formulation d'une hypothèse de départ, et les données nécessaires à la confirmation ou l'infirmerie de l'hypothèse peuvent ne pas être disponibles. Si les données existent en très grand nombre, les hypothèses ne sont plus indispensables pour mener l'analyse par corrélation, et ce d'autant plus qu'on ne se limite pas à une recherche de relations linéaires et qu'on peut identifier aisément des relations non linéaires entre les données. Cela explique le débat souvent évoqué entre une approche « *hypothesis-driven* » et une approche « *data-driven* ». Cependant, on ne peut pas considérer que l'approche « *data-driven* » n'a pas de base théorique. Certes, elle n'utilise pas une approche causale, comme l'est la gravité par exemple pour expliquer des phénomènes mécaniques, mais elle s'appuie sur des théories statistiques, mathématiques et informatiques. Par ailleurs, toute analyse à partir des *Big Data* commence par une façon de choisir les données. Ce peut être pour des raisons pratiques, la facilité d'obtenir les données, pour des raisons économiques, l'obtention des données à faible coût, mais aussi des *a priori* d'analyse, considérer par exemple que ce qui est important, ce sont les variants du génome et choisir les données génomiques ou les

²⁵⁴ Terry Shinn, *Research-Technology and Cultural Change : Instrumentation, Genericity, Transversality*, Oxford, The Bardwell Press, 2007.

interactions entre protéines et choisir les données du protéome et du métabolome. Par ailleurs, on utilise, pour obtenir des données, des outils dont la conception repose sur certaines théories, et l'accumulation de données et leur utilisation de plus en plus poussée, ne suppriment pas l'étape la plus importante, celle de leur interprétation, et celle-ci est souvent associée aux savoirs pré-existants et s'appuie, elle aussi, sur des théories. Notons enfin que l'accumulation de plus en plus de données fait qu'elles seront disponibles pour des questionnements ultérieurs qui permettront de construire *a posteriori* des modèles de recherche non dépendants d'hypothèses de départ. L'accent mis sur la génomique n'est pas le seul à prendre en considération. Ainsi que le rappelle Richard Lewontin, chaque être vivant est à la fois le lieu et le produit de ses interactions. Les éléments de l'environnement, les éléments constitutifs de la personne, les microbes que nous hébergeons²⁵⁵, constituent une singularité évolutive et multiple.

L'opposition *hypothesis-driven research* versus *data-driven research* ne tient pas. Le fait d'acheter tel instrument ou tel autre oriente la recherche. La science est ainsi autant « *technology-driven* »²⁵⁶ que « *hypothesis-driven* » ou « *data-driven* ». L'activité technique est aussi une activité de connaissance, elle n'a pas qu'une perspective utilitaire. On retrouve ce que défendait Gilbert Simondon²⁵⁷ en écrivant « *opérateur* n'est pas synonyme de *pratique* ». Le rôle de la technologie et de l'instrumentation a été largement abordé par des philosophes comme Ihde²⁵⁸ et Pitt²⁵⁹ qui

²⁵⁵ Rappelons que nous hébergeons des centaines de milliers de milliards de bactéries, ce qui constitue notre microbiome dont la composition est singulière chez chaque personne. Ce microbiome a un effet sur notre métabolisme, nos réactions inflammatoires et sur la manière dont fonctionne notre système immunitaire.

²⁵⁶ W.Bains, « Technological speculations and science », *Bioscience Hypotheses*, 1, 2008, p : 177-178.

²⁵⁷ Gilbert Simondon, *Du mode d'existence des objets techniques*, Paris, Aubier 1989.p.255-256.

²⁵⁸ D. Ihde, « Has the philosophy of technology arrived ? A state-of-the-art review », *Philosophy of Science*, 71, 2004, p : 117-131.

²⁵⁹ J. C. Pitt, *Thinking about technology : Foundations of the philosophy of technology*, Seven Bridges, New York, 2000.

ont notamment relevé que la technologie n'intervient pas que comme un outil mais est constitutive de la démarche de recherche et de développement. Avant eux, Carl Mitcham²⁶⁰ avait relevé que la technologie est à la fois objet, savoir, activité et la marque d'une volonté. Tout comme la biologie, la biomédecine apparaît exemplaire par la sophistication des technologies développées pour produire, stocker et disséminer les données, par exemple les données génomiques. Les technologies haut débit, notamment les nouvelles techniques de séquençage et les *micro-arrays* font que les activités sont de plus en plus automatisées et « *technology-driven* » et que les infrastructures concernant les données prennent de plus en plus d'importance.

IX-2 La modélisation

En fait, l'attention se porte de plus en plus sur la complexité des réseaux d'interactions moléculaires et sur les facteurs épigénétiques associés notamment aux environnements auxquels est confronté un individu. La modélisation est la nouvelle forme de la connaissance scientifique. À la notion de modélisation est associée la nécessité d'expérimentation progressive, de séquences de prise de décision [décisions intermédiaires qui exigent une évaluation multi-critères (économique, éthique, ...)], ce qui peut amener des compromis. Il y a par ailleurs eu une évolution : jusqu'à présent, les modèles décrivaient les données, la modélisation consistant à identifier des dépendances logiques entre les données ; désormais, les données écrivent les modèles et apprennent d'elles-mêmes, par auto-apprentissage par exemple. Enfin, la modélisation a la grande force de pouvoir être modifiée/améliorée en permanence. On peut parler de modélisation

²⁶⁰ Carl Mitcham, *Thinking through technology*, Chicago, University of Chicago Press, 1994.

progressive, voire de modélisation en boucle, en allant des modèles aux données et vice versa.

La modélisation et la simulation permettent de dépasser les approches épistémologiques trop théorie-centrées.

La simulation est le calcul informatique d'une réalisation particulière d'un modèle, tandis que le modèle est un ensemble d'hypothèses ayant pour but de décrire un phénomène cible, avec un niveau de généralité moins élevé qu'une théorie.

Elles présentent l'avantage de ne pas être mono-disciplinaires et de prendre en compte des données hors des champs disciplinaires²⁶¹. Elles permettent de mettre en rapport des connaissances hétérogènes avec des données connues ou supposées et ainsi d'envisager de « découvrir de nouveaux phénomènes et non seulement de construire des hypothèses ou des théories en accord avec des phénomènes déjà existants »²⁶² ainsi que le souhaitait le biogéochimiste russe Vladimir Ivanovitch Vernadsky (1863-1945), également historien des sciences, qui a inscrit toute sa réflexion dans une tension continue entre l'observable, le déjà là et l'inobservable. La compréhension des phénomènes complexes, qui sont transdisciplinaires par nature, ne peut pas s'appuyer que sur une démarche théorico-centrée impliquant un centrage disciplinaire. Il ne faut pas avoir peur d'une « situation épistémologique dans laquelle nous devons pouvoir tenir compte de modèles multiples et dont les hypothèses parfois se contredisent »²⁶³.

²⁶¹ Le réel est toujours plus riche que ce qu'en disent les diverses disciplines.

²⁶² Léo Coutellec & Anne-Françoise Schmid, « Modélisation, simulation, expérience de pensée : la création d'un espace épistémologique – Regards à partir des œuvres de Vernadsky et de Poincaré », in *Modéliser & Simuler* sous la direction de F. Varenne, M. Silberstein, S. Dutreuil, P. Huneman, Tome 2, Paris, Editions Matériologiques, 2014, page 24.

²⁶³ Ibid. page 41.

La question de la sous-détermination des modèles

Sans être un spécialiste du climat, on sait que le climat est une affaire complexe et fait intervenir de nombreux paramètres. Par ailleurs, tous les systèmes complexes (en particulier si on ne peut pas les expérimenter) ont cette propriété de présenter des modèles sous-déterminés par rapport aux observations. En d'autres termes, le problème n'est pas de réaliser un bon modèle ; le problème est qu'il existe trop de bons modèles capables d'expliquer les mêmes observations. Si l'intérêt des modèles est de suggérer de nouvelles expériences, encore faut-il qu'elles soient réalisables. Si les modèles sont bons du point de vue du modélisateur puisqu'ils expliquent des phénomènes connus, ils ne décrivent pas forcément la réalité et il est nécessaire de réaliser des expériences pour tenter de réduire leur sous-détermination.

Pour comprendre des phénomènes il est essentiel de développer des interprétations à plusieurs niveaux, ce que peut faciliter l'exploitation de données en très grand nombre obtenues à des niveaux d'analyses différents, sans oublier, cela a déjà été relevé, que les modèles, les interprétations, les théories sont toujours contextualisées. On remarque que ces dernières décennies ont été marquantes dans le développement des savoirs en biologie, laissant présager une rupture épistémologique imminente du paradigme du « tout génétique » pour une vision plus complexe et flexible, sous le paradigme interactionniste.

Dans ce cadre, l'approche de la biologie systémique, comme celle proposée par Weston et Hood²⁶⁴, les promoteurs de la médecine 4P, est une approche *hypothesis-driven* puisqu'elle s'appuie sur un modèle de fonctionnement des cellules intégrant

²⁶⁴ A. D. Weston & L. Hood, « Systems Biology, Proteomics and the Future of Health Care : Toward Predictive, Preventive and Personalised Medicine », *Journal of Proteome Research*, 3(2), 2004.

différents niveaux dont les niveaux génomique, transcriptomique, protéomique, métabolomique.

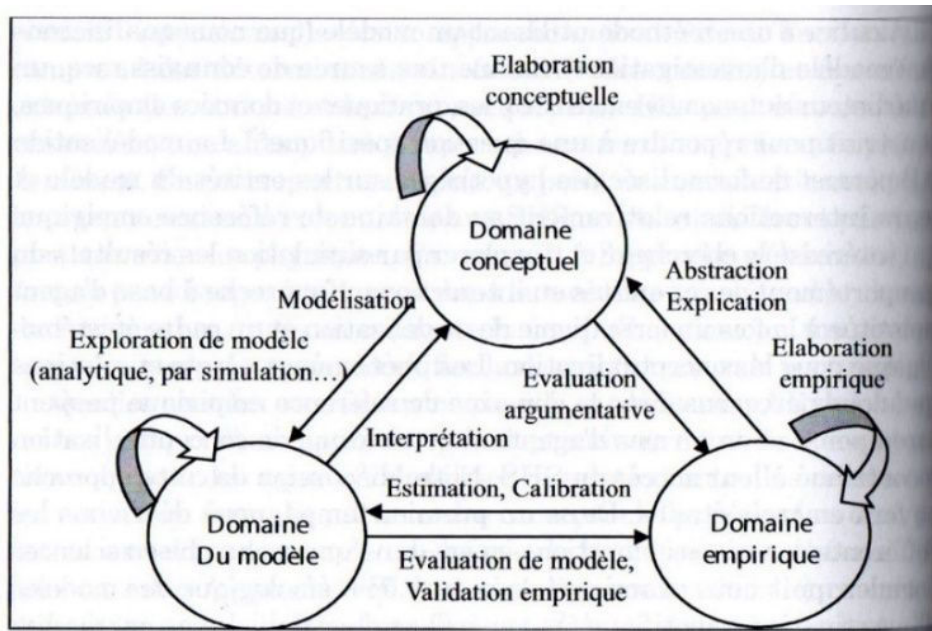
Un modèle n'explique pas tout, il est généralement incomplet. Il faut accepter la question de l'incomplétude, la signification biologique des corrélations entre des variations génétiques et des traits phénotypiques étant loin d'être évidente. Si l'on se réfère à Bergson²⁶⁵ qui distingue l'intelligence et l'intuition, l'activité de l'intelligence consiste à simplifier le réel pour ne retenir de lui que ce qui sert l'action sur les choses, tandis que l'intuition, délivrée des nécessités de l'action, est un rapport au réel qui cherche à en saisir l'inépuisable richesse. On peut penser que la pratique médicale utilise les deux.

Pour synthétiser, Pierre Livet et ses collègues²⁶⁶ proposent trois domaines, représentés sur la figure ci-dessous, qui doivent s'articuler dans toute étude comportant des aspects humains et sociaux

- Le domaine empirique concerne les informations qui proviennent des phénomènes observés. Dans ce domaine, on privilégie l'observation et la mesure, le repérage des régularités et des similitudes.
- Dans le domaine conceptuel, on cherche à identifier des facteurs pertinents pour expliquer les phénomènes du domaine empirique auxquels on s'intéresse.
- Le domaine du modèle correspond à une formalisation des théories formulées dans le domaine conceptuel et/ou des observations réalisées dans le domaine empirique.

²⁶⁵ Bergson, *l'évolution créatrice*, Paris, Quadrige/PUF, 2008.

²⁶⁶ P. Livet et al., « Ontology, a mediator for Agent-Based Modeling in Social Science », *The Journal of Artificial Societies and Social Simulation*, 13(1), 2010.



La pratique médicale représente la combinaison particulièrement difficile à mettre en œuvre d'un savoir sur la maladie, de propriétés encore inconnues de la maladie et de l'individu malade dans ses dispositions, et de l'action du médecin. La médecine ne peut donc pas être simplement guidée par des données et par là par des techniques et des algorithmes. Elle l'est aussi par des savoirs.

En se référant à Simondon, il faut rappeler que penser l'individuation, c'est penser l'apparition de l'individu et de son « milieu associé »²⁶⁷. C'est dire l'importance des relations de l'individu au milieu, mais dire aussi que l'individu et son milieu ne sont pas donnés antérieurement à leur mise en relation. L'individu fait autant le milieu que le milieu fait l'individu : « ce qui est premier, c'est la relation »²⁶⁸. Dans son ouvrage *Du mode d'existence des objets techniques*²⁶⁹, Gilbert Simondon distingue trois modes d'existence : l'élément, l'individu, l'ensemble. Le statut d'élément technique correspond au stade de l'outil ou de l'instrument. Le

²⁶⁷ Jean-Hugues Barthélémy, *Simondon*, Paris, Les Belles Lettres, 2014, page 31.

²⁶⁸ Ibid, page 48.

²⁶⁹ Gilbert Simondon, *Du mode d'existence des objets techniques*, Paris, Aubier 1989

statut d'individu technique est le stade de la machine qui correspond à l'articulation finalisée de divers éléments. Au stade d'ensemble technique, la machine prend place dans un réseau d'individus techniques densément interreliés, dont les relations sont indispensables à la réalisation des tâches visées. Dans ce sens, les pratiques médicales mettent en œuvre un ensemble technique. On peut dire que cela illustre l'intégration de la technique dans la culture et de la technique médicale dans la culture médicale.

Chapitre X

La mise en œuvre des techniques

Si l'on suit Georges Canguilhem²⁷⁰, la maladie n'est pas le résultat automatique d'un simple écart à des normes²⁷¹ biologiques que les progrès techniques permettraient de mieux en mieux de mettre en évidence, mais un nouvel état ressenti par le patient. Il n'y a ainsi pas de distinction franche entre normal et pathologique, pas de lien direct entre un facteur causal et une expression clinique. Par ailleurs, pour lui, c'est le vivant lui-même qui définit ses normes : « La vie est en fait une activité normative. [...] Au sens plein du mot, normatif est ce qui institue des normes »²⁷².

« S'il existe des normes biologiques, c'est parce que la vie, étant non seulement soumission au milieu mais institution de son milieu propre, pose par là même des valeurs non seulement dans le milieu, mais aussi dans l'organisme même. C'est ce que nous appelons la normativité biologique »²⁷³.

L'organisme est un tout, et la maladie affecte ce tout. Il ne peut donc y avoir de critères objectifs de la maladie. L'interaction permanente entre un organisme, les gènes qui le constituent et l'environnement qui l'entoure, fait que l'identité biologique, qui fait d'un être vivant un être unique, différent de tous les autres êtres vivants, et une individualité ne peut se concevoir que dans un contexte, un gène

²⁷⁰ Georges Canguilhem, *Le Normal et le pathologique*, Paris, PUF/Quadrige, 2011.

²⁷¹ Une norme est de l'ordre de la réglementation alors que la normativité dont parle Georges Canguilhem est de l'ordre de la régulation.

²⁷² Georges Canguilhem, *Le Normal et le pathologique*, Paris, PUF/Quadrige, 2011, page 77.

²⁷³ Ibid., page 155.

n'acquérant un pouvoir causal que dans un environnement cellulaire. La médecine anatomo-clinique permettait « un fondement enfin objectif, réel et indubitable d'une description des maladies »²⁷⁴. Elle était associée à un savoir sur l'individu et au concept de « dégénération »²⁷⁵. Avec la biomédecine, les formes de visibilité changent. Il n'y a pas qu'un changement d'échelle, de l'échelle tissulaire à l'échelle moléculaire, il y a une modification de la notion même de maladie en ce sens que la maladie devient détectable au niveau d'un déterminisme moléculaire, même si ce déterminisme est un déterminisme probabiliste, et peut même ne pas être encore apparue.

On peut rejoindre Gaston Bachelard qui, dans *La formation de l'esprit scientifique*²⁷⁶, explique que la science expérimentale progresse grâce à des ruptures de dogmes, grâce notamment à l'apparition de techniques ou d'instruments mieux adaptés et à des méthodes plus performantes et mieux appropriées. Comme le dit Daniel Cerezulle, « chaque avancée technique multiplie les possibilités d'applications des autres techniques »²⁷⁷. On ne peut cependant pas mesurer tout ce qu'on veut mesurer. Il peut y avoir des difficultés techniques, comme l'absence des technologies adaptées, mais aussi y avoir des difficultés conceptuelles comme lorsqu'on veut savoir comment mesurer l'éducation ou le mode de vie par exemple. La mise en œuvre des techniques dans la pratique médicale amène ainsi à faire face à plusieurs problématiques qui ont une composante éthique mais qu'on ne peut pas séparer des approches techniques. Relevons en cinq :

- Une des premières problématiques concernant la médecine, notamment la médecine prédictive ou la médecine

²⁷⁴ Michel Foucault, *Naissance de la clinique*, Paris, PUF, Quadrige, 2012, page 182.

²⁷⁵ Ibid., page 217.

²⁷⁶ Gaston Bachelard, *La formation de l'esprit scientifique*, Paris, 1938.

²⁷⁷ Daniel Cerezulle, *La technique et la chair, Essais de philosophie de la technique*, Lyon, Parangon/Vs, 2011, page 127.

anticipative, est la compréhension et la perception par les individus de la question du risque. La difficulté repose sur la distance entre risque et certitude, d'autant que les corrélations se situent dans le domaine des probabilités.

- La deuxième problématique concerne la validité du consentement et de l'information reçue. La liberté de réaliser un test comme un test génétique implique une compréhension de ce qu'il représente, de ce qu'il permet de réaliser par la suite.
- La troisième problématique concerne les différents acteurs impliqués par ces tests. En particulier, au sein du modèle traditionnel de la relation médecin-patient, basée sur le colloque singulier et le respect du secret médical, les pratiques de tests génétiques introduisent une série d'acteurs supplémentaires potentiellement intéressés par les informations ainsi produites. C'est notamment le cas des membres de la famille du patient, lorsque les données concernent une maladie génétique héréditaire ou une prédisposition génétique familiale pour une maladie déterminée.
- La quatrième problématique soulevée par les évolutions des pratiques médicales pour prédire ou anticiper une maladie concerne la limite de plus en plus ténue qui existe entre soin et recherche.
- La cinquième problématique concerne la responsabilisation excessive que cette nouvelle médecine doit éviter de mettre en place. En effet, cette médecine se fonde en partie sur un pari sur les progrès de la science. Si, aujourd'hui il est possible de surveiller, voire d'anticiper, il est rarement possible de guérir, si bien que cette médecine se base beaucoup sur l'espoir que demain un traitement existera.

La médecine occidentale tire son statut de scientificité essentiellement de son articulation aux sciences biologiques. Pour la médecine, l'homme est alors un organisme pluricellulaire complexe et toute pathologie est un dysfonctionnement de cet organisme. Mais ce réductionnisme méthodologique en jeu dans la pratique médicale tend à négliger complètement la spécificité de l'humain. Or, selon Catherine Draperi, « le savoir acquis sur le corps organique vient s'inscrire dans une pratique qui s'exerce dans la relation à l'autre. L'autre n'est donc pas seulement objet de l'action mais aussi sujet ; il n'est pas seulement porteur d'une maladie, mais cet autre porteur d'une demande »²⁷⁸. La question de la nature des savoirs se pose dès l'instant où ils sont imparfaits. À nouveau, face à une connaissance imparfaite, on peut la considérer soit comme un risque soit comme une incertitude, ce qui conduit à se demander ce que cela signifie de remplacer les causalités par les corrélations.

X-1 Ruptures, continuités et limites des pratiques

L'activité diagnostique relève d'une médecine de prévision, puisqu'elle peut consister à identifier une maladie avant que les symptômes n'apparaissent ; mais la maladie qu'il s'agit d'assumer est d'ores et déjà présente, ce qui pose la question de son évolution, de sa continuité ou des bifurcations créées par des événements extérieurs.

On utilise alors des techniques mais la technique est-elle neutre ou ne l'est-elle pas ? Si elle l'est, elle ne fait qu'obéir aux lois de la nature. Si elle ne l'est pas, elle incarne des valeurs spécifiques, elle peut devenir autonome, et alors, les moyens et les fins ne peuvent être séparés. Les lois sont des abstractions théoriques. Comme l'écrit Georges Canguilhem, « le vivant ne vit

²⁷⁸ Catherine Draperi, *La médecine réfléchie au miroir des sciences humaines*, Paris, Ellipses Editions, 2010, page 5.

pas parmi des lois, mais parmi des êtres et des événements qui diversifient ces lois. Ce qui porte l'oiseau c'est la branche et non les lois de l'élasticité »²⁷⁹.

Si les technologies permettent d'accéder à une meilleure compréhension des phénomènes, certains considèrent aussi les technologies comme une boîte à outils au service de l'humain. Ainsi, le Docteur Théodore Berger, directeur du *Center for Neural Engineering* à l'Université de Caroline du Sud travaille sur une micro-prothèse susceptible d'agir sur l'hippocampe²⁸⁰. Rappelons que l'hippocampe est la région du cerveau impliquée dans la création de nos souvenirs.

Le Docteur Théodore Berger développe une prothèse agissant sur l'hippocampe, une région du cerveau responsable de la mémoire à long terme. Il cherche ainsi, d'une part à comprendre comment l'hippocampe encode l'information pour donner la naissance à la mémoire, d'autre part à décrire mathématiquement ce processus de codage pour réaliser un modèle prédictif de la façon dont l'hippocampe répond à un événement. Il cherche à réaliser un modèle biomimétique de la dynamique non linéaire de l'hippocampe qui peut jouer la même fonction que l'hippocampe et ainsi corriger les dommages causés par des troubles neurologiques, comme dans le cas de la maladie d'Alzheimer. Il ne s'agit pas de restaurer des souvenirs perdus, mais de restaurer la capacité de mémorisation de nouveaux souvenirs.

En fait, utiliser une technique, c'est adopter un point de vue particulier et donc laisser d'autres points de vue dans l'obscurité. C'est donc « en multipliant les expériences et les techniques que le biologiste [et le médecin] peut aborder l'inconnu sans le réduire au connu »²⁸¹. La complémentarité des méthodes et des techniques s'impose.

²⁷⁹ Georges Canguilhem, *Le Normal et le pathologique*, Paris, PUF/Quadrige, 2011, page 131.

²⁸⁰ http://schedule.sxsw.com/2016/events/event_PP50701.

²⁸¹ Claude Debru, *Penser l'inconnu ? La recherche en biologie*, Paris, Hermann, 2009, page 6.

Un objet technique concret est un objet qui est défini par ses relations. C'est un objet dont les relations avec le milieu extérieur font partie du fonctionnement même de l'objet. Être en position critique, c'est d'abord interroger l'objet, connaître ses caractéristiques et le type de relation qu'on a avec l'objet. Selon Simondon, il est impossible de définir les objets techniques en tant qu'objets de connaissance, sans introduire la question des relations qu'ils entretiennent avec un milieu associé composé d'autres objets techniques, de réseaux, d'infrastructures et des hommes qui les utilisent. Si l'on prend l'exemple de la tomographie par émission de positrons, il faut faire un lien entre ce que l'on détecte et une activité du cerveau. Cela passe par la nécessité d'injecter de l'oxygène marqué dans le sang, l'hypothèse étant que le flux sanguin augmente dans les parties du cerveau en activité. Mais le lien n'est pas établi et rien ne dit que l'augmentation du flux sanguin ne soit pas associée à un processus d'inhibition. En fait, l'évaluation d'une nouvelle technologie est souvent basée, d'une part sur l'accord avec les théories existantes, d'autre part avec la cohérence des résultats issus d'autres approches, alors que, le plus souvent, la comparaison est difficile car les différentes approches apportent des informations complémentaires.

Il y a un clivage à dépasser entre les potentialités de la technique et les attentes des personnes. Il faut repenser le statut de la technique en considérant qu'elle n'est pas neutre (c'est une grande différence entre les ingénieurs et les médecins, biologistes, philosophes, etc.). Elle ne l'est qu'au niveau de l'outil, mais il faut considérer qu'avec l'outil, c'est tout un système qui lui est

associé²⁸². Cela repose la question de l'autonomie des personnes. L'*empowerment* peut y aider :

- Au niveau de l'individu, il augmente sa capacité d'action ;
- Au niveau de la relation de soins, il prend en compte les réalités différentes entre le soigné et le soignant ;
- Au niveau du système de santé, il permet de passer d'une vision paternaliste des soins à une vision plus émancipatrice.

Il articule deux dimensions²⁸³, une dimension de pouvoir et une dimension du processus d'apprentissage pour y accéder. L'environnement diagnostique ne doit pas seulement reposer sur une économie de la promesse ou induire une injonction à savoir. La question de la validité d'un résultat est cruciale. Pour cela, on utilise généralement des tests statistiques, mais les tests statistiques ne sont pas suffisants²⁸⁴ dès l'instant où ils ne s'appliquent qu'à une cohorte particulière. Cela pose la question de l'extrapolation, soit sur une autre population, soit dans des circonstances différentes, soit dans le temps. Par ailleurs, si les probabilités peuvent bien caractériser des dispositions, l'intensité d'une disposition peut varier d'une personne à l'autre en fonction de ses caractéristiques propres.

D'un autre côté, qu'en est-il des gènes ? La biologie de la médecine des systèmes n'a pas évacué le génocentrisme (même si le génocentrisme est de plus en plus contesté). Le facteur de risque génétique est, de fait, traité différemment des autres facteurs

²⁸² A titre d'illustration, un aspirateur autonome est un outil, mais en supprimant une corvée, on change l'humain.

²⁸³ Marie-Hélène Bacqué et Carole Biewener, *L'Empowerment, une pratique émancipatrice*, Paris, la Découverte, 2013.

²⁸⁴ Le but du test statistique est souvent de rejeter l'hypothèse nulle, c'est-à-dire de conclure que les deux échantillons sont statistiquement différents.

de risque. C'est le concept de l'exceptionnalisme génétique qui repose sur un ensemble de caractéristiques :

- La permanence des gènes ;
- Le fait que l'analyse génétique permet l'identification personnelle ;
- Le fait qu'on peut l'associer à un risque familial ;
- L'importance que prennent les probabilités.

Cependant, la prédiction doit être plus complexe, c'est-à-dire pas une prédiction génétique, mais une prédiction basée sur les gènes dans un environnement. La meilleure façon de caractériser un individu est-il toujours le génome ? C'est une tendance forte, d'autant que les technologies comme les séquenceurs de nouvelle génération facilitent cette caractérisation. Il faut associer des technologies permettant d'appréhender, à l'échelle moléculaire, l'exposition²⁸⁵ à un environnement tant extérieur (comme l'exposition à des polluants) que personnel (comme le mode de vie). Cela revient à dire que les biomarqueurs génétiques doivent être complétés par des biomarqueurs corrélés à la fois avec des expositions et avec la maladie.

X-2 Problème de l'hétérogénéité

Outre le fait de faire face à des phénomènes complexes, la difficulté est que la biologie est une science constamment en prise avec l'hétérogénéité. De fait, plusieurs laboratoires travaillant en parallèle sur un même questionnement produisent fréquemment

²⁸⁵ L'exposition fait intervenir le contexte. Elle peut créer de petites causes, créant de petits effets, mais ayant de grandes conséquences dans les réseaux d'interactions.

des résultats non superposables. On peut au moins identifier trois grands types d'hétérogénéité :

- Une hétérogénéité liée aux phénomènes biologiques (notamment aux différentes échelles) ;
- Une hétérogénéité liée aux données, à leur stockage, à leur sélection, etc. ;
- Une hétérogénéité liée aux traitements et à la modélisation.

De façon générale, les *data scientists* ou bio-informaticiens, qui n'ont pas été formés à la physiologie et à la biologie, sont convaincus que les modèles qu'ils développent sont la réalité et non sa représentation. Rendre les données diffusables et accessibles n'a pas pour objet l'interprétation des données en tant que représentation des phénomènes. Les processus de visualisation peuvent révéler des phénomènes qui n'apparaissent pas nécessairement quand les données sont très nombreuses, mais seules les statistiques permettent de simplifier les données. Ainsi que le fait remarquer Sabina Leonelli²⁸⁶, d'une part la dissémination des données fait qu'elles subissent des modifications qui peuvent les rendre significativement différentes de celles issues de l'expérimentation originelle, d'autre part les modèles de données, c'est-à-dire des modèles « conçus pour intégrer toute l'information sur une expérience susceptible d'être utilisée dans des tests statistiques pour vérifier l'adéquation à une théorie »²⁸⁷, ne peuvent avoir les mêmes attributs que les données elles-mêmes.

Par ailleurs, le modèle n'est pas universel : il ne vaut qu'à une échelle donnée et sous certaines conditions. Il y a les modèles

²⁸⁶ Sabina Leonelli, *Data-centric biology, a philosophical study*, Chicago MI, The University of Chicago Press,

²⁸⁷ Patrick Suppes, « Models of Data », in *Logic, Methodology and Philosophy of Science : Proceedings of the 1960 International Congress*, E. Nagel, P. Suppes, A. Tarski (Ed.) Stanford, CA, Stanford University Press, 1962, Page 258.

mathématiques mais aussi les modèles computationnels. Ces derniers construisent, à partir des données disponibles et sans aucun *a priori* sur les mécanismes, des modèles empiriques qui représentent l'enchaînement des processus biologiques sous la forme d'algorithmes. Mais il ne faut pas oublier que, derrière un modèle, il y a toujours un modélisateur et, s'il est important de spécifier l'ensemble des hypothèses qui sous-tendent le modèle, il ne faut pas occulter l'histoire, le vécu et l'expérience du modélisateur.

Ajoutons que si la probabilité est une notion qui a connu un développement considérable dans de nombreuses activités, il faut cependant pouvoir interpréter la signification de l'information donnée par la probabilité. La probabilité pose la question de son rapport au possible, et, sans doute faut-il distinguer les notions de possibilité et de probabilité, la probabilité étant une qualification ou une quantification de la possibilité²⁸⁸.

X-3 La question de l'apprentissage

L'accès aux technologies, aux données, à l'information ne dispense pas de l'apprentissage. Si avoir une encyclopédie médicale sur la table suffisait pour devenir médecin, on aboutirait rapidement à des consensus²⁸⁹ et à des connaissances partagées.

L'apprentissage est d'abord un apprentissage sur les technologies et les instruments. Connaître la machine ne suffit pas. Il faut aussi savoir comment prélever et préparer les échantillons à analyser et comment exploiter les résultats. Cela passe par des protocoles et des standards et montre l'importance de coupler

²⁸⁸ Le possible est sans doute ce qui est peu prévisible, tandis que le probable exprime une qualification des connaissances.

