

# Coordination et partage de données de cohortes

**E**n 2020, 75 % des décès au niveau mondial seront dus à des maladies chroniques complexes. Les progrès au niveau des technologies de haut débit, notamment en génomique, ont permis des avancées importantes, notamment en termes de définitions de sous-groupes de pathologies. De nombreuses recherches sont articulées autour d'études de caractéristiques, de différences et de singularités génétiques. Dans le champ thérapeutique, les travaux visent à développer une médecine personnalisée qui répond efficacement et rapidement aux besoins des patients. Dans ce cadre, les échantillons biologiques et leurs données associées sont essentiels pour :

- Élucider les interactions entre des facteurs génétiques et environnementaux responsables de pathologies, comprendre les mécanismes des maladies complexes et des maladies rares.
- Développer des programmes médicaux ajustés aux particularités des patients et savoir, en fonction de facteurs de risque identifiés, adapter la prévention, diagnostiquer précocement et orienter vers un traitement adapté.
- Déterminer les influences des facteurs génétiques sur les effets et les résultats des traitements, prévenir les effets délétères, fournir des médicaments sûrs, efficaces et adaptés en fonction des particularités individuelles.
- Identifier de nouvelles cibles thérapeutiques, développer de nouveaux médicaments ou améliorer ceux existants.

La recherche utilisant ces ressources, telles que la génomique, la protéomique ou encore l'imagerie, a accompli des avancées majeures dans la compréhension des facteurs physiopathologiques qui sous-tendent les maladies complexes et rares. Les études d'associations génétiques (GWAs) ont permis l'identification d'associations génétiques dans le diabète type 1 et type 2, la pathologie coronarienne, le cancer colorectal, du sein, ou de la prostate, la dégénérescence maculaire, la maladie de Crohn, l'autisme, les maladies neurodégénératives. Malgré ces avancées dans la détection des associations génétiques avec les pathologies complexes, les déterminants génétiques ne représentent qu'un des facteurs qui jouent un rôle dans leur développement. L'influence du mode de vie, des facteurs environnementaux et sociaux a été reconnue et explorée depuis bien longtemps. Cela implique que l'interaction entre le gène et l'« environnement » joue un rôle essentiel dans la chaîne causale. Il est alors important que les sciences biomédicales aient accès non seulement aux études génétiques, mais également aux données épidémiolo-

giques qui incluent le mode de vie, et les informations socio-économiques et environnementales.

Par ailleurs, la recherche biomédicale rencontre différents obstacles qui limitent le développement de la recherche étiologique, de la recherche translationnelle ou du développement de nouvelles molécules thérapeutiques. Parmi ces obstacles figure l'accès aux échantillons biologiques provenant de cohortes d'individus bien définis sur le plan clinique et, en particulier dans le domaine de l'épidémiologie génétique, la nécessité d'accès à un grand nombre de sujets pour identifier des associations génétiques significatives sur le plan statistique dans les maladies complexes. Pour relever ces défis, il est indispensable de s'assurer de la qualité des échantillons, de la standardisation ou de l'harmonisation de leur collecte, transformation et conservation. De plus, il est nécessaire d'élargir le champ des pathologies concernées, d'aboutir à une interopérabilité des bases de données, et bien entendu de garder la confiance qu'accordent les citoyens à ces activités.

## Biobanques : un outil de mutualisation des bases de données biologiques

Le succès de cette entreprise repose principalement sur la qualité de l'échantillon et celle des informations qui lui sont associées. Il est donc important que la communauté de la recherche biomédicale coordonne mieux ses efforts pour assurer le recueil et la conservation d'échantillons biologiques dans des conditions techniques qui en garantissent la qualité et la possibilité de mise en commun pour des analyses groupées. C'est l'objectif majeur de l'infrastructure nationale Biobanques retenue dans le cadre du programme national « Investissements d'avenir ».

Biobanques<sup>1</sup> mobilise, au sein d'une infrastructure distribuée sur tout le territoire français, les biobanques, tumorothèques, centres de ressources microbiologiques (mBRCs) et cohortes, ainsi que des expertises associées aux collections d'échantillons biologiques. L'infrastructure s'appuie sur les acquis d'un solide réseau structuré depuis 2005, et couvre tous les champs de la recherche biomédicale. Le projet fédère 72 biobanques, dont 8 mBRCs, et implique de nombreuses équipes de recherche à travers des projets utilisant les collections d'échantillons biologiques.

