

## Préface

Si on remonte au plus loin dans l'histoire de l'humanité, il est possible que le premier geste du premier chercheur ait été de collecter des objets prélevés dans la nature. Réunir, classer et conserver ces objets pour les observer et les étudier sont probablement à l'origine de la première démarche scientifique de notre ancêtre, doué de conscience et d'abstraction, et de l'envie irrésistible de connaître le monde. Plus récemment, dans les cabinets de curiosité du XVII<sup>e</sup> siècle il s'agissait d'émerveiller le visiteur, mais aussi d'affirmer son rang et sa puissance. Nos rois l'avaient bien compris en créant les muséums et autres jardins royaux qui rassemblaient des collections d'animaux et de végétaux extraordinaires prélevés aux confins de leur royaume. Ces collections ont été à l'origine du bouleversement de nos connaissances en sciences de la vie. Que l'on songe à la monumentale *Histoire naturelle* de Buffon au siècle des Lumières !

Les collections biologiques humaines se sont constituées avec l'avènement des hôpitaux au XIX<sup>e</sup> siècle, devenus non plus lieux de bienfaisance mais établissements de soins et de transmission du savoir. Pour décrire et tenter de comprendre des maladies ou des anomalies congénitales, il s'agissait de prélever, conserver et classer les organes atteints et des fragments de tissus biologiques ; ainsi seront constituées les célèbres collections anatomiques de Jean-Martin Charcot à la Salpêtrière, de Marc-Antoine Petit, premier chirurgien-major de l'Hôtel-Dieu de Lyon, ou encore celles des écoles de médecine de Montpellier et de Strasbourg. C'est aussi la naissance de l'anatomopathologie qui étudie et classe les lésions macroscopiques et microscopiques des tissus lésés, avec Jean-Baptiste Morgagni à Padoue, Rudolf Virchow à Berlin et Jean Lobstein à Strasbourg. D'importantes collections se sont ainsi accumulées pendant des décennies au gré des passions de ces médecins-chercheurs ; collections qui échouaient le plus souvent dans les sombres sous-sols des hôpitaux quelque temps après la disparition de leur créateur, suivie de près par celle du cahier noirci d'annotations cabalistiques compréhensibles par son seul auteur.

À l'aube du XXI<sup>e</sup> siècle, avec le développement des techniques d'analyse à haut débit, l'accès à des échantillons biologiques de qualité est devenu un enjeu majeur. En effet, il est rapidement apparu que l'industrialisation croissante de la recherche en biologie-santé exigeait que les objets biologiques soient de qualité irréprochable et pourvus d'annotations fiables pour les analyser. Les premières bases de données issues de l'exploitation des échantillons humains étaient bien souvent entachées d'erreurs et génératrices d'informations scientifiques erronées. Un autre enjeu majeur était celui de la valorisation économique. Le développement des biotechnologies, la montée en puissance des médicaments biologiques et l'identification de biomarqueurs associés aux classifications moléculaires des lésions tissulaires rendaient indispensable un accès ordonné aux échantillons biologiques humains, devenus précieux et rares.

La France a été pionnière sur le plan international dans l'organisation et la définition des missions des centres de ressources biologiques, ou biobanques, nouvelles infrastructures de biologie-santé destinées à collecter, stocker et mettre à disposition des échantillons biologiques humains. Elle a également été le premier pays à instaurer une norme de fonctionnement des biobanques, récemment internationalisée en norme ISO et à établir des règles éthiques d'accès aux échantillons biologiques issus du corps humains, inscrites dans le Code de la santé publique. Les progrès normatifs ont été considérables, la production scientifique améliorée et les services rendus à l'avancement de la médecine indiscutables. Les biobanques fonctionnelles sont aujourd'hui celles qui associent assurance-qualité des échantillons biologiques et de leurs annotations, expertises techniques et scientifiques, transparence de la gouvernance, innovation, formation et insertion dans le tissu de la recherche qu'il soit local, national ou international.

Plus de vingt ans après leur création, le chantier des biobanques n'est pourtant pas clos [CLE 19]. Le modèle économique est encore instable faute d'une véritable stratégie nationale pour structurer et financer un réseau national des biobanques qui s'appuierait sur des évaluations indépendantes par les pairs. Bien souvent, les stocks augmentent ainsi indéfiniment, dans l'espoir d'une hypothétique valorisation. De plus, l'accès aux données épidémiologiques, biocliniques et génomiques associées aux collections demeure mal aisé, en l'absence d'harmonisation des systèmes d'informations hospitaliers et d'une réelle volonté de partage de ces données. Malgré des avancées importantes au niveau européen, le cadre législatif et réglementaire tarde à s'adapter à l'évolution des biobanques et à la constitution de grandes cohortes qui impliquent un stockage des échantillons biologiques et des données associées pour des recherches futures.