²⁸⁹ Le consensus dit un état du savoir, mais c'est une construction momentanée, c'est une commodité pour agir.

plusieurs activités et de rassembler de nombreuses compétences différentes. Il ne faut pas sous-estimer le fait qu'il est difficile de tout préciser dans les moindres détails et que les médecins, les techniciens, les opérateurs réagissent et interprètent en fonction de leur propre expérience.

L'apprentissage est ensuite un apprentissage informatique, un apprentissage des algorithmes eux-mêmes. Ce peut être un apprentissage supervisé au cours duquel on se focalise entièrement sur la caractérisation d'une variable cible. C'est le cas par exemple quand on l'utilise comme aide au diagnostic en médecine ou pour caractériser des profils de consommateur en marketing. Ce peut être un apprentissage non supervisé qui peut aboutir à de nouvelles segmentations, ce qui permet de créer une typologie de patients, de clients ou de produits par exemple, et de définir des politiques adaptées à chaque segment ainsi découvert. Cet apprentissage non supervisé utilise des règles d'association permettant de découvrir l'ensemble des relations probabilistes directes présentes au sein des données. La segmentation de variables qui en résulte permet de découvrir de nouveaux concepts et de synthétiser des variables cachées.

X-4 Conflit ou consensus entre les chercheurs et les cliniciens

La recherche translationnelle a déjà été évoquée, mais, dans le domaine de la santé, il est important et nécessaire d'introduire les essais cliniques. Ils nécessitent de nombreuses actions, auprès d'une population nombreuse et diversifiée, dans des centres géographiquement dispersés, de façon à rencontrer le maximum de cas différents susceptibles d'avoir accès au traitement étudié pour une pathologie déterminée. Leur objectif premier est de quantifier le rapport bénéfice/risque du traitement. Ils comportent trois phases, complétées par une phase de pharmacovigilance après l'autorisation de mise sur le marché. La première phase vise à

analyser la toxicité du produit, la deuxième à étudier l'efficacité du produit, la troisième à établir le service médical rendu.

Trois méthodologies peuvent être envisagées pour les essais cliniques en biomédecine, notamment en médecine personnalisée :

- un protocole dont la randomisation n'est pas stratifiée par les biomarqueurs et qui lie l'efficacité du médicament et les résultats des tests,
- un protocole faisant appel à la stratification, en différenciant les marqueurs négatifs et positifs,
- un protocole d'enrichissement, qui évince les marqueurs négatifs qui ne pourront pas répondre au traitement.

Le recours à la « randomisation », qui consiste à répartir au hasard les patients et les thérapies testées, afin de réduire l'interférence des variables autres que celles étudiées, a été une étape essentielle. Dans ce type d'étude, les patients sont regroupés en deux groupes préalablement constitués et reçoivent deux traitements différents assignés par tirage au sort. Le test peut être réalisé en aveugle, le sujet ne sachant pas le traitement qui lui est administré, ou en double aveugle, l'investigateur ne le sachant pas également, ce qui permet d'éviter tout biais dû à l'investigateur. La loi des grands nombres démontre que, lorsque l'effectif est suffisant, une répartition aléatoire des sujets dans les deux groupes assure une homogénéité correcte des déterminants au sein des groupes. Il est alors possible d'attribuer la différence d'effet observée aux traitements. Ainsi, la loi des grands nombres, et les outils statistiques associés, permettent de traiter des données subjectives pour en tirer des résultats objectifs, mais il est important de souligner qu'elle ne s'applique que si la taille de l'échantillon est beaucoup plus grande que le nombre de paramètres à prendre en compte.

Les systèmes étant complexes et leurs dynamiques « multi-échelles », ils ne peuvent être étudiés qu'*in vivo*. Il faut rappeler l'importance de la recherche translationnelle qui assure le transfert et l'application rapide des avancées de la recherche au bénéfice des malades. Elle assure le continuum entre la recherche et les soins et permet aux patients de bénéficier plus rapidement des innovations diagnostiques et thérapeutiques. Elle facilite le passage de la science à la clinique.

Mais, ainsi qu'il a été mentionné, il ne faut cependant pas occulter le conflit qui peut exister entre l'acquisition de nouvelles connaissances par les chercheurs et la résistance à leur utilisation par les praticiens comme l'a révélé la réticence à la prise en compte du rôle du gène ApoEε4 dans les tests génétiques utilisés en clinique. Rappelons qu'à la prédiction est associée l'incertitude, dès lors notamment que la maladie est difficile à saisir. Sans doute faut-il s'interroger sur le fait de savoir si les chercheurs et les cliniciens doivent avoir des points de vue indépendants ou des points de vue qui convergent vers un consensus.

Troisième partie

Trois questions épistémologiques

Chapitre XI

La démarche explicative dans la connaissance médicale

Quand on aborde la question de la causalité et du mécanisme, deux approches sont complémentaires. D'un côté, il y a ceux qui cherchent à construire des théories pouvant expliquer, du moins en partie, la réalité observée. De l'autre, il y a ceux qui cherchent à gérer cette réalité sans pour autant l'expliquer. Il y a ainsi un paradoxe. La médecine peut simplement prédire sans se mettre en tête de chercher à expliquer. Elle calcule alors des trajectoires prévisibles à partir des corrélations issues de données recueillies sur des cohortes de patients.

Le terme de trajectoire²⁹⁰ fait référence non seulement au développement physio-pathologique de la maladie, mais également à l'ensemble des organisations utilisées pour suivre le cours de cette maladie. Ce concept de trajectoire est adapté chaque fois qu'un sujet cherche à contrôler son état. Comme l'écrit Anselm Strauss²⁹¹, ce concept « est nécessaire à une appréhension sociologique de la gestion de la maladie ».

²⁹⁰ On pourrait dire aussi que la trajectoire d'une maladie est un phénomène changeant, parfois de manière inattendue, que c'est un processus.

²⁹¹ Anselm Strauss (1992) « Maladies et trajectoires » dans *La Trame de la négociation*, Paris, L'Harmattan, page 144.

Mais cette approche prédictive repose sur le principe du déterminisme : elle vise à mettre en évidence les causes en présence, et les effets qu'on peut en attendre. Or, voir les causes qui vont déterminer un état futur, ce n'est pas prédire cet état.

La médecine fondée sur des faits (EBM ou *Evidence Based Medicine*) promeut une médecine de la preuve, censée déjouer les failles d'une médecine de l'autorité²⁹², et facilite la mise en place de systèmes automatisés et l'accès aux connaissances médicales. La preuve est la matérialisation d'un fait, elle peut n'en être que la trace, mais elle apporte la certitude. Comme le souligne Anne Fagot-Largeault²⁹³, l'EBM « doit permettre d'assurer la cohérence du comportement médical » en cherchant à construire une méthodologie de la pratique médicale. Son fondement repose donc sur une méthodologie de constitution du savoir. Cette preuve peut être de nature expérimentale ou de nature statistique, même si cette distinction est assez floue²⁹⁴. Mais il faut l'articuler avec une médecine fondée sur des valeurs (VBM ou *Value Based Medicine*) car il s'agit, en effet, non pas de traiter une maladie, mais plutôt de soigner une personne, en tenant compte de sa singularité, de ses priorités, de ses attentes. Cela conduit, de fait, à envisager deux approches, une approche médicale de la maladie, basée sur des statistiques, et une approche personnelle de la maladie. Il y a une tension, généralement croissante, entre le devenir scientifique de la maladie d'une part, et la disparition progressive de la singularité et de la subjectivité du malade d'autre part.

²⁹² D. L. Sackett et al., « Evidence-Based Medicine : What it is and What it isn't », *British Medical Journal*, 1996, 312, pp. 71-72.

²⁹³ Anne Fagot-Largeault, *Philosophie des sciences biologiques et médicales*, cours au Collège de France 2001-2002, https://www.college-de-france.fr/media/anne-fagot-largeault/UPL8928_AFL_res_0102.pdf.

²⁹⁴ Par exemple, les essais cliniques, considérés comme des preuves expérimentales, sous prétexte qu'ils sont contrôlés, conduisent, de fait, à des résultats établis statistiquement.

Progressivement, on est aussi passé de l'*evidence-based medicine* à la *precision-based medicine* qui fait appel à la multiplication des paramètres d'analyse d'un sujet, notamment par l'utilisation de toutes les technologies 'omics. En devenant scientifique, la médecine tient compte de la singularité de la personne mais est aussi porteuse du paradoxe épistémique et éthique d'être « la médecine de moins en moins centrée sur la personne ».

La médecine n'est pas que l'application d'un savoir, fut-il biologique, elle est plus une pratique qui a trait à un savoir sur l'homme, dans la complexité de son contexte. Devant la nécessité de prendre en compte la complexité du problème médical et les nombreuses interactions entre un individu et son milieu, avec leurs composantes dynamiques à travers le temps, la notion d'information mérite d'être discutée. L'information est un processus. Elle doit être répétée, disponible sur des supports variés, en cohérence avec les besoins des patients. On peut distinguer « information pour » et « information au sujet de », la première étant plus probablement l'information effective pouvant justifier ou expliquer l'effet de différences. Cette information, qui, nous avons déjà insisté, doit toujours être associée à un contexte, permet d'aborder la question des enjeux, qu'ils soient épistémiques²⁹⁵, éthiques ou économiques et sociaux, des objets techniques que sont les biomarqueurs, en différenciant :

- Une donnée, qui est un fait brut.
- Une information, qui est une forme d'interprétation de la donnée.

²⁹⁵ Federica Russo, « Science without technology ? The case of biomarkers in molecular epidemiology », *Congrès 2013 de la Society for the Philosophy of Technology*, Lisbonne. Marianne Boenink, « The technological constitution of medicine's future : Ontological and epistemological issues concerning 'biomarkers ' », Ibid.

- Une connaissance, qui est une mise en sens et en cohérence d'informations.
- Un savoir ou un savoir-faire, qui est la capacité à organiser un ensemble de connaissances pour agir dans un contexte donné.

Avec les *Big Data*, on accepte de travailler avec des données parcellaires, disponibles et rapidement accessibles, en recherchant – le plus souvent instantanément – des informations supposées essentielles et traitées dans leur globalité sans *a priori*. Mais il n'existe pas de corrélation simple et évidente entre un diagnostic physiopathologique, lié notamment à l'usage de biomarqueurs, et un diagnostic clinique.

XI-1 Des données aux connaissances

Se posent au préalable les questions sur le statut et la valeur des données, sur l'incertitude associée à ces données, sur l'interprétation que l'on peut faire de ces données en abordant les questions de causalité et de mécanismes, et sur les conséquences que le choix de l'acquisition de données en masse a sur les pratiques de recherche, notamment de recherche clinique. La question de savoir comment se construisent et se disséminent les données, leur format, leur qualité se pose ainsi avant d'analyser comment elles produisent de la connaissance, sans « oublier que toute production des connaissances est toujours déjà située »²⁹⁶.

La production massive de données a l'ambition de mettre à jour de nouvelles connaissances scientifiques et de conforter ou de faire évoluer les pratiques médicales. Il y a deux types de raisonnements. Dans le premier, une accumulation d'énoncés

²⁹⁶ Dominique Pestre, *Science, argent et politique. Un essai d'interprétation*, Paris, INRA-éditions, 2003.

singuliers conduit à un énoncé de type général : c'est l'inférence inductive. Dans le second, le raisonnement déductif relie des énoncés généraux à des énoncés particuliers. Appliqués au fonctionnement de la science, on peut considérer qu'on a une conception inductiviste du processus de constitution du savoir scientifique et une conception déductiviste de l'explication scientifique. Ainsi que nous l'avons relevé en abordant la question de la démarche, la méthodologie traditionnelle hypothético-déductive n'est plus la seule. On passe d'une science de l'hypothèse à une science de la découverte par comparaison. L'analyse des immenses masses de données que constituent les *Big Data* se fait principalement par la mise en évidence de corrélations. Les *Big Data* procèdent plus par induction et génèrent du sens en passant des faits observés à une connaissance générale alors que la déduction est un processus qui permet de conclure une affirmation à partir d'hypothèses.

Les *Big Data* et les technologies associées ouvrent à une très grande libération de la connaissance et à un nouvel âge pour les capacités prédictives²⁹⁷ :

- Jusqu'à la Renaissance, les prédictions étaient basées sur l'interprétation de signes (astrologie, augures), ces signes étant supposés associés à des plans divins ;
- A l'âge moderne, à partir de Pascal, les prédictions ont été basées sur un travail mathématique en interprétant les faits par des probabilités. La prédiction n'est alors vraie qu'en moyenne, elle suppose de disposer de données en masse (un grand échantillon pour peu de paramètres), et cette prédiction est inopérante individuellement (l'espérance de vie des gens nés en 1945 ne signifie pas que vous, né en 1945, vivrez jusqu'à tel âge, par exemple) ;
- Avec l'arrivée des *Big Data*, apparaît une capacité de prédiction individuelle. La prédiction repose alors sur l'isolement de l'individu de la masse par l'utilisation de profils par exemple. C'est ainsi que dans le domaine médical, il ne faut pas considérer qu'il y a une maladie (le cancer par exemple) et des malades, mais qu'il y a autant de maladies que de malades.

²⁹⁷ François Ewald, « Big data, chance ou malédiction », Conférence à la fondation Calouste-Gulbenkian, Paris le 5 juin 2014

Induction et déduction sont deux procédures de raisonnement. Elles ne sont pas en réelle opposition²⁹⁸. En fait, les hypothèses ne disparaissent pas, elles changent plutôt de statut en n'étant plus la source première d'une investigation, mais plutôt des moyens d'ajustement théorique à la suite de l'exploitation des données.

L'abduction, proposée par Charles S. Peirce²⁹⁹ qui la définit comme la « conjecture sans force probante, fondée sur une hypothèse tirée de l'expérience » est, avec la déduction et l'induction, l'un des trois types majeurs de l'inférence³⁰⁰. Elle consiste, lorsque l'on observe un fait dont on connaît une cause possible, à conclure à titre d'hypothèse que le fait est probablement dû à cette cause-ci. Une méthode de recherche peut être modélisée selon l'enchaînement élémentaire « abduction – induction – déduction ». Traduite pour le traitement des *Big Data*, l'abduction signifie que les données vont « parler » d'elles-mêmes sans qu'il y ait besoin d'avoir de modèles préconçus pour les analyser. Dans cette perspective, l'abduction est le processus de formation d'une hypothèse générale sans l'assurance qu'elle réussisse.

Selon Boyd et Crawford, l'accélération de la collecte et l'accumulation de données poussent à un mode de raisonnement

²⁹⁸ On peut considérer cependant que l'induction ne fait passer que du vrai au probable tandis que la déduction va du vrai au vrai.

²⁹⁹ I. Mourral, L. Millet, *Petite encyclopédie philosophique*, Editions Universitaires, 1995, page 7.

³⁰⁰ L'inférence, c'est passer d'une proposition tenue pour vraie à une autre qu'on juge en conséquence l'être aussi.

de type empirico-déductif³⁰¹. Cependant, les *Big Data* ont permis une libération de la connaissance selon plusieurs points de vue :

- une libération par la production des données en quantité, et on peut rappeler que la génétique d'aujourd'hui n'existerait pas sans les technologies d'acquisition des données qui ont accompagné le programme du séquençage du génome humain et ses suites ;
- une libération par la capacité de faire apparaître entre ces données des relations surprenantes, pas forcément expliquées, mais constatées avec une certaine régularité. Ces corrélations, pas nécessairement explicatives, sont bien de nature prédictive, et le domaine médical est bien bouleversé par ces approches ;
- une libération par le côté dynamique de la production de données, puisque les données engendrent de nouvelles données.

Ces auteurs mettent en avant « six provocations » :

- L'automatisation de la recherche change la définition du savoir.
- Les revendications d'objectivité et d'exactitude sont trompeuses.
- De plus grosses données ne sont pas toujours de meilleures données.
- Toutes les données ne sont pas équivalentes.
- Accessible ne veut pas dire éthique.
- L'accès limité aux *Big Data* crée de nouvelles fractures numériques.

³⁰¹ D. Boyd, K. Crawford, « Critical Questions for Big Data. Provocations for a cultural, technological, and scholarly phenomenon », *Information, Communication and Society*, 15 : 5, 2012.

En contrepoint, la vision positive sur les *Big Data* est caractérisée par les avantages suivants :

- La donnée est le moteur de la découverte.
- La totalité des données peut être prise en compte.
- La recherche de corrélations est la méthode de choix.
- Les *Big Data* permettent la prise de décision.
- Les *Big Data* restent utiles même en l'absence de cadre théorique.

Mais ces avantages suscitent à leur tour des critiques quant à l'utilisation des *Big Data* telles que :

- L'hypothèse est nécessaire comme moteur de la découverte.
- La totalité des données n'est pas nécessaire, un échantillon de données suffit.
- La causalité est ce qu'il faut absolument rechercher.
- La prise de décision ne peut s'appuyer que sur la connaissance.
- Un cadre théorique est toujours nécessaire.

Ces deux points de vue coexistent.

Si la médecine veut être véritablement une science de la personne humaine – la question doit être argumentée, est-elle une science ? Est-elle une technique comme le voulait Georges Canguilhem³⁰² ? Est-elle un savoir, savoir de la personne humaine, comme disait Michel Foucault ? - elle doit chercher à comprendre le sens de la maladie pour la personne qui en souffre, « bien au-delà du mécanisme causal que le diagnostic est susceptible d'établir pour en trouver la cause et les moyens de la soigner »³⁰³.

³⁰² Selon Georges Canguilhem, la médecine n'est pas une science, c'est une technique utilisant des process au service de la vie.

³⁰³ Marie Gaille, *Philosophie de la médecine - Frontière, savoir, clinique*, Paris, Vrin, 2011, page 31.

Ainsi, la médecine scientifique s'est-elle développée sous la forme d'une introduction des mathématiques probabilistes, l'outil probabiliste se répandant massivement à partir du XIX^e siècle et jouant un rôle essentiel dans l'« *Evidence Based Medicine* » contemporaine. Comme le dit Michel Foucault, la médecine, à travers l'usage des probabilités, peut s'accepter comme savoir incertain, sans faire de l'incertitude une marque de non-savoir³⁰⁴. Ce faisant, elle donne en outre une place à l'individu malade, tout en cherchant à s'extraire de la particularité : « la certitude médicale se constitue non pas à partir de l'individualité complètement observée, mais d'une multiplicité entièrement parcourue de faits individuels »³⁰⁵.

XI-2 La question de l'incertitude

Dire que tout devient possible et parler de possibilité signifie qu'il y a une incertitude à la base. Cela signifie que décider présuppose l'incertitude, dont on peut distinguer deux types³⁰⁶ :

- L'incertitude de type épistémique due à l'imperfection de nos connaissances.
- L'incertitude ontologique due à la variabilité inhérente aux systèmes complexes.

À chaque degré de complexité apparaît un niveau d'incertitude, ce qui montre, de fait, la limite du déterminisme scientifique. Mais l'incertitude est aussi une valeur sociale, et, dans le traitement des données, il faut faire la différence entre la variabilité qui fait la singularité et ce qui est interprétable en termes de maladies.

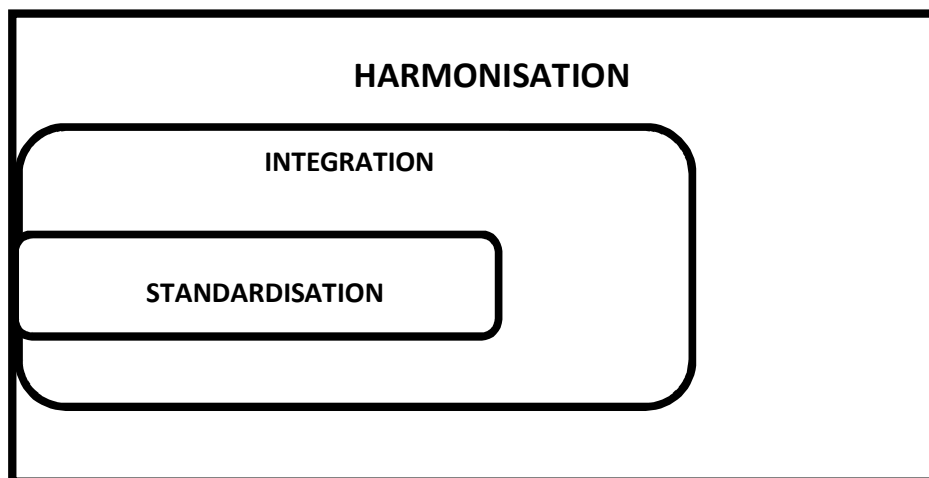
³⁰⁴ Michel Foucault, *La naissance de la clinique*, Paris, PUF/Quadrige, 2015, page 139.

³⁰⁵ Ibid., page 145.

³⁰⁶ W. E. Walker et al., Defining Uncertainty. A Conceptual Basis for Uncertainty Management in Model-Based Decision Support", *Integrated Assessment*, 4:1, 2003.

L'incertitude est aussi quelque chose qui affecte les données. De fait, ce ne sont pas des données brutes qui sont utilisées. Toute une série de programmes en contrôle la qualité, en les nettoyant et en s'assurant que les données agrégées, si elles sont hétérogènes par nature et présentent des variabilités tant interindividuelles qu'intra-individuelles, ont été obtenues selon des procédures normalisées. Ajoutons que si nous avons à considérer des données, qu'elles soient issues de l'expérimentation ou de l'observation, nous ne devons pas oublier que ces données doivent être validées, avant de les utiliser pour déterminer les causes d'un phénomène. Les statistiques y aident.

Les technologies autour des biomarqueurs doivent satisfaire des critères de standardisation, d'intégration et d'harmonisation³⁰⁷ :



La standardisation concerne la collecte des données, leur analyse pour produire de l'information, ainsi que les procédures pour créer de la connaissance. L'intégration suppose la possibilité de prendre en compte des données provenant de sources diverses pour considérer le vivant comme un tout et le couplage entre les données

³⁰⁷ Harvey and al., "The future of technologies for personalised medicine", *New Biotechnology*, Vol. 29, N° 6, 2012.

scientifiques et les données de santé. Quant à l'harmonisation, elle concerne l'harmonisation des pratiques dans les différents centres, l'harmonisation par la communication entre le monde scientifique et le monde médical, l'harmonisation des réglementations et l'harmonisation des systèmes de santé.

Ces critères suffisent-ils ? On peut s'interroger en considérant à titre d'exemple, l'utilisation de techniques comme les techniques d'imagerie. Elles donnent à voir, mais, la question a déjà été posée, voir, est-ce savoir ? Le fait de voir le fonctionnement du cerveau, notamment l'activité de certaines zones dans un contexte donné, permet-il de mieux comprendre notre manière d'agir et de raisonner ? Nous pouvons plutôt considérer que le médecin, quand il analyse une image biomédicale, ne voit que parce qu'il sait – il voit par exemple les lésions caractéristiques de la maladie d'Alzheimer.

Parler de preuve, c'est-à-dire d'une information utilisée pour croire ou non à une proposition, nécessite de se préoccuper de la qualité de l'information. Il ne faut pas oublier qu'un diagnostic, un diagnostic génétique par exemple, ne se justifie que par les avantages que l'on peut en attendre, et il ne faut pas cacher les incertitudes quant à la qualité des résultats obtenus, et à la nécessité qu'ils ont à être interprétés. Dans les démarches diagnostiques, la prise en compte de l'incertitude reste centrale car les tests ne délivrent que des probabilités et non des certitudes, et c'est encore plus marquant pour les maladies évolutives. Par ailleurs, il ne faut pas oublier que les tests, s'ils sont personnels, sont néanmoins probabilistes, ce qui nécessite de passer d'une information issue d'un grand nombre de données obtenues sur des cohortes d'individus très différents, à des données valables pour un individu singulier³⁰⁸.

³⁰⁸ C'est le cas aussi des essais cliniques, généralement considérés comme des preuves expérimentales parce qu'ils sont particulièrement contrôlés, mais dont le résultat est

L'interrogation qui concerne la définition de la vérité – à partir de quand un savoir établit-il qu'une chose est vraie ? - est certes centrale. « Mais, autant elle peut alimenter les réflexions sur sa place et son usage dans les pratiques scientifiques, autant elle ne doit pas être normative et confondue avec celle qui consisterait à chercher si la science atteint « la » vérité sous sa forme désincarnée et surplombante. La vérité qui nous intéresse ici est plutôt définie par l'ensemble des règles et des procédures qui permettent de prononcer ou de faire jouer des énoncés qui sont considérés, à un moment donné, comme vrais par les personnes³⁰⁹. Ces règles et ces procédures s'inscrivent plus largement dans un « régime de vérité », qui recouvre à la fois les types de discours qu'une société accueille et fait fonctionner comme vrais, les techniques et les procédures qui sont valorisées pour l'obtention de la vérité, et le statut de ceux et celles qui ont la charge de dire ce qui fonctionne comme vrai »³¹⁰. La traduction actuelle du collectif de pensée cher à Ludwig Fleck³¹¹, réside dans les consortiums qui se mettent en place et qui reflètent un « style de pensée ». Ce style de pensée est le fait de scientifiques, ce qui pourrait rapprocher cette notion de la notion de paradigme développée par la suite par Thomas Kuhn³¹², mais il n'est pas uniquement le fait des scientifiques puisqu'il est consolidé ou conforté par le contexte social :

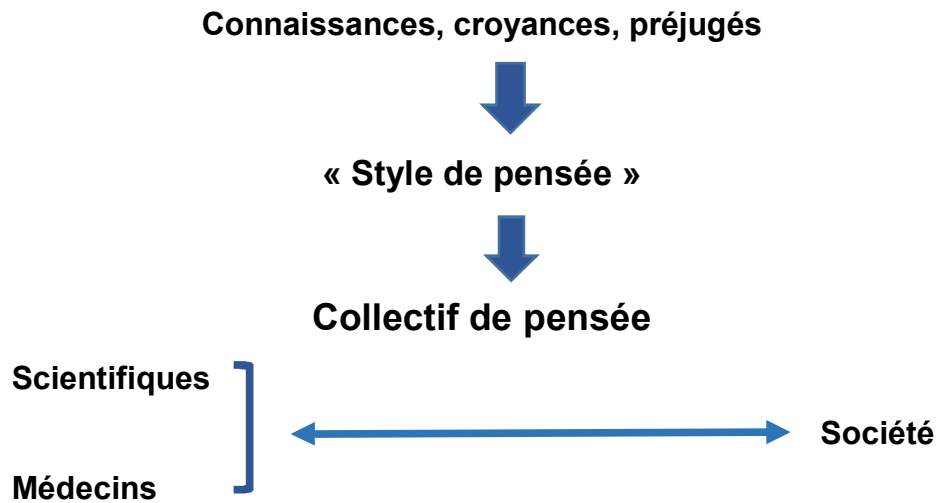
toujours établi statistiquement.

³⁰⁹ Michel Foucault, *Dits et écrits II, 1976-1988* (1994), Paris, Gallimard, 2001.

³¹⁰ Joëlle Vailly, *Naissance d'une politique de la génétique – Dépistage, biomédecine, enjeux sociaux*, Paris, Vrin, 2011, pp. 79-80.

³¹¹ Ludwig Fleck, *Genèse et développement d'un fait scientifique*, Trad. Française, Paris, Les Belles lettres, 2005.

³¹² Thomas Kuhn, *la structure des révolutions scientifiques*, Paris, Flammarion, 2008.



Pour Kuhn, tout se passe dans le monde scientifique et ses paradigmes sont théoriques, la question de la pratique n'intervenant pas. Fleck introduit la composante sociale et pour lui, la pratique a toute son importance, les concepts n'intervenant que dans leur mise en œuvre opérationnelle.

Augmenter le volume des données ne règle pas la question de la causalité, mais ouvre plutôt la porte à l'inexactitude et laisse une impression de désordre, d'une part parce que plus il y a de données, plus il y a des risques d'avoir des données erronées, d'autre part parce qu'on associe des données provenant de sources très diverses, qui ne concordent pas toujours facilement. Cela caractérise une évolution : au lieu de chercher quelque chose d'exact, on cherche plutôt à mettre en évidence une tendance, ce qui conduit à modifier la conception des bases de données. Alors qu'aujourd'hui, elles n'acceptent que des données parfaitement formatées et en principe exactes, demain elles accepteront des données qui ne seront pas enregistrées dans des champs préétablis. On considère en effet que, toutes sources confondues, seules 5% de toutes les données numériques sont structurées.

XI-3 La question de la causalité est-elle pertinente ?

L'exigence d'action est au cœur de la connaissance médicale même si la causalité est un thème majeur de la médecine, l'étiologie étant la recherche de la cause. Elle a pour finalité d'intervenir sur des cas pathogéniques et non sur les symptômes, tout en notant qu'il faut distinguer les mécanismes de la cause. De façon générale, comme le relève Emile Meyerson dans son ouvrage *Identité et réalité*³¹³, la science ne se contente pas de décrire les phénomènes, mais tente d'exhiber leur cause. Ce qui est important, c'est l'analyse du chemin causal, l'identification des passages obligatoires permettant de définir des actions. Les sciences ont, en effet, la particularité d'envisager leur savoir comme un pouvoir-faire. La conception selon laquelle l'événement naturel relève de lois mathématiquement quantifiables tend à isoler les rapports de cause à effet et permet, ce faisant, d'offrir à l'homme la possibilité d'agir et de contrôler précisément ses interventions. On peut ainsi relever que le problème de la causalité peut s'analyser sous différents angles :

- L'inférence, ou relation causale entre un phénomène X et un phénomène Y ;
- La prédiction qui revient à se demander « que va-t-il advenir ? » ;
- L'explication qui vise à comprendre un phénomène et, notamment, à s'interroger sur la question de savoir jusqu'à quelle mesure un phénomène peut être expliqué par des analyses statistiques ;
- Le contrôle qui cherche à savoir ce qui peut être fait face à un phénomène ou une maladie ;
- La modélisation qui vise à utiliser un raisonnement pour construire un modèle représentatif du phénomène.

³¹³ Emile Meyerson, *Identité et réalité*, Félix Alcan, 1908.

Platon dans le *Timée*³¹⁴ énonce le principe de causalité qui stipule que tout fait a une cause et que, dans les mêmes conditions, la même cause produit les mêmes effets. Ce n'est cependant pas du déterminisme absolu qui supposerait une continuité des séries causales dans le temps, alors que l'indéterminisme suppose simplement une pluralité et une discontinuité des chaînes causales. Aristote, dans *La Physique*³¹⁵, distingue plusieurs causes, la cause matérielle, la cause formelle, la cause motrice, la cause finale. Appliqué à la maladie d'Alzheimer, cela donne :

- Cause matérielle : les gens sont atteints de déficiences mnésiques, voire de démences ;
- Cause formelle : on nomme la maladie, maladie neurodégénérative ;
- Cause motrice ou efficiente : la maladie vient d'une altération caractérisée par la cascade amyloïde ;
- Cause finale : la maladie mène à la mort.

De son côté, Ernst MAYR³¹⁶ distingue la causalité proximale de la causalité distale :

- La causalité proximale est de nature réductionniste, c'est du ressort de la biologie fonctionnelle ;
- La causalité distale est de nature historique, c'est du ressort d'une approche évolutionniste.

