



## Biobanques tumorales et gestion des données complexes enjeux actuels et futurs

Paul Hofman, G. Dagher, Pierre Laurent-Puig, C. Hugo-Marquette, F.  
Barlesi, F. Bibeau, B. Clément

### ► To cite this version:

Paul Hofman, G. Dagher, Pierre Laurent-Puig, C. Hugo-Marquette, F. Barlesi, et al.. Biobanques tumorales et gestion des données complexes enjeux actuels et futurs. *Annales de Pathologie*, 2019, 39 (2), pp.137-143. 10.1016/j.annpat.2019.01.017 . hal-02094608

**HAL Id: hal-02094608**

**<https://univ-rennes.hal.science/hal-02094608>**

Submitted on 22 Oct 2021

**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.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

## **Biobanques tumorales et gestion des données complexes : Enjeux actuels et futurs**

### **Tumor banks and complex data management: Current and future challenges**

Paul Hofman (1, 2, 3), Georges Dagher (4), Pierre Laurent-Puig (5), Charles Hugo-Marquette (3, 6),  
Fabrice Barlesi (7), Frédéric Bibeau (8) et Bruno Clément (9)

<sup>[1]</sup> Laboratoire de Pathologie Clinique et Expérimentale, Hôpital Pasteur B.P. 69 – 30 avenue de la voie romaine 06001 Nice cedex 01, Centre Hospitalo-Universitaire de Nice, Université Côte d’Azur, Nice, France

<sup>[2]</sup> Biobanque Hospitalière (BB-0033-00025), Hôpital Pasteur, B.P. 69 – 30 avenue de la voie romaine 06001 Nice cedex 01, Centre Hospitalo-Universitaire de Nice, Université Côte d’Azur, Nice, France

<sup>[3]</sup> FHU OncoAge, Hôpital Pasteur, Université Nice Côte d’Azur, B.P. 69 – 30 avenue de la voie romaine 06001 Nice cedex 01, Centre Hospitalo-Universitaire de Nice, Université Côte d’Azur, Nice, France

<sup>[4]</sup> Biobanques, Inserm US013, Hôpital de la Salpêtrière, 47, boulevard de l’Hôpital, 75651 Paris, France

<sup>[5]</sup> UMR-1138 Université Paris Descartes; INSERM; Assistance Publique-Hôpitaux de Paris Département de Biologie, Hôpital européen Georges Pompidou, Paris France

<sup>[6]</sup> Service de Pneumologie, Hôpital Pasteur, B.P. 69 – 30 avenue de la voie romaine 06001 Nice cedex 01, Centre Hospitalo-Universitaire de Nice, Université Côte d’Azur, Nice, France

<sup>[7]</sup> Aix-Marseille université, Assistance publique-Hôpitaux de Marseille, service d'oncologie multidisciplinaire et innovations thérapeutiques, 13920 Marseille cedex 15, France

<sup>[8]</sup> Laboratoire de Pathologie, CHU de Caen, avenue de la Côte de Nacre, 14000 Caen

<sup>[9]</sup> Inserm, Inra, Université de Rennes, Nutrition Métabolismes et Cancer, CRB-Santé, rue Henri Le Guilloux, 35033 Rennes

Titre courant : Biobanque et gestion des données

Correspondance : Paul Hofman, adresse ci-dessus; tel : 04 92 03 88 55; fax : 04 92 03 87 50; E-mail : [hofman.p@chu-nice.fr](mailto:hofman.p@chu-nice.fr)

## **Résumé**

Les biobanques tumorales sont sollicitées dans le cadre du développement de la recherche translationnelle et clinique en cancérologie. Elles participent à l'évaluation puis à la validation des biomarqueurs à visée diagnostique, pronostique ou prédictive. L'évolution progressive de ces structures a permis de professionnaliser leur fonctionnement et de les identifier comme des rouages incontournables en oncologie par les acteurs du monde académique et industriel. Les progrès technologiques et thérapeutiques modifient l'impact et le fonctionnement de ces biobanques, ces dernières devant relever plusieurs défis afin notamment d'être pérennisées. Parmi les enjeux, l'intégration des données (cliniques et biologiques) de plus en plus complexes et massives associées aux échantillons biologiques, conduit à avoir une réflexion urgente pour une organisation optimisée des biobanques en France. Le but est d'être attractif aux yeux des partenaires face à la compétition internationale. Cette revue aborde rapidement les principales évolutions actuelles des biobanques, puis les défis actuels et futurs, et enfin la place que le pathologiste peut jouer dans ces nouveaux challenges de l'oncologie.