Les biobanques sont à la croisée des chemins. Comme toutes les infrastructures essentielles en biologie-santé, elles doivent s'adapter aux progrès de la recherche et

des techniques. À l'ère du numérique et de la production de données massives par pratiquement tous les laboratoires qui ont accès à des plateformes technologiques à très haut débit et à bas coûts, les biobanques deviennent les garantes de la reproductibilité des expériences et de la qualité globale de la recherche en biologie-santé. Elles étendent leurs périmètres d'intervention, automatisent les processus et explorent de nouvelles techniques d'analyse sur des échantillons précieux, tout en les préservant pour les générations futures. Elles diversifient leurs capacités de stockage ; outre les informations relatives aux objets biologiques, les données épidémiologiques, cliniques, biologiques, génomiques et d'imagerie rejoignent leurs bases informatiques. Les biobanques, par nature infrastructures d'interface, effacent le clivage artificiel entre les activités de soin et de recherche. Un exemple frappant est celui de leur rôle central dans la génomique médicale qui s'appuie sur le déploiement intensif du numérique pour produire de l'information, utile simultanément à la prise en charge des patients et aux chercheurs qui étudient la maladie.

Cet ouvrage fait le point sur l'actualité des biobanques. Il explore des pistes nouvelles, en particulier le bouleversement lié au déploiement massif de la santé numérique et de la médecine de précision. Les auteurs montrent que l'avenir de la biologie-santé passe par un déploiement des biobanques dans des champs qui restent encore à défricher, à l'avant-garde de l'aventure extraordinaire de la recherche du XXI<sup>e</sup> siècle.

Bruno CLÉMENT  
Directeur de recherche à l'Inserm  
Institut de nutrition, métabolismes et cancer  
Centre de ressources biologiques humaines – CRB Santé de Rennes

# Les biobanques en 2022 : pourquoi et comment ?

La recherche repose sur plusieurs socles : les concepts fondamentaux, les modèles cellulaires et animaux, la validation des découvertes obtenues dans les laboratoires par une approche « translationnelle », puis, selon les secteurs d'intérêt, la recherche clinique ou appliquée ou réalisée *in silico* à partir des bases de données. Selon les thématiques de recherche, il est nécessaire de réaliser un *continuum* entre la découverte de nouveaux mécanismes cellulaires et une application dans le monde du vivant, que celui-ci soit le monde végétal, microbien ou animal et humain.

Un des maillons incontournables des liens entre la recherche fondamentale et ses différentes applications est l'utilisation de différentes ressources biologiques, que ces dernières soient issues des végétaux, des microbes, des animaux ou des êtres humains sains ou malades [ALD 19, VAU 19]. De tout temps, les chercheurs ont eu besoin d'analyser ces différentes ressources biologiques pour valider leur hypothèse ou des observations réalisées sur des modèles cellulaires. Plusieurs finalités sont progressivement apparues, comme une meilleure compréhension de la biologie du développement ou des mécanismes de mort, de croissance ou de transformation cellulaire. L'un des exemples les plus aboutis de l'utilisation des ressources biologiques est la possibilité de découvrir des biomarqueurs des maladies humaines, ces derniers pouvant être des biomarqueurs à visée diagnostique, pronostique ou prédictive d'une réponse thérapeutique [BAR 20, HEW 11]. Si l'utilisation de ces ressources biologiques s'est faite initialement d'une façon peu contrôlée et sans véritable règle établie, plusieurs politiques de réflexion ont conduit progressivement à l'établissement d'un fonctionnement contrôlé et rigoureux. Ainsi la notion de **collection** est née, associée à la nécessité de maîtriser leur utilisation et les espaces de stockage nécessaires. Au cours des ces dernières décennies, puis plus récemment, ces différentes collections ont

été associées à la création des **centres de ressources biologiques** (CRB), encore appelés aujourd'hui « biobanques », dont la structuration s'est faite progressivement, pour devenir aujourd'hui des « outils » incontournables au monde de la recherche [WAT 19].