Découvrir «la cause de» est le but ultime des recherches. En pratique, ces investigations s'attachent à découvrir essentiellement les causes immédiates ou proximales, notamment lorsqu'elles s'effectuent selon des approches réductionnistes. Cependant le

³¹⁴ Platon, *Timée* 28a

³¹⁵ Aristote, *Physique*, Livre II, Chap. 3, 194b – 195a

³¹⁶ Ernst Mayr, « Proximate and ultimate causations », *Biology and Philosophy*, 1993, 8, pp 93_94.

pourquoi des maladies, en particulier dans le domaine des pathologies comportementales et psychiatriques, a très souvent des origines lointaines, voire historiques. S'interroger sur une causalité distale ouvre des perspectives nouvelles, en rapport avec la lente transformation des sociétés. La biologie humaine a des contraintes, certaines issues de l'évolution.

Quand on parle de causalité, on cherche à connaître différentes choses : quelle est la cause d'un phénomène et comment ? Quelle est la part d'un paramètre ou d'un événement ? Quelle différence il y a entre les hommes ? Ainsi, on peut chercher à répondre à deux types de question, « comment ça marche ? » d'une part- c'est la causalité proximale, « pourquoi ça met en jeu tels éléments ? » d'autre part – c'est la causalité distale. Il faut, de plus, examiner les causes des causes, c'est-à-dire les conditions sociales qui accroissent des styles de vie délétères. Prendre en considération les causes distales est une nécessité pertinente dans les sciences de la vie et de la santé. C'est d'ailleurs dont cherche à rendre compte l'équation

Genome * Exposome = Phenome

puisqu'elle tente de regrouper l'ensemble des facteurs, l'exposome visant à agréger les causes proximales et les causes distales.

Enfin, la cause explique l'effet, non parce qu'elle l'implique (logiquement), mais parce qu'elle contribue à le produire (réellement). L'ordre causal n'est pas logique, il est factuel.

La causalité a très longtemps été envisagée sous l'angle du déterminisme. La génétique est au cœur de cette question puisqu'elle est marquée par une tension entre le gène, issu de l'hérédité, et l'environnement qui intervient sur l'expression du gène, voire ses variations accidentelles. Si l'on donne au génome

la place première, « l'humain est déterminé par une nature fixe, bien définie, inaliénable et atemporelle »³¹⁷. On a alors affaire à un déterminisme essentialiste. C'est le cas pour les maladies monogéniques, comme la mucoviscidose par exemple. Par contre, le plus souvent, on ne peut pas parler de certitude causale, mais au mieux de prédisposition génétique. Il ne faut pas confondre prédictabilité et déterminisme, ni assimiler déterminisme et causalité.

On recherche toujours la causalité d'un phénomène, d'une maladie³¹⁸, ne serait-ce que parce que la causalité est une catégorie de l'action – trouver une thérapie efficace par exemple – avant d'être une catégorie de la connaissance. Il faut s'interroger sur la signification philosophique des causes, et notamment s'interroger sur la causalité dans un contexte de multifactorialité³¹⁹, sans se limiter à une relation naïve entre cause et effet, et ne pas oublier qu'une revendication causale est toujours faite dans un contexte. « La cause est alors un facteur qui contribue dans une certaine mesure à l'apparition de l'effet, sans être pour autant un élément déterminant nécessairement l'apparition de l'effet »³²⁰. La cause n'est pas alors que ce qui opère, que ce qui donne des résultats. La cause se rapproche plus du concept grec d'*aition* (αἴτιον)³²¹ que Heidegger, dans son article sur *la question de la technique*³²² définit comme « l'acte dont on répond »³²³ ou « ce qui

³¹⁷ Laurence Perbal, « L'humain, cet être interactivement prédisposé ? Une histoire du déterminisme de l'inné en génétique humaine », Lille, *Revue Mosaïque* N°4, juillet 2010.

³¹⁸ Cause vient du latin *causa*. Le grec « αἴτια » a survécu dans étiologie, la science des causes.

³¹⁹ Qui dit multifactoriel dit à la fois l'intervention de plusieurs gènes et la nécessité d'associer l'environnement de l'organisme.

³²⁰ Raffaella Campaner, « Causalité médicale et déterminants de la santé », dans B. Fantini et L. L. Lambrichs (Dir.), *Histoire de la pensée médicale contemporaine - Evolutions, découvertes, controverses*, Paris, Seuil, 2014, page 27.

³²¹ En grec τὸ αἴτιον désigne la cause ou la responsabilité.

³²² M. Heidegger (1954) « La question de la technique », dans *Essais et conférences*, Paris, Gallimard, 2014.

³²³ Ibid., page 15.

répond d'une autre chose » dans l'apparition d'une chose quelconque, qu'elle soit naturelle ou artificielle. Une cause est ainsi ce qui répond de sa conséquence. Dans de nombreux contextes médicaux, l'idée « même cause, même effet » ne convient pas, tant de nombreuses causes agissent en même temps.

De fait, de nombreuses maladies ont une origine multifactorielle associant des facteurs de prédisposition génétiques, épigénétiques ou environnementaux. C'est un ensemble de facteurs qui provoquent la maladie ou favorisent sa progression sans qu'aucun d'entre eux puisse le faire à lui seul. Ces situations complexes peuvent être beaucoup mieux analysées que dans le passé grâce aux progrès de la génétique et de la biologie, mais aussi parfois de l'imagerie ou d'autres disciplines. Néanmoins, bien souvent, une grande incertitude persiste, d'autant plus regrettable qu'il est toujours préférable d'agir sur la cause que sur les symptômes. Plus un système est complexe, plus il y a de traits qui peuvent varier, plus par conséquent sont nombreuses les combinaisons possibles de ces variations. Lorsque nous définissons une maladie comme génétique, cela ne revient évidemment pas à dire que le gène est le seul facteur causal de la maladie, puisque par principe, il est toujours possible de concevoir une cause environnementale. Cela signifie plutôt que nous résolvons le problème de la sélection causale en faveur des facteurs génétiques : en définissant une maladie comme génétique, nous affirmons que le gène est le facteur causal le plus important – ce qui n'empêche pas de reconnaître l'intervention de facteurs environnementaux

Dans le cas de la maladie d'Alzheimer, la causalité est difficilement saisissable. Ce sont en fait des phénomènes multicausaux dont les déterminations ne sont ni complètement génétiques, ni complètement biologiques, ni complètement socio-

environnementales. Cela explique l'émergence des *Big Data*. Les promoteurs des études d'association (GWAS) soutiennent que ces recherches accélèrent la mise en évidence des causalités de certaines pathologies, permettant ainsi aux individus d'anticiper les risques, et ainsi d'acquérir une certaine autonomie dans la gestion de leur santé, ne serait-ce qu'en terme de prévention. Mais cela peut aussi accentuer les sentiments de vulnérabilité, notamment dans le cas de la maladie d'Alzheimer pour laquelle il n'y a pas d'avancées thérapeutiques significatives validées. Par ailleurs, ne faut-il pas se demander s'il peut ne pas y avoir de cause.

En fait, avec les maladies complexes, on passe d'une causalité linéaire et unique à une causalité multifactorielle et probabiliste où « les causes sont identifiées aux facteurs qui augmentent la probabilité d'apparition d'une pathologie, sans qu'elles la provoquent avec certitude ni la rendent nécessairement plus probable »³²⁴. On ne recherche plus guère la cause des phénomènes, la mise en relation systématique et de type probabiliste et statistique suffit.

XI-4 Causalité et corrélation

Avec les *Big Data*, le point de vue change. Les données permettent d'établir des corrélations qui révèlent des tendances mais qui ne disent pas précisément, ni même obligatoirement, le pourquoi de quelque chose : « Les corrélations statistiques ne vont plus de la cause vers les conséquences, mais remontent des conséquences vers une estimation des causes probables »³²⁵. Toute corrélation n'est pas causale. Tout au plus nous alerte-t-elle sur un fait qui se produit. En médecine, il est fréquent de ne

³²⁴ Rachel A. Ankeny, « Maladie et santé : concepts et représentations », dans B. Fantini & L. Lambrichs (dir.), *Histoire de la pensée médicale contemporaine*, Paris, Seuil, 2014, page 29.

³²⁵ Dominique Cardon, *A quoi rêvent les algorithmes, nos vie à l'heure des Big Data*, Paris Seuil, , 2015, p. 39.

disposer que de corrélations dont la signification causale reste mystérieuse car, si causalité implique nécessairement corrélation, corrélation n'implique pas nécessairement causalité. En exploitant de vastes quantités de données, et non des échantillons, même s'ils sont choisis de façon pertinente, la compréhension des phénomènes change. On renonce en quelque sorte à la recherche d'une causalité pour tirer parti de corrélations.

Dès 1967 est formalisée la distinction entre les modèles de connaissances (c'est-à-dire l'identification de variables et la mise en évidence d'une imputation causale) et les modèles d'aide à la décision, souvent qualifiés de pragmatiques³²⁶. Mais, avec les *Big Data*, un nouveau rapport à la causalité se met en place, la statistique « bayésienne » prenant la place de la statistique « fréquentiste ».

En statistique bayésienne :

- on interprète les probabilités comme un degré de croyance plutôt que comme la fréquence limite d'un phénomène
- on modélise les paramètres du modèle par des lois de probabilité
- on infère des paramètres devenant d'autant plus plausibles à mesure qu'on affine cette distribution de probabilité au fur et à mesure que sont connus de nouveaux résultats d'observation.

À partir d'un ensemble de données, un algorithme apprend son modèle à partir d'un premier sous-ensemble de données et l'utilise pour prédire ce qui se passera pour un deuxième sous-ensemble de données. La prédiction n'est en fait qu'une estimation statistique et ne présente aucune certitude mais, selon Isabelle Drouet, « le bayésianisme se présente comme la meilleure façon de représenter

³²⁶ D. Schwartz et J. Lellouch, « Explanatory and pragmatic attitudes in therapeutic trial », *Journal of Chronic Disease*, 20, p.637, 1967.

l'incertitude dans un contexte décisionnel »³²⁷. La statistique bayésienne est une approche statistique fondée sur l'inférence bayésienne, une méthode d'inférence permettant de déduire la probabilité d'un événement à partir de celles d'autres événements déjà évalués. Elle est surtout utilisée lorsque l'on n'a que de petits échantillons.

Renoncer à la recherche d'une causalité pour tirer parti de corrélations n'est cependant pas accepté par l'ensemble des acteurs. Ainsi Russo et Williamson³²⁸, en s'appuyant sur le lien entre le tabac et les maladies cardiaques, soutiennent que l'explication causale est plurielle et qu'elle doit être soutenue à la fois par un mécanisme causal et par une évidence statistique, défendant ainsi la complémentarité des approches corrélacionistes et mécanistes. Leur thèse est que pour établir des revendications causales en médecine, on a besoin, d'une part d'établir des faits sur ce qui crée des différences, d'autre part de connaître un mécanisme qui crée l'événement. Disposer seulement d'une corrélation ne permet jamais d'affirmer à soi seul qu'un premier fait en cause un second. Une telle corrélation peut faciliter un diagnostic, une prédiction, voire une action, mais elle ne suffit pas à expliquer.

En fait, les volumes considérables de données amènent à se demander s'il faut privilégier la recherche de causes de phénomènes ou la recherche de corrélations utiles. Les corrélations ne donnent aucune certitude, uniquement de la probabilité. Elles ne renseignent donc pas sur le fonctionnement réel d'un système, elles ne permettent que d'identifier une (des) donnée(s) utile(s) et elles

³²⁷ Isabelle Drouet, « Le bayésianisme : éléments de définition et mutations récentes », in I. Drouet (dir.) *Le bayésianisme aujourd'hui*, Paris, Editions Matériologiques, 2016, page 14.

³²⁸ F. Russo et J. Williamson, « Interpreting Causality in Health Sciences », *International Studies in the Philosophy of Science*, 2007, 21 (2), pp : 157-170.

peuvent n'être le résultat que d'une pure coïncidence³²⁹, même si on peut considérer que si une corrélation est forte, la probabilité qu'il existe un lien augmente. De fait il peut y avoir confusion. Une corrélation peut ne pas être due à une connexion causale entre deux paramètres, mais plutôt à l'intervention d'un troisième paramètre qui agit sur les deux premiers. Dans ce cas, la corrélation n'implique pas la causalité.

Par ailleurs, la disparition de la moyenne marque la différence essentielle entre les *Big Data* et les statistiques. La statistique vise la moyenne alors que l'objectif des *Big Data* est l'exception. Elles débusquent les « signaux faibles ».

Cela conduit aussi à s'interroger sur l'approche pour appréhender un phénomène. Faut-il partir d'une hypothèse ou l'utilisation de données variées et multiples suffit-elle ? La question a été posée. Certes, pour mener une analyse par corrélation, les hypothèses ne sont plus indispensables, mais pour autant est-ce que « le déluge de données rend la méthode scientifique obsolète » comme le prétend Chris Anderson³³⁰ qui proclamait la fin de la théorie et n'hésitait pas à écrire : « la connaissance issue des *Big Data* sera produite à partir de statistiques agnostiques » ? Dans son article, qui a fait grand bruit, ce journaliste annonçait la « fin de la théorie », argumentant que les *Big Data* permettent d'avoir des corrélations sans avoir à se préoccuper d'un modèle leur donnant une explication. C'est sans doute aller vite en besogne, ne serait-ce, et cela a déjà été relevé, que parce qu'il y a une façon de choisir les données, pour des raisons pratiques - des données facilement disponibles -, ou des raisons économiques - un faible coût d'obtention des données (ce qui peut expliquer le changement de

³²⁹ Nassim Nicholas Taleb, *Le Hasard sauvage. Comment la chance nous trompe*, Paris, Les Belles Lettres, 2009.

³³⁰ Chris Anderson, « The End of Theory : The Data Deluge Makes the Scientific Method Obsolete », *Wired*, juin 2008.

la prise en compte des données génétiques, le séquençage du génome entier devenant accessible). Face à l'avalanche de données se pose néanmoins la question des démarches des scientifiques et des médecins. Une démarche « *data-driven* » sans hypothèse préalable pourrait sembler favorisée, tant de nombreuses publications ne concernent que l'acquisition de données. Cependant on peut considérer que les circonstances dans lesquelles les données sont collectées affectent profondément les voies à travers lesquelles elles sont interprétées et que les outils interdisciplinaires, associés aux modèles mathématiques et statistiques, ont un rôle essentiel. La visée d'une recherche et d'une pratique sans hypothèses pourrait ainsi être illusoire. Les données ne font pas sens par elles-mêmes, le savoir fait nécessairement intervenir le regard du chercheur ou du clinicien qui interroge les bases de données à partir des connaissances dont il dispose déjà. La technique aide le médecin à dire ce qu'il sait déjà. On peut rappeler ce qu'écrivait Georges Canguilhem : « la médecine est la science des limites des pouvoirs que les autres sciences prétendent lui conférer »³³¹.

Ainsi, si la biomédecine apparaît plutôt comme pilotée par les données, et par là par des technologies, et leur analyse, que par des hypothèses, il ne faut pas oublier que le choix des données que l'on accumule, des données génomiques ou des données d'expression par exemple, suppose que l'on ait retenu un niveau d'analyse que l'on considère pertinent, et cette pertinence s'appuie sur une hypothèse théorique, le choix que tout est génétiquement déterminé ou que l'expression des gènes est fortement associée à un état ou à un environnement par exemple. Différents niveaux d'analyse coexistent et, c'est le cas pour la maladie d'Alzheimer, les syndromes cliniques peuvent être mieux déterminés que les agents

³³¹ Georges Canguilhem, « Le statut épistémologique de la médecine », in *Etudes d'histoire et de philosophie des sciences concernant les vivants et la vie*, Paris, Vrin, 1994

causaux. Qui dit *Big Data* dit accès à des connaissances, mais ces connaissances ne peuvent être associées qu'à des théories ou des hypothèses. Nous avons déjà relevé, en nous référant notamment aux travaux de Ronan Le Roux³³², que la recherche en médecine, particulièrement en médecine de précision, ne se faisait pas sans *a priori* théoriques. En fait, comprendre la façon dont les données sont produites est essentiel et il ne faut pas oublier que, pour que les données acquièrent et conservent toute leur valeur, il faut connaître leurs conditions de production, informations nécessaires pour être en mesure de comparer les technologies. C'est le cas, par exemple pour les mutations ponctuelles. Les SNPs identifiés par les études d'association sont compris comme des marqueurs biologiques qui identifient une région du génome humain supposée accroître le risque pour la maladie sous investigation, mais les SNPs n'indiquent ni ne précisent une cause génétique de la maladie.

La prédiction n'est jamais qu'une extrapolation du passé. Or, si les conditions changent, les mêmes relations de causalité n'induiront pas les mêmes conséquences. La question de la précision technologique se pose. Dans quelle mesure les outils disponibles permettent-ils de détecter avec une bonne fiabilité la présence de mutations ? Par ailleurs la causalité est difficile à prouver et bien souvent elle ne peut l'être qu'à un certain niveau élevé de probabilité³³³.

On aboutit à un paradoxe. Les nouvelles connaissances, la prise en compte des différents niveaux d'incertitude et de l'aspect essentiellement probabiliste des processus, entraînent une scientification de la médecine, qui peut entraîner une

³³² Ronan Le Roux, « Nanotechnologies for Cancer Detection, Instruments, Paradigms and the Two Dogmas of Molecular Diagnostics », article à paraître.

³³³ Judea Pearl, *Causality: Models, Reasoning and Inference*, Cambridge, Cambridge University Press, 2009.

molécularisation du patient. Parler de médecine personnalisée ne fait qu'amplifier cette position et conduit à s'interroger sur la signification des biomarqueurs, qui peuvent témoigner d'un processus normal ou anormal ou de la présence d'un trouble ou d'une maladie, mais qui ne présentent généralement aucun lien de causalité direct avec une maladie et pour lesquels on ne peut définir une origine.

Il faut donc s'interroger à nouveau sur le concept de biomarqueurs. Ils ont changé la façon d'appréhender les phénomènes puisqu'au mode d'explication fondé sur la mise en évidence de mécanismes moléculaires causaux, les biomarqueurs ont amené une information sans qu'il y ait de liens de causalité directs entre un biomarqueur et une maladie. Ils jouent cependant un rôle privilégié³³⁴. Rappelons quelques caractéristiques déjà évoquées :

- Les biomarqueurs sont des molécules biologiques présentes dans le sang, dans des liquides corporels, des tissus ou dans des tumeurs, qui peuvent témoigner d'un processus, normal ou anormal, ou de la présence d'un trouble ou d'une maladie.
- Le biomarqueur est une altération, qui a une signification biologique, mais il ne renvoie pas une situation préalable non altérée.
- Qui plus est, les biomarqueurs peuvent être très divers et être constitués d'une entité complexe impliquant la mesure simultanée d'un nombre important de composants, composants qui, par nature, sont variables quel que soit l'état de l'individu.

³³⁴ J. M. Meyer , G. S. Ginsburg, « The path of personalized medicine », *Current Opinion in Chemical Biology*, 2002, 6(4), pp. 434-438.

Que les biomarqueurs soient particulièrement utiles en recherche ou pour le développement de médicaments ou de nouvelles voies thérapeutiques, c'est largement admis, mais que ces biomarqueurs soient utilisés en médecine, que ce soit à des fins de diagnostic ou de personnalisation de traitements, cela ne va pas de soi car le biomarqueur n'est pas plus un symptôme que le signe d'un état pathologique.

Il faut sans doute tout repenser sous l'angle des interactions que tout être entretient avec son milieu, c'est-à-dire « en situation » ou, si l'on préfère, « en contexte », un être ne pouvant pas être traité comme une entité isolée. Cela suppose d'associer des modèles de connaissance, le plus souvent disjoints, à savoir celui des sciences dures et de la technologie et celui des sciences humaines. On peut, à ce stade, rappeler l'analyse faite par C. Ginzburg³³⁵. D'un côté, il rappelle ce qu'est le « paradigme galiléen ». Dans la tradition galiléenne, seuls comptent les figures, les nombres et les mouvements, seuls les phénomènes quantifiables, transposables dans le langage des mathématiques, peuvent être objets de science, et on ne peut pas parler de ce qui est individuel. De l'autre il introduit ce qu'il appelle le « paradigme indiciaire ». Il caractérise un autre versant de la connaissance, celle qui s'intéresse à des indices marginaux, qui cherche des traces et des indices, avec une perspective individualisante, mettant à profit des analogies, des comparaisons. On peut considérer que la médecine occupe une position quelque peu intermédiaire entre ces deux approches extrêmes puisque les données mesurables, certes centrales et nécessaires, ne constituent qu'une partie des connaissances utilisées par le médecin. Pour le médecin, les deux modes de connaissance s'avèrent indispensables et complémentaires. Les données mesurables donnent des informations sur l'état de la

³³⁵ C. Ginzburg, *Mythes, emblèmes, traces. Morphologie et histoire*, Paris, Verdier, 2010.

personne vue statistiquement, alors que les autres données sont plus tournées vers la personne en tant qu'être singulier, notamment vers la personne dans son environnement. Le développement des maladies chroniques a probablement amplifié la nécessité de cette complémentarité.

En abordant la question des biomarqueurs et de leur nature (molécules, réseaux, interactions, etc.), nous pouvons nous interroger et nous demander si ces biomarqueurs sont des objets donnés ou des objets construits. À n'en pas douter le biomarqueur est un objet de la biologie et de la médecine. Il existe, mais, nous référant à Gilbert Simondon, la question de son mode d'existence et la question de savoir comment le caractériser se posent.

Ce qui a précédé montre que le biomarqueur, à des exceptions singulières près³³⁶, est un objet complexe. Ce n'est pas une molécule isolée, c'est un réseau d'interactions, c'est un algorithme. Ainsi, dans le cas de la maladie d'Alzheimer, les biomarqueurs sont très divers, ce peuvent être des biomarqueurs biologiques (comme la baisse de concentration de l'Abéta42 ou l'augmentation de concentration des protéines Tau), des biomarqueurs d'imagerie (IRM, PET-Scan), des biomarqueurs génétiques (comme APOE4) ou encore des marqueurs neurophysiologiques ou cognitifs. Mais ces biomarqueurs pris isolément ne peuvent apporter que des informations très partielles. En ce sens, un biomarqueur est en fait un ensemble de biomarqueurs reliés par des modèles, une véritable signature de la pathologie, et on ne peut pas lui enlever une de ses caractéristiques sans lui modifier sa nature. On peut ajouter que le biomarqueur, en plus d'être complexe, est :

- d'abord un objet de connaissance et non une représentation directe de la réalité ;

³³⁶ Un exemple d'exception singulière est le cas où le biomarqueur est une molécule dont la concentration dans un milieu biologique constitue le marqueur.

- ensuite un objet entouré d'incertitude ;
- enfin un objet dont on n'a pas la maîtrise totale, ce qui le rend non synthétisable.

Le biomarqueur peut avoir plusieurs voies explicatives, tout en n'étant jamais complètement donné. On peut considérer qu'il laisse ouvertes des zones d'incertitudes, voire d'imprévisibilité. Ce n'est pas, à proprement parler, un objet réel. On pourrait dire que c'est une inconnue sans définition, toujours en devenir !

XI-5 Les mécanismes

L'explication causale en médecine a évolué. En se limitant à l'époque moderne, on peut considérer qu'elle a démarré avec Claude Bernard³³⁷ qui affirmait que l'objectif des recherches expérimentales était d'identifier les « causes prochaines ou immédiates » des maladies.

Cette approche était essentiellement centrée sur les maladies infectieuses, avec une vision mono-causale : un micro-organisme, une maladie. Dans les années 1950, la notion de causalité multifactorielle est apparue, notamment en raison du développement des recherches sur les maladies cardio-vasculaires et les cancers. Ainsi que l'a relevé Anne Fagot-Largeault, « le même syndrome clinique pouvait être dû à plusieurs agents différents [...] ; le même agent causal pouvait induire des syndromes différents [...] ; l'agent « principal » d'un syndrome donné pouvait varier selon les âges, les saisons, les groupes de population »³³⁸ faisant apparaître que ce qui est nommé cause est en fait un facteur qui contribue dans une certaine mesure à la manifestation de l'effet et non pas un élément nécessaire à l'apparition de l'effet. Les

³³⁷ Claude Bernard, *Introduction à l'étude de la médecine expérimentale*, Paris, Editions Flammarion, « les livres qui ont changé le monde » pour le journal Le Monde, 2010

³³⁸ A. Fagot-Largeault, *Médecine et philosophie*, Paris, PUF, 2010, page 91.

statistiques sont alors intervenues et Austin B. Hill³³⁹ a proposé de retenir une association statistique comme association causale sous plusieurs critères, notamment :

- il faut que la relation entre la variable supposée être la cause et celle supposée être l'effet soit forte ;
- il faut que le facteur supposé être la cause précède dans le temps le facteur supposé être l'effet ;
- il faut une base théorique, rationnelle ;
- il faut une certaine spécificité ;
- il faut une relation de proportionnalité linéaire dose-effet.

A. B. Hill considère ainsi que le soupçon de causalité est d'autant plus intense que l'association statistique est plus forte. Il y ajoute la nécessité d'une plausibilité chronologique et de données biologiques ou cliniques pouvant justifier une certaine spécificité. Mais la satisfaction des critères de Hill ne permet pas d'inférer qu'une association statistique est une relation causale. Elle permet plutôt « de considérer l'hypothèse causale d'une association statistique, si elle est vraie, comme la meilleure explication possible »³⁴⁰.

Puis, dans les années 1960, il a été considéré qu'une pathologie pouvait être due à différents groupes de facteurs causals, chaque groupe, pris dans son ensemble, pouvant être suffisant pour qu'apparaisse la maladie³⁴¹. Enfin, la causalité en termes de causes nécessaires et suffisantes a progressivement laissé la place à une notion de causalité multifactorielle caractérisée par une vision probabiliste. Apparaissait ainsi une identification des causes aux facteurs qui augmentaient la probabilité d'apparition

³³⁹ A. B. Hill, « The Environment and Disease : Association or Causation ? », *Proceedings of the Royal Society of Medicine*, Vol. 58, 1965, pp : 295-300.

³⁴⁰ A. Laplège, P. Bizouarn & J. Coste (Dir.), *De Galton à Rothman, Les grands textes de l'épidémiologie au XX^e siècle*, Paris, Hermann, 2011, page 111.

³⁴¹ K. Rothman, *Modern Epidemiology*, Boston, 1986.

d'une pathologie, sans qu'elles la provoquent avec certitude ni la rendent nécessairement plus probable. L'apparition d'une maladie est alors le résultat d'un ensemble de facteurs causals multiples, impliqués dans un mécanisme complexe constitué de nombreux facteurs et de leurs interactions mutuelles³⁴².

Il faut bien distinguer les mécanismes des causes. Qui dit mécanisme dit système complexe et interactions nombreuses entre différents éléments du système, mais, dans le domaine des sciences du vivant, mécanisme peut avoir plusieurs sens³⁴³ :

- Mécanisme au sens du mécanicisme, l'organisme vivant étant considéré comme une machine pouvant être complètement déterminée en termes de structures et d'interactions. Il lui correspond une approche réductionniste et les premiers succès de la biologie moléculaire ont largement consolidé cette approche.
- Mécanisme au sens machinisme, le système vivant étant décrit en termes mécaniques, en termes explicites à l'époque de Descartes, sous un habillage thermodynamique au XIX^e siècle, sous la forme d'un ordinateur définissant son propre programme et traitant l'information en fonction d'un environnement à la fin du XX^e siècle.
- Mécanisme au sens de mécanisme causal, qui vise à mettre en évidence les processus causaux donnant naissance à un phénomène particulier, sens qui devient premier à l'époque actuelle.

Si le mécanisme au sens de mécanisme causal est prédominant dans la biologie d'aujourd'hui, ne doit-on pas cependant s'interroger de savoir si ces mécanismes sont des mécanismes réels,

³⁴² P. Thagard, *How Scientists Explain Disease*, Princeton (NJ), 1999.

³⁴³ Daniel J. Nicholson, « The concept of mechanism in biology », *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43 (2012) pp. 152-163.

autonomes, ou simplement des modèles heuristiques qui facilitent la compréhension ou l'explication de phénomènes ?

Aujourd'hui, l'accent est mis sur le caractère complexe des pathologies et les mécanismes possibles sont analysés à plusieurs niveaux (molécules, cellules, tissus, organes, ...), organisés hiérarchiquement et entre lesquels peuvent exister plusieurs types d'interactions. Au cours du XX^e siècle, les progrès de la biologie moléculaire et de la biochimie ont permis des avancées décisives mais, ainsi que Michel Morange l'a souligné³⁴⁴, les connaissances issues de niveaux de complexité inférieurs ne peuvent pas expliquer certaines propriétés à des niveaux supérieurs de complexité. Il reste beaucoup à faire pour comprendre comment des facteurs génétiques interagissent avec des facteurs comme les facteurs socio-économiques et psychosociaux et comment, par leurs interactions, ces facteurs peuvent faire partie des causes d'une pathologie.

Le vrai défi est d'avoir suffisamment de connaissances pour justifier d'une action, même dans le cas d'une connaissance imparfaite. Cela aboutit aux entreprises de modélisation qui impliquent certains degrés de simplification et d'abstraction. Il en résulte, d'une part que les modèles sont faillibles, d'autre part qu'on peut construire différents modèles pour expliquer le même phénomène.

Cela conduit à une mathématisation de la médecine et des sciences de la vie, mathématisation qui se heurte néanmoins à un certain nombre d'obstacles :

- Le premier est d'ordre philosophique et pose le problème du déterminisme ;

³⁴⁴

Michel Morange, « The Gene : Between Holism and Generalism », in M. Van Regenmortel et D. Hull (dir.), *Promises and Limits of Reductionism in the Health Sciences*, Chichester, 2002, pp : 179-187.

- Le deuxième est d'ordre éthique, les essais cliniques randomisés supposant qu'une population témoin ne dispose que d'un traitement moins performant, voire d'un placebo ;
- Le troisième est d'ordre biologique avec le problème de la complexité et de la variabilité biologique³⁴⁵.

Les algorithmes prédictifs ont pour objectif de modéliser des comportements et donc d'anticiper les actions futures, en neutralisant l'incertitude au maximum. Il faut pour cela rassembler massivement des données, les traiter pour produire des connaissances, et, dans la mesure du possible, en déduire des actions³⁴⁶. Quant aux traitements, ils sont fondés sur des statistiques. En fait, la personne est une mosaïque de caractéristiques biologiques. On a tendance à aller vers une formalisation mathématique par la définition de réseaux.

Un modèle n'est cependant pas un mécanisme. Ce peut simplement être un modèle descriptif, parfaitement utile pour faire de la prédiction. C'est pourquoi est intervenue la notion de fonction biologique, un mécanisme étant, selon William Bechtel³⁴⁷, une structure réalisant une fonction de par ses composants, leurs opérations et leur organisation. Si l'on s'intéresse à une fonction biologique, on peut relever qu'un organisme vivant est un réseau organisé de processus de production, de transformation et de régénération. L'organisme maintient son identité dans le temps en régulant, réparant et régénérant ses divers composants et l'organisme, comme un tout, influence le comportement de ses parties. La biologie des systèmes est clairement une approche

³⁴⁵ La statistique peut apporter une réponse à la variabilité biologique, en considérant que la moyenne renvoie au collectif, la variance réintroduisant l'individuel.

³⁴⁶ Antoinette Rouvroy, *Politique des algorithmes*, Paris, La Découverte, 2013.

