

RAPPORT 12-01

(Au nom de la Commission I) *

Techniques d'analyse du génome et de son expression : applications médicales

MOTS-CLÉS : SPECTROMÉTRIE DE MASSE. IMAGERIE PAR RÉSONANCE MAGNÉTIQUE. SÉQUENCE NUCLÉOTIDIQUE. TUMEURS DU CÔLON. TUMEURS DU RECTUM. MÉLANOME. TUMEURS DU SEIN.

Methods of analyzing the genome and its expression : medical applications

KEY-WORDS (Index medicus) : MASS SPECTROMETRY. MAGNETIC RESONANCE IMAGING. BASE SEQUENCE. COLONIC NEOPLASMS. RECTAL NEOPLASMS. MELANOMA. BREAST NEOPLASMS

L'auteur déclare ne pas avoir de liens d'intérêt en relation avec le contenu de cet article

Jean-Yves LE GALL *

RÉSUMÉ

L'exploration biologique médicale est passée récemment d'une approche ponctuelle (détermination d'un ou de quelques paramètres) à une approche globale c'est-à-dire la prise en compte simultanée de l'ensemble du génome ou de ses différents niveaux d'expression ; ces explorations à grande échelle sont qualifiées alors de « Génomique », de « Transcriptomique », de « Protéomique »...Ce rapport traite de quatre méthodes analytiques globales (spectrométrie de masse, résonance magnétique nucléaire, séquençage de l'ADN et puces à ADN) et aborde ensuite à titre

* Membre de l'Académie nationale de médecine ; e-mail : legall56@orange.fr

* Membres de la Commission I

Membres titulaires : M^{mes} ADOLPHE, MARCELLI, MM. ARDAILLOU, BAULIEU, CABANIS, CAZENAVE, DENIS, DREUX, GALIBERT, HAUW, LAUNOIS, LE GALL J.Y. (Président), MILGROM, MONTAGNIER, NETTER, NEZELOF, NICOLAS J.P., PARODI, PESSAC, RONCO, ROSSET, STRAER, TIOLLAIS, VINCENT.

Membres correspondants : M^{mes} DEJEAN-ASSEMAT, EVAIN-BRION, MOREL, MM. BASTIDE, BRICE, DEBRE (Secrétaire), DELMAS, DELPECH, DOUAY, DUSSAULE, GRIEDLANDER, JEANTEUR, LE BOUC, MAQUART, SOUBRIER, STOLTZ, SWYNGHDAUW, VIGNERON.

Membres invités : M^{me} LECOMTE, MM. CAEN, CHOUARD, ROCHEFORT.

d'exemples quatre chapitres de l'oncologie, domaine où les retombées attendues sont sans doute les plus prometteuses. Les applications de la résonance magnétique nucléaire restent limitées à la biologie fondamentale en particulier structurale ; en biologie médicale cette technique se heurte en effet à deux inconvénients : sa faible sensibilité et ses durées d'analyse relativement longues. Avec l'évolution des appareillages et de leur support informatique, la spectrométrie de masse est en pleine expansion dans les laboratoires hospitaliers ; ses applications sont multiples : analyse de stérols, stéroïdes, acides biliaires, prostaglandines, glyco et sphingolipides.... ; détermination de SNPs et de mutations, génotypages à haut débit, diagnostic rapide d'infections bactériennes, dépistage et diagnostic de maladies héréditaires, analyses toxicologiques etc. Les puces oligonucléotides atteignent une densité de 2 500 000 SNPs, elles sont utilisées dans les études GWAS (« Genome Wide Association Studies ») menées dans le but d'identifier les facteurs génétiques en cause dans les maladies communes dites multifactorielles, mais les résultats de ces études sont pour le moment relativement décevants. La technologie de « CGH array » utilisant des puces à grands fragments d'ADN (BAC ou PAC) a conduit à la notion de variations à grande échelle du génome ou CNV (« Copy Number Variants »), impliquant environ 4,8 % du génome et dont les anomalies sont en cause dans différentes situations pathologiques héréditaires ou non. L'évolution la plus importante concerne le séquençage de l'ADN : les nouvelles technologies à haut débit laissent penser que dans un avenir proche le séquençage d'un génome individuel sera possible en quelques heures et pour un coût de revient modique ; leurs performances sont en particulier utilisées aujourd'hui pour l'identification des gènes en cause dans les maladies héréditaires très rares et pour la caractérisation des anomalies génomiques des leucémies et des tumeurs. Dans ce dernier cas, différents programmes internationaux ont pour objectifs d'améliorer les classifications histologiques, d'identifier les gènes « critiques », de définir des critères moléculaires de pronostic et de traitements ciblés, dont l'un des exemples est l'inhibiteur sélectif (PLX4032) du gène BRAF muté (V600E) dans la moitié des cas de mélanome.