Les biobanques ont évolué ces dernières années et il s'agit aujourd'hui de structures professionnalisées et complexes dont le fonctionnement est soumis aux réglementations nationales et internationales. Ces biobanques peuvent fonctionner comme des espaces de stockages sécurisés de collections biologiques mettant à disposition des échantillons pour des projets de recherche. Elles peuvent aussi être organisées comme de véritables **centres experts** et proposer à la fois des prestations aux chercheurs (grâce à l'existence de plateaux techniques) et une collaboration scientifique (grâce à des connaissances uniques). Les biobanques peuvent aussi développer leur propre projet de recherche sur des thématiques spécifiques notamment au contrôle, à la gestion et à l'utilisation des échantillons biologiques, essentiellement dans le but d'optimiser leur fonctionnement et leur offre de service [WAS 18a].

Les biobanques et la politique de collection ont évolué au cours du temps. Un des exemples est typiquement celui de l'évolution des biobanques collectionnant des échantillons issus de patients. Dans un premier temps, de nombreux échantillons ont été accumulés et congelés, en particulier des échantillons de tumeurs. S'est alors posée la question de la qualité de ces milliers d'échantillons stockés. En effet, la découverte et/ou la validation des différents marqueurs s'est avérée totalement dépendante d'une reproductibilité des analyses sur des échantillons parfaitement conservés. Finalement, la nécessité d'acquérir des données associées de plus en plus nombreuses ou ciblées a modifié le fonctionnement des biobanques en imposant le développement de bases de données multiples et complexes. En effet, l'un des points clés associés à la mise à disposition des échantillons biologiques pour des projets de recherche est maintenant la nécessité d'y associer des données très précises et pertinentes. Le monde du *bio-banking* a donc évolué progressivement ces dernières décennies, ce domaine devenant une **spécialité médicale à part entière**. Ainsi, schématiquement, l'activité des biobanques est passée par plusieurs étapes, de type « biobanque 1.0 » (activité de collectes massives souvent sans *a priori* et basée sur du « quantitatif »), puis de type « biobanque 2.0 » (où la qualité des échantillons a été considérée comme cruciale avec la mise en place de la maîtrise des paramètres de la phase pré-analytique), et enfin de type « biobanque 3.0 » (intégrant plusieurs impératifs, en particulier la gestion maîtrisée des données clinicobiologiques, l'anticipation des demandes et le contrôle du modèle économique) [DOU 17, ELL 15, LIN 20, SIM 14]. Ainsi la création et la construction d'une structure de type biobanque doivent répondre aujourd'hui à des exigences strictes, notamment associées à une stratégie de collecte et une planification de l'utilisation des échantillons stockés [BAI 16, HOF 13].

La multiplicité des expertises nécessaires pour générer les données associées aux échantillons, leur intégration et leur analyse nécessitent de repenser l'organisation et la structuration des biobanques pour s'associer les services des spécialistes dans différents domaines, notamment médical, biologique, de l'imagerie, des statistiques, de la bio-informatique et des mathématiques. Cette évolution doit concerner, au niveau national ou international, des biobanques impliquées dans la collecte, l'intégration et l'analyse des données complexes. Petit à petit apparaît le concept des **biobanques de nouvelle génération** ou *next generation biobank* (NGB). Une NGB doit être capable d'associer de gigantesques bases de données hébergeant des données phénotypiques, comportementales, familiales, d'imagerie, -omiques, radiomiques, biologiques, etc. provenant de différents centres. Le fonctionnement d'une NGB nécessite de relever plusieurs défis, le premier étant celui d'infrastructures de calcul capables de traiter des téra, penta et exabits. À ceci s'ajoute le développement d'une suite de méthodes d'analyses et d'algorithmes appropriés. De plus, cette évolution s'accompagne maintenant d'un changement de paradigme de l'analyse de données provenant d'un grand nombre de patients à l'analyse d'un grand volume de données provenant d'un seul sujet (*Big Data versus Fat Data*). Un aspect capital pour des biobanques est leur pérennisation, compte tenu du fait que le modèle économique de ces structures est souvent fragile [RAO 19, VAU 11]. En effet, le fonctionnement d'une biobanque nécessite aujourd'hui un personnel dédié et qualifié, une infrastructure sécurisée et coûteuse et des équipements que l'on doit renouveler régulièrement afin de maintenir la qualité des collections.

Il existe de plus en plus de biobanques à travers le monde et le nombre de collections pouvant être à la disposition des chercheurs augmente, pouvant conduire progressivement à une baisse des demandes pour certaines biobanques compte tenu d'une mise en concurrence. Dans ce contexte, pour être compétitive, une biobanque doit faire des choix d'organisation et de stratégie. Il est difficile de maintenir un niveau d'expertise dans plusieurs domaines d'activité, et s'orienter vers un nombre restreint de pathologies permet certainement de se concentrer sur l'exhaustivité des collections et d'y associer des données cliniques et biologiques complètes, ces dernières devenant de plus en plus complexes.