³⁴⁷ William Bechtel and Adele Abrahamsen, « Explanation : A machanist alternative », *Studies in the History and Philosopshy of the Biological and Biomedical Sciences*, 36 (2005), 421-441.

intégrative et interdisciplinaire. D'un côté, elle utilise la quantité d'informations disponibles, tant au niveau moléculaire qu'au niveau cellulaire, en génomique, protéomique ou métabolomique, ce qui appuie l'approche moléculo-mécaniste, de l'autre, elle s'appuie sur des modèles mathématiques permettant d'approcher le fonctionnement dynamique du système. Les équations mathématiques sont, de fait, le meilleur outil pour représenter le comportement dynamique de systèmes complexes, notamment par le biais des équations différentielles, si seul le temps est pris en compte, ou des équations aux dérivées partielles si le temps et l'espace, et donc la localisation, sont retenus. Elles peuvent, par exemple, aboutir à la concentration d'un métabolite suite à des interactions biochimiques ou donner le niveau de transcription d'un gène régulé par d'autres gènes. En biologie des systèmes, deux approches peuvent être envisagées :

- une approche *bottom-up* qui part d'information sur un nombre limité de composés et cherche à expliquer, ou à prédire, les effets résultant de leurs interactions,
- une approche *top-down* qui part du comportement global d'un système et cherche à l'expliquer par la fonctionnalité de différentes parties du système.

On peut aussi se demander si, comme le suggère John Collier³⁴⁸, la connexion causale est une connexion informationnelle, auquel cas les mécanismes seraient alors les canaux à travers lesquels circule l'information.

³⁴⁸ John Collier, « Information, causation and computation », in *Information and Computation : Essays on Scientific and Philosophical Understanding of Foundations of Information and Computation*, (ed. G. D. Crnkovic and M. Burgin) Singapore, World Scientific, 2011.

XI-6 De la sollicitation à l'opportunité : dispositions et aptitudes

Avec les études d'association, des investigations systématiques de la totalité du génome pour chercher des régions susceptibles d'être associées significativement à une pathologie, on a la conviction que, si des gènes sont explicatifs, ils seront détectés. Ainsi, la causalité génétique est couramment utilisée mais on en a montré les limites. On peut en plus se demander si on peut parler de causalité lorsqu'on ne peut pas expliquer les mécanismes mis en jeu. La technologie pourrait y aider. Par exemple, un progrès significatif dans la qualité des mesures, tel que l'annoncent les nanotechnologies, réduirait considérablement le bruit instrumental au point de laisser apparaître en pleine lumière les mesures faites en relation avec le protocole expérimental.

Les technologies, et les données qu'elles permettent d'accumuler et de traiter, font apparaître la notion de facteurs de risque, même s'il faut rappeler l'importance de l'incertitude et de l'aspect « constructiviste » des *Big Data*. Les facteurs de risque ne sont pas nécessairement des variables causales. Par nature, le statut de facteur de risque impose une démonstration statistique effectuée à partir de constatations empiriques, sans mécanismes de causalité. Mais, selon Anne Fagot-Largeault, « la notion de facteur de risque, c'est-à-dire de facteur qui a le pouvoir d'augmenter la probabilité d'occurrence de l'effet, reflète la notion intuitive d'efficacité, et la mesure du risque est une mesure de l'efficacité de la cause »³⁴⁹. Par ailleurs, en toute rigueur, on ne peut parler de « risque » que pour des mesures qui donnent une probabilité, c'est-à-dire un nombre compris entre 0 et 1, ce qui revient à estimer un risque individuel par le risque moyen de nombreux autres individus présumés identiquement représentatifs,

³⁴⁹ Anne Fagot-Largeault, *Médecine et Philosophie*, Paris, PUF, 2010, page 78.

alors que pour un individu, le risque de contracter une maladie ne prend que deux valeurs, 0 ou 1. On pourrait également introduire la notion de risque relatif qui mesure le risque de survenance d'un événement dans un groupe par rapport à un autre.

Ainsi si le risque d'avoir la maladie d'Alzheimer est $R_1 = 5\%$ (donnée arbitraire) chez des personnes de 70 à 75 ans et de $R_2 = 10\%$ chez les personnes de 80 à 85 ans, le risque relatif $RR = R_2/R_1 = 10/5 = 2$ dit que le risque d'avoir la maladie d'Alzheimer est deux fois plus élevé chez les personnes âgées de 80 à 85 ans que chez les personnes âgées de 70 à 75 ans. Dans cet exemple, le risque relatif est la mesure du rôle étiologique du facteur de risque « âge de la personne ».

Enfin, un facteur de risque rentre dans un processus dynamique, dans une certaine continuité.

La notion de facteur de risque est apparue à partir des années 1950 suite à l'application des statistiques à l'épidémiologie ainsi que l'a présenté Elodie Giroux dans sa thèse³⁵⁰. L'épidémiologie des facteurs de risque admet le principe d'une causalité multifactorielle à la source des pathologies et se focalise surtout sur les maladies chroniques. Elle est fondamentalement populationnelle, et associe variables individuelles et biologiques et variables environnementales et sociales. Un facteur de risque désigne une variable individuelle associée à l'augmentation de probabilité de survenue d'une maladie spécifique. Alors qu'une cause est suffisante pour expliquer une maladie, un facteur de risque n'est qu'un caractère contributif qui rend la maladie plus probable mais pas pour autant réelle. Sa détermination repose d'ailleurs sur un paradoxe puisqu'il est établi à partir de la pathologie, ce qui revient à utiliser *a priori* un diagnostic rétrospectif.

³⁵⁰ GIROUX Elodie, *Epidémiologie des facteurs de risque : Genèse d'une nouvelle approche de la maladie*, Thèse soutenue le 8 décembre 2006 à l'Université Paris I Panthéon-Sorbonne, Villeneuve d'Ascq, Diffusion ANRT (Atelier national de reproduction des thèses).

Néanmoins, un nouveau changement de paradigme peut émerger de cette acquisition et de ces traitements d'un nombre toujours plus importants de données, variables ou paramètres. Tout au long de la vie d'un individu, le métabolisme subit des évolutions, évolutions qui se traduisent par des dommages. Ces dommages peuvent entraîner à terme – l'importance que prennent les maladies chroniques le démontre amplement – des pathologies difficiles, voire impossibles actuellement, à soigner. Plutôt que de développer des thérapies de soins, en se focalisant sur les pathologies, ne faut-il pas mieux développer une stratégie de « maintenance », au sens de la maintenance industrielle ? Cette métaphore peut surprendre en laissant supposer qu'elle nous ramène vers une conception mécaniste du vivant. En fait, il ne s'agit pas d'un retour à Descartes avec une vision mécaniste du vivant conçu comme une machine qui peut être complètement expliquée en termes de structures et d'interactions entre ses parties, même si les succès de la biologie moléculaire ont consolidé le « mécanicisme » et si le développement de la biologie de synthèse va dans ce sens. Il s'agit plutôt de considérer les mécanismes causals comme des modèles heuristiques qui facilitent la compréhension des phénomènes et de considérer le concept de fonction biologique, l'organisme vivant étant vu comme un ensemble intégré qui maintient son identité dans le temps par des régulations, des réparations et de la régénération de ses composants et parties³⁵¹. La fonction est omniprésente dans les sciences de la vie. Dans le cas d'une machine, une fonction est une condition nécessaire à la réalisation d'un effet, tandis qu'un biologiste ou un médecin ne s'intéresse pas seulement à l'effet réel d'un processus, « il ne s'intéresse pas seulement à ce qu'il fait, mais aussi à ce qu'il est censé faire »³⁵². Nous pouvons ajouter, comme l'écrit Henri Atlan, qu'« entre les

³⁵¹ Daniel J. Nicholson, « The concept of mechanism in biology », *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43 (2012), pp : 152-163.

³⁵² Jean Gayon, Les biologistes ont-ils besoin du concept de fonction ? Perspective philosophique, C. R. Palevol, 5, 2006, page 482.

gènes et l'environnement, l'organisme contrôle l'activité des gènes autant que ceux-ci contrôlent l'organisme »³⁵³ ou que l'organisme se construit lui-même en même temps qu'il construit son environnement, ce qui n'est pas le cas d'une machine.

Ce peut être l'occasion de mettre en tension le « *cure* » et le « *care* ». Ce sont deux stratégies pour lutter contre une maladie telle que la maladie d'Alzheimer, le *cure* visant le traitement médical, le *care* proposant des stratégies alternatives, y compris sous forme de soutien de la personne malade. Ces deux stratégies sont au cœur de l'activité du médecin puisque, si son rôle premier est de porter un diagnostic et de proposer des thérapeutiques, son rôle est également d'accompagner le patient. Au soin médical (*cure*) qui vise à la guérison et qui est souvent partiel et dissociatif, prenant en compte élément par élément, et qui revient à soigner quelque chose, mais qui a une réelle dimension scientifique et technique, est associé un soin plus global et intégratif (*care*) qui présente une dimension relationnelle qui vise au bien-être de la personne en réponse à sa vulnérabilité. Le *care* désigne ainsi une capacité à percevoir l'autre dans sa singularité tout en prenant soin de lui. D'un côté le *care* relève de la sollicitude, de l'autre, il relève d'un travail effectif, la personne étant l'acteur de sa santé tandis que le médecin ne cache pas son impuissance à guérir.

Ainsi que cela a été relevé, les lésions neuropathologiques de la maladie d'Alzheimer se développent bien avant l'apparition des signes cliniques, mais certains assurent une maintenance contre la maladie à l'aide de processus cognitifs et neuronaux de compensation, tandis que d'autres disposent d'un niveau élevé de la protéine REST qui permet de résister à la maladie d'Alzheimer³⁵⁴.

³⁵³ Henri Atlan, *Le vivant post-génomique, ou qu'est-ce que l'auto-organisation ?*, Paris, Odile Jacob, 2011, page 51.

³⁵⁴ T. Lu & al, « REST and stress resistance in ageing and Alzheimer's disease », *Nature*, 2014, 507 (7493), p.448-454.

Ces deux exemples illustrent une stratégie qui consiste à renforcer les systèmes de défense du cerveau.

Mais la complémentarité entre le *cure* et le *care* permet d'inverser le processus de compréhension de la maladie. Classiquement, on repère des symptômes, on découvre ensuite les lésions responsables et on recherche les processus explicatifs de la maladie. On peut envisager l'approche inverse qui consiste, non pas à partir des symptômes, mais à « maintenir » la personne pour éviter le processus déclenchant la maladie. Se pose alors la question de l'origine dont nous reparlerons dans le chapitre suivant. Mais le *care* et le *cure* peuvent être vus comme deux pôles s'articulant, de manière parfois souple, parfois rigide, selon les moments et les situations.

Parler de maintenance, c'est ouvrir des perspectives de prévention en proposant aux personnes à risque, détectées notamment par des diagnostics précoces alors qu'il n'y a pas encore de symptômes, des mesures médicamenteuses ou des règles hygiéno-diététiques susceptibles d'éviter, à tout le moins de retarder, l'arrivée de la maladie.

Ainsi, face à la maladie d'Alzheimer, un certain nombre d'actions d'anticipation permettent de retarder, à défaut d'éviter, la maladie. C'est important car il faut rappeler que, selon les experts, retarder l'apparition des symptômes de la maladie d'Alzheimer de cinq ans diminuerait de moitié le nombre de cas dans la population³⁵⁵. Pour reprendre la notion de normativité défendue par Georges Canguilhem, on pourrait considérer qu'un sujet atteint de la maladie d'Alzheimer, mais en phase asymptomatique, c'est-à-dire en phase silencieuse sans le moindre signe clinique, est un

³⁵⁵ La maladie d'Alzheimer survient dans la majorité des cas après 80 ans et met dix ans ou plus à s'installer, si bien que de nombreuses personnes atteintes mourraient avant de présenter les premiers symptômes de la maladie.

sujet qui a encore la possibilité d'une normativité, c'est-à-dire qui a encore la possibilité à changer de norme. C'est le sens que l'on pourrait donner à une stratégie de maintenance.

D'après l'étude coordonnée par Gill Livingston de l'*University College of London*³⁵⁶ sur la prévention des démences et le care, on peut relever certains points :

- La formation et les études longues favoriseraient la constitution et le maintien de la réserve cognitive, c'est-à-dire du nombre de neurones et de leurs connexions.
- Les relations sociales évitent l'isolement des personnes et leur renfermement sur elles-mêmes. En ce sens, il faut surveiller la perte d'audition qui pousse les personnes à l'isolement et prévenir toute dépression.
- Tout ce qui maintient une bonne santé cardio-vasculaire, en protégeant les vaisseaux, préserve le cerveau. Il faut donc éviter l'hypertension et l'excès de cholestérol par une hygiène de vie adaptée, et éviter de fumer.
- Pratiquer une activité physique régulière est toujours positif.

Mais face à une maladie complexe comme la maladie d'Alzheimer, les effets réels de ces différentes actions sont difficiles à évaluer. Il faudrait mener une étude sur une dizaine d'années et des dizaines de milliers de personnes pour obtenir des résultats clairs et significatifs, ce qui n'est ni facile à financer, ni facile à organiser. Ces actions agissent tout au long de la vie, dès l'enfance grâce à des études prolongées jusqu'à un âge avancé en maintenant des liens sociaux et en continuant des activités physiques régulières, en passant par l'âge adulte en ayant une bonne hygiène de vie.

³⁵⁶ Gill Livingston and al., « Dementia prevention, intervention, and care », *The Lancet*, 2017 Jul 19.

Le schéma présenté ci-après, tiré de l'étude de Gill Livingston et de ses collègues, résume, en les chiffrant, les actions qu'on peut avoir sur différents facteurs de risque. La conclusion qu'en tirent les auteurs est que 35 % des cas de démence, dont la maladie d'Alzheimer, pourraient être théoriquement évités :

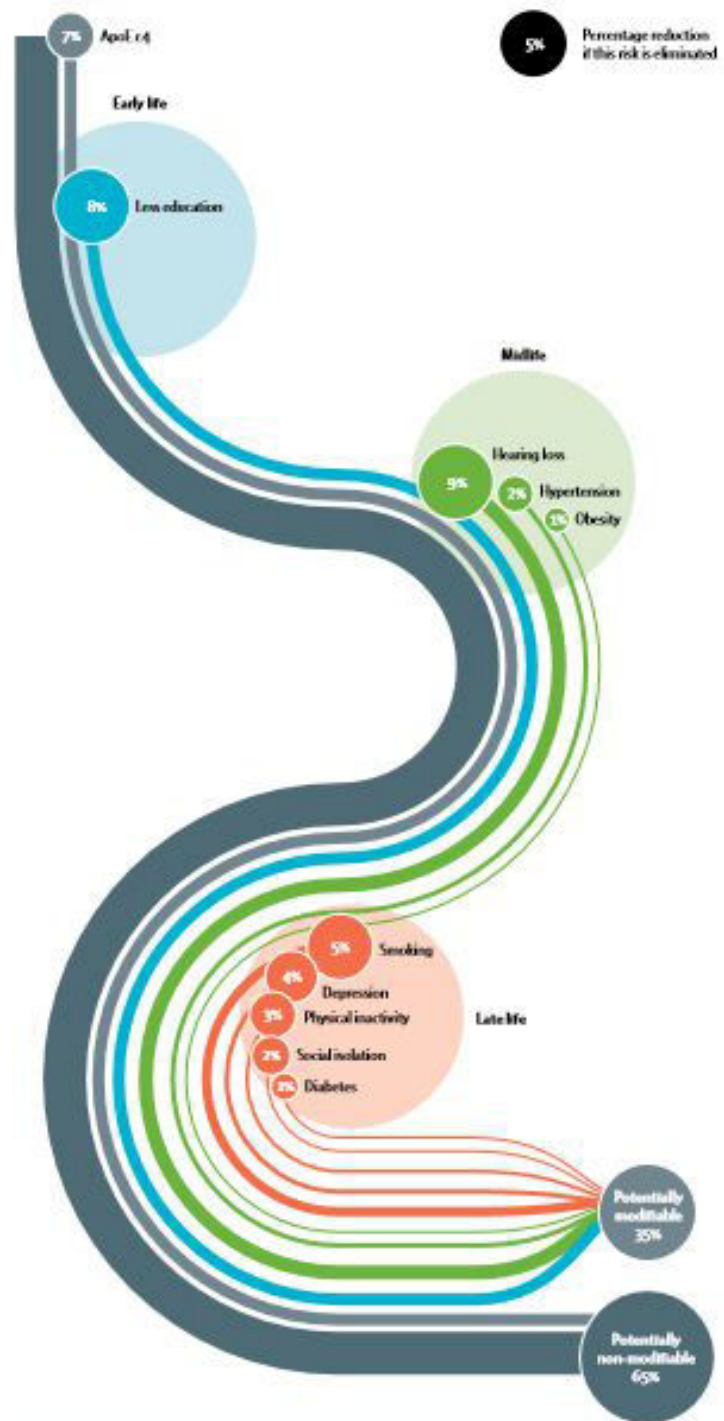


Figure 2. Life course model of contribution of modifiable risk factors to dementia

En ce sens, le *care* fait partie de la maintenance. C'est d'abord une prise de conscience de ce qui compte pour les personnes, c'est ce qui permet d'entretenir la vie en la nourrissant en énergie, ce que manifestent l'alimentation, l'hydratation, la luminosité par exemple. C'est ensuite un comportement qui ne s'oppose à la nécessité d'utiliser des traitements pour limiter la maladie³⁵⁷. C'est enfin la possibilité d'analyser la vitesse de progression de la maladie, sans avoir à attendre l'apparition des premiers symptômes caractéristiques, en utilisant des marqueurs d'évolutivité du processus biologique, comme la mesure du ratio dans le liquide céphalo-rachidien de l'amyloïde et de la protéine Tau, ou des clichés d'imagerie cérébrale par exemple. Nous reviendrons sur l'importance de prendre en compte les processus dynamiques et les caractéristiques d'évolution.

Il s'agit d'envisager une stratégie de soin pour maîtriser l'évolution des dommages avant qu'ils ne se traduisent en pathologies, c'est-à-dire s'intéresser à l'être dans son ensemble. Sur le plan de la connaissance, cela suppose de s'interroger sur les éléments et paramètres qui transforment des facteurs de risque en une pathologie. Il faut comprendre la période de latence – et elle est importante dans le cas de la maladie d'Alzheimer – l'intervalle de temps entre l'action d'une cause et la manifestation de la maladie. Une cause peut être présente, mais les autres composantes requises pour compléter cette cause peuvent ne pas être présentes. On peut alors envisager une possibilité d'action, faire en sorte que la cause en question ait une action non prolongée, de façon à ce qu'elle disparaisse

³⁵⁷ C'était le cas du solanezumab, un anticorps monoclonal du laboratoire Eli Lilly, qui réduit le déclin cognitif provoqué par la dégénérescence cérébrale. Mais son efficacité est contestée. Cela illustre la difficulté des essais cliniques auprès de personnes ne présentant aucun symptôme perceptible, les risques de tels essais étant moins acceptables dans le contexte d'une maladie potentielle que dans celui d'une maladie avérée.

avant que les autres composantes voient leur effet s'ajouter au fur et à mesure que le temps passe. On retrouve l'importance d'une reconnaissance précoce de la cause. Ce n'est pas, à proprement parler, une lutte rationnelle contre la maladie, mais, comme l'a écrit Georges Canguilhem³⁵⁸ « la prise en charge d'un malade ne relève pas de la même responsabilité que la lutte rationnelle contre la maladie ». Georges Canguilhem parle « d'allures de vie » préférant cette expression « au terme de comportement pour mieux faire sentir que la vie est polarité dynamique »³⁵⁹, et soulignant que même asymptomatique, une maladie est pour le sujet une autre allure de la vie.

Les facteurs de risque, qui ont accompagné le développement de l'épidémiologie moderne, ne résultent pas nécessairement d'une analyse concernant la causalité des maladies, mais sont plutôt vus comme une cause au sens probabiliste, cause qui n'entraîne pas nécessairement son effet mais en augmente simplement la probabilité. Il faut considérer qu'il y a une différence entre attribuer causalement un effet potentiel à un facteur de risque et prédire que la neutralisation de ce facteur réduira le risque. On touche aux fondements de la prévention. Les résultats d'une étude sur une cohorte d'individus sont généralisés à une population, puis appliqués à un individu en particulier.

Une évolution est apparue avec la notion d'actionnabilité (*actionability*) proposée par Nelson, Keating et Cambrosio³⁶⁰. Dans l'analyse qu'ils font des travaux en oncologie, il s'agit non plus d'identifier des facteurs de risque mais des cibles

³⁵⁸ Georges Canguilhem (1978) « Puissance et limites de la rationalité en médecine », dans *Etudes d'histoire et de philosophie des sciences*, Paris, Librairie Philosophique J. Vrin, 1994, page 408

³⁵⁹ Georges Canguilhem, *Le Normal et le pathologique*, Paris, PUF/Quadrige, 2011, page 137.

³⁶⁰ N. C. Nelson, P. Keating, A. Cambrosio, « On Being 'Actionable' : Clinical Sequencing and the Emerging Contours of a Regime of Genomic Medicine in Oncology », *New Genetics and Society*, Vol. 32, Issue 4, 2013, pp : 405-428.

moléculaires sur lesquelles on pourra agir techniquement et ainsi définir des thérapies ciblées. Cette approche peut être étendue aux stratégies de maintenance et de réparation : identifier ce sur quoi il faut agir pour maintenir le système biologique et l'organisme avant qu'ils n'évoluent irréversiblement vers la maladie. Les stratégies de maintenance et de réparation se heurtent à la complexité du vivant caractérisée par un foisonnement de variables et une extraordinaire imbrication de causalités diverses sur lesquelles il est difficile d'agir de façon cohérente. Ainsi, près de 40 millions de cellules meurent dans le corps chaque minute, et cependant l'ensemble des cellules échangent en continu pour produire des organismes extrêmement complexes. Cette complexité se caractérise ainsi par une certaine incertitude amplifiée par l'hétérogénéité des situations et des populations et par la multicausalité (génétiques, environnementales, individuelles, ...) des phénomènes. Cela redonne de l'importance aux modélisations qui ne sont que des représentations partielles, une recherche de simplicité dans la complexité en quelque sorte, comme moyens d'actions.

Au concept d'actionnabilité, on peut associer le concept d'*affordance*, parfois traduit par « invite »³⁶¹. *To afford*, en anglais, signifie à la fois fournir, procurer et être en mesure de faire quelque chose. Le mot *affordance* renvoie ainsi indissociablement aux propriétés de la technique et aux capacités de la personne. Il désigne quelque chose qui renvoie à la fois à l'environnement, au sens large, et à la personne. Il implique la complémentarité de la personne et de l'environnement. La notion d'*affordance* n'appartient ni à l'environnement ni à l'individu mais plutôt à la relation qu'a un individu de la perception de son environnement, et c'est cette relation qui crée un lien direct entre la perception et l'action.

³⁶¹ James J. Gibson, *Approche écologique de la perception visuelle*, Paris, Editions Dehors, traduction d'Olivier Putois, 2014.

Ainsi, l'*affordance* n'est pas une propriété du milieu considéré indépendamment de l'individu, ni de l'individu considéré indépendamment de son environnement, mais de l'individu dans son environnement.

Le couplage *affordance* / *actionability* amène à faire une distinction entre dispositions et *abilities*³⁶² Il faudra préciser les distinctions conceptuelles qu'elles recouvrent, mais on peut introduire une catégorie nouvelle, la préemption, tenir pour acquis quelque chose qui n'est que potentiel. Par ailleurs, la catégorie de la possibilité devient centrale par rapport à la catégorie de la réalité. C'est ainsi que l'approche par les *capabilities* d'Amartya Sen³⁶³ met en avant la liberté de choisir entre différentes façons de vivre. L'important, ce n'est pas ce qu'une personne effectue, mais ce qu'elle peut accomplir en choisissant vraiment selon ses caractéristiques, ses préférences, ses aspirations. Dans *capability*, il y a le pouvoir de faire quelque chose, en ajoutant la dimension de liberté.

Paul Ricoeur, dans *Parcours de la reconnaissance*³⁶⁴, reprend le concept de *capabilities* proposé par Amartya Sen. Il envisage la personne sous l'angle de son action, en reconnaissant et en respectant sa capacité de concevoir des buts, des engagements, des valeurs, etc, et en mettant en avant son *bien-être*, qui exige une certaine attention, ce qui permet de faire un lien avec le *care*.

En reprenant ce que disait Bernard Giraudeau³⁶⁵ en parlant de son cancer, on peut dire « il y a une nécessité pour le patient de se prendre en charge, de faire connaissance avec lui-même ». Plutôt que de faire disparaître définitivement la pathologie - et cela n'apparaît pas faisable, aujourd'hui, dans la

³⁶² A. Chemero (2003), "An outline of a theory of affordances", *Ecological Psychology*, 15(2), 181-195.

³⁶³ A. K. Sen (1999) *Development as Freedom*, Oxford University Press. Traduction française : *Un nouveau modèle économique*, Paris, Editions Odile Jacob, 2005.

³⁶⁴ Paul Ricoeur (2004), *Parcours de la reconnaissance*, Paris, Editions Stock.

³⁶⁵ Eric Favereau, « Entretien avec Bernard Giraudeau », *Libération* 10 mai 2010.

phase silencieuse de la maladie d'Alzheimer -, il s'agit plutôt d'en contrôler les évolutions et les conséquences dans la vie des personnes.

Toutes ces approches s'adressent à l'intérieur du corps. Or de nombreux facteurs externes interviennent. Il ne faut pas, par exemple, ignorer la question de l'influence du niveau socio-économique. Il faut donc intégrer l'intérieur du corps à une démarche plus grande, intégrer les analyses sur la personne à des analyses qui sont plus du ressort de la santé publique, même si l'exposomique, en intégrant les facteurs d'exposition, internes ou externes, dans des modèles diagnostics ou prédictifs est plus du ressort de la personne. La médecine ne se limite pas aux données sur une personne, mais doit prendre en compte les données de toutes les personnes pour repositionner la personne singulière dans l'ensemble de santé publique. C'est ainsi que « dans la majorité des cas, le raisonnement médical est un raisonnement sous incertitude »³⁶⁶.

Face à la réalité complexe qui accompagne les mutations technologiques – et l'impact du numérique sur la médecine ne peut être nié – plusieurs perspectives peuvent être retenues pour la médecine et la santé ainsi que le propose Mark Hunyadi³⁶⁷ :

- La santé, c'est d'abord la santé vécue par une personne.
- La santé, c'est ensuite la santé dans la relation avec le personnel médical, notamment le médecin.
- La santé, c'est enfin la santé vue sous le regard des politiques publiques ou des industriels de la pharmacie.

³⁶⁶ Anne Fagot-Largeault, *Médecine et Philosophie*, Paris, PUF, 2010, page 152.

³⁶⁷ M. Hunyadi (2017) « Sortir la médecine du carcan de la petite éthique », Intervention lors de l'Université d'été 2017 *Éthique, Alzheimer et Maladies Neurodégénératives*, Lyon, le 12 septembre 2017.

Ces trois dimensions ne sont pas indépendantes puisque la santé vécue peut être perçue différemment suite aux informations reçues d'un médecin, informations elles-mêmes dépendantes du savoir médical et des moyens techniques utilisés.

On retrouve ce qu'on a déjà mis en évidence, l'importance du contexte. Mais il faut s'interroger sur le rôle de ces trois dimensions dans le cadre de l'anticipation :

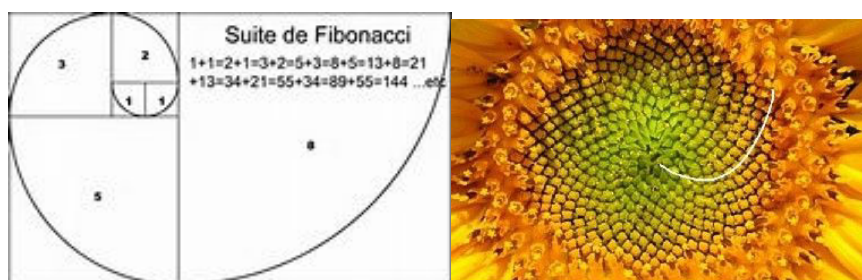
- Les médecins ne sont-ils pas dépossédés de leur savoir dès lors que la dimension relationnelle de la médecine s'estompe et que la téléconsultation, par exemple, se développe ?
- Représenter le corps comme un assemblage de briques élémentaires et d'organes n'est-ce pas considérer que tout est quantifiable et qu'une gestion de l'alimentation, de l'hygiène, des activités physiques ou intellectuelles suffit ?

La tension entre l'autonomisation du soin médical et le souci du sujet malade, ou potentiellement malade, est toujours présente. Mais il faut souligner la responsabilité du sujet. La comparaison entre la pratique d'un instrument de musique et l'utilisation d'une chaîne hifi peut y aider. D'un côté, il y a un engagement actif, de l'autre un comportement passif. Or, pour reprendre ce qu'écrit Matthew Crawford dans *Eloge du carburateur*³⁶⁸, quand on parle de maintenance, « il s'agit d'une activité orientée vers une fin qui est affirmée comme bonne par l'agent, mais cette affirmation n'a rien d'arbitraire ou de strictement privé. Elle découle plutôt de l'appréhension des caractéristiques réelles de son

³⁶⁸ M. B. Crawford, *Eloge du carburateur, Essai sur le sens et la valeur du travail*, traduction française par Marc Saint-Upéry, Paris, Editions La Découverte, 2010.

environnement. [...] Cela peut être quelque chose qui exige un certain discernement »³⁶⁹. Il s'agit de prendre les choses en main par nous-même.

Il faut ajouter que parler de maintenance quand on parle d'un sujet humain, ce n'est pas « corriger » de façon immédiate, mais plutôt mener des actions pour réaliser une série « d'oscillations-autocorrectrices ». L'analogie³⁷⁰ pourrait être faite avec la suite de Fibonacci³⁷¹ qui montre que le quotient entre deux termes successifs de la suite converge bien vers une certaine valeur connue comme le nombre d'or, symbole de l'harmonie parfaite, mais qu'on n'atteint pas de façon immédiate ce nombre d'or :



Rappelons que la suite de Fibonacci est très présente dans la nature et que l'on retrouve le nombre d'or dans de nombreux domaines comme l'architecture ou la peinture par exemple.

³⁶⁹ Ibid., page 238.

³⁷⁰ Cette analogie est également utilisée par Matthew Crawford pour parler de son travail de mise au point et de réglage du moteur d'une moto.

³⁷¹ Les deux premiers termes de la suite de Fibonacci sont 0 et 1, et ensuite chaque terme successif est la somme des deux termes précédents, ce qui donne 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, 144, 233, ... Cette suite croît très vite et on montre que le rapport de deux nombres successifs de cette suite tend vers le nombre d'or $\varphi = (5^{1/2} + 1) / 2$.

Chapitre XII

La question de la représentation de la maladie d'Alzheimer

Les choix épistémologiques d'employer telle ou telle technologie, ainsi que l'organisation des recherches, qu'elles soient biologiques ou cliniques, conduisent à des représentations différentes de la maladie d'Alzheimer. Pour certains, c'est la génétique qui prime. Pour d'autres, il faut associer ensemble le plus d'indicateurs possibles. Pour d'autres enfin, il se peut que la maladie n'ait pas d'origine.