**Mot-clés :** Biobanque, données massives, défis, oncologie

## **Summary**

Tumor banks are asked to clinical and translationnal research project development in oncology. They strongly participate to the assessment, then to the validation of diagnostic, prognostic and predictive biomarkers. The progressive change of these structures leads to induce a professionalization of their functioning and to identify them as key actors in oncology by the stakeholders of the public and private worlds. The progresses made in biotechnologies and therapeutics are rapidly modifying the impact and the proper functioning of the biobanks. These latter are now facing different challenges, in particular for their sustainability. Among the major issues, the integration of the clinical and biological data becoming increasingly complex leads to urgently consider an optimization of the role of different biobanks in France. Their goal is to be an attractive counterpart face to the international competition. The purpose of this review is to briefly describe the current evolution of the biobanks, then their present and future challenges, and finally the role made by the pathologists in these new issues in oncology field.

**Key-words:** Biobanks, big data, challenges, oncology

## 1. Introduction

La recherche en cancérologie nécessite depuis toujours un transfert des connaissances issues d'expériences réalisées au laboratoire (analyses *in vitro* sur des cultures cellulaires, analyses *in vivo* sur des modèles animaux, analyses *in silico*) vers une validation chez le patient. Un des axes majeurs de cette recherche dite « translationnelle » a pour but la découverte de nouveaux biomarqueurs et molécules thérapeutiques. Les nouvelles stratégies thérapeutiques visent à développer des traitements personnalisés orientés sur des anomalies génomiques ou sur des cibles immunologiques. Dans ce contexte, les molécules ne sont actives que si des marqueurs cellulaires sont exprimés au niveau des tissus tumoraux. C'est le fondement même de la médecine dite personnalisée ou de précision. Dans ce contexte la validation des découvertes réalisées au laboratoire passe par des essais cliniques qui comparent des groupes contrôles (patients n'exprimant pas les marqueurs d'intérêt) avec des populations de patients chez qui ces marqueurs sont mis en évidence avec un niveau d'expression variable. Ces approches permettent ainsi d'évaluer la sensibilité et la spécificité de biomarqueurs à visée diagnostique, pronostique ou prédictive.

Les analyses effectuées à partir des échantillons biologiques (tissus ou biofluides) sont prospectives et/ou rétrospectives. Les études de recherche translationnelle passent souvent par une évaluation sur un groupe d'apprentissage (« training set » comportant un nombre souvent restreint d'échantillons) puis par une deuxième étape sur un nombre plus important d'échantillons conduisant ou non à la validation d'un biomarqueur (« validation set »). Il faut ici distinguer les échantillons biologiques associés aux essais cliniques qui en principe ne peuvent être exploités que

dans le cadre restreint de l'étude, et les échantillons collectés systématiquement chez des patients hospitalisés, souvent non inclus dans les essais. C'est dans ce contexte que les biobanques trouvent leur place et leur importance en oncologie [1-4]. En effet la gestion au long cours des échantillons à visée de recherche nécessite une organisation optimale compte tenu des retombées thérapeutiques possibles reposant sur les conclusions issues des analyses biopathologiques. Le circuit des échantillons, leur conservation dans des zones de stockage sécurisées et dédiées, impliquent une équipe et une structure hautement spécialisées à ce type de réalisation, dans un cadre professionnel adapté et règlementé.

L'utilisation des informations obtenues à partir de l'analyse des biomarqueurs tissulaires et/ou des biofluides n'a de valeur que si des données cliniques et biologiques de qualité y sont associées. Ces données sont enregistrées sur des entrepôts informatiques sécurisés. La nature et le nombre de ces données associées peuvent être extrêmement variables selon la nature des collections et du projet de recherche. Ainsi au-delà d'un jeu minimal de données ou « *minimun data set* », comportant notamment l'identité, l'âge, les codes CIM-10 (ou codes des classifications internationales des maladies) et la nature de l'échantillon, les données associées peuvent être plus ou moins complexes et s'accumuler au fil des projets scientifiques. Cette complexité devient croissante avec l'intégration des informations issues des analyses à grande échelle, génomiques notamment. Ainsi, les biobanques doivent avoir la capacité de maîtriser l'intégration des données complexes et la gestion physique d'échantillons biologiques d'origines variées [5].

Après avoir brièvement décrit les principales évolutions associées au fonctionnement des biobanques, cette revue aborde les enjeux auxquels elles vont faire

face, avec un intérêt particulier pour le rôle des (bio) pathologistes dans ces nouveaux challenges de l'oncologie.