Un point incontournable est l'obtention des **consentements** des patients pour utiliser leur échantillon à des fins de recherche. Ce point clé doit être associé à une démarche interne propre à chaque biobanque qui doit être parfaitement maîtrisée et menée en concertation avec les services cliniques. La **visibilité** extérieure d'une biobanque et sa reconnaissance sont accrues lorsque les partenaires d'une biobanque ont accès à différentes expertises (par exemple diagnostic histologique réalisé par des pathologistes seniors dans le domaine de pathologie concernée ; expertise en biologie moléculaire et banque de donnée génomique accessible). Le fait de se focaliser sur une ou peu de pathologies permet aussi d'associer pour les mêmes patients des collections diversifiées

de tissus (fixés et congelés) et de biofluides (plasma, culot leucocytaire, sang total, urines, autres). Les données cliniques et notamment le suivi des patients en fonction des différents traitements successifs ou le recueil des événements (progression, métastase, décès) peuvent être plus aisément intégrés aux bases biologiques [WAT 17]. Dans ce contexte, il est souvent plus facile de s'associer à un réseau national ou international d'experts d'un même domaine.

Une **certification** ou une **accréditation** de la biobanque selon des normes nationales ou mieux, internationales, est aussi essentielle à la robustesse de son fonctionnement et à la qualité du service rendu. Une biobanque compétitive doit aussi développer des projets innovants, notamment en lien avec l'industrie des groupes pharmaceutiques ou des biotechnologies. Différents projets peuvent permettre de transférer les résultats de l'innovation ainsi acquis vers la pratique clinique, après une phase de validation réalisée avec des échantillons biologiques. Ainsi la visibilité d'une biobanque vis-à-vis de différents partenaires et/ou demandeurs d'échantillons biologiques peut bénéficier de la mise en place d'**indicateurs de performance** qui doivent s'adapter selon la structure concernée et ses ambitions [HOF 13].

Plusieurs défis se dessinent dans un avenir proche pour les biobanques et ces structures doivent répondre à un certain nombre d'objectifs communs. Ainsi il va falloir rassembler (ou utiliser) des données issues de différentes bases de données ou présentes dans différents documents (anatomie pathologique, biologie moléculaire, imagerie, données cliniques et thérapeutiques) dans une seule source accessible. Il va falloir aussi intégrer, stocker et traiter des volumes de données complexes et souvent très hétérogènes. Un des défis est le partage et l'accès à l'information qui doit être sécurisé et réalisé à partir de données parfaitement dé-identifiées. Cet accès pourra être envisagé pour des partenaires extérieurs désireux de connaître les échantillons disponibles et les données complexes associées. Compte tenu de ces évolutions rapides et d'une modification des activités relatives au métier des biobanques, il est donc crucial de continuer ou de développer des formations universitaires adaptées et de recruter ainsi dans les biobanques des personnes ayant bénéficié d'un enseignement pratique et théorique consacré à ce domaine. La maîtrise (saisie et exploitation) des données complexes et massives associées aux échantillons biologiques collectés chez des patients atteints d'un cancer est devenue un enjeu majeur en oncologie. Cette maîtrise passe par plusieurs actions combinées. La première étape est le contrôle de **la qualité des données** récoltées. Dans le monde des données massives, le risque de générer des données secondaires biaisées par des données primaires incomplètes ou fausses est un risque majeur. Quels que soient les outils sophistiqués que l'on peut déployer, notamment basés sur des approches de type *deep learning*, *machine learning* et d'intelligence artificielle, il est impossible de transformer « du plomb en or », c'est-à-dire de générer des données de qualité à partir de données initiales insuffisantes ou

erronées. Le rôle du pathologiste est ici primordial et seule son expertise peut vérifier la qualité du diagnostic et des analyses réalisées sur les coupes tissulaires. L'intégration des données de biologie moléculaire associées est aussi une des étapes clés. Ces données sont de complexité variable, pouvant ainsi concerner les données de séquençage entier du génome pour une tumeur, mais aussi de transcriptomique, de protéomique et de métabolomique. À nouveau, la qualité des données (liées aussi à la qualité des échantillons et donc des étapes pré-analytiques) est le point essentiel à contrôler.