Les « objets » se définissent beaucoup plus par ce qu'ils font que par ce qu'ils sont. En ce sens, quand on s'intéresse à l'épistémologie des biomarqueurs, il faut s'interroger sur ce qu'ils ont en commun : est-ce une propriété (comme croître continuellement ou bifurquer brutalement par exemple) ? Est-ce une proposition d'ordre relationnel (une relation entre les plaques amyloïdes et d'autres composants des neurones par exemple) ? Quant à l'origine, elle peut se définir sous deux aspects si l'on en retient la traduction en langue allemande :

- L'origine en tant que *Ursprung*, qui désigne alors le point source.
- L'origine interprétée comme *Entstehung*, qui désigne le processus menant à l'émergence de quelque chose.

« L'idée, dominante en médecine, selon laquelle la maladie peut être définie objectivement et scientifiquement comme une simple déviation par rapport à une normalité statistique ou physiologique serait illusoire. »³⁷². En fait, les notions de santé et de maladie sont multiples, ambiguës et complexes, même si la maladie, en s'appuyant sur des termes anglais³⁷³, peut avoir :

- Un sens médical et objectif (disease)
- Un sens subjectif, existentiel et clinique (illness)
- Un sens subjectif et social (sickness),

Le malade au sens *illness* est confronté paradoxalement au médecin qui devient « progressivement un conseiller »³⁷⁴, sa compétence scientifique le conduisant à devenir un expert en prédiction, alors que la personne, qui, à ce stade, est un individu sain, ne peut pas réfuter le diagnostic de la prédisposition qu'on lui présente puisqu'il n'a pas l'expérience subjective de la maladie.

La maladie d'Alzheimer n'est pas plus une simple déviation par rapport à une normalité statistique qu'un dysfonctionnement biologique. La technoscience, que l'on peut synthétiser sous le vocable des *'omics* dans les sciences de la vie, a fait de chaque corps un objet stochastique dont la caractéristique principale est qu'il serait désormais repérable avant tout comme une association de facteurs de risque simultanés ou successifs. Cette association n'est pas simple. La

³⁷² Elodie Giroux, *Après Canguilhem, définir la santé et la maladie*, PUF, Paris, 2010, p. 8.

³⁷³ B. Hofmann, « On the Triad Disease, Illness and Sickness », *Journal of Medicine and Philosophy*, 2001, n° 6, pp ; 651-673.

³⁷⁴ Jean Dausset, *Clin d'œil à la vie, La grande aventure HLA*, Paris, Odile Jacob, 1998, p. 264.

vitesse de progression de la maladie n'est pas la même chez tous les malades. Elle est influencée par différents facteurs génétiques et environnementaux, ainsi que par la plasticité plus ou moins grande du cerveau qui permet de compenser l'impact de la progression de la maladie. Le niveau d'éducation est un des facteurs. L'hypothèse biologique sous-jacente est que l'éducation, le niveau de stimulation cérébrale et les liens sociaux stimuleraient le développement d'un réseau neuronal dense qui permettrait de compenser plus longtemps les lésions induites par la maladie d'Alzheimer. C'est ce que l'on appelle la réserve cognitive et stimuler son cerveau pour favoriser le développement de cette réserve cognitive permettrait de retarder les premiers signes de la maladie. En fait, avec la maladie d'Alzheimer, on est très loin du concept classique de maladie, avec des lésions causales et des symptômes en rapport. Dans l'état actuel des connaissances, on n'est pas en mesure de déterminer un enchaînement causal précis et fiable. Le contexte est essentiel et cela conduit à repenser l'idée de diagnostic en incluant des données psychologiques et sociales.

Sans doute faut-il poser la question de la pertinence des repères cognitifs de discrimination comme facteur de risque, susceptibilité, prédisposition, etc, et des tests qui peuvent leur être associés :

- En médecine, la susceptibilité désigne la tendance d'un organisme à développer certaines pathologies. On parle aussi de diathèse. Avec les maladies multifactorielles, on ne parle plus de terrain génétique mais de susceptibilité génétique, ce qui explique la « chasse aux gènes »³⁷⁵ qui

³⁷⁵ A titre d'illustration, une étude internationale de grande ampleur, impliquant des chercheurs français de l'Unité mixte de recherche Inserm-Institut Pasteur Lille-Université Lille Nord de France « Santé publique et épidémiologie moléculaire des maladies liées au vieillissement », dirigée par Philippe Amouyel, a mis à jour un gène de susceptibilité sur une région du

entoure toutes les études au niveau moléculaire, tout en relevant dans la notion de susceptibilité génétique l'importance de la notion d'éventualité.

- La prédisposition est une disposition, naturelle, constitutionnelle ou acquise, due souvent à une fragilité particulière de l'organisme ou du psychisme à contracter certains troubles ou certaines maladies. C'est donc l'ensemble de facteurs qui augmentent le risque chez une personne de développer une maladie donnée. Un gène de prédisposition à une maladie est un gène qui augmente la probabilité de développer la maladie, et l'intervention d'autres facteurs, environnementaux et/ou génétiques, est nécessaire dans le cas des maladies multifactorielles.
- Un facteur de risque est une caractéristique liée à une personne, à son environnement, sa culture ou son mode de vie et qui entraîne pour elle une probabilité plus élevée de développer une maladie sans causer en soi la maladie. Ils font état d'un risque accru, mais pas d'une certitude que la maladie d'Alzheimer, par exemple, se développera. Mais, de la même manière, si une personne est peu ou n'est pas du tout exposée aux facteurs de risque connus, elle n'est pas pour autant protégée contre la maladie.

L'idée de facteurs de risque a été introduite comme des potentialités évaluées au niveau des populations. En la déplaçant au niveau d'une personne, on donne au risque une réalité. L'épidémiologie clinique³⁷⁶ s'appuie sur un ensemble d'individus étudiés dans des conditions cliniques bien précises.

chromosome 6, dans laquelle des mutations du gène TREM2 sont associées à un risque 5 fois plus élevé de développer une maladie d'Alzheimer à début tardif (*The New England Journal of Medicine* datée du 14 novembre 2012). Ce résultat a permis notamment de mieux comprendre l'implication du système immunitaire dans la maladie d'Alzheimer.

³⁷⁶ L'épidémiologie clinique a été introduite dans les années 1980 par Alvan Feinstein. A. R. Feinstein, *Clinical Epidemiology, The Architecture of Clinical Research*, Philadelphia, W. B. Saunders, 1985. Le mot « épidémiologie » peut porter à confusion dès l'instant où une

Elle s'appuie souvent sur les facteurs de risque, limités à des facteurs aisément quantifiables, plus faciles à prendre en compte que des données environnementales ou comportementales. Ce n'est pas le cas de la cohorte du projet INSIGHT dont l'objet est notamment de mettre au service de la décision clinique des savoirs obtenus grâce à des techniques statistiques, intégrant des données qualitatives³⁷⁷.

La coexistence de plusieurs facteurs de risque ou de plusieurs causes, se traduit souvent par la sommation des facteurs, car leurs effets sont considérés indépendamment les uns des autres. C'est très réducteur car les effets peuvent être antagonistes, inhibiteurs ou en synergie³⁷⁸, et ce, d'autant plus qu'il existe des outils statistiques permettant de prendre en compte des interactions entre paramètres. Relevons cependant ce que dit Michel Morange³⁷⁹, « La médecine a beaucoup progressé dans ses méthodes diagnostiques, permettant une objectivation de la maladie, et un diagnostic de plus en plus précis, jusqu'au niveau moléculaire. [...] C'est de la diversification des tests biologiques et de leur précision qu'on peut attendre la prise en compte de plus en plus fine des différences individuelles ». Aucun moyen n'est décisif, c'est pourquoi il faut utiliser une authentique approche multidisciplinaire.

étude épidémiologique porte en principe sur une population et non sur une cohorte bien définie d'individus.

³⁷⁷ En matière de facteur de risque, la plupart des essais portent sur cinq années, ce qui ne relève pas de la même échelle de temps qu'une prévention qui doit porter sur des décennies.

³⁷⁸ La synergie peut être définie comme la relation entre des facteurs dont l'effet conjoint excède la somme des effets pris séparément.

³⁷⁹ A. Fagot-Largeault, C. Debru, et M. Morange (dir.), Hee-Jin Han (éd.), *Philosophie et Médecine, en hommage à Georges Canguilhem*, Paris, Vrin, 2008, page 229.

XII-1 Vers une nouvelle compréhension de la maladie d'Alzheimer

L'évolution des connaissances, notamment celles associées aux marqueurs, que ce soit des biomarqueurs ou des marqueurs comportementaux, change la conception que l'on peut avoir de la maladie d'Alzheimer. D'un côté, elle permet de prendre en compte les temporalités en jeu, notamment la phase asymptomatique. De l'autre, elle change la nature du réseau causal en accentuant le lien entre corrélations et causalités.

La maladie d'Alzheimer, en dehors de sa forme familiale, est un phénomène multi-causal dont les déterminations ne sont ni complètement génétiques, ni complètement biologiques, ni complètement socio-environnementales. Elle n'est pas qu'un dysfonctionnement biologique. Dans un premier temps, le déclin de la personne est lent grâce à des mécanismes de compensation biologiques, psychiques, environnementaux. Dans un second temps, le processus pathologique l'emporte sur la compensation et la progression s'accélère. Il est ainsi crucial de comprendre cet équilibre entre compensation et progression pathologique, ce qui redonne de l'importance à la stratégie de maintenance car les actions sont plus efficaces lors de la première phase de déclin, avant l'effondrement des mécanismes compensateurs. Face à la maladie d'Alzheimer, de nombreuses incertitudes demeurent, même s'il est de plus en plus fréquemment admis que l'efficacité d'un traitement ralentissant le développement de la maladie est d'autant plus grande qu'il est administré tôt. Par ailleurs, la maladie d'Alzheimer touche la personne en profondeur et de façon globale, et conduit à réfléchir à la place de l'autre dans les relations interpersonnelles.

C'est dans ce contexte qu'a émergé la problématique des *Big Data*. Le phénomène *Big Data*, la production, le recueil et stockage de données en grand nombre, est ainsi investi de grandes ambitions dans le champ biomédical : mettre à jour de nouvelles vérités scientifiques, comprendre l'étiologie des maladies, identifier de nouvelles réponses thérapeutiques ou préventives. Avec les statistiques et les *Big Data*, on est à l'âge des « algorithmes de décision ». Mais il faut y ajouter la façon dont les personnes racontent leur histoire, c'est-à-dire un aspect narratif. Face à la maladie d'Alzheimer, ne faut-il pas abandonner toute structure argumentative pour proposer des récits, des contes permettant par des touches successives de proposer une vision de ce qu'en est la connaissance ?

Malgré des avancées réelles, nombre d'incertitudes demeurent sur la maladie d'Alzheimer, notamment sur les causes, les rôles respectifs joués par les dépôts amyloïdes et la neuro-défibrillation n'étant pas encore compris. Il persiste une forte incertitude sur les mécanismes causaux de la maladie d'Alzheimer, par essence multifactorielle. On peut aussi se demander si la majorité des programmes n'échouent pas car les concepts sur lesquels ils sont construits sont faux en se basant sur les affirmations de leaders d'opinion, qui laissent prévaloir le concept *A beta* et *Tau* comme causes de la maladie d'Alzheimer alors qu'il s'agit peut-être de conséquences non spécifiques. Si la cascade amyloïde est encore le modèle classiquement reconnu, il est envisageable d'accueillir d'autres hypothèses explicatives concurrentes qui peuvent être incompatibles. C'est un enjeu auquel il faut sans doute faire face.

Le champ thérapeutique de la maladie d'Alzheimer a rapidement évolué depuis 2007, dès l'instant où a été envisagée la possibilité de faire un diagnostic de la maladie au stade prodromal, c'est-à-dire alors que les symptômes sont limités à la mémoire récente sans atteinte fonctionnelle significative. Il était alors proposé³⁸⁰ de combiner des critères cliniques, comme l'atteinte de la mémoire de type épisodique, et des critères biologiques, comme des examens du liquide céphalorachidien, des imageries ou des tests génétiques.

Si le critère majeur reste la présence d'un trouble de la mémoire épisodique, le Professeur Bruno Dubois propose d'ajouter de nouveaux critères diagnostiques :

- Critère morphologique : atrophie de l'hippocampe à l'IRM.
- Critère biologique : les biomarqueurs de la maladie d'Alzheimer (protéine Tau et peptide A bêta) dans le liquide céphalo-rachidien ou dans le sang.
- Critère fonctionnel : diminution de l'activité cérébrale dans les régions temporales ou pariétales par imagerie fonctionnelle PET.

Plus récemment, des essais cliniques³⁸¹ visent à modifier les facteurs dits « de style de vie » comme le niveau d'éducation, le risque vasculaire, les activités physiques et intellectuels, pour améliorer les fonctions cognitives à court terme et, possiblement, retarder le déclin cognitif à long terme causé par la maladie d'Alzheimer.

³⁸⁰ B. Dubois et al., « Research criteria for the diagnosis of Alzheimer's disease : revisiting the NINCDS-ADRDA criteria », *Lancet Neurology*, 2007, 6, pp : 734-746.

³⁸¹ M. C. Carillo, H. R. Brashear, V. Logovinsky et al ;, « Can we prevent Alzheimer's disease ? Secondary « prevention » trials in Alzheimer's disease », *Alzheimer Dement.* 2013 ; 9 : 123-131.

Toutes les études s'expliquent par le souci d'aller vers une stratégie de prévention de la maladie d'Alzheimer. Elles ont pour objet de comprendre comment les gènes et leurs produits interviennent tout au long de la vie des individus, c'est-à-dire de documenter en détail les voies métaboliques impliquées au cours du temps. En ce sens, la dimension temporelle de la maladie prend en compte les activités de dépistage, mais aussi les activités d'anticipation, voire de prédiction et il faut s'interroger sur la façon de se penser en fonction de facteurs de risque alors que le concept de risque est peut-être trop pauvre ou trop étroit. Le risque, et les facteurs qui lui sont associés, ne sont généralement ni visibles, ni tangibles pour les personnes qui y sont exposés, ce qui leur confèrent un caractère paradoxal, puisqu'ils sont à la fois réels et irréels. Ils apparaissent en quelque sorte a-temporels. Le diagnostic ne peut pas être qu'une information sur les risques auxquels serait exposé un individu, compte-tenu de son génome par exemple. Il ne peut être indissociable d'un soin, pas nécessairement thérapeutique, mais d'accompagnement de la personne, tout en étant conscient qu'il faut se poser la question de savoir accompagner sans comprendre. Si l'approche en termes de « facteurs de risque » est sans doute nécessaire pour commencer, elle doit être prolongée par une approche dynamique, c'est-à-dire en leur associant le paramètre temps et évolution. Par ailleurs, logiquement, on peut considérer qu'un diagnostic doit être d'autant plus précocité que des facteurs de risque existent, et on ne peut justifier de l'absence d'un traitement curatif pour disqualifier la démarche du diagnostic précocité. Il faut considérer qu'une prévention non médicamenteuse est essentielle.

Mais il faut noter et rappeler que les données de départ doivent être contextualisées si on veut espérer que les gènes ainsi mis en évidence soient réellement associés à la maladie d'Alzheimer, dont la complexité n'est plus à démontrer. Il y a plusieurs processus impliqués et les gènes identifiés codent pour des molécules qui ont plusieurs fonctions. C'est ainsi que certains gènes peuvent suggérer que le système immunitaire et une inflammation peuvent avoir une implication causale dans la maladie d'Alzheimer, ce qui pourrait laisser supposer que l'inflammation peut précéder le dépôt amyloïde et remettre en cause le modèle dominant. La prise en compte de l'APOE a d'ailleurs été combattu initialement car le modèle présenté³⁸² suggérait que les allèles APOEε2 et APOEε3 étaient des allèles protecteurs, protégeant la protéine *Tau* des dommages pouvant conduire à la mort neuronale, et ainsi à la maladie d'Alzheimer, tandis que l'allèle APOEε4 supprimait ce rôle protecteur. Un tel modèle n'était pas conforme au modèle dominant pour expliquer la maladie d'Alzheimer, modèle basé sur la bêta-amyloïde et la cascade amyloïde. On retrouve l'importance de la notion de « collectif de pensée » proposée par Ludwig Fleck³⁸³.

Treize sites proposent des tests pour la maladie d'Alzheimer. A titre d'exemple, la société américaine 23andMe propose un test basé sur des variants du gène APOE³⁸⁴, utilisant deux mutations ponctuelles (rs7412 et rs429358) permettant d'identifier les variants ε2, ε3 ou ε4. Ils affirment cependant que leur test ne peut en aucun cas être un test de diagnostic et que le rôle de l'apolipoprotéine dans la maladie d'Alzheimer n'est pas clair. Au niveau européen, le projet NADINE (*Nanosystems for the early diagnosis of neurodegenerative diseases*)³⁸⁵ met en

³⁸² A. D. Roses, « Apolipoprotein E alleles as risk factors in Alzheimer's disease », *Annual Review of Medicine* 47 (1996) : 387-400.

³⁸³ Ludwig Fleck, *Genèse et développement d'un fait scientifique*, Trad. Française, Paris, Les Belles lettres, 2005.

³⁸⁴ https://www.23andme.com/en-gb/health/i_alzheimers/techreport/.

³⁸⁵ <http://www.fp7nadine.eu>.

œuvre des nanoparticules magnétiques couplées à des anticorps pour détecter des biomarqueurs associés aux maladies neuro-dégénératives à partir d'un très petit échantillon sanguin.

Autre exemple, le *LeARN program* (Leiden Alzheimer Research Netherlands) a eu pour objet d'évaluer les conséquences d'un diagnostic moléculaire précoce de la maladie d'Alzheimer. On peut en retenir les points suivants :

- Toutes les technologies ont leur propre trajectoire (notamment de développement et de validation) et sont en compétition. Dès qu'il est validé, un test (ou une combinaison de tests) change la pratique du diagnostic. Ainsi l'analyse du liquide céphalo-rachidien, dès l'instant où on a prélevé l'échantillon, est facile et bon marché tandis qu'une imagerie PET-Scan est beaucoup plus complexe à réaliser et beaucoup plus onéreuse.
- Dans le cas d'un diagnostic précoce, la première question à se poser est de savoir ce qu'on fait de l'information fournie par les tests. En particulier il faut passer de la recherche à la pratique clinique et cela nécessite d'avoir analysé les questions de qualité, de sensibilité, de spécificité et d'apport d'un test par rapport à d'autres tests. Il faut faire un choix entre ce que l'on retient et ce que l'on exclut avant « d'explorer la complexité ». Un test par lui-même ne fait rien, le plus important est de pouvoir intervenir.
- S'il est assez clair qu'un diagnostic précoce peut améliorer la qualité de vie des personnes, il n'en demeure pas moins qu'un diagnostic précoce entraîne des coûts supplémentaires pour les systèmes de santé.

Parler de diagnostic précoce n'est pertinent que si la personne a des capacités d'adaptation et que s'il peut être associé à une organisation adaptée des parcours de prise en charge. De fait, certaines voies s'élèvent pour remettre en cause les diagnostics précoces sur des personnes ne se sentant pas malades, l'objection principale reposant sur la question de savoir pourquoi vouloir détecter précocement une maladie pour laquelle il n'existe aucun traitement efficace. Face à la question du diagnostic précoce :

- Certains argumentent que le jour où un traitement efficace sera disponible, il faudra l'administrer le plus tôt possible et donc savoir à qui le proposer.
- Certains considèrent que le diagnostic précoce se justifie, même en l'absence de traitement, si la personne peut bénéficier d'une prise en charge adaptée à sa situation.
- D'autres, enfin, mettent en avant que l'on est entré dans une ère de responsabilisation des personnes au regard des maladies qu'elles pourraient développer, tout en relevant que le diagnostic précoce de la maladie d'Alzheimer pose problème car, pour reprendre le vocabulaire de Martine Bungener³⁸⁶, le diagnostic précoce est encore trop souvent l'annonce d'une « sentence de mort sociale ».

Même si le dépistage de la maladie d'Alzheimer n'est pas facile car il nécessite des moyens lourds (analyse du liquide céphalo-rachidien et imagerie)³⁸⁷ et coûteux, poser un diagnostic précoce, c'est pouvoir bénéficier rapidement des traitements validés ou participer aux programmes de recherche sur la

³⁸⁶ Enquête Lecma-Vaincre Alzheimer – <http://vaincrealzheimer.org>. L'enquête sur l'importance du diagnostic précoce Alzheimer révèle que près de 45% des Français négligeraient leurs troubles de la mémoire s'ils en avaient

³⁸⁷ Il est impossible de faire une IRM cérébrale tous les deux ans aux français de plus de 60 ans, et il n'est pas recommandé de faire une ponction lombaire tous les cinq ans.

maladie, c'est anticiper et prendre soi-même les décisions qui concernent son projet de vie.

Il faut souligner que le souci du diagnostic précoce s'accompagne de développements de nouveaux tests faciles à mettre en œuvre et peu coûteux. Ainsi les chercheurs de l'Université d'Oakland (USA) et de l'Université d'Alberta (Canada) travaillent sur un test salivaire de détection précoce de la maladie d'Alzheimer. En s'appuyant sur la métabolomique, ils ont identifié un ensemble de 22 biomarqueurs, présents dans la salive, permettant d'identifier une signature biologique caractéristique du déclin cognitif et de la maladie d'Alzheimer. Les résultats ne sont que préliminaires mais peuvent constituer une preuve de concept d'un test salivaire de détection précoce de la maladie d'Alzheimer³⁸⁸.

Dans la même veine, une société californienne, *Neurovision Imaging*³⁸⁹, travaille sur la détection de la maladie d'Alzheimer par observation de la rétine. La méthode repose sur l'hypothèse que les plaques amyloïdes apparaissent dans les yeux avant de coloniser le cerveau. Elle consiste à injecter une substance fluorescente, puis à observer et à analyser, par rétinographie (cliché photographique de la rétine) les plaques amyloïdes. Cette méthode, issue des travaux du Cedars-Sinai Medial Center, devrait être validée, ou pas, d'ici trois ans et constitue une méthode peu invasive et peu coûteuse.

Mais on peut faire un diagnostic précoce pour des personnes se plaignant de leur mémoire ou présentant des troubles cognitifs légers lors de test mémoire. Faute de traitements curatifs, certains s'y refusent. Mais, cela a déjà été souligné, ne plus porter le diagnostic précoce d'une maladie sous prétexte d'une absence de traitement empêchera d'en trouver un à ce stade de déficience cognitive légère. Par ailleurs, la justification capitale du diagnostic anticipé, au-delà de l'accès à

³⁸⁸ *Journal of Alzheimer's Disease* May 11 2017 DOI : 10.3233/JAD-161226 "Diagnostic Biomarkers of Alzheimer's Disease as Identified in Saliva using 1H NMR-Based Metabolomics".

³⁸⁹ <http://www.neurovisionimaging.com>.

la recherche, réside dans l'espoir d'une amélioration de la qualité de vie de la personne qui, à ce stade, ne se sent pas malade.

Si les médicaments sont qualifiés d'inefficaces, c'est parce qu'on voudrait qu'ils guérissent la maladie d'Alzheimer³⁹⁰, alors qu'aujourd'hui, s'ils apportent des effets bénéfiques, ils n'empêchent pas la progression des lésions cérébrales de la maladie. Ces lésions font lentement disparaître des neurones ayant l'acétylcholine et le glutamate comme neuromédiateurs, ce qui explique les deux grandes familles de médicaments qui doivent être utilisés dès le stade léger de la maladie d'Alzheimer, voire de la déficience cognitive légère³⁹¹.

Les anticholinestérasiques bloquent l'enzyme qui dégrade l'acétylcholine. C'est le cas de l'Aricept (donépézil) et de l'Exelon (rivagstimine) ou encore du plus récent le Réminyl (galantamine). Le Professeur Dubois mène actuellement des essais sur le donépézil³⁹² pour démontrer son efficacité dans le traitement des malades diagnostiqués avant l'apparition des premiers symptômes.

Les antiglutamates, comme l'Ebixa (mémantine) réduisent les effets neurotoxiques du glutamate.

XII-2 Vers une nouvelle nosologie

Ainsi qu'il a déjà été esquissé, le développement des nouvelles techniques d'imagerie, notamment l'IRM et la TEP, qui permettent de mettre en évidence la présence de plaques amyloïdes et de visualiser les activités des tissus, ainsi que l'identification de biomarqueurs spécifiques de la maladie d'Alzheimer au niveau du liquide céphalorachidien (diminution de

³⁹⁰ Il n'existe, à ce jour, pas de traitement curatif de la maladie d'Alzheimer, tout comme rien ne guérit le diabète, les troubles du cholestérol ou l'hypertension artérielle.

³⁹¹ B. Winblad & al., "3-year study of donepezil therapy in Alzheimer's disease : Effects of early and continuous therapy", *Dement. Geriatr. Cogn. Disord.*, 2006, 21 (5-6), p.353-363.

³⁹² La société américaine Pfizer, qui commercialise le donépézil, soutient cet essai et participe au financement de l'étude Insight.

la protéine bêta-amyloïde et augmentation de la protéine *Tau* ou de *Tau* phosphorylée) permettent de mettre en évidence un profil clinique et de revoir la terminologie classiquement attachée à la maladie d'Alzheimer :

- Si ces marqueurs sont normaux ou négatifs, la maladie est clairement absente ;
- Si ces marqueurs sont positifs alors que la personne n'a pas de déclin cognitif documenté, cela signifie simplement qu'il est à risque de développer la maladie, mais en aucun cas on ne pose le diagnostic de la maladie d'Alzheimer sur la seule base des biomarqueurs.

En particulier, aujourd'hui, on ne se limite pas aux phases cliniques de la maladie, mais on en prend en compte le spectre complet depuis ses phases précliniques jusqu'au stade démentiel, avec notamment, le stade préclinique³⁹³ :

- Un stade asymptomatique à risque de maladie d'Alzheimer, où les biomarqueurs sont positifs mais le fonctionnement cognitif normal.
- Un stade de maladie d'Alzheimer pré-symptomatique qui concerne les porteurs d'une mutation génétique familiale autosomique ne présentant pas encore de symptômes.

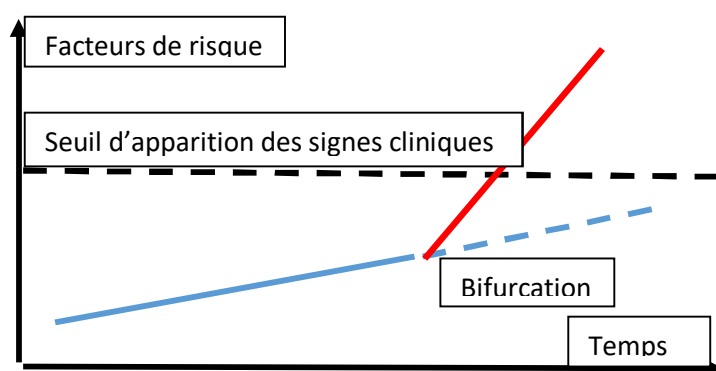
Il y a ainsi une avancée dans la nosologie de la maladie d'Alzheimer en dissociant les aspects syndromiques comme les déficits cognitifs légers, et l'étiologie spécifique de la maladie. Cette distinction est importante car, en identifiant de façon précoce des individus présentant un tableau spécifique de la maladie d'Alzheimer, elle permet la mise en place d'essais

³⁹³ B. Dubois et al., "Revising the definition of Alzheimer's disease : a new lexicon", *Lancet Neurology* 2010, 9 (11), pp : 1118-1127.

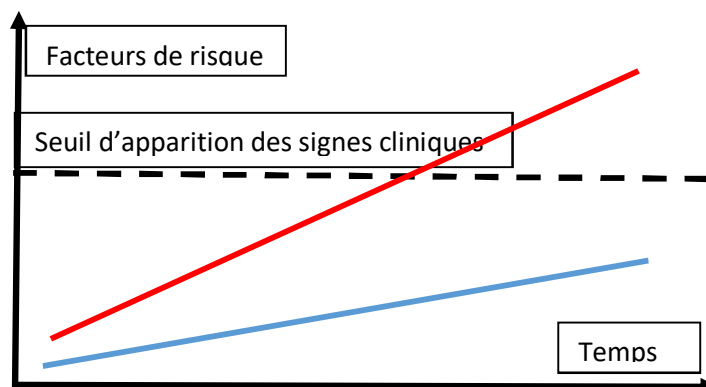
cliniques pour évaluer des traitements modificateurs de la maladie. Néanmoins, les règles actuelles ne recommandent pas de dépistage chez des personnes asymptomatiques. On dispose pourtant de biomarqueurs sophistiqués de lésions dont on pense qu'elles sont en lien avec la maladie d'Alzheimer et qui peuvent être positifs avant le déclenchement des premiers symptômes. De plus, il n'existe pas encore de règles de bonne pratique pour l'évaluation clinique de routine pour des personnes présentant des symptômes cognitifs légers.

Si l'on comprend mieux aujourd'hui comment se constituent les lésions cérébrales et leurs conséquences sur la survie des neurones, on ignore encore précisément ce qui déclenche le mécanisme pathologique. On ne connaît pas un ou des facteur(s) déclenchant(s) de la maladie et comme la maladie se déclare de façon silencieuse 10 à 20 ans avant l'apparition des signes cliniques, les recherches sur les causes ne sont guère facilitées, ainsi que nous l'avons déjà relevé.

Se pose néanmoins la question de l'origine : est-ce que des événements particuliers, le deuil d'un conjoint par exemple, vont créer une bifurcation dans l'évolution des facteurs de risque, qui va provoquer, quelques temps plus tard, l'apparition des signes cliniques de la maladie



Ou bien est-ce que la survenue de la maladie après un délai de 15 à 20 ans n'a pas d'origine particulière



Dans le premier cas, on peut considérer trouver des éléments « actionnables », notamment ce qui peut provoquer une expression différentielle de gènes, tout en étant conscient qu'il faut détecter cette bifurcation avant toute apparition de signes cliniques, tandis que dans le second c'est la complexité des réseaux d'interactions et des facteurs épigénétiques qui est prépondérante, c'est-à-dire l'ensemble des événements auxquels peut être soumis un individu et leur perpétuel devenir.

Dans le premier cas, on trouve la possibilité d'une intervention technique ou thérapeutique tandis que dans le second cas, c'est l'aspect relationnel entre la personne et son médecin qui sera important. Mais dans les deux cas on a affaire à une maladie dont on ne sait pas qu'elle a commencé et c'est une grande caractéristique de la maladie d'Alzheimer par rapport aux autres grandes maladies, les cancers par exemple.

Cette double possibilité permet un lien entre les deux concepts de biomarqueur que défend Xavier Guchet, le biomarqueur comme image de la cible sur laquelle agir, ou plus

précisément le biomarqueur comme indice de quelque chose qui renvoie à un événement passé, et le biomarqueur comme trace, signe de « quelque chose qui n'est ni révélé, ni dissimulé » dont le propre est que « ce dont elle est la trace se dérobe toujours »³⁹⁴.

La trace décrit, peut-être de manière paradoxale, comment s'annonce et se montre ce qui n'a jamais été présent et ne le sera jamais, un passé qui n'a jamais été présent.

Cela permet de considérer deux points de vue, d'une part l'existence d'un premier événement moléculaire au niveau de la bifurcation, entraînant par la suite un enchaînement d'autres événements pouvant *in fine* s'expliquer selon des schémas mécanistes, d'autre part l'hypothèse qu'il n'y a pas d'événement premier mais « des renvois d'altération à altération, de variation à variation »³⁹⁵.