## **2. Evolution des biobanques**

Le monde du « biobanking » a évolué progressivement ces dernières décennies et s'est professionnalisé, ce domaine devenant en fait une spécialité biomédicale à part entière. Ainsi schématiquement, l'activité des biobanques est passée par plusieurs périodes, de type « biobanque 1.0 » (activité de collectes souvent sans à priori et basée sur du « quantitatif »), puis de type « biobanque 2.0 » (ou la qualité des échantillons a été considérée comme cruciale), et enfin de type « biobanque 3.0 » (intégrant plusieurs impératifs, en particulier la gestion maîtrisée des données clinico-biologiques, et le contrôle du modèle économique) [6, 7].

Une des évolutions majeures a concerné la nature des échantillons collectionnés. Les premiers projets soutenus par l'Institut National du Cancer (INCA) à travers les programmes de soutien aux tumorothèques étaient en effet orientés sur la collecte des échantillons tumoraux congelés. En fait, le budget consacré à chaque tumorothèque était relativement proportionnel à l'activité déclarée et donc au volume d'échantillons collectés. Ceci a eu parfois comme effet une activité non contrôlée de collecte, le plus souvent sans aucune stratégie pré définie sur le plan scientifique. La conséquence a été l'accumulation de plusieurs centaines de milliers d'échantillons tumoraux congelés de qualité médiocre et/ou mal annotés dans l'espoir infondé de les valoriser ultérieurement. Pour essayer d'optimiser l'offre disponible et établir aussi des partenariats scientifiques, des catalogues virtuels affichant les

différentes collections et les différents échantillons disponibles ont été élaborés [8]. Malgré ces efforts, la mise à disposition de ces milliers d'échantillons congelés n'a pas été particulièrement efficace. Un autre écueil a été les sources de financement variables pour les biobanques et la difficulté de distinguer les tumorothèques financés par l'INCA et les Centres de Ressources Biologiques (ou CRB) définis par l'Organisation de Coopération et de Développement Economique), financés par l'Inserm. Ainsi les CRB pouvaient aussi collectionner des tissus tumoraux congelés, mais aussi des biofluides (échantillons sanguins, urines, selles...). Cette organisation a créé une certaine confusion car les lieux de stockage des différentes collections congelées pouvaient être différents selon les institutions, i) tissus à proximité des laboratoires d'anatomie-pathologique, ii) biofluides dans les laboratoires de biologie, comme les laboratoires de biochimie, d'hématologie ou plus rarement de génétique. Progressivement, l'intérêt des tissus fixés et inclus en paraffine a pris une place majeure dans les projets de médecine personnalisée en oncologie [9]. En effet, la grande majorité des biomarqueurs à visée prédictive ou pronostiques peuvent être évalués par des analyses immunohistochimiques et de biologie moléculaire réalisées à partir des protéines et des acides nucléiques analysés sur ces tissus fixés. Ainsi, une demande croissante et brutale de ces blocs de paraffine a pris le pas sur les demandes de tissus tumoraux congelés. Ces échantillons fixés étant dédiés à des projets de recherche, il est apparu naturel de les intégrer dans les biobanques et de leur faire suivre le circuit consacré aux échantillons biologiques consacrés à la recherche. En réalité, un nombre considérable de ces blocs sont stockés dans les laboratoires d'anatomie pathologique et non dans des biobanques, pour plusieurs raisons, i) la préparation distinctes de blocs réservés au sanitaire et de blocs réservés à la recherche n'est pas réaliste au quotidien (disponibilité du matériel biologique

exploitable, circuits automatisés, stockage etc..) ; ii) les pathologistes souhaitent conserver ces blocs dans leur service, certainement pour des raisons médico-légales ; iii) le contrôle de qualité d'un bloc (qualité de l'échantillon, diagnostic, pourcentage de cellules tumorales) assuré par le pathologiste demande un temps de vérification à prendre en considération avant un éventuel déstockage vers la biobanque ; iv) l'absence de consentement pour la recherche signé du patient est constatée dans la grande majorité des cas. L'utilisation de ces blocs avec une non opposition du patient a été une solution souvent exploitée pour utiliser ces blocs à des fins de recherche. Les « tissue micro arrays » (TMA) construits à partir des blocs tissulaires peuvent être aussi collectionnés dans une biobanque et constituent une importante ressource biologique pour la validation de certains biomarqueurs, en permettant de réaliser rapidement des analyses sur plusieurs centaines de tumeurs. Hormis ces échantillons tissulaires bruts, certaines biobanques ont constitué des collections d'ADN ou d'ARN tissulaires à partir de tissus congelés ou plus rarement fixés. Les biofluides constituent également une source très importante de collections dans les biobanques tumorales et ces collections prennent une importance exponentielle. Ainsi, les collections de plasma, de sérum), de sang total, de culots leucocytaires, voire plus exceptionnellement de plaquettes sont aujourd'hui en pleine évolution. Hormis les échantillons sanguins, d'autres collections de biofluides sont parfois réalisées à partir d'urines, de selles (microbiote), de liquides pleuraux, voire plus exceptionnellement de liquides céphalo-rachidiens.