Les biobanques sont soumises à une **législation** rigoureuse et l'utilisation des échantillons pour réaliser des projets de recherche ne peut se faire qu'en respectant la loi. Ainsi les partenariats public-public ou public-privé ne s'établissent que grâce à des contrats et à des procédures signées gérant des accords de transfert de matériel biologique (ou *material transfert agreement*, MTA) [CLE 14, HOF 14, SIM 20]. Ces contrats doivent impérativement bien spécifier le but de la recherche réalisée à partir des échantillons et l'identité de l'utilisateur final. En effet, certaines compagnies privées peuvent servir de plateforme de recrutement d'échantillons biologiques, notamment à partir des laboratoires académiques et fournir des utilisateurs sans que l'on connaisse la finalité de la recherche de ces derniers.

Le monde des biobanques n'échappe pas à l'arrivée de l'**intelligence artificielle** (IA) [LIN 20, MAR 20, NAR 20]. L'IA est en fait un terme très général pour désigner « l'ensemble des théories et des techniques mises en œuvre en vue de réaliser des machines capables de simuler l'intelligence ». Souvent classée dans le groupe des sciences cognitives, elle fait appel à la neurobiologie computationnelle, à la logique mathématique et à l'informatique. Depuis sa naissance, il y a 70 ans, avec quelques abstractions simples du fonctionnement d'un neurone dans le cerveau humain, l'IA est devenue un outil indispensable pour les applications de vision par ordinateur. Plus particulièrement, les **réseaux de neurones artificiels** (*artificial neural networks*, ANN) ont gagné une grande popularité en raison de leur capacité de reconnaissance impressionnante lorsqu'ils sont mis en œuvre avec de nombreuses couches de neurones artificiels (unités de traitement qui peuvent effectuer une simple agrégation des valeurs synaptiques entrantes provenant d'autres unités). Compte tenu de la somme des données intégrées dans une biobanque, la mise à disposition des logiciels d'intelligence artificielle afin de pouvoir sélectionner les collections en fonction des données complexes, intégrant notamment le suivi des patients selon les traitements, paraît incontournable à l'avenir. Il peut paraître illusoire de continuer à collectionner des échantillons, alors que les biobanques sont actuellement presque saturées et que les demandes ne permettent de faire qu'un déstockage très faible en général par rapport aux stocks disponibles [RUS 19].

On peut penser aussi qu'il existe suffisamment de données omiques accessibles, notamment sur les bases de données publiques, qu'il sera de moins en moins utile d'accumuler les différents produits biologiques. En réalité, c'est sans compter sur l'évolution, les modifications climatiques et environnementales, et leur impact sur l'émergence des maladies humaines. L'un des exemples récents concerne la collection des échantillons biologiques chez des patients atteints par la [Covid-19](#). Plusieurs collections se sont ainsi formées rapidement, grâce à des infrastructures déjà existantes et performantes permettant d'organiser différentes collections de produits sanguins ou cellulaires associées à des bases de données clinicobiologiques [HUA 20, PEE 20, ROV 20, TAN 20]. Ces collections ont été parfois établies de façon prospective et adossées à des projets de recherche. Ainsi, c'est grâce aux biobanques que la gestion de ces nouvelles collections a pu être parfaitement maîtrisée selon les règles sanitaires et de protection du personnel et en assurant la traçabilité et la qualité des échantillons pour les futurs utilisateurs. Toutefois, certaines situations particulières, comme ici l'apparition de maladies infectieuses ou des crises économiques, peuvent avoir de lourdes circonstances sur la pérennisation d'une biobanque. En effet, il est aussi possible, compte tenu de l'impact d'une infection émergente sur le fonctionnement des hôpitaux, que les interventions chirurgicales ou les hospitalisations des patients non infectés diminuent brutalement, limitant alors le flux des échantillons vers les services biologiques et de pathologie et ainsi les possibilités de collections [HOF 20, VAU 20].

L'avenir des biobanques est lié à l'évolution des technologies, notamment du monde biomédical, mais aussi des connaissances issues d'autres disciplines comme les sciences sociales et mathématiques [HAR 19, KIN 20]. Il est fort probable que les bases de données de plus en plus complexes conduisent à limiter la collecte et le stockage des échantillons biologiques. Toutefois, « l'infobésité » conduit à saturer progressivement les espaces de stockage informatique avec ces différentes données massives, leur coût de sécurisation va croissant et de nombreuses informations pourraient en réalité être conservées au sein des acides nucléiques gérés par des biobanques. Enfin l'un des enjeux certains à atteindre dans le futur sera que les différentes biobanques puissent travailler en commun à travers le monde sur des sujets communs, en assurant les mêmes critères de qualité et d'accessibilité aux données afin de pouvoir faciliter des flux d'échanges des échantillons et permettre des projets de recherche ambitieux [VAU 14].