Il existe ainsi une relation complexe entre les différentes représentations de la maladie d'Alzheimer, mais ce qui ressort, c'est la nécessité de maîtriser l'évolution des facteurs de risque de la maladie d'Alzheimer, et donc de maîtriser la vitesse de la progression de la maladie avant l'apparition des premiers symptômes, c'est-à-dire soit maîtriser le point de « bifurcation », soit maîtriser la « pente » de l'évolution des facteurs de risque dans le temps. On retrouve que le *cure* et le *care* sont deux stratégies qui se complètent et qu'il faut mettre la personne au centre. On retrouve que le diagnostic précoce, en identifiant les traces de la maladie d'Alzheimer dans le cerveau, fait passer la maladie des symptômes à la pathologie et ouvre la possibilité de

³⁹⁴ Xavier Guchet, *La médecine personnalisée, un essai philosophique*, Paris, Les Belles Lettres, 2016, page 389.

³⁹⁵ Ibid., page 288.

suivre les paramètres de ce diagnostic au fil du temps chez une personne. Cela permet, d'une part de meilleures approches pour la recherche et le développement de médicaments³⁹⁶, d'autre part un meilleur suivi des personnes.

³⁹⁶ Même lorsque le traitement manque, il est important d'anticiper sa disponibilité future.

Chapitre XIII

L'épistémologie, l'éthique et les enjeux économiques et sociaux

Travailler sur l'épistémologie des biomarqueurs, avec une attention particulière sur la complexité, vise, en premier lieu, à comprendre comment se constituent une connaissance et une pratique. Mais, qui détient le savoir, détient le pouvoir, telle pourrait être l'approche du lien entre épistémologie, éthique, enjeux sociaux et politiques.

Selon Spinoza, le processus de la connaissance articule au moins deux genres :

- La connaissance de premier genre exploite l'expérience et les signes. Elle se développe à travers des idées.
- La connaissance de deuxième genre procède de démonstrations et produit des vérités générales sur lesquelles tous puissent s'entendre.

Kant, de son côté, dans *la Critique de la raison pure*, s'interroge sur les possibilités de la connaissance et distingue :

- Les phénomènes, choses que l'on peut découvrir à travers notre sensibilité,
- Les noumènes, choses en soi, telles qu'elles sont en vérité, indépendamment de notre perception, qui demeurent inaccessibles à l'entendement humain.

Pour lui, seule la dimension phénoménale est accessible à la connaissance. La sensibilité, lieu d'interaction de l'humain avec son environnement physique, fournit les éléments de base pour la connaissance. Puis le sujet humain, au moyen de son entendement, va structurer les données de la sensibilité pour aboutir à la connaissance de l'objet phénoménal. Enfin, la raison parachève le travail de l'entendement en appliquant notamment des catégories abstraites telles que la quantité ou la causalité.

Plus précisément, dans le domaine médical, auquel appartiennent les biomarqueurs, la connaissance porte sur l'humain qui est, de fait, trinitaire : 100 % individu, 100 % social, 100 % biologique, et ainsi particulièrement complexe. Cette complexité peut se traduire en modèles partiels et incomplets, comme nous l'avons dit, mais il faut se protéger de la position de considérer que tout ce qu'on ne comprend pas ne sert à rien³⁹⁷ car un tel rejet empêche l'accumulation de connaissances.

Les changements technologiques influencent les pratiques médicales, les professions impliquées, l'organisation des systèmes de santé. Ils amènent les « frontières » à être floues. Par ailleurs, on ne peut pas oublier que la construction du savoir est orientée par les modalités de financement. Enfin, la santé connectée oriente la médecine vers le prédictif. Elle transforme le médecin en référent et le patient en fabrique de signaux. Tout cela au détriment de l'échange humain, pourtant essentiel. Les technologies numériques interviennent dans la perception que nous avons du temps, de l'espace et des autres, et aussi dans la construction de nos identités. Il n'y a sans doute pas de risque technologique en soi mais des contextes qui

³⁹⁷ On peut rappeler que les gènes qui ne codaient pas ont été appelés les gènes poubelles alors qu'ils ont une fonction importante dans la conformation et l'expression du génome.

favorisent ou non une appropriation plus ou moins clairvoyante des objets technologiques.

XIII-1 L'anticipation

La dimension temporelle de la médecine doit être prise en compte, notamment dans le cadre des maladies chroniques. Aussi, l'anticipation est au cœur de notre problématique. La question de l'anticipation se pose, elle prend des formes multiples, elle peut notamment être prédictive ou adaptative, et les activités d'anticipation et de prédiction deviennent des outils de contrôle. Nous sommes invités à apprendre à nous surveiller, à nous penser en fonction de nos facteurs de risque.

L'anticipation s'appuie sur des techniques et les données associées. Le développement du *Big Data* pourrait cependant menacer le déploiement de la vie dans son imprévisibilité. Or le développement de la vie veut que rien ne soit sûr tout en considérant que si on a tous besoin de vivre au présent, on a aussi besoin d'anticiper selon trois points de vue :

- Anticipation créatrice : demain peut être fort différent d'aujourd'hui.
- Anticipation adaptative : on crée les conditions de l'imprévisible.
- Anticipation prédictive (renoncement) : on lâche sur certains points.

Trois figures de l'anticipation se mélangent :

- L'anticipation comme prédiction, avec un aspect préemption (diagnostic/pronostic).

- L'anticipation comme projection (élaborer étape par étape un futur).
- L'anticipation comme adaptation (le futur est indéterminé, on fait confiance).

La prédiction et la projection sont indispensables, tandis que l'adaptation est insuffisante pour affronter le futur. La prédiction est plus du domaine médical. Basée sur les probabilités, elle annonce ce qui va se passer à partir des conditions du passé. Elle anéantit les capacités d'anticipation car prédire quelque chose à quelqu'un, c'est lui dire « les choses vont arriver ainsi ». La projection est plus du registre personnel. Elle se rapproche de l'utopie en imaginant ce qui ne peut pas être imaginé, en mettant l'accent sur le souhaitable. Quant à l'adaptation, elle nécessite un espace d'évolution de l'individu (voire de la société ou des institutions) qui permette la prise en main de ce qui se passera, pour faire face à l'imprévisible.

Par anticipation, on retiendra plutôt un mouvement adaptatif qui ne fait jamais appel à un scénario précis du futur, mais plutôt, du moins abstraitement, à la diversité des scénarios possibles, ce qui permet d'affronter la réalité. En ce sens, et contrairement à la prédiction, l'anticipation mobilise tout à la fois le présent et le futur, mais également le passé.

La notion de prédiction doit être interrogée. La prédiction est-elle un service rendu à la personne ? Ne sur-évaluons-nous pas l'interprétation des tests génétiques ? Ne risque-t-on pas d'écraser l'avenir ? Par ailleurs, en médecine, la prédiction ne s'appuie-t-elle pas trop sur l'analyse du génome humain par des

tests génétiques ? Rappelons que les tests génétiques dans le monde médical peuvent avoir pour objectif de :

- confirmer le diagnostic d'une maladie mono-génétique³⁹⁸ (un seul gène touché). Ce résultat permettra de mieux prendre en charge le malade et de donner une indication sur les évolutions possibles de la maladie : test diagnostique (ex : ce sont des maladies telles que la mucoviscidose, l'hémophilie, les myopathies ...).
- détecter chez un individu sain une mutation mono-génétique à l'origine d'une maladie à révélation tardive présente dans sa famille : test pré symptomatique. (ex : maladie de Huntington).
- réaliser chez un individu sain l'identification d'anomalie génétique mono-génique récessive liée à une maladie qui ne s'exprimera donc pas chez lui mais qu'il pourrait transmettre à sa descendance (ex : la bêta-thalassémie, la mucoviscidose) : test d'identification.
- détecter les porteurs de gènes impliqués dans des maladies multifactorielles à révélation tardive.

Le résultat obtenu sera à nuancer, en fonction des études épidémiologiques antérieures ayant pu être réalisées, permettant de fournir un résultat probabiliste de risque de développer la pathologie sans en préciser le moment ni l'intensité: on parle alors de test de susceptibilité pour les faibles risques et des tests de prédisposition qui concernent les maladies polygéniques dont un gène dit majeur.

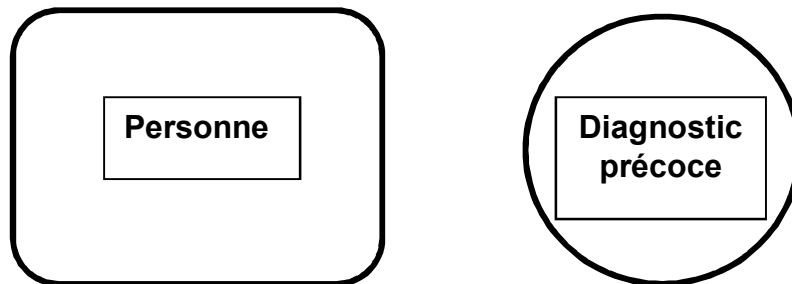
³⁹⁸ Environ 3000 maladies génétiques de ce type ont été identifiées à ce jour.

Si l'anticipation est l'objectif, la surveillance est le moyen. Mais quand l'anticipation est prédictive, la surveillance devient une fin en soi et non plus un moyen. Par ailleurs, le registre de la probabilité n'existe que dans le domaine scientifique. Quand on annonce une probabilité à un individu, ce n'est plus une probabilité et il faut faire attention à ne pas transformer un danger fondé sur le calcul probabiliste (« quelque chose peut arriver ») en un danger déjà présent (« quelque chose m'est arrivé »). Anticiper les risques, sous réserve qu'une susceptibilité génétique a été mise en évidence, peut entraîner un assujettissement à un état de veille et de contrôle sans relâche susceptible de générer une forme d'angoisse, la pensée de la survenue, supposée inéluctable, de la maladie étant permanente.

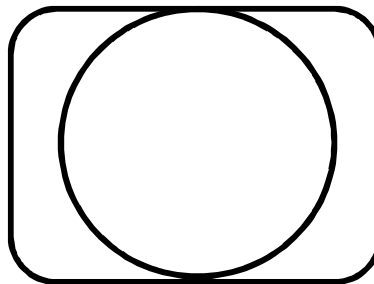
La chronicité de la maladie nécessite de l'inscrire dans une temporalité. Cela implique, nous l'avons déjà noté, non seulement de prendre en compte l'état présent, mais aussi l'état antérieur et l'état futur « anticipé ». Le risque de l'annonce de la maladie est de précipiter le sujet dans un avenir catastrophique et de lui couper toute possibilité d'agir alors que ce qui peut justifier cette annonce c'est de permettre au sujet de devenir acteur de sa maladie. *L'empowerment* accroît les capacités d'une personne à gérer au mieux sa situation dans un contexte d'une meilleure connaissance de soi.

Trois situations peuvent être schématisées :

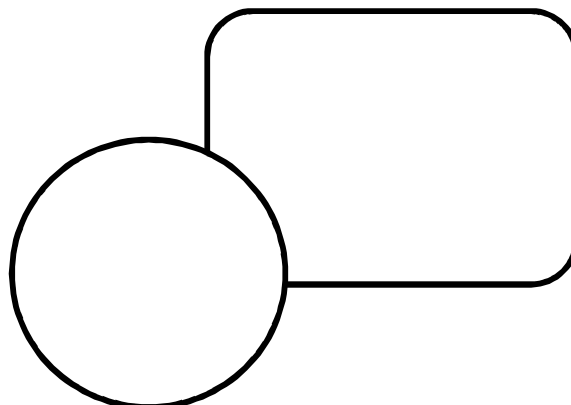
- Le déni du diagnostic



- L'obnubilation, tout tourne autour de la maladie potentielle



- L'annonce vécue comme un défi qui vaut la peine d'être relevée



On retrouve le paradoxe du diagnostic précoce ou anticipé dont la probabilité est associée à de l'incertitude. D'un côté, l'incertitude préserve, la personne considérant qu'elle n'est pas atteinte par la maladie. De l'autre, l'incertitude peut être un facteur d'immobilisme, la personne pouvant considérer qu'il n'y a rien à faire tant que le diagnostic de la maladie n'est pas avéré.

La maladie d'Alzheimer ayant une longue phase silencieuse et son évolution, après les premiers signes cliniques pouvant être très différente selon les personnes et se traduire par des comportements inattendus, faut-il, avec le diagnostic précoce, voire l'anticipation, annoncer une catastrophe inéluctable à venir ? Anticiper est-ce envisager le pire qui puisse arriver ? Comme l'écrit Marie Gaille³⁹⁹, « une interrogation philosophique sur la médecine qui serait aveugle de l'impact de l'évolution du savoir et des technologies sur les questions éthiques, politiques et juridiques posées par la médecine en manquerait une dimension essentielle. [...] Il convient [...] de voir dans la médecine autre chose que l'application d'un savoir biologique, mais une pratique qui a trait à un savoir sur l'homme, et un homme qui n'est pas considéré indépendamment de son contexte social et politique ».

XIII-2 L'épistémologie et l'éthique

L'éthique peut être avant tout une démarche prescriptive, donnant des repères, rappelant des valeurs et des principes fondamentaux. Mais elle peut être aussi une démarche réflexive, interrogeant sur les valeurs qui doivent gouverner l'action, questionnant les principes qui commandent les choix individuels

³⁹⁹ Marie Gaille, *Philosophie de la Médecine – Frontière, savoir, clinique*, Paris, Vrin, 2011, pp. 24-25.

et les actions collectives. L'éthique a, à la fois, une composante normative, qui cherche à élaborer des normes morales universelles sur lesquels fonder les comportements, et une composante méthodologique, qui vise à élaborer des méthodes permettant de guider les choix en cas de conflits de valeurs. Les normes ont souvent été premières et se sont déclinées à travers les principes d'autonomie, de non-malfaisance et de bienfaisance, de liberté. La discussion est une condition de l'éthique, les principes en étant les matériaux. Ce faisant l'éthique, dans sa dimension bioéthique, favorise « l'acceptabilité sociale »⁴⁰⁰ des approches et technologies médicales.

La discussion éthique présuppose d'être capable non seulement d'identifier des tensions, mais d'en discuter ensemble. Paul Ricoeur⁴⁰¹ situe bien l'éthique dans une « tension » entre les différents acteurs. Cette tension s'exprime dans les trois pôles de l'agir : « ma liberté, ta liberté, la règle ». J'ai ma liberté, mais existe aussi la liberté de l'autre et, à un moment ou à un autre, il va falloir arriver à un consensus sur une règle. Ainsi, les examens génétiques ne doivent être prescrits que lorsqu'ils ont une utilité clinique et sont souhaités par la personne. Le seul fait qu'un examen soit disponible et réalisable ne justifie ni sa prescription, ni sa réalisation. Paul Ricoeur invite à maintenir en cohérence ce qu'il appelle :

- Le niveau prudentiel, niveau de la relation immédiate à la souffrance ;
- Le niveau déontologique, niveau des normes qui régissent la profession médicale ;
- Le niveau réflexif, niveau de l'éthique générale.

⁴⁰⁰ B. Bensaude-Vincent et V. Nurock, « Éthique des nano-technologies », i, E. Hirsch (éd.), *Traité de bioéthique I. Fondements, principes, repères*, Toulouse, Éres, 2010, p. 355-369.

⁴⁰¹ Paul Ricoeur, « Les trois niveaux du jugement médical », *Esprit* N° 12, 1996, p. 21-33.

L'éthique est un questionnement. Ce questionnement porte d'abord sur les valeurs, les valeurs étant ce qui est mobilisé pour donner un sens aux actions. L'éthique est alors comprise comme axiologie. Ce questionnement porte ensuite sur les finalités, l'éthique étant alors comprise comme une téléologie. C'est répondre à la question du pourquoi et du comment d'une action, les finalités précédant ainsi l'action. Enfin le troisième questionnement de l'éthique porte sur les conséquences, qui ne peuvent être assimilées aux finalités puisqu'elles arrivent après l'action et sont du registre de l'anticipation.

Dans une perspective de médecine génomique, la croissance spectaculaire des données disponibles soulève la question des modalités de la prise de décision médicale. Au préalable, la question de la protection des données se pose. Si elle trouve d'ores et déjà des réponses techniques en matière de cryptage et de sécurité de l'information, cette question en cache une autre, plus fondamentale, plus vaste, c'est celle de savoir comment définir ce qu'est une donnée de santé. La production des données à l'occasion de prises en charge médicales impose que leurs utilisations soient compatibles avec le respect du secret médical et, plus généralement, du secret professionnel. Des outils pour extraire l'information pertinente utile au médecin, et disponible dans un délai acceptable, sont nécessaires. Le séquençage du génome entier pose la question de savoir ce qu'on fait des données qui n'étaient pas directement recherchées pour le problème concerné et qui sont acquises en plus, sans le moindre consentement éclairé puisqu'il n'est donné que dans un cadre précis. La définition de la donnée actionnable est la vraie question, notamment pour les données secondaires non recherchées initialement. Cela repose aussi la signification de ce qu'est un biomarqueur. C'est un indicateur objectif, qui

peut être mesuré précisément et de façon reproductible, et qui est le plus souvent employé à des fins de diagnostic. De fait, il devient moins cher de séquencer le génome entier que de chercher une séquence particulière (conseil génétique). On peut ainsi se trouver dans une situation paradoxale : conclure qu'on n'a rien trouvé sur la séquence particulière, mais dire qu'on a trouvé, à cette occasion, ceci ou cela. Il y a ainsi trois options :

- Ne pas tout lire dans le séquençage ;
- Tout lire mais ne pas tout dire ;
- Tout lire et tout dire.

Il y a une discordance entre ce que l'on peut interpréter et ce que l'on peut changer en intervenant. Se pose alors la question de savoir comment communiquer cette information ou qu'en faire ? Faut-il tout lire et choisir ensuite ce que l'on va dire ? Faut-il ne pas tout lire ? La disponibilité d'informations plus ou moins productives pose, de fait, plusieurs questions :

- La question de l'interprétation,
- La question de la responsabilité individuelle de la personne vis-à-vis de la gestion de sa santé et vis-à-vis de sa famille,
- La question de la disponibilité et de l'accès aux informations.

Il est délicat de communiquer sur une éventuelle prédisposition génétique à la maladie ou sur des facteurs de risques. On peut se demander si informer une personne de la probabilité d'une maladie, pour laquelle aucune réponse thérapeutique efficace n'est disponible au moment de l'annonce, ne risque pas de lui porter préjudice, alors qu'une personne « à risque » peut découvrir de nouvelles possibilités dans une telle

situation, la signification de sa maladie potentielle lui appartenant⁴⁰². Schématiquement, la population a trop tendance à assimiler le facteur de risque au risque lui-même. La diffusion de la culture scientifique n'est pas chose simple. Il faut se demander si telle connaissance doit être gérée comme un risque ou comme une incertitude, en particulier dans un contexte d'absence de clarté causale. On utilise des valeurs de pénétrance, c'est-à-dire la probabilité de tomber malade pour les personnes porteuses de certaines mutations. Or des individus porteurs de ces mutations, parfois en double dose, auraient dû être considérés comme malades alors qu'ils n'ont pas développé la maladie. Il faut donc s'interroger sur notre capacité à bien mesurer les risques. Les risques mesurés sont toujours des risques qui ne sont absolument pas déterministes⁴⁰³ et communiquer sur un risque présente un vrai piège si on évoque une moyenne sans avoir une idée des paramètres de dispersion, alors que l'importance de ces paramètres de variance peut souvent être supérieure à la valeur de la moyenne proprement dite.

Tous les savoirs méritent-ils d'être connus et diffusés ? La question se pose, notamment dans le cas d'un savoir diagnostique précoce d'une maladie comme la maladie d'Alzheimer. Est-ce que savoir c'est pouvoir agir comme le pense Benasayag⁴⁰⁴ ou n'est-ce qu'un objet de responsabilité dans un contexte d'absence de clarté causale ? Le fait que le malade dépende presque entièrement du savoir et des compétences de ceux qui le soignent crée inévitablement un déséquilibre qui, en quelque sorte, est compensé

⁴⁰² On pourrait paraphraser Fritz Korn en disant : « Depuis que je suis potentiellement malade, je vais beaucoup mieux » dans Fritz Korn (1977) *Mars. Je suis jeune et riche et cultivé, et je suis malheureux, névrosé et seul*, trad. G. Lambrichts, Paris, Gallimard, 1979, « Folio » 2005 : 22.

⁴⁰³ Rappelons que dans le cas de la maladie d'Alzheimer pour laquelle une personne aura un risque moyen, si l'on ne fait pas d'analyse génétique, de 10 % environ, son risque variera de 8 % ou de 51 % selon son génotype pour le gène ApoE.

⁴⁰⁴ M. Benasayag, *Connaître et agir. Paysages et situations*, Paris, La Découverte, 2006.

en France dans la loi du 4 mars 2002, dite « loi Kouchner » selon laquelle « une personne prend, avec le professionnel de santé [...], les décisions concernant sa santé. [...] Aucun acte médical ni aucun traitement ne peuvent être pratiqués sans le consentement libre et éclairé de la personne ». Dans cet esprit, la loi du 22 avril 2005, dite « loi Léonetti », révisée en 2016 pour autoriser la sédation profonde d'un patient jusqu'à son décès pour lui éviter de souffrir, permet à toute personne de rédiger ses volontés sur les conditions de limitation ou d'arrêt de traitement pour le cas où elle serait hors d'état de pouvoir s'exprimer. Ajoutons que les effets d'un diagnostic sur l'existence humaine sont très largement imprévisibles. D'un côté le savoir libère puisque l'on peut planifier sa vie. De l'autre la maladie, même la maladie potentielle, détruit des possibles.

Si l'on suit Canguilhem, « la logique du probable que la médecine doit prendre en compte est une science de l'espérance et du risque. Sous ce rapport, elle est authentiquement une science de la vie »⁴⁰⁵. Cela peut nous amener à envisager une éthique du risque assumé et du dépassement face à une éthique bien souvent fondée sur la peur comme celle de Hans Jonas. « Dans le Principe de Responsabilité, [...] la responsabilité est d'emblée définie comme une relation : elle désigne la formation d'un lien entre deux êtres, l'un vulnérable, l'autre assez puissant pour agir en réponse à cette vulnérabilité »⁴⁰⁶. « La responsabilité [...] fait entrer le futur dans la sphère de l'éthique, comme engagement, parole qui se donne pour le temps à venir »⁴⁰⁷. Mais, ainsi que le relève Edgar Morin⁴⁰⁸, « connaître ce n'est pas arriver à établir des vérités absolument certaines, c'est savoir qu'il y a de l'inconnaissable et de l'inconcevable, c'est dialoguer avec l'incertitude ».

⁴⁰⁵ Georges Canguilhem, *Etudes d'histoire et de philosophie des sciences concernant les vivants et la vie*, Paris, Vrin, 7^e édition, 1994.

⁴⁰⁶ Roland Schaer, *Répondre du Vivant*, Paris, Editions le Pommier, 2013, page 36.

⁴⁰⁷ Ibid., page 38.

⁴⁰⁸ Edgar Morin, *Mes philosophes*, Paris, Germina, 2011, page 44.

Il n'en demeure pas moins que dire à des personnes parfaitement normales qu'elles ont « une maladie d'Alzheimer asymptomatique » sur la base de biomarqueurs nécessite de la prudence et de l'expérience clinique, notamment en clinique génétique. Il y a en effet une incertitude du diagnostic, une difficulté à prédire l'évolution et une absence de traitements efficaces. Cette prudence fait partie intégrante des formulaires de consentement destinés aux personnes sans symptômes ou avec des symptômes minimes. Cela pose la question du rapport entre la recherche biomédicale et le soin ou, plus généralement, l'accompagnement des personnes. Communiquer une incertitude est une responsabilité forte dans la transmission des savoirs.

Lorsqu'on fait de la médecine de précision ou de la médecine statistique, on s'adresse à tout autre chose que la personne. Il faut faire l'effort de réintroduire la personne, une question restant des plus saillantes, celle de l'autonomie du malade. Est-elle l'indépendance absolue ou est-elle cette capacité inaliénable en l'homme d'agir selon sa raison, capacité qu'aucune circonstance, pas même la maladie, ne peut supprimer ? La question de l'autonomie en situation de grande vulnérabilité se pose, même si l'on peut considérer qu'avec le diagnostic précoce, et alors que la personne ne se sent pas malade, la véritable question est plus celle de sa vulnérabilité que celle de son autonomie. Il faut sans doute distinguer les notions de vulnérabilité et de dignité, deux points forts sur lesquels s'appliquent l'éthique et la sociologie médicales. L'individu ne peut pas être pensé abstraitement comme un être rationnel et autonome, ainsi que le considère John Rawls dans *La Théorie de la justice*⁴⁰⁹, mais plutôt comme une personne ayant des capacités, comme le suggère Amartya Sen. Une maladie comme

⁴⁰⁹ John Rawls, *La justice comme équité, Une reformulation de Théorie de la justice*, Traduction française, Paris, Editions la Découverte, 2008.Paris,

la maladie d'Alzheimer est le révélateur d'un nouvel équilibre à trouver, d'un point de vue épistémologique, entre savoirs hétérogènes et, d'un point de vue éthique, entre espérance et responsabilité.

XIII-3 Les enjeux économiques et sociaux

La philosophie de la technologie ne prend son sens que dans un contexte socio-politique. Les technologies ne sont pas neutres. Toute invention a des conséquences sociales et politiques, mais ces conséquences sont complexes. La production de connaissances est contextualisée et dépend des pratiques.

Faciliter la détection précoce de la maladie d'Alzheimer, malgré l'absence de traitements efficaces, rentre dans ce cadre. Elle comporte certes un avantage thérapeutique car les traitements en cours de tests montrent une certaine efficacité sur les malades précoces et offre la possibilité aux personnes malades et à leur entourage d'anticiper leur avenir. C'est tout l'objet des nouveaux critères de diagnostic de la maladie d'Alzheimer, introduits en 2007, qui sont, selon le Professeur Bruno Dubois, particulièrement importants. Avec ces critères des chercheurs se sont aperçus que 36% de leurs patients inclus dans un essai thérapeutique sur la base d'anciens critères cliniques n'avaient pas la maladie d'Alzheimer. Ajoutons une autre donnée significative : 97% des dépenses de santé pour la réparation, 3% pour la prévention⁴¹⁰.

Par ailleurs, existe-t-il une opposition entre le paternalisme médical et le patient acteur de son parcours, l'émergence de la médecine de précision impactant tout l'écosystème de la santé ?

⁴¹⁰ Jean-Claude Ameisen durant la première journée Numérique et Médecine de la Fédération des spécialités médicales, Paris, le 7 juin 2016.

En fait, il y a un jeu entre le malade, le médecin et l'institution qui le prend en charge. Il faut introduire les déterminants socio-économiques de la santé. Si l'annonce du diagnostic ou la qualité de vie par exemple sont de vrais problèmes éthiques liés à la maladie d'Alzheimer, il ne faut pas occulter les problèmes sociaux, voire sociétaux, et politiques qui lui sont associés. Dans tous les cas, la connaissance des situations est indispensable. L'identité d'une personne, du point de vue médical, permet de considérer sa responsabilité morale, la prise en compte de ses directives anticipées, et son état, sans oublier le lien avec l'épigénétique, nécessaire quand on cherche à identifier une personne. L'épigénétique est directement liée à son phénotype et les processus qui régulent l'expression des gènes caractérisent la persistance de la personne.

La pratique médicale vit une époque charnière associée à trois grandes tendances :

- Une tendance scientifique avec le développement de toutes les '*omics*', et notamment la généralisation de l'utilisation des données génétiques dans la pratique clinique.
- Une tendance sociologique, avec l'informatisation des dossiers médicaux, la multiplication des données personnelles et l'interopérabilité entre les systèmes.
- Une tendance économique avec une explosion des coûts de santé qui crée une pression toujours plus forte sur les prix et les conditions de remboursement.

De fait, les changements technologiques influencent non seulement les pratiques médicales, mais également l'organisation des systèmes de santé et les relations entre une personne/patient et les professionnels de santé. Le développement des technologies a un impact considérable sur la vie économique et sociale. Ayant introduit la nécessité de l'anticipation, il faut rappeler que nous

sommes maîtres de la technologie et devons la définir en formulant le projet de société que nous voulons. C'est effectivement le cas pour la maladie d'Alzheimer qui constitue un problème de santé publique majeur d'autant plus préoccupant qu'il n'existe à ce jour aucun remède. Sans doute faut-il mieux caractériser ce que veut dire anticiper et s'interroger sur le fait de savoir si anticiper c'est toujours surveiller. Dans une société démocratique l'autonomie de la personne⁴¹¹ est une des premières valeurs et pourtant la société est une société de l'hyper-contrôle (la vidéosurveillance se développe partout, chaque individu apparaît en moyenne dans 400 fichiers). C'est un véritable paradoxe qui nécessite de trouver le type de partage entre ce qui est pertinent et ce qui atteint la vie personnelle de l'individu (sans oublier que, dès qu'il y a automatisation de la collecte des informations, il n'y a plus de consentement). S'y ajoute la question de l'anonymat des données, car on peut facilement les dé-anonymiser⁴¹². De la même façon, prédire et surveiller, ce n'est pas la même chose et il convient d'introduire une dimension temporelle dans la biomédecine. Le vivant acquiert une valeur d'échange en devenant *Big Data*, c'est-à-dire masse de données collectées par des technologies très haut débit, stockées dans des *data bases* et traitées par des algorithmes. Sur un plan technique, les premières étapes d'un travail de recherche consistent à transférer, stocker, sécuriser et analyser des données. La notion de donnée utile est d'une importance capitale. S'il faut rappeler les principes de la CNIL, toujours d'actualité :

- Principe de finalité ;
- Principe de pertinence des données ;
- Principe du droit à l'oubli ;
- Principe du respect des personnes ;
- Principe de sécurité.

⁴¹¹ L'autonomie de la personne ne peut être appréhendée sans lien avec les valeurs promues pour le vivre ensemble.

⁴¹² M. Gymrek et al., 2013, *Science*, 339 (6117) : 321-324.

Il faut néanmoins noter qu'ils ne sont pas toujours faciles à respecter avec les *Big Data* car :

- La finalité du rassemblement des données n'est pas possible ;
- La pertinence non plus (dans le cas du *data mining*, on sort quelque chose de la fouille, mais on ne sait pas à l'avance que les données sont pertinentes).

Le seul principe qui demeure de façon ferme est le principe de sécurité. On peut aussi ajouter que, compte-tenu des investissements que nécessitent la biologie contemporaine et la biomédecine, les questions sur la propriété des données et sur leur auteur, ainsi que les responsabilités qui sont associées à l'utilisation des données, deviennent de plus en plus significatives. De fait, les données ont une valeur financière, voire une valeur affective. Mais le fait que les données soient accessibles ne signifie nullement qu'elles soient analysées. L'accès à l'information ne dispense pas de son apprentissage. Ainsi, disposer d'une encyclopédie médicale ne fait pas un médecin pour autant. Dans une perspective de médecine génomique, les médecins ont besoin d'outils capables d'extraire l'information pertinente de la profusion de données qui sont ou seront disponibles.

Enfin, ajoutons que, bien que statistiquement soit un mot abominable quand on parle de relations humaines, on peut noter que chaque année, pas moins de 1,2 milliard de feuilles de soins, cinq cents millions d'actes médicaux et onze millions de séjours hospitaliers transitent par le système national d'information inter-régimes de l'assurance maladie (*Sniliram*). Cette base comprend des données à caractère personnel d'une part, des données

agrégées⁴¹³ (*open data*) qui ne représentent aucun risque pour la vie privée d'autre part.