Certaines biobanques ont continué à fonctionner comme des entrepôts sécurisés d'échantillons, assurant leur enregistrement, leur stockage et leur mise à disposition. La plupart des biobanques proposent désormais des prestations plus ou

moins variées selon les plateaux techniques associés comme, des plateformes d'immunohistochimie, de biologie moléculaire, de construction de TMA à façon, etc.

La prise en compte et l'évaluation des coûts de stockage, de la cession des échantillons, et des prestations sont devenues une nécessité pour assurer, du moins en partie, la pérennisation du fonctionnement des biobanques [10]. Initialement totalement budgétisé par l'INCA (tumorothèque) et par l'Inserm (CRB), le fonctionnement des biobanques tumorales est assuré depuis quelques années par une enveloppe financière délivrée par la Direction Générale de l'Offre des Soins (DGOS) (au prorata de la déclaration du nombre d'échantillons collectés et cédés, et selon l'obtention d'une certification). Cette évolution conduit à intégrer de plus en plus un budget associé au travail des biobanques dans les projets de recherche associant une utilisation d'échantillons biologiques. Cependant, la pérennité des biobanques n'est envisageable qu'à la condition qu'un budget socle garantisse le fonctionnement au long cours de l'infrastructure, après évaluation de son activité et des retombées sur la recherche biomédicale.

Enfin, les démarches qualité sont aujourd'hui indispensables et conduisent à la certification de la quasi-totalité des biobanques en France (selon les normes ISO 9001 et/ou NFS96-900). Les biobanques s'orientent aujourd'hui vers l'obtention d'une norme internationale spécifique ISO 20 387, dans le but de faciliter les échanges européens et internationaux, notamment avec des partenaires industriels. Ces démarches qualité ont le mérite de mettre en place des indicateurs de suivi d'activité permettant de définir des actions stratégiques d'optimisation [11-13].

### **3. Comment être une biobanque compétitive?**

Faute de régulation par la plupart des états membres de l'OCDE, le nombre des biobanques à travers le monde est croissant, entraînant ainsi entre elles une compétition selon les projets et les clients académiques et industriels. La nécessité pour une biobanque d'être compétitive répond aux besoins suivants, *i*) un besoin de visibilité scientifique ; les échantillons biologiques étant destinés à des projets de recherche, l'activité des biobanques est évaluée sur le nombre et l'importance des projets et publications associés, *ii*) un besoin d'attractivité, notamment dans le cadre de collaboration publique-privée et pour le développement de projets d'innovation, *iii*) un besoin financier ; les projets étant une source de revenus, *iv*) un besoin de reconnaissance par les organismes hébergeant la biobanque [14].

Pour être compétitive une biobanque doit faire des choix d'organisation et de stratégie [15-17]. En particulier, il est difficile de maintenir un niveau d'expertise dans plusieurs domaines d'activité et s'orienter vers un nombre restreint de pathologies permet de se concentrer sur l'exhaustivité des collections et d'y associer des données cliniques et biologiques complètes, pouvant être complexes [18]. 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 mené en concertation avec les services cliniques concernés [19]. La visibilité extérieure est accrue lorsque les partenaires d'une biobanque ont accès à différents expertises (diagnostic anatomopathologique réalisé par des pathologistes seniors dans le domaine concerné ; 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, le recueil des événements (progression, métastase, décès) peuvent être plus aisément intégrés aux bases biologiques. Dans ce contexte, il est souvent plus facile de s'associer à un réseau national ou international d'experts d'un même domaine. De plus, une certification de la biobanque est aussi essentielle à la robustesse de son fonctionnement et la qualité du service rendu.

Une biobanque compétitive doit développer des projets innovants, notamment en lien avec l'industrie pharmaceutique ou des biotechnologies. Ces projets permettent de transférer les résultats d'innovation ainsi acquis vers la pratique clinique, après une phase de validation réalisée avec des échantillons biologiques provenant des biobanques [20].

Bien d'autres points permettent à une biobanque d'être compétitive, et on citera comme exemple la possibilité de développer des xénogreffes, à partir de tissus tumoraux frais implantés dans des souris. Ces xénogreffes constituent des modèles extrêmement utiles pour étudier les processus de carcinogénèse mais aussi pour tester l'action de nouveaux traitements [21, 22].