La mise en place de l'infrastructure Biobanques se déroule en deux phases. Une phase de construction (2011-2016), destinée à coordonner et harmoniser les

**Georges Dagher**  
Infrastructure  
nationale  
Biobanques,  
Inserm US 13, Paris

1. <http://www.crbfrance.fr/>



différentes activités existantes ; cette phase assurera la mise en place et la préparation des services communs et des plates-formes technologiques qui seront opérationnelles progressivement au cours des prochaines années. Ultérieurement, une phase opérationnelle sera mise en œuvre (2017-2019), durant laquelle l'infrastructure fournira services et plates-formes à l'ensemble de la communauté scientifique.

L'objectif de Biobanques est d'intégrer les collections d'échantillons biologiques, les technologies et les expertises associées afin de les pérenniser et de les enrichir, cela dans le respect des cadres éthiques et juridiques français et européens. Par ailleurs, l'infrastructure assure une coordination plus efficace des actions des biobanques et des mBRCs, ainsi que la valorisation des collections au sein de projets de recherche d'excellence, tant académiques que privés. Pour atteindre ces objectifs, l'infrastructure mutualisera les moyens et les compétences en développant des services communs tels que : affaires réglementaires et éthiques, qualité de l'échantillon, bioinformatique, unité de méthodologie et de biostatistique, ainsi que des plates-formes technologiques destinées à la communauté scientifique.


Parmi les services communs figure le service bioinformatique et bases de données, clés de voûte de la recherche biomédicale, notamment génomique et protéomique. Biobanques développera, dans un premier temps, un projet pilote afin d'équiper et de tester l'interopérabilité des bases de données cliniques et analytiques provenant de projets portant sur une sélection restreinte de maladies complexes et de maladies rares. Ce projet pilote permettra d'identifier la fiabilité et la validité des modules informatiques proposés ainsi que les goulots d'étranglement et les difficultés de leur mise en œuvre. Parmi celles-ci figure l'accès aux données cliniques pseudonymisées ou anonymisées des patients. Il faudrait développer une solution sécurisée permettant le transfert, à partir des services hospitaliers, d'une partie de ces données à des fins de recherche et cela dans le respect de la confidentialité et de la volonté du patient.

### Harmoniser les données des cohortes de pays européens

Par ailleurs, l'harmonisation par pathologie du contenu de ces bases est une nécessité pour collecter les données d'un nombre suffisant de patients, dépassant souvent les 10 000 sujets, afin d'atteindre la puissance statistique suffisante pour l'analyse des polymorphismes génétiques des maladies complexes. Une telle harmo-

nisation peut être également utile dans la synthèse des informations provenant de différentes études ou encore dans le développement de nouveaux projets. Plusieurs initiatives dans cette perspective ont été entreprises au niveau international. Par exemple, l'harmonisation des données de l'étude Epic, qui associe des cohortes de plusieurs pays européens, est un élément essentiel de son succès ; il en est de même pour différentes autres études telles que GenomeEUTwin, Euralim ou encore Engage. De plus, le consortium P3G (Public Population Projects in Genomics) a entrepris depuis plusieurs années une harmonisation des bases épidémiologiques en association avec le projet européen Phoebe. Il a ainsi développé un outil pratique dénommé Datashaper<sup>2</sup>, qui inclut deux séries de données, des variables primaires nécessaires à toute étude épidémiologique et des variables d'ajustement, et qui décline ces variables primaires en une série d'informations. Datashaper est un travail collaboratif impliquant plus de 25 cohortes (études longitudinales de population) internationales provenant de 14 pays. Biobanques, en concertation avec P3G, encouragera l'accès à Datashaper et l'utilisation des outils informatiques annexes développés par le consortium.

Biobanques a pour objectifs de porter les collections d'échantillons biologiques d'origine humaine et les collections microbiologiques à un nouveau niveau de coordination, de qualité et de valorisation en développant une infrastructure pérenne et reconnue à l'échelle nationale et internationale. L'infrastructure visera en particulier à atteindre les objectifs suivants :

- Accroître l'excellence scientifique et l'efficacité de la recherche française dans les sciences de la vie.
- Atteindre une masse critique suffisante en termes de recherche et d'investissements, et éviter la duplication des efforts en établissant des liens avec d'autres projets européens et internationaux, en améliorant la qualité et la standardisation des bioressources et de données associées, et en mettant en place l'interopérabilité des bases de données. Cela en étroite collaboration avec les infrastructures européennes BBMRI<sup>3</sup> et Embarc<sup>4</sup>.
- Faciliter l'accès des chercheurs académiques et privés aux ressources biologiques et aux données associées afin de favoriser l'innovation et la compétitivité, accélérer la mise en place de partenariats public-privé. 

2. <http://www.datashaper.org/>

3. <http://www.bbMRI.eu/>

4. <http://www.embarc.eu/>