SUMMARY

Medical biology is rapidly evolving from the study of one or several individual analytes to a more global approach in which the genome and its expression are taken into account as a whole, by means of genomic, transcriptomic or proteomic methods. This report focuses on four such analytical methods (mass spectrometry, nuclear magnetic resonance, DNA sequencing, and DNA chips) and four aspects of oncology, the field in which these methods offers most promise. Nuclear magnetic resonance is restricted to applications in fundamental biology, and particularly structural studies, being relatively slow and poorly sensitive. With new devices and greater computing power, mass spectrometry is finding increasing uses in hospital laboratories, for the analysis of sterols, steroids, bile acids, prostaglandins, glyco- and sphingolipids, etc., detection of SNPs and mutations, high-throughput genotyping, rapid diagnosis of bacterial infections, screening and diagnosis of hereditary diseases, toxicology, etc. Oligonucleotide microarrays, some now reaching a density of 2 500 000 SNPs, are used for genome-wide association studies (GWAS) to identify genetic factors underlying common multifactorial diseases, although the results have so far proven rather disappointing. Results obtained with CGH

arrays, using chips bearing large DNA fragments (BAC or PAC), have revealed large scale genomic variations, or copy number variants, involving some 4.8 % of the genome and being implicated in a variety of hereditary and non hereditary disorders. The most impressive developments concern DNA sequencing : new high-throughput technologies will be able to sequence the entire genome in a few hours at near-negligible cost ; they are currently used to identify culprit genes in very rare diseases and to characterize genetic anomalies in leukemia and solid tumors. Several international programs are seeking to improve histological classifications, to identify " critical " genes, to identify molecular prognostic indicators and to develop targeted treatments. One example is a selective inhibitor (PLX4032) of the BRAF gene, which is mutated (V600E) in about 50 % of melanomas.

L'analyse biologique dans le domaine expérimental et dans ses applications médicales est passée récemment d'une approche ponctuelle limitée à la détermination d'un ou d'un petit nombre de paramètres, à une approche globale c'est-à-dire à la prise en compte simultanée de l'ensemble du génome ou de ses différents niveaux d'expression. Cette évolution technique a vu l'apparition d'un nouveau vocabulaire qualifié globalement d' « omique » : à la « génomique », éventuellement limitée à l'« exomique » c'est-à-dire à la seule étude des séquences codantes et des séquences immédiatement environnantes, sont venues s'ajouter la « transcriptomique » (analyse des ARN messagers ou transcriptome), la « protéomique » (analyse des protéines), la métabolomique (« analyse des métabolites du métabolisme intermédiaire »)..... Cette évolution est le résultat d'une augmentation considérable et rapide des performances de la technologie analytique, elle-même largement sous-tendue par les progrès de l'informatique ; elle permet aujourd'hui des analyses non seulement globales mais aussi très rapides, qualifiées de « haut débit » ; elle devrait également conduire à une profonde révision des concepts médicaux, en particulier nosologiques et diagnostiques, ainsi qu'à une amélioration importante des outils pharmacologiques. Ce rapport a pour objet de dresser un état, certes partiel, de cette évolution scientifique en traitant de quatre méthodes analytiques « globales » : la spectrométrie de masse, la résonance magnétique nucléaire, le séquençage de l'ADN et les puces ADN, et en abordant ensuite quatre chapitres de l'oncologie, domaine médical ou les retombées pratiques, en particulier thérapeutiques, sont sans doute les plus prometteuses.

Spectrométrie de masse

Le principe de la spectrométrie de masse (MS) repose sur l'éventuelle fragmentation de la ou des grosses molécules (protéines) à analyser en ions qui sont ensuite séparés en fonction de leurs masses et de leurs charges. Les appareils correspondants sont habituellement utilisés comme systèmes de détection à la sortie d'un chromatographe en phase gazeuse ou d'un chromatographe liquide à haute performance. La technique ne pouvant analyser que