#### **4. Les enjeux actuels: l'intégration des « big » ou des « fat » data**

Le nombre de données accumulées par patient peut être considérable [23]. On peut facilement imaginer que les analyses tissulaires morphologiques, protéiques (biomarqueurs immunohistochimiques sur les différents compartiments tissulaires),

transcriptomiques, épigénétiques et génétiques vont générer de plus en plus d'informations. Les analyses réalisées à partir des « biopsies liquides » conduisent aussi à de nombreuses données (génomiques à partir de l'ADN libre circulant ou des culots leucocytaires, transcriptomiques, épigénétiques, protéiques, etc.). A ces données biopathologiques s'ajoutent l'ensemble des informations cliniques, radiologiques et thérapeutiques. Il faut aussi intégrer les points suivants, *i*) souvent la survie des patients est améliorée au prix de changement de ligne de traitement, *ii*) les stratégies thérapeutiques s'adaptent à l'apparition des résistances aux différentes molécules, ce qui nécessite un monitoring biologique. Ce monitoring est donc associé à de nouveaux échantillons parvenant à des temps différents ensuite dans la biobanque, ce qui produit de nouvelles données pour un seul patient. Ainsi il faut distinguer deux types de données massives:

- celles qui concernent un nombre assez restreint de patient (au maximum 3000 à 5000 patients par exemple). Ici pour chaque patient, un nombre considérable de constantes et d'informations est enregistré. Ces données dites « fat data » sont de plus en plus présentes au sein d'une biobanque tumorale dédiée à un nombre restreint de pathologie et qui a les capacités de gérer les résultats des analyses décrites plus haut.

- celles qui concernent des cohortes de plusieurs centaines de milliers de patients. Ici ces données sont issues de réseaux parfaitement structurés réunissant plusieurs dizaines de biobanques ou bien de centres experts dans lesquels les échantillons ont tous été collectés. Ces organisations peuvent générer des « big data » [23-25].

L'enjeu est de pouvoir utiliser ces données massives et de pouvoir faire aussi des requêtes informatiques parfois complexes et obtenir des résultats assez rapidement. Idéalement, un même logiciel devra intégrer des données provenant de sources différentes. Ce logiciel doit donc posséder des capacités de stockage et d'analyse très

puissants. Il s'agit ici d'un défi important pour les départements des systèmes d'information des établissements de santé en France, car non préparés à ce type de problématique. La sécurisation de ces données est aussi un enjeu majeur, et la mise en place des systèmes de sauvegarde et de « back up » devra être anticipée. Dans ce contexte le projet national de « Health Data Hub » a pour objectif de répondre à ces contraintes (26)

Un des éléments à prendre en compte est l'amélioration rapide de la technologie et les méthodes d'analyse génomique qui deviennent de plus en plus sensibles. Ainsi des résultats moléculaires déjà intégrés pour un patient dans une base de données associée à une biobanque peuvent s'avérer incomplets ou partiels si l'ADN déjà analysé est à nouveau étudié en utilisant de nouvelles approches de séquençage. C'est dans ce contexte que l'on pourrait considérer que la meilleure façon de stocker de l'information (et sans occuper des espaces de stockage informatique particulièrement volumineux) est de garder celle-ci disponible au sein de l'ADN extrait. Dans cet optique, deux écueils sont possibles, i) un stockage du matériel biologique dans des congélateurs ou des cuves d'azote et/ou, ii) une collection centralisée dans un seul espace et qui pourrait disparaître en cas de sinistre. La possibilité d'encapsuler de l'ADN lyophilisé et de garder les capsules à température ambiante permet d'éviter les accidents de dégradation de l'ADN liés à une rupture de la chaîne du froid et donne aussi l'opportunité de créer des collections miroirs dupliquées sur plusieurs lieux de stockage [27-29].

## **5. Quel rôle pour le pathologiste?**

Le pathologiste joue un rôle central dans l'organisation des biobanques tumorales. Cette mission vise à faciliter les collaborations du laboratoire avec la biobanque. Dans ce contexte, il existe deux situations distinctes, soit le pathologiste est le responsable de la biobanque, soit il s'agit d'un des partenaires de la biobanque. Dans le premier cas de figure les secteurs de pathologie et de la biobanque ont des liens privilégiés. Dans le deuxième cas de figure, ces liens doivent être maîtrisés et le secteur de stockage doit être alimenté régulièrement par des échantillons tissulaires en provenance du secteur de pathologie clinique. Le travail du pathologiste est ici plus important avec le choix des blocs tissulaires et des informations à transmettre. Les échantillons tissulaires congelés ne font alors en général que transiter par le laboratoire de pathologie. Dans tous les cas c'est bien le pathologiste qui seul décide de l'orientation de tout ou partie de l'échantillon biologique vers une utilisation aux fins de recherche. De plus, le pathologiste participe activement à la gestion de la base de données en recueillant les informations sur les analyses morphologiques et immunohistochimiques, voire aussi de biologie moléculaire selon l'organisation du laboratoire.