Si les données ont une valeur scientifique dès l'instant où elles sont considérées comme indispensables à la connaissance, elles ont aussi une valeur politique dès l'instant où elles sont utilisées pour légitimer des politiques ou des accords entre gouvernements ou institutions.

La médecine met en jeu des subjectivités, des intérêts politiques, économiques, éventuellement idéologiques, ainsi que des représentations. Sans occulter la prégnance actuelle de la sphère économique dans toutes les décisions et actions, il est du devoir des professionnels de santé de démontrer l'intérêt de leur démarche et son utilité pratique. Le contenu des savoirs, – et l'épistémologie des biomarqueurs en est un élément essentiel dans les pratiques médicales –, est une composante de la responsabilité médicale, indépendamment de l'usage que l'on pourra faire de ces savoirs. Sans préjuger si tout ce qui est connaissable doit être connu, il faut s'interroger sur les savoirs d'anticipation et sur l'économie de la promesse qui accompagne toute incertitude.

⁴¹³ Les données agrégées sont des données traitées afin d'obtenir des informations anonymes sur des groupes d'individus ayant des caractéristiques communes, par exemple des hommes de cinquante ans présentant un diabète de type 2.

CONCLUSION ET PERSPECTIVES

Conclusion et perspectives

« Si tu ne sais pas où tu vas, tu termineras à coup sûr à un autre endroit »⁴¹⁴ interpellation à laquelle on peut associer la métaphore de l'horizon⁴¹⁵, le propre de l'horizon, étant qu'il pose une direction d'une part, qu'il s'en éloigne au fur et à mesure que l'on s'en rapproche d'autre part. Cette interpellation est d'actualité dans un contexte où nous nous interrogeons sur la signification des biomarqueurs et sur les retombées qu'ils impliquent dans la pratique médicale, en mettant en avant une accumulation de données diversifiées, non structurées, indépendantes les unes des autres, acquises sans hypothèse préalable. Les biomarqueurs sont aujourd'hui considérés, d'une part dans le domaine de la recherche pour comprendre des mécanismes biologiques ou des pathologies, d'autre part dans le domaine médical pour caractériser une maladie, la prédire, l'anticiper, c'est-à-dire pour une pratique médicale moléculaire. Ces biomarqueurs sont des composantes fondamentales pour faire face au vrai défi auquel est confrontée l'évolution des pratiques médicales, à savoir le transfert des connaissances scientifiques en biologie à la clinique.

⁴¹⁴ Interpellation de Lawrence « Yogi » Berra, célèbre joueur de baseball américain à ses coéquipiers, in CURY Philippe et PAULY Daniel, « Mange tes méduses ! Réconcilier les cycles de la vie et la flèche du temps », Paris, Editions Odile Jacob, 2013, page 135.

⁴¹⁵ J. Ladrière, *Les enjeux de la rationalité. Le défi de la science et de la technologie aux cultures*, Paris, Aubier, Unesco, 1977.

Cette interrogation fait ainsi partie de la problématique de l'homme, de sa santé, de son bien-être pour retenir la définition de l'OMS⁴¹⁶. En fait, la personne est une mosaïque de caractéristiques biologiques, et plus encore, qui amène à prendre en compte la question de la complexité, en introduisant notamment la notion de réseaux. Le vivant est une totalité dynamique qui n'est pas réductible à un ensemble de fonctions. On peut rappeler ce qu'écrit Henri Atlan⁴¹⁷ : « le vivant résout des problèmes, non pas seulement en s'adaptant, c'est-à-dire en modifiant sa relation au milieu (comme une machine peut le faire), mais en se modifiant lui-même, en inventant des structures internes nouvelles ». Comment, dans ce contexte, prendre soin de ces patients incurables dont la maladie progresse inexorablement, si ce n'est en détectant suffisamment tôt la pathologie afin de retarder le plus possible l'apparition de signes cliniques ? Ces patients assistent impuissants à la perte de leur autonomie mais ont-ils perdu leur normativité propre (normativité au sens de Canguilhem, à savoir capacité à créer un régime de vie), même si dans le cas de la maladie d'Alzheimer on peut considérer que la perte de normativité propre du patient est considérable ?

La médecine considère la personne comme un tout, prenant en compte non seulement ses caractéristiques moléculaires, mais aussi ses composantes individuelles, familiales, sociales, culturelles et professionnelles. C'est dire que caractériser une personne nécessite d'utiliser des marqueurs présentant une très grande hétérogénéité qui dépasse largement le seul domaine

⁴¹⁶ <http://who.int>. Pour l'Organisation mondiale de la santé, « La santé est un état de complet bien-être physique, mental et social et ne consiste pas seulement en une absence de maladie ou d'infirmité ». Toutes les composantes d'un sujet sont donc prises en compte et, suivant cette définition qui date de 1948, personne ne peut réellement se dire en bonne santé.

⁴¹⁷ Henri Atlan, *Le vivant post-génomique, ou qu'est-ce que l'auto-organisation ?*, Paris, Editions Odile Jacob, 2011, page 10.

des sciences de la vie. Ce n'est cependant pas considérer que la médecine ne peut être qu'une médecine holiste. C'est de la diversification des tests et de leur précision que l'on peut attendre une prise en compte de plus en plus fine des différences individuelles. Derrière le terme "environnemental" sont regroupés tous les déterminants non génétiques dans une maladie "complexe". Le terme "environnemental" est donc bien à prendre au sens très large : à la fois dans son sens écologique (l'environnement que subit ou a subi le patient) mais aussi dans un sens plus individuel (ce que le patient vit ou a vécu, compte tenu de son comportement présent et passé).

Le vrai défi pour l'évolution des pratiques médicales est de transférer les connaissances scientifiques en biologie à la clinique et les biomarqueurs y participent. Leurs utilisations sont variées. Les biomarqueurs peuvent être des biomarqueurs de pronostic pour prédire l'évolution d'une maladie complexe, des biomarqueurs de diagnostic, notamment pour les diagnostics précoces, ou encore des biomarqueurs permettant de suivre les réponses à des traitements thérapeutiques. Ils présentent aussi une très grande diversité. Ces biomarqueurs peuvent être de simples variants génétiques, mais aussi le résultat de la sur- (ou sous-) expression d'un gène, tout autant que la concentration de métabolites, considérée comme anormale par rapport à une concentration de référence.

Mais le biomarqueur, qu'est-il ? C'est un objet, un objet défini par ses relations, un objet dont les relations avec le milieu extérieur font partie du fonctionnement même de l'objet. Le biomarqueur est plus qu'une molécule. C'est une altération, qui a une signification biologique, mais il ne renvoie pas à une situation préalable non-altérée. Il est impossible d'en repérer l'origine. C'est plutôt un signal dont les variations présentent une corrélation statistique

significative avec une pathologie et ces variations doivent pouvoir être détectées grâce à un test validé et utilisable en clinique. Cela implique la mesure simultanée d'un nombre important d'entités, suivant l'état de l'individu et une analyse faisant appel aux statistiques, tout en soulignant que, ce qui est important, c'est d'intégrer⁴¹⁸ toutes les informations, qu'elles soient quantitatives ou qualitatives.

De plus, ainsi que le propose Xavier Guchet⁴¹⁹, le biomarqueur présente une ambivalence certaine puisqu'il peut avoir une double signification :

- Ce peut être un biomarqueur « actionnable » en ce sens qu'il permet le diagnostic, le pronostic et éventuellement l'intervention thérapeutique en identifiant des cibles moléculaires sur lesquelles agir.
- Ce peut être un biomarqueur « signe ou trace » plus associé au mode de vie et à la qualité de vie des personnes, qui intervient dans des cascades d'interactions moléculaires, suivies ainsi à la trace, qui résultent tant de phénomènes internes que de phénomènes externes liés à des événements naturels ou sociaux.

À cette dernière dimension, qui intègre les notions de probabilité et d'incertitude, peut être associée la notion de surveillance, ce qui peut aboutir à l'émergence d'une société complètement contrôlée, comme le développement de la traçabilité des activités quotidiennes (déplacements, achats, consultations sur internet, etc.) des individus en est déjà une illustration.

⁴¹⁸ S. Huang, « The practical problems of post-genomic biology », *Nature Biotechnology*, 2000, 18, pp. 471-472.

⁴¹⁹ Xavier Guchet, *La médecine personnalisée, un essai philosophique*, Paris, Les Belles Lettres, 2016.

En abordant la question des biomarqueurs et de leur nature (molécules, réseaux, interactions, etc.), nous pouvons nous demander si ces biomarqueurs sont des objets donnés ou des objets construits. À n'en pas douter le biomarqueur est un objet de la biologie et de la médecine. Il existe, mais la question de son mode d'existence et la question de savoir comment le caractériser se posent. Le biomarqueur est un objet complexe. Ce n'est pas une molécule isolée, c'est un réseau d'interactions, c'est un algorithme. En ce sens, on ne peut pas lui enlever une de ses caractéristiques sans en modifier sa nature. Par ailleurs le biomarqueur, en plus d'être complexe, est :

- d'abord un objet de connaissance et non une représentation directe de la réalité ;
- ensuite un objet entouré d'incertitude ;
- enfin un objet dont on n'a pas la maîtrise totale, ce qui le rend non synthétisable.

Le biomarqueur peut avoir plusieurs voies explicatives, tout en n'étant jamais complètement donné. Ce n'est pas un objet réel. On pourrait dire que c'est une inconnue sans définition, toujours en devenir !

Quoi qu'il en soit, avant de voir ce que les biomarqueurs apportent dans la connaissance, il faut les identifier et c'est là qu'interviennent les approches et les techniques.

La première approche, plus orientée vers la recherche, se focalise essentiellement sur les aspects génétiques. L'approche qui consiste à déterminer, à partir des mesures d'expressions différentielles de gènes, quels gènes pourraient être impliqués dans les processus pathologiques comme celui de la maladie

d'Alzheimer avant l'apparition des symptômes cliniques, nécessite de construire des modèles de réseaux d'interactions entre tous les constituants moléculaires d'une cellule, notamment des réseaux de régulation entre gènes ou des réseaux d'interactions entre protéines.

La deuxième approche, plus orientée vers la pratique clinique, cherche à intégrer beaucoup plus de données, en associant des données génétiques, biologiques et d'imageries, mais aussi des données sur le mode de vie et l'histoire des personnes. Cette approche multimodale, associée à des modèles biophysiques pour faire des analyses de corrélation, permet d'étudier les interactions entre les différents biomarqueurs susceptibles d'être utilisés pour la prédiction de la survenue de la maladie d'Alzheimer, tout en prenant en compte des marqueurs associés à la nutrition ou au mode de vie par exemple.

Quant aux technologies d'acquisition et d'analyse de grandes masses de données, elles permettent de rassembler les fameuses *Big Data* caractérisées par le volume de données, leur vitesse d'accès, leur diversité, leur variabilité, ainsi que par leur valeur et leur véracité. Les données issues des techniques apportent de nouvelles connaissances et créent, de fait, sans cesse de nouveaux pouvoirs. Avec les *Big Data*, on peut être moins exigeant sur l'exactitude des données, les probabilités l'emportant sur l'exactitude. Ainsi que le relèvent Viktor Mayer-Schönberger et Kenneth Cukier en écrivant « [c'est] l'équivalent d'une peinture impressionniste, où chaque coup de pinceau est déstructuré quand on l'observe de près, mais dont l'ensemble révèle un tableau majestueux lorsqu'on prend du recul pour le regarder »⁴²⁰, on peut

⁴²⁰ V. Mayer-Schönberger, K. Cukier, *Big Data, la révolution des données est en marche*, Paris, Ed. Robert Laffont, 2014, page 63.

accepter une certitude imparfaite si on bénéficie d'un sens plus complet de la réalité.

Les techniques d'apprentissage permettent de passer d'un ordre de la réalité, noyé dans les *Big Data*, à un autre ordre de réalité, basé sur des variables cachées. Leur traduction en médecine est qu'elles permettent de passer d'un individu singulier vu au travers d'un individu standard, à un individu singulier vu dans sa singularité du point de vue médical. Il ne faut cependant pas limiter les *Big Data* à leur dimension technologique. Si l'on suit Boyd et Crawford⁴²¹, les *Big Data* reposent sur l'interaction de technologies, d'approches analytiques et de méthodes faisant émerger de nouvelles données, en comparant de grands ensembles de données hétérogènes, en permettant l'identification de tendances et la génération de nouvelles idées, impossibles à trouver sans ces données en masse. Ajoutons que la majorité des approches s'adressent à l'intérieur du corps. Or de nombreux facteurs externes comme l'influence du niveau socio-économique interviennent. Il faut donc intégrer les analyses de la personne à des analyses plus du ressort de la santé publique.

Mais les biomarqueurs apportent-ils quelque chose à la connaissance ?

D'un côté, on cherche à aborder les questions autour de la causalité et du mécanisme. La question qui se pose est de savoir si l'on peut inférer la structure et les mécanismes de régulation des systèmes biologiques à partir de données portant sur leur comportement, auquel cas l'accumulation de données sur l'expression des gènes, les interactions entre protéines, etc., suffirait pour construire des modèles représentant les mécanismes

⁴²¹ Danah Boyd & Kate Crawford, "Critical questions for Big Data", *Information, Communication & Society*, 15 : 5, pp. 662-679, 2012.

moléculaires sous-tendant le fonctionnement des cellules et des organismes. Pour choisir entre différents modèles possibles et déterminer les paramètres du modèle, il faudrait une quantité de données beaucoup plus grande que ce que l'on produit actuellement dans un contexte *Big Data*. La solution la plus souvent adoptée est en fait de construire des modèles extrêmement simplifiés de ces réseaux sans prendre en compte la nature et les détails des interactions moléculaires. Reconnaître l'incertitude quand on parle des biomarqueurs, c'est reconnaître l'impossibilité de synthétiser leur enchaînement, et cette impossibilité de synthèse tient à l'absence de clarté causale.

Si la biomédecine apparaît plutôt comme pilotée par les données, et par là par des technologies, et leur analyse, que par des hypothèses, il ne faut cependant pas oublier que le choix des données que l'on accumule, des données génomiques ou des données d'expression par exemple, suppose que l'on ait retenu un niveau d'analyse que l'on estime pertinent, et cette pertinence s'appuie sur une hypothèse théorique, le choix que tout est génétiquement déterminé ou que l'expression des gènes est fortement associée à un état ou à un environnement par exemple. Il faut tout repenser sous l'angle des interactions que tout être entretient avec son milieu, c'est-à-dire « en situation » ou, si l'on préfère, « en contexte », un être ne pouvant pas être traité comme une entité isolée. Cela suppose d'associer des modèles de connaissance, le plus souvent disjoints, à savoir celui des sciences dures et de la technologie et celui des sciences humaines. Il faut mettre en relation de la science et de l'humain, tenir compte des hétérogénéités et ne pas privilégier une discipline particulière. Il semble plus utile de ne pas considérer une discipline centrale mais de décentrer toutes les disciplines et de mettre au centre l'objet complexe que l'on cherche à comprendre.

D'un autre côté, au lieu de chercher une causalité, on peut chercher à gérer la réalité qu'est une maladie complexe comme la maladie d'Alzheimer sans pour autant l'expliquer. La médecine veut alors simplement prédire sans se mettre en tête de chercher à expliquer. Elle calcule alors des trajectoires prévisibles à partir des corrélations issues de données recueillies sur des cohortes de patients et des biomarqueurs qui en résultent. Il y a les savoirs – et les biomarqueurs en sont - mais aussi l'organisation des savoirs, et celle-ci est mise en œuvre dans un certain contexte qui, d'une part participe à la définition des questions posées et des moyens d'observation retenus, et qui, d'autre part, influence les valeurs à l'aune desquels les résultats sont interprétés, sans oublier la tendance à surévaluer ce qui est faisable. Clairement, l'information génétique ne peut pas être le seul type d'information utilisé pour fonder une décision clinique, même si on peut admettre qu'un test génétique soit utile sans qu'on ait besoin d'avoir une connaissance complète de la fonction biologique des différents variants génétiques. Ajoutons que la forte prévalence des effets environnementaux et de ceux de la vie sociale, collective et personnelle doit être intégrée dans les analyses sur les maladies complexes ou chroniques.

Les biomarqueurs apportent quelque chose à la connaissance, mais il ne faut pas oublier les liens qui existent entre l'épistémologie, l'éthique et les retombées économiques et sociales. Si la médecine s'appréhende souvent dans « le colloque singulier » entre le médecin et le patient, il ne faut pas oublier de la considérer comme une pratique sociale à part entière, en lien avec une politique de santé publique, un mode spécifique de prise en charge collective des problèmes de santé rencontrés par ses membres, une organisation hospitalière, un appareil législatif et le plus souvent aujourd'hui des instances éthiques. Les enjeux épistémologiques associés aux biomarqueurs et à leur association

dans des indicateurs de « conversion » vers la maladie ne peuvent qu'être associés aux enjeux politiques et industriels, aboutissant à une combinaison de technologies ayant des retombées tant en clinique que dans les politiques publiques de la santé. Au couple médecin/patient, se substitue le couple personne/sa maladie (déclarée ou prévue). Le couple patient/maladie fait que la personne devient acteur et, à ce titre, responsable des coûts des systèmes de santé.

En premier lieu, les problèmes éthiques ne sont pas les mêmes selon le degré d'évolution de la maladie. Il faut donc être attentif au contexte, notamment quand on se place dans une phase asymptomatique d'une maladie à évolution lente, sans solution thérapeutique affirmée. On peut être inquiet de la volonté de prendre en compte les « malades » jeunes en bonne santé. On retrouve des mots-clés : solidarité, vulnérabilité, respect de la personne, exigence que les malades soient acteurs de leur propre histoire, accompagnement, consentement, etc., à la base d'un engagement éthique contribuant à la création collective d'un sens, de finalités communes. Le registre de la probabilité n'existe, de fait, que dans le domaine scientifique. Face à une information ou à une décision thérapeutique, l'individu n'attend pas une probabilité ou une statistique. Il faut donc s'interroger sur l'influence qu'une probabilité peut avoir sur les trajectoires de vie. Dans le cas de la maladie d'Alzheimer, la certitude quant à la manifestation clinique de la maladie et à son évolution est loin d'être avérée, mais, même dans une maladie comme la maladie de Huntington, dont la pénétrance génétique est de 100 %, la revendication des porteurs de la mutation est que ça ne détermine pas leurs trajectoires de vie.

Les approches techniques ne doivent pas occulter qu'en médecine on s'adresse à un individu singulier et que cet individu souhaite être partie prenante de ce qui le concerne. Est ainsi

apparue la notion de « médecine participative », coopération entre le(s) patient(s) et le(s) médecin(s). Cette évolution s'accompagne d'une diversification des intervenants dans le processus de médicalisation. Rappelons ce qu'écrit Joëlle Vailly, « [La « biomédicalisation »] désigne le processus d'hybridation des approches technoscientifiques liées aux sciences de la vie d'une part et des pratiques cliniques d'autre part. [...] Il n'implique évidemment pas la disparition de toutes les approches plus anciennes de la médecine, mais plutôt une série d'ajouts de nouvelles pratiques, de transformations et de remplacements. L'une des caractéristiques de la biomédicalisation est donc liée à l'émergence de nouveaux savoirs, avec un effacement de la frontière entre sciences, techniques et médecine »⁴²², sans pour autant se dégager « de toute contingence et conséquences sociales »⁴²³. La médecine participative est plus centrée sur le patient que sur la maladie et s'appuie plus sur des facteurs de risques environnementaux. C'est plus une médecine « préventive » qui retient que, de façon générale, tout être humain est en relation avec son milieu, que ce soit son milieu naturel ou son milieu social, et que cette relation est dynamique.

En complément, les biomarqueurs, et les nouvelles pratiques qui leur sont associées, ont des effets sur les politiques des systèmes de santé et suscitent des interrogations morales. Passer du droit d'être soigné au devoir de ne pas être malade peut être vu comme un changement radical. Ce qu'écrit Marie Gaille a déjà été souligné, « une interrogation philosophique sur la médecine qui serait aveugle à l'impact de l'évolution du savoir et des technologies sur les questions éthiques, politiques et juridiques posées par la médecine en manquerait une dimension essentielle. [...] Il convient [...] de voir dans la médecine autre chose que l'application d'un

⁴²² Joëlle Vailly, *Naissance d'une politique de la génétique – Dépistage, biomédecine, enjeux sociaux*, Paris, Vrin, 2011, page 8.

⁴²³ Ibid., page 67.

savoir biologique, mais une pratique qui a trait à un savoir sur l'homme, et un homme qui n'est pas considéré indépendamment de son contexte social et politique »⁴²⁴. En complément ; la métaphore du « pouvoir pastoral », empruntée à Michel Foucault, illustre bien un type de pouvoir orienté à la fois vers l'ensemble du troupeau dont le berger prend soin et vers chacune des brebis qui le composent. Elle est complétée en disant que la médecine est de moins en moins une médecine clinique reposant sur des symptômes et qu'elle se transforme de plus en plus en une médecine préventive orientée vers un dépistage de la conversion de facteurs de risque vers le démarrage de la maladie. Dans le cas de la maladie d'Alzheimer, par exemple, il s'agit de passer de la notion de risques de développer la maladie, risques associés à des biomarqueurs de la bêta-amyloïde ou de la protéine Tau, à la notion de démarrage de la maladie, de façon précoce bien avant que la personne ne le soupçonne.

Nous avons placé notre travail sous l'angle de l'épistémologie des biomarqueurs en nous appuyant sur la maladie d'Alzheimer et cette approche permet d'aborder des questions éthiques et sociales à partir des programmes de recherche, des technologies, de l'épistémologie des données. Cela revient à positionner la médecine du point de vue de la philosophie des techniques. Mais, pour poursuivre le travail, ne devons-nous pas renverser la question et nous demander en quoi la maladie d'Alzheimer nous oblige à repenser la philosophie des techniques. Les choix techniques sont des choix de société, une société de confiance, une confiance loyale et vigilante.

Qu'il s'agisse des biomarqueurs ou des technologies à usage médical, comme les nanobiotechnologies ou les techniques

⁴²⁴ Marie Gaille, *Philosophie de la médecine – Frontière, savoir, clinique*, Paris, Vrin, 2011, pp. 24-25.

d'imagerie par exemple, il faut se méfier de « l'inflation des promesses »⁴²⁵, tout en considérant le rôle des promesses dans l'acceptation de nouvelles approches ou de nouvelles technologies. Il faut aller vers la définition d'indicateurs personnalisés. On peut faire sien ce que dit Edgar Morin : « je crois profondément qu'il faut espérer l'inespéré, ou du moins espérer l'improbable. [...] Le probable n'est pas le certain »⁴²⁶ et dépasser la contradiction entre l'espérance d'un Ernst Bloch⁴²⁷ et la responsabilité d'un Hans Jonas⁴²⁸, la contradiction entre l'heuristique de la peur et les pensées utopistes, pour entrevoir une espérance avec responsabilité.

Cette thèse a développé une réflexion épistémologique autour de la notion de biomarqueurs dans le cas de la maladie d'Alzheimer. Elle est centrée sur le transfert des connaissances du domaine de la biologie à celui de la pratique médicale et clinique, avec une attention particulière aux techniques de diagnostic précoce et aux *Big Data*. Ce travail n'est cependant pas une fin en soi, mais une étape dans un parcours à continuer.

Nous avons insisté sur le fait que la maladie n'est pas que cellules, chiffres, data ou autres données objectives, mais qu'elle est aussi « exposée » à l'environnement ; dont le sujet, par son mode de vie, ses valeurs ou encore la perception qu'il a de la maladie, fait partie. Cela nécessite de pousser plus loin la réflexion sur la médecine participative.

⁴²⁵ Marc Audétat (Dir.), *Sciences et technologies émergentes : Pourquoi tant de promesses ?*, Paris, Hermann, 2015.

⁴²⁶ Edgar Morin, *Les philosophes*, Paris Germina, 2011, page 30.

⁴²⁷ Ernst Bloch, *Le Principe d'espérance ou l'Esprit de l'utopie*, Gallimard

⁴²⁸ Hans Jonas, *Le Principe de responsabilité*, Paris, Les éditions du Cerf, 1990.

Les biomarqueurs et leur statut épistémique font partie des « solutions » d'avenir à creuser dans le contexte d'une médecine participative, mais l'anticipation est au cœur de notre réflexion. L'anticipation est une action, une décision ou un comportement orienté vers le futur sur la base d'une prévision, qu'elle soit implicite ou explicite. Elle se différencie de la prédiction qui est la représentation d'un événement futur précis, nécessitant néanmoins de rendre plus nette la place de la probabilité versus la statistique au sein d'une éthique épistémologiquement réévaluée des biomarqueurs appliquée aux maladies évolutives, notamment les maladies neuro-évolutives.

Cela nous conduira à prolonger nos réflexions selon d'autres perspectives. Comment articuler une épistémologie de l'anticipation à une éthique de l'anticipation ? Comment définir le devenir d'un malade, qui n'est pas encore un patient, par anticipation ? Comment mieux comprendre les enjeux entre complexité et incertitude ? Telles sont des questions que nous continuerons à aborder, et la philosophie nous y aidera.

GLOSSAIRE

GLOSSAIRE

ADN : acide désoxyribonucléique, molécule ayant la structure d'une double hélice et représentant le support chimique de l'hérédité.

Allèle : variation d'un locus (position sur le chromosome où l'on trouve un gène) génétique à l'intérieur de l'espèce ; variante d'un même gène.

Anamnèse : renseignements fournis par le patient ou son entourage, destinés au médecin et concernant le début de sa maladie.

Anthropocentrisme : conviction selon laquelle l'existence de l'homme est la raison de l'existence de l'univers.

ARN : acide ribonucléique, de structure très proche de l'ADN, molécules jouant un rôle clé dans la première phase de la synthèse des protéines.

Cellule : unité structurelle la plus petite des organismes vivants.

Chromosome : pelote d'ADN portant les gènes.

Clonage : obtention d'un être vivant génétiquement identique à l'original qui lui donne naissance.

Clone : famille de cellules ou d'organismes génétiquement identiques.

Complexité : caractéristique d'un système dont le comportement est potentiellement imprévisible.

Contextualisme : série de conceptions qui prétendent qu'un phénomène ne peut être correctement compris que dans le contexte dans lequel il se déroule.

Dédictionisme : vision selon laquelle la connaissance scientifique doit être acquise d'abord en proposant une hypothèse, puis en essayant de prouver qu'elle est fausse par des tests empiriques sur des données observables.

Déterminisme : proposition selon laquelle tous les faits, y compris les pensées humaines, ont une cause déterminée par une chaîne ininterrompue d'événements qui les ont précédés.

EBM (*Evidence-Based Medicine*) : approche qui s'efforce de fonder, autant que possible, les décisions cliniques sur les données actuelles les plus probantes issues de la recherche médicale, ces données étant issues d'études cliniques systématiques, en particulier les essais contrôlés, randomisés, et les méta-analyses.

Empirisme : doctrine scientifique disant que toute connaissance provient en dernière analyse de l'expérience sensorielle et de l'évidence observable, plutôt que de l'intuition ou des idées préconçues.

Épigénétique : branche de la biologie qui étudie les relations de cause à effet entre les gènes et leurs produits, faisant apparaître le phénotype. L'épigénétique est le domaine qui étudie comment l'environnement et l'histoire individuelle influent sur l'expression des gènes.

Exposome : ensemble des expositions auxquelles est soumis un individu, que ce soit dans son environnement, son alimentation ou son mode de vie.

Expression des gènes : synthèse de l'ARN messager, suivie par la synthèse des protéines. L'expression d'un gène représente le fonctionnement de ce gène.

Gène : unité moléculaire d'information, constituée d'un fragment d'ADN, qui sera traduite par la synthèse d'une protéine, et que les cellules consultent en permanence pour savoir comment fonctionner de façon adaptée.

Génome : ensemble du matériel génétique (environ 25 000 gènes) contenu dans une cellule.

Génotype : information portée par le génome d'un organisme.

Holisme : théorie selon laquelle les propriétés d'un système ne peuvent pas être comprises uniquement par la somme des parties qui le constituent, mais par la manière dont le système se comporte comme un tout.

Inductivisme : vision selon laquelle le progrès scientifique est guidé par différentes observations effectuées lors d'expériences antérieures, et que les manifestations de la vérité doivent être basées sur l'ensemble des informations.

Instrumentalisme : idée selon laquelle la vérité de la connaissance est non pertinente, et que la connaissance doit être jugée à l'aune de son utilité.

Mécanicisme : théorie selon laquelle tous les phénomènes naturels peuvent être expliqués par des causes physiques et peuvent être compris par les éléments qui les composent et leurs fonctions, plutôt que par une interprétation holistique.

Métabolome : ensemble des interactions et des changements physico-chimiques qui se déroulent dans un organisme vivant.

Métadonnées : données fournissant des renseignements sur d'autres données.

Mutation : modification structurelle de l'ADN.

Mutation ponctuelle (SNP - *Single Nucleotide Polymorphism*) : mutation de l'ADN ne portant que sur une base nucléotidique).

Ontologie : «étude de la nature réelle de cui nous entoure et du sens de la vie, domaine philosophique qui se concentre sur l'étude de l'être.

Pharmacogénomique : prise en compte des mécanismes généraux permettant d'expliquer les variations individuelles de la réponse aux thérapies.

Phénotype : ensemble des caractères observables d'un individu.

Polymorphisme : mot indiquant que les gènes existent sous différentes formes.

Pragmatisme : affirmation selon laquelle le côté pratique doit être le premier composant de la vérité et que la pratique et la théorie doivent être mises en relation.

Prévalence : nombre de cas d'une maladie dans la population.

Protéines : macromolécules élémentaires qui donnent à toute cellule sa forme, sa structure et ses fonctions.

Protéome : ensemble des protéines présentes dans une cellule, dans un état donné et à un instant donné.

Réductionnisme : idée que la nature des choses complexes peut toujours être réduite et expliquée par des vérités plus simples et plus fondamentales.

Systémique : science des systèmes qui vise à l'élaboration et au développement des méthodes de modélisation des phénomènes complexes.

Téléologie : étude de la finalité.

Transcriptome : ensemble des ARN messagers, photocopies des gènes, selon une structure continue, avant leur traduction en protéines.

Vitalisme : doctrine selon laquelle des « forces vitales », souvent assimilées à l'âme, sont actives dans les organismes vivants et que

la vie ne peut pas être expliquée uniquement par la matière ou le mécanisme.

BIBLIOGRAPHIE

Philosophie et sciences humaines

Abastado P (2007) *L'impasse du savoir, Essai d'épistémologie médicale*, Sèvres, Editions EDK.

Anderson C (2008) "The End of Theory : The Data Deluge Makes the Scientific Method Obsolete", *Wired*, june 2008.

Atlan H (1996) *La fin du tout génétique, vers de nouveaux paradigmes en biologie*, Paris, INRA.

Atlan H (2011) *Le vivant post-génomique, ou qu'est-ce que l'auto-organisation ?*, Paris, Editions Odile Jacob.

Atlan H, Vergely B (2012) *Sommes nous libres ?*, Paris, Salvator.

Atlan M, Droit R-P (2012) *Humain, une enquête philosophique sur ces révolutions qui changent nos vies*, Paris, Flammarion.