des molécules ionisées et en phase gazeuse, son champ d'application ne concernait à l'origine que de petites molécules (sucres, acides aminés, stéroïdes, dérivés de l'acide arachidonique....) ; il a été ensuite considérablement élargi par la mise au point de techniques permettant de vaporiser et de ioniser les protéines ; encore plus récemment des développements technologiques ont étendu la méthode à des échantillons solides ou à des coupes de tissus. Le spectromètre de masse comporte ainsi schématiquement trois parties : chambre d'ionisation/ vaporisation de l'échantillon, système séparateur et capteur des signaux couplé à un système d'analyse informatique. Deux méthodes d'ionisation/ vaporisation sont utilisées : — la méthode MALDI (« Matrix Associated Laser Desorption Ionization ») où l'échantillon est co-cristallisé avec une matrice puis soumis à un rayon laser puissant ; — la méthode ESI (« Electro-Spray Ionisation ») où l'échantillon est injecté dans un micro-capillaire porté à un potentiel très élevé. Les systèmes séparateurs sont de trois types : — le système de mesure du temps de vol (« Time Of Flight » ou TOF) basé sur le fait que la vitesse d'une molécule ionisée dans un champ électrique est fonction du rapport de sa masse sur sa charge (m/z) ; — la trappe à ions constituée de trois électrodes soumises à des différences de potentiel alternatif, les molécules ionisées étant éjectées en fonction de leur rapport m/z ; — le quadripôle constitué de quatre cylindres reliés deux à deux et subissant simultanément des différences de potentiel continues et alternatives ; seules les molécules ayant un rapport m/z leur permettant d'être en résonance à un instant donné peuvent traverser ce quadripôle. Toutes les combinaisons entre les différentes méthodes d'ionisation/vaporisation et de séparation sont théoriquement possibles, en pratique deux sont principalement utilisées : le MALDI-TOF et l'ESI-MS en tandem (ESI-MS-MS) ; la résolution analytique de ces systèmes et la puissance des logiciels informatiques qui leur sont couplés sont suffisantes pour établir la composition en acides aminés et même leur séquence dans le polypeptide analysé. En biologie, l'identification des protéines nécessite souvent qu'elles soient préalablement purifiées par couplage du système MS avec un système électrophorétique (gel de polyacrylamide en deux dimensions) ou avec une séparation chromatographique (chromatographie liquide haute performance ou HPLC). Enfin le système SELDI-TOF, mis au point pour l'analyse directe des échantillons de plasma, urines, LCR..., effectue un tri préalable par dépôt de l'échantillon sur différents types de barrettes (échangeuses d'anions, de cations...) suivi d'une technologie de type MALDI-TOF.

Dans le domaine des sciences du vivant, les principales applications de la spectrométrie de masse concernent la biologie structurale, la « protéomique », la « métabolomique », et l'analyse des structures complexes. La MS est un moyen très utile d'analyse fine des structures moléculaires permettant par exemple de démontrer que l'inhibition de la ribonucléotide réductase par le monoxyde d'azote est due à la fixation spécifique de ce composé (ou plus exactement de son dérivé, le peroxydinitrite) sur le site actif de l'enzyme ;

cependant en raison de difficultés expérimentales importantes ce type d'activité relève encore des laboratoires académiques. Par définition, le protéome est l'ensemble des protéines synthétisées par une cellule ; aux 25 000 gènes humains correspondent environ un million de protéines différentes (en raison des phénomènes d'épissage alternatif et des multiples modifications post-traductionnelles possibles) exprimées de façon différentielle d'un type cellulaire à l'autre. En « protéomique », les applications classiques de la MS (qui donne accès à environ 10-15 % de ces protéines) correspondent à l'identification des spots protéiques obtenus après électrophorèse bidimensionnelle ; elles concernent au premier plan la biologie fondamentale : c'est ainsi qu'il a été montré que les cryptes intestinales de souris CFTR^{-/-} étaient dépourvues d'annexine 1 et que ce déficit était retrouvé dans les cellules nasales des patients atteints de mucoviscidose, cela quelle que soit la mutation du gène CFTR en cause (l'annexine 1, présente principalement dans le colon, les poumons, le pancréas et les polynucléaires neutrophiles, a une activité anti-inflammatoire car elle inhibe la phospholipase A2 cytosolique et en définitive la libération d'acide arachidonique précurseur des prostaglandines et des leucotriènes ; son absence au cours de la mucoviscidose relève d'un mécanisme post-transcriptionnel car le taux de son ARN messager est sensiblement normal) . En biologie clinique la MS se prête à l'analyse de la composition protéique des liquides biologiques (sang, urine, LCR...) et des tissus en particulier tumoraux ; de nombreux travaux se consacrent ainsi à l'identification de biomarqueurs diagnostiques et/ou pronostiques spécifiques et sensibles, détectables de façon précoce et reproductible, susceptibles de conduire à la mise au point de tests simples et automatisables [1, 2]. En pathologie rénale, l'urine (qui peut être obtenue en grande quantité et conservée à -20 degrés ou lyophilisée pendant plusieurs années) est bien entendu un matériel biologique de choix. Les composants déterminés en routine servent à explorer la fonction rénale globale (créatinine, serum-albumine...), mais n'ont ni spécificité ni valeur pronostique, ce qui explique l'intérêt potentiel de la « protéomique » urinaire, tout en sachant que 70 % des protéines (protéines solubles et exosomes, c'est-à-dire fragments de membranes) proviennent non pas des reins mais du tractus urinaire. Aucun des nombreux travaux consacrés à ce sujet n'a encore obtenu de validation définitive et leurs résultats doivent être considérés comme préliminaires : c'est ainsi que dans les néphropathies à IgA est retrouvée une excrétion accrue d'endorepelline (extrémité C-terminale du perlecan, protéoglycane de la membrane basale) ; dans les néphropathies lupiques une « signature » est constituée de 170 pics dont deux correspondent à l'hepcidine (peptide connu par ailleurs pour être synthétisé par le foie et comme hormone régulatrice de l'absorption intestinale du fer) ; au cours des différents stades évolutifs de l'insuffisance rénale aiguë deux marqueurs semblent prometteurs : NGAL (« Neutrophil Associated Gelanilase Lipocalin ») et KIM1 (« Kidney Injury Molecule ») [3] ; chez le nouveau-né, en cas de néphropathie obstructive par dysplasie de la jonction urétéropelvienne, une signature de