## **6. Quels défis pour le futur ?**

Plusieurs défis se dessinent dans un avenir plus ou moins proche et les biobanques doivent répondre à un certain nombre d'objectifs : i) Comment 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 ? ii) Comment intégrer, stocker et traiter des volumes de données complexes et hétérogènes ? [30-32]. Il est important de lier les images aux échantillons tissulaires congelés ou fixés en paraffine. Ainsi les images virtuelles (lames d'histologie ou imageries radiologiques) doivent être intégrées dans les bases de données, ce qui nécessite des espaces de stockages adaptés, qui seront probablement délocalisés à distance des établissements de santé [33]. Un point difficile est de réaliser d'une façon dynamique et rapide l'intégration d'un flux de données provenant de plusieurs épisodes cliniques survenant à différents temps de la maladie. Ainsi se pose la question de gérer des informations « dynamiques » provenant de milliers (ou de centaines de milliers) de patients ? L'implication du grand public est aussi importante afin de maintenir et d'afficher la transparence des actions et des projets réalisés à partir des échantillons de patients [34, 35]. Au-delà de ces quelques défis, l'arrivée de nouvelles approches technologiques et les nouveaux domaines de la science appliqués à la médecine vont aussi impacter l'avenir des biobanques. Ces nouveaux domaines vont probablement conduire à faire diminuer considérablement la nécessité de collecter massivement des échantillons biologiques, et d'être beaucoup plus sélectif sur le choix des patients chez qui ces échantillons devront être prélevés. Ainsi les programmes d'intelligence artificielle seront alimentés et utilisés en lien avec les bases de données clinico-biologiques et les données de l'imagerie. L'exploitation des données complexes pourra aboutir à l'utilisation de nouveaux biomarqueurs ou de nouvelles associations de biomarqueurs robustes en lien avec des traitements uniques ou combinés, en particulier dans le domaine prometteur de l'immunothérapie des cancers [36-40]. Le domaine de la « radiomic » est aussi une source majeure d'information à considérer en lien avec les données de la pathologie et des biomarqueurs tissulaires [31].

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 [41]. Cet accès pourra être envisagé pour des partenaires extérieures désireux de connaître les échantillons disponibles et les données complexes associées.

Enfin, compte tenu de ces évolutions rapides et d'une modification des activités relatives au métier des biobanques, il sera 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 [42-44]

## **Conclusion**

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. Quelques 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 », soit 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 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, de métabolomique. A nouveau la qualité des données (liées aussi à la qualité des échantillons et donc des étapes pré analytiques) est le point essentiel à contrôler [45]. L'apport des données cliniques et d'imagerie (radiologie, histologie...) représente une autre étape clé, rendu délicate à gérer compte tenu de leurs caractères dynamiques (hospitalisation multiple, épisodes de la maladie, différents traitements, ...).

Les ministères chargés de la santé et de la recherche et de l'innovation, les Centre Hospitalo-Universitaires et de Lutte contre le Cancer ainsi que les organismes de recherches publiques doivent déployer tous les moyens nécessaires afin d'être à la hauteur de ce défi dans un contexte international très compétitif. En particulier, les goulots d'étranglement liés à des systèmes d'information parfois obsolètes doivent être levés. Finalement, certains freins consécutifs aux nouvelles réglementations régissant les règles de la confidentialité et de l'anonymisation des données personnelles (règles générales de la protection des données) peuvent aussi ralentir la mise en place des bases informatiques performantes, ainsi que l'exploitation et les transferts de ces données vers la recherche [41, 46]. Le déploiement du projet national « Health Data Hub » représente un espoir important pour la recherche biomédicale dont la compétitivité dépend en grande partie de l'accès à des données biocliniques et génomiques associées à des échantillons biologiques dont l'assurance-qualité et la traçabilité sont garanties [26, 47].

**Conflit d'intérêt:** Les auteurs ne déclarent pas de conflit d'intérêt

**Remerciements :** l'Inserm, IBiSA, la Fondation ARC pour la Recherche sur le Cancer, la Ligue Départementale 06 de Lutte contre le Cancer, le Conseil Départemental 06, et le Cancéropôle PACA

## Références

1. Hewitt RE. Biobanking: the foundation of personalized medicine. *Curr Opin Oncol* 2011;23:112-9.
2. Liu A, Pollard K. Biobanking for Personalized Medicine. *Adv Exp Med Biol* 2015;864:55-68.
3. Vaught J. Biobanking Comes of Age: The Transition to Biospecimen Science. *Annu Rev Pharmacol Toxicol* 2016;56:211-28.
4. Zatloukal K, Hainaut P. Human tissue biobanks as instruments for drug discovery and development: impact on personalized medicine. *Biomark Med* 2010;4:895-903.
5. Bycroft C, Freeman C, Petkova D, Band G, Elliott LT, Sharp K, et al. The UK Biobank resource with deep phenotyping and genomic data. *Nature* 2018;562:203-9.
6. Doucet M, Becker KF, Björkman J, Bonnet J, Clément B, Daidone MG. Quality Matters: 2016 Annual Conference of the National Infrastructures for Biobanking. *Biopreserv Biobank* 2017;15:270-6.