Audétat M (dir.) (2015), *Sciences et technologies émergentes : pourquoi tant de promesses ?*, Paris, Editions Hermann.

Bacqué et Biewener C (2013), *L'empowerment, une pratique émancipatrice*, Paris, Editions La Découverte.

Barberousse A, Morange M and Pradeu T, editors (2009) *Mapping the Future of Biology, Evolving Concepts and Theories*, Boston studies in the philosophy of science, Springer.

Barthélémy J-H (2014) *Simondon*, Paris, Les Belles Lettres.

Beauchamp T L, Childress J F (2008), *Les principes de l'éthique biomédicale*, Paris, Les Belles Lettres.

Bechtel W & Abrahamsen A (2005) "Explanation : a mechanist alternative", *Studies in the History and Philosophy of the Biological and Biomedical Sciences* 36 : 421-441.

Beck U (1986/2001 pour la trad. française), *La société du risque, sur la voie d'une autre modernité*, Paris, Editions Flammarion.

Benasayag M, Gouyon P-H (2012), *Fabriquer le vivant ?*, Paris, Editions La Découverte.

Bensaude-Vincent B (2009), *Les vertiges de la technoscience – Façonner le monde atome par atome*, Paris, Editions La Découverte.

Bensaude-Vincent B, Benoit-Browaeys D (2011), *Fabriquer la vie, Où va la biologie de synthèse ?*, Paris, Editions du Seuil.

Bergé C (2010), *L'Odyssée de la mémoire*, Paris, Editions La Découverte.

Bergson H (1932), *Les deux sources de la morale et de la religion*, Paris, Presses Universitaires de France.

Bergson H (2008) *L'évolution créatrice*, Paris, Quadrige/Presses Universitaires de France.

Boyd D & Crawford K (2012) "Critical Questions for Big Data. Provocations for a cultural, technological, and scholarly phenomenon", *Information, Communication and Society* 15 (5) : 662-679.

Canali S (2016), « Big Data, epistemology and causality : Knowledge in and knowledge out in Exposomics », *Big Data & Society*, July-December 2016 : 1-11.

Canguilhem G (1965), *La connaissance de la vie*, Paris, Librairie Philosophique J. Vrin.

Canguilhem G (1966), *Le normal et le pathologique*, Paris, Presses Universitaires de France.

Carosella E D, Pradeu T (2010), *L'identité, la part de l'autre – Immunologie et philosophie*, Paris, Editions Odile Jacob.

Cerezuelle D (2011) *La technique et la chair, Essais de philosophie de la technique*, Lyon, Editions Paragon.

Chalmers A F (1976), *What is this thing called science ? An assessment of the nature and status of science and its methods*, St Lucia, University of Queensland Press.

Chemero A (2003) "An outline of a theory of affordances", *Ecological Psychology* 15 (2) : 181-195.

Chneiweiss H (2006), *Neurosciences et neuroéthique, Des cerveaux libres et heureux*, Paris, Alvik Editions.

Cooper M (2012) « The pharmacology of distributed experiment – user-generated drug innovation », *Body & Society* 18 : 18-43.

Coutellec L (2015), *La science au pluriel, Essai d'épistémologie pour des sciences impliquées*, Versailles, Editions Quae.

Coutellec L et Weill-Dubuc P-L (2016), « Les figures de l'anticipation, ou comment prendre soin du futur », *Revue Française d'Ethique Appliquée* 2 : 14-18.

Crawford M B (2010), *Éloge du carburateur, Essai sur le sens et la valeur du travail*, Traduction française M Saint-Upéry, Paris, Editions La Découverte.

Debru C (2009), *Penser l'inconnu ? La recherche en biologie*, Paris, Editions Hermann.

Desplan J (2014) *La médecine de santé, Prédictive, Préventive, Personnalisée, Participative*, Paris, Editions Frison-Roche.

Dewey J (1938) *Logic : The Theory of Inquiry*, New York, Holt, Rinehart and Winston.

Draperi C (2010) *La médecine réfléchie au miroir des sciences humaines*, Paris, Ellipses Editions.

Engel P (2007), *Va savoir ! De la connaissance en général*, Paris, Editions Hermann.

Ewald F (2014) *Big data, chance ou malédiction*, conférence à la fondation Calouste-Gulbenkian, Paris le 5 juin 2014.

Fagot-Largeault A (1989) *Les causes de la mort : histoire naturelle et facteurs de risques*, Paris, Librairie Philosophique J. Vrin.

Fagot-Largeault A (2002) *Philosophie des sciences biologiques et médicales*, Cours au Collège de France 2001-2002, https://www.college-de-france.fr/media/anne-fagot-largeault/UPL8928_AFL_res_0102.pdf.

Fagot-Largeault A (2010), *Médecine et philosophie*, Paris, Presses Universitaires de France.

Fagot-Largeault and al. (dir) (2008) *Philosophie et Médecine, en hommage à Georges Canguilhem*, Paris, Librairie Philosophique J. Vrin.

Fantini B et Lambrichs L (dir.) (2014) *Histoire de la pensée médicale contemporaine. Evolutions, découvertes, controverses*, Paris, Editions du Seuil

Fleck L (2005) *Genèse et développement d'un fait scientifique*, Traduction française par N Jas, Paris, Les Belles Lettres.

Foucault M (1963), *Naissance de la clinique*, Paris, Presses Universitaires de France.

Foucault M (1966), *Les mots et les choses*, Paris, NRF Gallimard.

Foucault M (1994) *Dits et Ecrits II, 1976-1988*, Paris, Editions Gallimard.

Fox Keller E (2003), *Le siècle du gène*, Paris, Gallimard.

Fox Keller E (2011), « Towards a science of informed matter », *Studies in History and Philosophy of Biological and Biomedical Sciences* 42 : 174-179.

Fruteau de Lacos F (2009), *L'épistémologie d'Emile Meyerson, une anthropologie de la connaissance*, Paris, Librairie Philosophique J. Vrin.

Fullford K W M, Peile E and Carroll H (2012), *Essential Values Based Practice*, Cambridge, Cambridge University Press.

Gaïlle M (2011) *Philosophie de la médecine – Frontière, savoir, clinique*, Paris, Librairie Philosophique J. Vrin.

Ginzburg C (2010) *Mythes, emblèmes, traces. Morphologie et histoire*, Paris, Verdier.

Giroux E (2006) *Epidémiologie des facteurs de risque : Genèse d'une nouvelle approche de la maladie*, Villeneuve d'Ascq, Diffusion ANRT (Atelier national de reproduction des thèses).

Giroux E (2010) *Après Canguilhem, définir la santé et la maladie*, Paris, Presses Universitaires de France.

Goffette J (2006), *Naissance de l'anthropotechnie, De la médecine au remodelage de l'humain*, Paris, Librairie Philosophique J. Vrin.

Griffiths P E (2001), « Genetic information : A metaphor in search of a theory », *Philosophy of Science* 68 (3) : 394-412.

Guchet X (2016) *La médecine personnalisée, un essai philosophique*, Paris, Les Belles Lettres.

Gzil F (2009), *La maladie d'Alzheimer : problèmes philosophiques*, Paris, Presses Universitaires de France.

Gzil F et Hirsch E (2012) *Alzheimer, éthique et société*, Paris, Editions Eres.

Hempel C (1966), *Philosophy of Natural Science*, Englewood Cliffs, Prentice Hall.

Hedgecoe A and Martin P (2003) "The Drugs don't work : Expectations and the Shaping of Pharmacogenetics", *Social Studies of Science* 33 (3) : 327-364.

- Hedgecoe A (2004) *The Politics of Personalised Medicine, Pharmacogenetics in the Clinic*, Cambridge, Cambridge University Press.
- Heidegger M (1954), *Essais et conférences*, Paris, Gallimard, 2014.
- Hey T and al. editors (2009) *The fourth paradigm : data-intensive scientific discovery*, Redmond, WA, Microsoft Research.
- Hill A B (1965) "The Environment and Disease : Association or Causation ?", *Proceedings of the Royal Society of Medicine* 58 : 295-300.
- Hofmann B (2001) "On the Trial Disease, Illness and Sickness", *Journal of Medicine and Philosophy* 6 : 651-673.
- Hooker C (ed) (2011) *Philosophy of complex systems*, Amsterdam, Elsevier.
- Hottois G (1988), *Evaluer la technique*, Paris, Librairie Philosophique J. Vrin.
- Hottois G (2004), *Philosophies des sciences, philosophies des techniques*, Paris, Editions Odile Jacob.
- Hottois G et Missa J-N (2001) *Nouvelle encyclopédie de bioéthique. Médecine, Environnement, Biotechnologie*, Bruxelles, De Boeck Université.
- Ihde D (2004) "Has the philosophy of technology arrived ? A state-of-the-art review", *Philosophy of Science* 71 : 117-131.
- Illari P, Russo F (2014), *Causality*, Oxford, Oxford University Press.
- Jonas H (1990), *Le principe de responsabilité, une éthique pour la civilisation technologique*, Paris, Les Editions du Cerf.
- Krämer S, Trad ; Chamayou-Kuhn (2012), « Qu'est-ce donc qu'une trace, et quelle est sa fonction épistémologique ? Etat des lieux », *Trivium*, [http ;//trivium.revues.org/4171](http://trivium.revues.org/4171).

- Kuhn T (1962), *The structure of Scientific Revolutions*, Chicago, The University of Chicago Press. Traduction française (2006) *La structure des revolutions scientifiques*, Paris, Editions Flammarion.
- Ladrière J (1977) *Les enjeux de la rationalité. Le défi de la science et de la technologie aux cultures*, Paris, Aubier/Unesco.
- Lakatos I (1978) *The methodology of scientific research programmes : Philosophical paper volume 1*, Cambridge, Cambridge University Press.
- Latour B (2010), *Cogitamus, Six lettres sur les humanités scientifiques*, Paris, La Découverte.
- Lemoine M (2017), *Introduction à la philosophie des sciences médicales*, Paris, Hermann.
- Leonelli S (2012) "Making sense of data-driven research in the biological and biomedical studies", *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 : 1-3.
- Leonelli S (2013) "Integrating data to acquire new knowledge : three modes of integration in plant sciences", *Studies in History and Philosophy of Biological and Biomedical Sciences* 44 : 503-514.
- Leonelli S (2016), *Data-centric biology, a philosophical study*, Chicago, The University of Chicago Press.
- Malherbe M (2015) *Alzheimer – la vie, la mort, la reconnaissance*, Paris, Librairie Philosophique Vrin.
- Mayr Ernst (1993) "Proximate and ultimate causations", *Biology and Philosophy* 8 : 93-94.
- Mitcham C (1994) *Thinking through technology*, Chicago, University of Chicago Press.
- Morin E (1992) *Science avec conscience*, Paris, Editions Fayard.
- Morin E (2005), *Introduction à la pensée complexe*, Paris, Editions du seuil.

- Morin E (2011) *Mes philosophes*, Paris, Germina.
- Monod J (1970), *Le hasard et la nécessité : essai sur la philosophie naturelle de la biologie moderne*, Paris, Editions du Seuil.
- Morange M (2005), *Les secrets du vivant, contre la pensée unique en biologie*, Paris, Editions La Découverte.
- Nagel E et al. (Ed.) (1962), *Logic, Methodology and Philosophy of Science : Proceedings of the 1960 International Congress*, Stanford, CA, Stanford University Press.
- Nelson R R (2008) "What enables rapid economic progress : what are the needs institutions", *Research Policy* 37 : 1-11.
- Nelson N C and al. (2013) "On Being 'Actionable' : Clinical Sequencing and the Emerging Contours of a Regime of Genomic Medicine in Oncology", *New Genetics and Society*, 32 (4) : 405-428.
- Nelson N C and al. (2014) "Testing devices or experimental systems ? Cancer clinical trials take the genomic turn", *Social Science & Medicine* 111 : 74-83.
- Nicholson D J (2012) "The concept of mechanisms sin biology", *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 : 152-163.
- Nowotny H, Scott P et Gibbons M (2003), *Repenser la science*, Paris, Editions Belin.
- O'Malley M A and al. (2012) "The role of integration in molecular systems biology", *Studies in History and Philosophy of Biological and Biomedical Sciences* 43 : 58-68.
- Pearl J (2009) *Causality : Models, Reasoning and Inference*, Cambridge, Cambridge University Press.
- Perbal L (2011), *Gènes et comportements à l'ère post-génomique*, Paris, Librairie Philosophique J. Vrin.

Perret-Gentil F (2013), *Le tact et le coup d'œil. Eléments d'épistémologie médicale*, Lyon, Jacques André éditeur.

Pestre D (2003) *Science, argent et politique. Un essai d'interprétation*, Paris INRA Editions.

Pitt J C (2000) *Thinking about technology : Foundations of the philosophy of technology*, New York, Seven Bridges.

Pommier E (2012), *Hans Jonas et le principe de responsabilité*, Paris, Presses Universitaires de France.

Popper K R (1968) *The logic of scientific discovery*, New York, Harper and Row.

Prochiantz A (2012), *Qu'est-ce que le vivant ?*, Paris, Editions du Seuil.

Ricoeur P (2004), *Parcours de la reconnaissance*, Paris, Editions Stock.

Rothman K (1986) *Modern Epidemiology*, Boston.

Rouvroy A (2013) *Politique des algorithmes*, Paris, Editions La Découverte.

Russo F and Williamson J (2007) "Interpreting Causality in Health Sciences", *International Studies in the Philosophy of Sciences*, 21 (2) : 157-170.

Salmon W C (1998), *Causality and Explanation*, Oxford, Oxford University Press.

Schaer R (2013) *Répondre du Vivant*, Paris, Editions Le Pommier.

Shinn T (2007) *Research Technology and Cultural Change : Instrumentation, Genericity, Transversality*, Oxford, The Bardwell Press.

Simondon G (1958), *Du mode d'existence des objets techniques*, Paris, Editions Aubier.

Tabery J (2009), « From a genetic predisposition to an interactive predisposition : rethinking the ethical implications of screening for gene-environment interactions », *Journal of Medicine and Philosophy* 34 : 27-48.

Taleb N N (2009) *Le Hasard sauvage. Comment la chance nous trompe*, Paris, Les Belles Lettres.

Thomas J-P (2013), *La médecine, nouvelle religion*, Paris, François Bourin Editeur.

Vailly J (2011) *Naissance d'une politique de la génétique. Dépistage, biomédecine, enjeux sociaux*, Paris, Presses Universitaires de France.

Van Regenmortel M & Hull D (dir) (2002) *Promises and Limits of Reductionism in the Health Sciences*, Chichester.

Varela F (1996), *Quel savoir pour l'éthique ? Action, sagesse et cognition*, Paris, Editions La Découverte.

Worms F (2010), *Le moment du soin. A quoi tenons-nous ?*, Paris, Presses Universitaires de France.

Sources scientifiques et médicales

Anand R, Gill K D and Mahdi A A (2014) "Therapeutics of Alzheimer's disease : Past, present and future", *Neuropharmacology* 76 : 27-50.

Anaya J-M, Duarte-Rey C and al. (2016) "Personalized medicine. Closing the gap between knowledge and clinical practice", *Autoimmunity Reviews* 15 : 833-842.

Bainbridge M N and al. (2011) "Whole-genome sequencing for optimized patient management", *Sci Transl. Med.* 3 (87) re 3.

Bertam L and Tanzi R E (2009) "Genome-wide Association Studies in Alzheimer's Disease", *Human Molecular Genetics* 18 (R2) : 270-281.

Berton P and Gerstein M (2001) « Integrative data mining : the new direction in bioinformatics », *IEEE Eng. Med. Biol. Mag.*, 20 (4) : 33-40.

Bradley P S and al. (1998) "Scaling clustering algorithms to large databases", *Proc. Knowledge Discovery and Data Mining* 1998 : 9-15.

Brechot C (2004) « La recherche translationnelle en santé, un nouveau paradigme », *Médecine/Sciences* 20 : 939-940.

Brookmeyer R, Johnson E, Ziegler-Graham K and Arright H M (2007) "Forecasting the global burden of Alzheimer's disease", *Alzheimer's & Dementia* 3 : 186-191.

Cardon D (2015) *A quoi rêvent les algorithmes, nos vies à l'heure des Big Data*, Paris, Editions du Seuil.

Chartier-Harlin and al; (2002) « Les formes tardives de la maladie d'Alzheimer : de la génétique à la biologie », *Médecine/Sciences* 18 : 709-716.

Chen R and al. (2012) "Personal omics profiling reveals dynamic molecular and medical phenotypes", *Cell* 148 : 1293-1307.

Chervitz S A and al. (2011) "Data standards for Omics data : the basis of data sharing and reuse", *Methods Mol. Biol.* 719 : 31-69.

Ciesielski T H, Aldrich M C, Marsit C J, Hiatt R A and Williams S C (2017) "Transdisciplinary approaches enhance the production of translational knowledge", *Translational Research* 182 : 123-134.

Claverie J-M and Notredame C (2003) *Bioinformatics for Dummies*, Hoboken NJ, Wiley.

Cox D (1972) « Modèles de régression et tables de survie », *Journal de la Société Statistique Royale, Séries B* 34 (2) : 187-220.

Croisile B (2014) *Alzheimer – Que savoir ?, Que craindre ?, Qu'espérer ?*, Paris, Editions Odile Jacob.

Czech C and al. (2012) "Metabolite profiling of Alzheimer's disease cerebrospinal fluid", *Plos ONE* 7 : e31501.

D'Arcy Hart R P (1999) "A change in scientific approach : from alternation to randomised allocation in clinical trials in the 1940's", *British Medical Journal* N° 319 : 572-573.

Dausset J (1998) *Clin d'œil à la vie, La grande aventure HLA*, Paris, Editions Odile Jacob.

Davies K (2010) *The \$ 1000 Genome, The Revolution in DNA Sequencing and the New Era of Personalized Medicine*, USA, Free Press.

Deam J and Ghemawat S (2008) "MapReduce : simplified data processing on large clusters", *ACM* 51 (1) : 107-113.

Deutsch J (2012) *Le gène, un concept en évolution*, Paris, Editions du Seuil.

Drouet I (dir.) (2016) *Le bayésianisme aujourd'hui*, Paris, Editions Matériologiques.

Dubois B and al. (2007) "Research criteria for the diagnosis of Alzheimer's disease : revisiting the NINCDS-ADRDA criteria", *Lancet Neurology* 6 : 734-746.

Dubois B and al. (2010) "Revising the definition of Alzheimer's disease : a new lexicon", *Lancet Neurology* 9 : 1118-1267.

Dubois B and al; (2014) "Advancing research diagnostic criteria for Alzheimer's disease : The TWG-2 criteria", *The Lancet Neurology* 13 (6) : 614-629.

Duits F H and al. (2014) "The cerebrospinal fluid 'Alzheimer profile' : Easily said, but what does it mean ?", *Alzheimer's & Dementia* 2014 : 1-11.

Eustache F, Chételat G, Desgranges B, de La Sayette V (2015) *Alzheimer : fatalité ou espoir*, Paris, Le Muscadier.

Fayyad U M and al. (1996) "From Data Mining to Knowledge Discovery in Databases", *AI Magazine* 17 (3) : 37-54.

Fields S (2001) "The Interplay of Biology and Technology", *Proceedings of the National Academy of Sciences, USA*, 98 : 10051-10054.

Flood D G, Marek G J and Williams M (2011) "Developing preictive CSF biomarkers – A challenge critical to success in Alzheimer's disease and neuropsychiatric translational medicine", *Biochemical Pharmacology* 81 : 1422-1434.

Garchon H J et Dieudé P (2011) « Les biomarqueurs en génomique, comment cela fonctionne-t-il ? », *Revue du Rhumatisme* 78 : 165-168.

Gaudillière J-P (2002) *Inventer la biomédecine*, Paris, Editions La Découverte.

Gibson W M (2001) "Can Personalized Medicine Survive ?", *Canadian Family Physician* 17 (8) : 29-88.

Gjuvsland A B and al. (2013) "Bridging the genotype-phenotype gap : what does it take ?", *The Journal of Physiology* 591 (Pt8) : 2055-2066.

Gzil F (2014) *La maladie du temps. Sur la maladie d'Alzheimer*, Paris, Presses Universitaires de France.

Hache J (2005) *Les enjeux des biotechnologies*, Colombelles, Editions EMS.

Hampel H and al ; (2012) "Development of biomarkers to chart all Alzheimer's disease stages : The royal road to cutting the therapeutic Gordian Knot", *Alzheimer's and Dementia* 8 : 312-336.

Hardy J A and Higgins G A (1992) "Alzheimer's Disease : The Amyloid Cascade Hypothesis", *Science* 256 : 184-185.

Holliday R (1990) "DNA methylation and epigenetic inheritance", *Philos. Trans. R. Soc. Lond. B Bio. Sci.* 326 : 329-338.

Huang S (2000) "The practical problems of post-genomic biology", *Nature Biotechnology* 18 : 4471-4472.

Hulette CM and al. (1998) "Neuropathological and neuropsychological changes in 'normal aging' : Evidence for preclinical Alzheimer's disease in cognitively normal individuals", *J. Neuropathol. Exp. Neurol.* 57 : 1168-1174.

Iqbal K and Grundke-Iqbal I (2010) "Alzheimer's disease, a multifactorial disorder seeking multitherapies", *Alzheimer's & Dementia* 6 (5) : 420-424.

Jablonka E and Lamb J (2002) "The changing concept of Epigenetics", *Annals of the New York Academy of Sciences* 981 : 82-96.

Jack C R Jr and al. (2010) "Hypothetical model of dynamic biomarkers of the Alzheimer's pathological cascade", *Lancet Neurol.* 9 : 119-128.

Jackson C E and Snyder P J (2008) "Electroencephalography and event-related potential as biomarkers of mild cognitive impairment and mild Alzheimer's disease", *Alzheimer's & Dementia* 4 (1 Suppl 1) : s 137-143.

Jain K K (2011) "Nanobiotechnology and Personalized Medicine", *Prog. Mol. Biol. Transl. Sci.* 104 : 325-354.

Katsnelson (2013) "Momentum grows to make 'personalized' medicine more 'precise'", *Nature Med* 19 (3) : 249.

Keevil S F (2012) "Physics and medicine : a historical perspective", *The Lancet* 379, april 21, 2012 : 1517-1524.

Klunck W E and al. (2004) "Imaging brain amyloid in Alzheimer's disease with Pittsburgh compound B", *Ann. Neurol.* 55 : 306-319.

Kupiec J-J, Gandrillon O, Morange M, Silberstein M, directors (2009) *Le hasard au cœur de la cellule. Probabilités, déterminisme, génétique*, Paris, Editions Syllepse.

Lahmani M and al ; (2007) *Les nanosciences – « Nanobiotechnologies et nanobiologie*, Paris, Editions Belin.

Laird P W (2010) "Principles and challenges of genomewide DNA methylation analysis", *Nat. Rev. Genet.* 11 : 191-203.

Lambert J-C and al. (2013) "Meta-analysis of 74,046 Individuals identifies 11 new susceptibility loci for Alzheimer's disease", *Nature Genetics* 45 (12) : 1452-1458.

Lewontin R C, Rose S, Kamin L J (1985) *Nous ne sommes pas programmés : génétique, hérédité, idéologie*, Paris, Editions La Découverte.

Livingstone G and al. (2017) "Dementia prevention, intervention and care", *The Lancet*, published on line july 20, 2017, [http://dx.doi.org/10.1016/50140-6736\(17\)31363-6](http://dx.doi.org/10.1016/50140-6736(17)31363-6).

Lock M (2013) *The Alzheimer Conundrum*, Princeton, Princeton University Press.

Mc Arthur D G and Monkol L (2012) "The uncertain road towards genetic medicine", *Trends in Genetics*, 28 (7) : 303-305.

Mc Clellan J and King M-C (2010) "Genetic heterogeneity in human disease", *Cell* 141 : 210-217.

Mc Kenna A and al. (2010) "The Genome Analysis Toolkit : a MapReduce framework for analysing next-generation DNA sequencing data", *Genome Research* 20 (9) : 1297-1303.

Manning C D, Raghavan P, Schütze H (2008) *Introduction to Information Retrieval*, Cambridge, Cambridge University Press.

Marks H M (1999) « L'irruption de la preuve statistique en médecine », *La Recherche* N° 316 : 76-81.

Massoud F et Robillard A (2013) *La maladie d'Alzheimer*, Montréal, Annika Parance Editeur.

Mayer-Schönberger et Cukier K (2014) *Big Data, la révolution des données est en marche*, Paris, Editions Robert Laffont.

Metzger M L (2010) "Sequencing technologies – the next generation", *Nat. Rev. Genet.* 11 : 31-46.

Meyer J M and Ginsburg G S (2002) "The path of personalized medicine", *Current Opinion in Chemical Biology* 6 (4) : 4434-4438.

Møldrup C (2009) "Beyond personalized medicine", *Personalized medicine* 6 : 231-233.

Morange M (1998) *La part des gènes*, Paris, Editions Odile Jacob.

Ozsolak F and Milos P M (2011) "RNA sequencing : advances, challenges and opportunities", *Nat. Rev. Genet.* 12 : 87-98.

Perneger T et Perrier A (2004) « Analyse d'un test diagnostique, Courbe ROC ou *receiver-operating characteristic* », *Revue des Maladies Respiratoires*, 21 (2) : 398-401.

Petersen R C and al. (1996) "Definition, Course, and Outcome of Mild Cognitive Impairment", *Aging, Neuropsychiatry and Cognition* 3 : 141-147.

Petersen R C and al; (1999) "Mild cognitive impairment : Clinical characterization and outcome", *Arch. Neurol.* 53 (3) : 303-308.

Phisicky E M and Fields S (1995) "Protein-protein interactions : methods for detection and analysis", *Microbiol. Rev.* 59 (1) : 94-123.

Powell J and Buchan I (2005) "Electronic health records should support clinical research", *J. Med. Internet Res.* 7 (1), Mar. 14.

Prifti and al. (2008) "FunNet : an integrative tool for exploring transcriptional interactions", *Bioinformatics* 24 (22) : 2636-2638.

Ray S and al. (2007) "Classification and prediction of clinical Alzheimer's diagnosis based on plasma signing proteins", *Nature Med.* 13 : 1359-1362.

Roses A D (1996) "Apolipoprotein E and Alzheimer's Disease : A Rapidly Expanding Field with Medical and Epidemiological Consequences", *Annals of the New York Academy of sciences* 802 (1) : 50-57.

Sackett D L and al. (1996) "Evidence-Based Medicine : What it is and what isn't", *British Medical Journal* 312 : 71-72.

Ruffié J (1993) *Naissance de la médecine prédictive*, Paris, Editions Odile Jacob.

Simel D L and al. (1991) "Likelihood ratios with confidence : sample size estimation for diagnostic test studies", *J. Clin. Epidemiol.* 44 (8) : 763-770.

Strimbu K and Travel J A (2010) "What are biomarkers ?", *Curr. Opin. HIV AIDS* 5 (6) : 463-466.

Strohmman R C (1997) "Epigenesis and complexity, The coming Kuhnian revolution in biology", *Nature Biotechnology* 15 : 194-200.

Templeton A R (1998) "The Complexity of Genotype-Phenotype Relationship and the Limitations of Using Genetic 'Markers' at the individual Level", *Science in Context* 11 (3) : 373-389.

Tuffery S (2007) *Data Mining et Statistique décisionnelle : l'intelligence des données*, Paris, Editions Technip.

Wanunu M (2012) "Nanopores : A journey toward DNA Sequencing", *Physics of Life Review* 9 : 125-158.

Weston A D & Hood L (2004) "Systems Biology, Proteomics and the Future of Health Care : Toward Predictive, Preventive and Personalised Medicine", *Journal of Proteome Research* 3 (2).

Wild C P (2005) "Complementing the genome with an 'exposome' : the outstanding challenge of environmental exposure measurement in molecular epidemiology", *Cancer Epidemiol. Biomarkers Prev.* 14 (8) : 1847-1850.

Wild C P (2012) "The exposome : from concept to utility", *J. Epidemiol.* 41 (1) : 24-32.

Woese C R (2004) "A new biology for a new century", *Microbiology and Molecular Biology Reviews* 68 : 173-186.

Zerbini D R and Bimey E (2008) "Velvet algorithms for de novo short read assembly using the de Bruijn graphs", *Genome Research* 18 (5) : 821-829.

Vers la prévention et l'anticipation dans la pratique médicale

Réflexions sur l'épistémologie des biomarqueurs dans le cas de la maladie d'Alzheimer

Résumé

Cette thèse développe une réflexion épistémologique autour de la notion de biomarqueurs dans le cas de la maladie d'Alzheimer. Elle est centrée sur le transfert des connaissances du domaine de la biologie à celui de la pratique médicale et clinique, avec une attention particulière aux techniques de diagnostic précoce et aux *Big Data*.

La maladie d'Alzheimer présente une temporalité particulière, son apparition étant insidieuse et sa phase asymptomatique longue. Elle se différencie des cancers en ne se prêtant pas à l'analyse génomique de cellules spécifiques, et permet ainsi d'aborder différemment le statut épistémique des biomarqueurs.

Le biomarqueur (que ce soit une molécule, un réseau d'interactions, voire un algorithme), est porteur d'information sans pour autant établir un lien de causalité directe avec la maladie. C'est essentiellement un indice et non la représentation de la condition réelle d'un sujet ; c'est ensuite un objet entouré d'incertitude ; c'est enfin un objet dont la maîtrise n'est pas totale, un objet qui n'est jamais complètement donné. Enfin, les relations entre le biomarqueur et le milieu extérieur font partie intégrante de son fonctionnement.

Les biomarqueurs sont essentiels dans la transformation des pratiques médicales vers l'anticipation et le suivi de l'évolution de la condition d'un sujet. En mettant en évidence des éléments transformant des facteurs de risque en une pathologie, les biomarqueurs invitent chacun à se surveiller et permettent un accompagnement des personnes dans une situation où elles n'ont encore aucun signe clinique d'une maladie évolutive.

Mots Clés : Causalité ; Clinique ; Corrélations ; Diagnostic ; Facteurs de risque ; Philosophie de la technique ; Philosophie de la médecine ; Techniques en biologie.

Towards prevention and anticipation in medical practice

Reflections on the epistemology of biomarkers in the case of Alzheimer's disease

Summary

This dissertation develops epistemological reflections on the notion of biomarkers in the case of Alzheimer's disease. It focuses on the challenge posed by the transfer of knowledge from the field of biology to medical and clinical practices, with a special attention to the techniques of early diagnosis and especially the role of Big Data.

Alzheimer's disease presents a particular temporality, its appearance being insidious with a long asymptomatic phase. It differs from cancer by not being amenable to genomic analysis of specific cells, and thus allows a different approach to the epistemic status of biomarkers.

The biomarker whether it be a molecule, network of interactions, or even an algorithm, sheds light on the disease in the absence of any direct causal links between the biomarker and the disease. It is primarily an indicator rather than the representation of a body condition. As a consequence, it is always surrounded by uncertainty and never fully mastered, nor fully given. The biomarker is an object whose relations with the environment are an integral part of its functioning.

Biomarkers are essential in transforming medical practices towards anticipating and monitoring the evolution of a subject's health condition. By highlighting elements that transform risk factors into a pathology, biomarkers invite everyone to monitor themselves and make it possible to support people well ahead of the appearance of clinical signs of an evolving disease.

Key words: Causality; Clinics; Correlations; Diagnosis; Philosophy of technique; Philosophy of medicine; Techniques for biology.