53 peptides (correspondant à des fragments de collagène de la matrice extra-cellulaire) permettrait de prédire la résolution spontanée ou la nécessité d'un acte chirurgical ; enfin de nombreux travaux ont pour objet la découverte d'un ou de marqueurs qui permettraient de juger du stade de non retour dans l'évolution de l'insuffisance rénale chronique, c'est-à-dire de l'irréversibilité des lésions de fibrose en réponse à la thérapeutique. La « métabolomique » a pour objet de caractériser les variations de concentration en métabolites entraînées par tout évènement biologique (alimentation, prise de médicament ou de xénobiotique.....) ou pathologique : démonstration que les spectres urinaires MS présentent des différences entre Anglais et Suédois en raison d'une alimentation de composition différente ; variations de composition lipidique entre cerveau normal et cerveau de sujet atteint de maladie d'Alzheimer ; démonstration du caractère prédictif précoce du diabète de type II par une augmentation des taux sériques à jeun de trois acides aminés ramifiés (leucine, isoleucine, valine) et de trois acides aminés aromatiques (phénylalanine, tyrosine, tryptophane) [4]. L'analyse directe des structures cytologiques ou histologiques (spectrométrie de masse TOF-SIMS où l'échantillon est bombardé par un faisceau d'agrégats de métaux lourds entraînant une ionisation de ses constituants) a de nombreuses applications : identification de bactéries pathogènes, profilage moléculaire de biopsies (diagnostic du cancer du poumon...), imagerie moléculaire de coupes tissulaires, localisation des classes et sous-classes lipidiques ou des métabolites osidiques dans les différentes régions du cerveau, caractérisation au cours de la stéatose hépatique d'une répartition différente des diacylglycérols, des acides gras à longue chaîne et de la vitamine E entre zones adipeuse ou non [5, 6]. ...L'interprétation de cette imagerie moléculaire nécessite bien entendu une collaboration étroite entre physicien et anatomopathologiste.

Dans le domaine de la biologie médicale et en raison à la fois de son très large spectre d'analyse, de sa capacité à doser simultanément un nombre important de composés et de sa très grande sensibilité (jusqu'à l'attomole dans certaines conditions), les applications potentielles de la MS sont multiples : analyse et dosage des stérols et stéroïdes, des vitamines D et de leurs métabolites, des amines et polyamines, des glyco et sphingolipides, des prostaglandines, des acides biliaires, des acides gras saturés et insaturés....[7]. Il est possible de réaliser par cette technique un diagnostic de la plupart des déficits de la β oxydation mitochondriale des acides gras, de quelques aciduries organiques, des principales amino-acidopathies et des déficits du cycle de l'urée, ce qui a amené certains à la proposer dans le dépistage systématique des affections néonatales [8]. La technique MALDI-TOF s'est également adaptée au génotypage à haut débit, détermination de SNPs ou de mutations ponctuelles (« Mass-array Sequenom ») ; elle a pris également un essor très important dans le diagnostic rapide des infections bactériennes. Enfin la MS couvre un très large champ de la pharmacologie et de la toxicologie.