7. Simeon-Dubach D, Watson P. Biobanking 3.0: evidence based and customer focused biobanking. Clin Biochem 2014;47:300-8.
8. Chabannon C, Honstettre A, Daufresne LM, Martin PM, Bonnetaud C, Birtwisle-Peyrottes I, et al. Publication of biological samples collections catalogues by tumor banks. Bull Cancer 2010;97:181-9.
9. Gaffney EF, Riegman PH, Grizzle WE, Watson PH. Factors that drive the increasing use of FFPE tissue in basic and translational cancer research. Biotech Histochem 2018;93:373-86.
10. Clément B, Yuille M, Zaltoukal K, Wichmann HE, Anton G, Parodi B, et al. EU-US Expert Group on cost recovery in biobanks. Public biobanks: calculation and recovery of costs. Sci Transl Med 2014;6:261fs45.
11. Furuta K, Allocca CM, Schacter B, Bledsoe MJ, Ramirez NC. Standardization and Innovation in Paving a Path to a Better Future: An Update of Activities in ISO/TC276/WG2 Biobanks and Bioresources. Biopreserv Biobank 2018;16:23-7.
12. Hofman V, Ilie M, Long E, Washetine K, Chabannon C, Figarella-Branger D, et al.. Measuring the contribution of tumor biobanks to research in oncology: surrogate indicators and bibliographic output. Biopreserv Biobank 2013;11:235-44.

13. McCall SJ, Branton PA, Blanc VM, Dry SM, Gastier-Foster JM, Harrison JH, et al. The College of American Pathologists Biorepository Accreditation Program: Results from the First 5 Years. *Biopreserv Biobank* 2018;16:16-22.
14. Hofman P, Bréchet C, Zatloukal K, Dagher G, Clément B. Public-private relationships in biobanking: a still underestimated key component of open innovation. *Virchows Arch* 2014;464:3-9.
15. Hofman P. The policies of tumor biobankers: Main strategies and an example of the policies adopted by the Nice Hospital biobank, France. In: *Public regulations of tumor banks. Establishment, heritage status, development and sharing of human biological samples*. X Bioy Ed, ISBN 978-3-319-90562-4, Springer Nature, Switzerland.
16. Patil S, Majumdar B, Awan KH, Sarode GS, Sarode SC, Gadgil AR, et al. Cancer oriented biobanks: A comprehensive review. *Oncol Rev* 2018;12:357.
17. Rush A, Matzke L, Cooper S, Gedye C, Byrne JA, Watson PH. Research Perspective on Utilizing and Valuing Tumor Biobanks. *Biopreserv Biobank* 2018 Dec 21.
18. Washetine K, Heeke S, Bonnetaud C, Kara-Borni M, Ilié M, Lassalle S, et al. Establishing a Dedicated Lung Cancer Biobank at the University Center Hospital of Nice (France). Why and How? *Cancers (Basel)* 2018;10.

19. Hofman V, Bonnetaud C, Gaziello MC, Ilie M, Lassalle S, Butori C, et al. The Nice CHU biobank experience to collect patients' informed consent for research context (2004-2009). *Ann Pathol* 2010;30:337-43.
20. Heeke S, Hofman V, Long-Mira E, Lespinet V, Lalvée S, Bordone O, et al. Use of the Ion PGM and the GeneReader NGS Systems in Daily Routine Practice for Advanced Lung Adenocarcinoma Patients: A Practical Point of View Reporting a Comparative Study and Assessment of 90 Patients. *Cancers (Basel)* 2018;10.
21. Ilie M, Nunes M, Blot L, Hofman V, Long-Mira E, Butori C, et al. Setting up a wide panel of patient-derived tumor xenografts of non-small cell lung cancer by improving the preanalytical steps. *Cancer Med* 2015;4:201-11.
22. Yada E, Wada S, Yoshida S, Sasada T. Use of patient-derived xenograft mouse models in cancer research and treatment. *Future Sci OA* 2017;4:FSO271.
23. Zisis K. Biobanking with Big Data: A Need for Developing "Big Data Metrics". *Biopreserv Biobank* 2016;14:450-1.
24. Cox N. UK Biobank shares the promise of big data. *Nature* 2018;562:194-5.
25. de Lecuona I, Villalobos-Quesada M. European perspectives on big data applied to health: The case of biobanks and human databases. *Dev World Bioeth* 2018;18:291-8.

26. [https://solidarites-sante.gouv.fr/IMG/pdf/181012\\_-\\_rapport\\_health\\_data\\_hub.pdf](https://solidarites-sante.gouv.fr/IMG/pdf/181012_-_rapport_health_data_hub.pdf)
27. Clermont D, Santoni S, Saker S, Gomard M, Gardais E, Bizet C. Assessment of DNA encapsulation, a new room-temperature DNA storage method. *Biopreserv Biobank* 2014;12:176-83
28. Washetine K, Kara-Borni M, Heeke S, Bonnetaud C, Félix JM, Ribeyre L, et al. Ensuring the Safety and Security of Frozen Lung Cancer Tissue Collections through the Encapsulation of Dried DNA. *Cancers (Basel)* 2018;10.
29. Washetine K, Heeke S, Ribeyre C, Bourreau C, Normand C, Blons H, et al. DNAsell protects DNA stored at room temperature for downstream NGS studies. *Biopreserv Biobank* 2019 sous presse
30. Chakraborty S, Hosen MI, Ahmed M, Shekhar HU. Onco-Multi-OMICS Approach: A New Frontier in Cancer Research. *Biomed Res Int* 2018;2018:9836256
31. Sun R, Limkin EJ, Vakalopoulou M, Dercle L, Champiat S, Han SR, et al. A radiomics approach to assess tumour-infiltrating CD8 cells and response to anti-PD-1 or anti-PD-L1 immunotherapy: an imaging biomarker, retrospective multicohort study. *Lancet Oncol* 2018;19:1180-91.

32. Wang L, Yu X, Zhang C, Zeng T. Detecting personalized determinates during drug treatment from omics big data. *Curr Pharm Des* 2018 Nov 5.
33. Hamilton PW, Bankhead P, Wang Y, Hutchinson R, Kieran D, McArt DG, et al. Digital pathology and image analysis in tissue biomarker research. *Methods* 2014;70:59-73.
34. Harati MD, Williams RR, Movassaghi M, Hojat A, Lucey GM, Yong WH. An Introduction to Starting a Biobank. *Methods Mol Biol* 2019;1897:7-16.
35. Mabile L, Dalglish R, Thorisson GA, Deschênes M, Hewitt R, Carpenter J, et al. BRIF working group. Quantifying the use of bioresources for promoting their sharing in scientific research. *Gigascience* 2013;2:7.
36. He J, Baxter SL, Xu J, Xu J, Zhou X, Zhang K. The practical implementation of artificial intelligence technologies in medicine. *Nat Med* 2019;25:30-6.
37. Koelzer VH, Sirinukunwattana K, Rittscher J, Mertz KD. Precision immunoprofiling by image analysis and artificial intelligence. *Virchows Arch* 2018 Nov 23
38. Salto-Tellez M, Maxwell P, Hamilton P. Artificial Intelligence - The Third Revolution in Pathology. *Histopathology* 2018 Oct 1.

39. Sherbet GV, Woo WL, Dlay S. Application of Artificial Intelligence-based Technology in Cancer Management: A Commentary on the Deployment of Artificial Neural Networks Anticancer Res.2018;38:6607-13.
40. Topol EJ. High-performance medicine: the convergence of human and artificial intelligence. Nat Med 2019;25:44-56.
41. Price WN 2nd, Cohen IG. Privacy in the age of medical big data. Nat Med 2019;25:37-43.
42. Gormally E, Hardy I, Caboux E, di Donato JH, Hainaut P, Hofman P. Training the Next Generation of Biobankers: A Two-Year Master's Course in the Management of Biobanks. Biopreserv Biobank 2017;15:438-50.
43. <https://univ-cotedazur.fr/archives/english/uca-education/diplomas/biobanks-complex-data>.
44. Williams RR, Gupta D, Yong WH. Orientation and Training of New Biobank Personnel. Methods Mol Biol 2019;1897:51-63.
45. Lee JE, Kim YY. Impact of Preanalytical Variations in Blood-Derived Biospecimens on Omics Studies: Toward Precision Biobanking? OMICS. 2017;21:499-508
46. <https://www.donnees-rgpd.fr>.

47. Dagher G, Lavitrano ML, Hofman P. Le “next generation biobanking”, un défi des données numériques. *Med Sci (Paris)* 2018; 34: 849–851.