Résonance magnétique nucléaire

La résonance magnétique nucléaire (RMN) est l'étude des propriétés magnétiques des noyaux atomiques ; ceux-ci constitués de protons, neutrons entourés d'électrons, possèdent un moment dipolaire magnétique et un moment cinétique, responsables du phénomène de « spin ». Le moment magnétique de l'atome dépend du nombre de protons et de neutrons ; si ces deux nombres sont pairs ce moment est nul, ce qui explique que les atomes ou leurs isotopes étudiés par RMN correspondent à des chiffres impairs dans la classification de Mendéléïev (en biologie : ^1H , ^{13}C , ^{15}N , ^{19}F , ^{31}P). Les noyaux magnétiquement actifs (présents dans un échantillon liquide de 50 à 500 μl) sont placés dans un champ très intense (10 à 20 Tesla) généré par un aimant supraconducteur plongé dans l'hélium liquide : le moment de spin subit un phénomène de précession donnant lieu à un signal qui est amplifié puis soumis à une transformation de Fourier. Dans le cas d'une macromolécule, la RMN de l'hydrogène donne ainsi une série de pics correspondant chacun à un atome mais avec des fréquences différentes en fonction de la nature des liaisons chimiques. La RMN à deux dimensions utilise un transfert d'aimantation à travers des liaisons covalentes ou à travers l'espace ; cette technique est utile pour les études de structure et de conformation d'une molécule, elle donne également des renseignements sur les associations entre les différents types d'atomes. La technologie HR-MAS permet de travailler sur des échantillons biologiques complexes non liquides (cellules, biopsies, tumeurs...) ou même sur des organismes entiers comme *C.Elegans*. Au total la RMN est spécifique d'un type d'atome ou d'isotope, elle donne un signal distinct pour chaque atome de la molécule, elle est quantitative et non destructrice de l'échantillon.

La RMN se prête en biologie à différents types d'application . — Identification et quantification de composés, par exemple l'ATP par RMN du ^{31}P afin de suivre l'évolution de sa synthèse dans les mitochondries. — Analyse du mécanisme de réactions chimiques ou enzymatiques (par exemple transformation du glucose-6-phosphate en 6 phosphogluconate sous l'action de la glucose-6-phosphate déshydrogénase). — Mesures de distances interatomiques et reconstitution de structures 3D. — Criblage de molécules pharmacologiques (par exemple la caractérisation d'inhibiteurs du site actif de la peptide déformylase, enzyme essentielle à la viabilité bactérienne car éliminant le groupe N formyl de la méthionine du codon d'initiation de la traduction ; ces composés, à constante de dissociation très élevée, ont des propriétés antibiotiques mais sont par ailleurs sans effet sur l'enzyme mitochondriale humaine). — Analyse de liquides biologiques (plasma, LCR, urine, bile....) qui se heurte à quelques inconvénients : faible sensibilité ($\geq 10\mu\text{M}$), durée d'analyse relativement longue (dix minutes), nécessité d'une infrastructure lourde et d'un traitement informatique complexe ; la sensibilité de la méthode peut cependant être améliorée par l'utilisation d'une sonde cryogénique (l'abaisse-

ment de la température entraînant une diminution du bruit de fond), par l'étude de deux atomes différents ou par la transformation chimique préalable des molécules à analyser (par exemple par marquage isotopique au ^{13}C ou par amidification des fonctions carboxyliques par l'éthanolamine ^{15}N qui permet l'identification de plus de 200 métabolites dans les urines ou le plasma, potentiellement utilisable pour le diagnostic de maladies héréditaires du métabolisme des acides aminés, des glucosaminoglycures...). — Métabolisme de xénobiotiques, la RMN étant utilisée soit pour le dosage spécifique de quelques dérivés, soit dans une approche globale dite « métabolomique ».

Détermination de la séquence de l'ADN

Le séquençage des acides nucléiques est sans aucun doute le domaine analytique ayant bénéficié de l'évolution la plus importante au cours des vingt dernières années. Deux méthodes principales ont d'abord été utilisées : la méthode chimique de Maxam et Gilbert, aujourd'hui pratiquement abandonnée, et surtout la méthode enzymatique de Sanger aux di-désoxynucléotides qui s'est généralisée en s'adaptant à l'évolution des outils de la génétique moléculaire (endonucléases de restriction, vecteurs de clonage, DNA polymérase, « Polymerase Chain Reaction » ou PCR, marqueurs isotopiques, fluorochromes...) et des moyens techniques (séparation des fragments par électrophorèse en gel de polyacrylamide puis par électrophorèse capillaire). L'analyse des ADN de grande taille procède de la technique de « Shotgun », c'est-à-dire de leur coupure au hasard en fragments de tailles limitées qui sont ensuite clonés (classiquement en phages M13) puis séquencés ; pour obtenir une couverture la plus large possible d'un génome, il est indispensable de produire un volume de séquences équivalent à huit à dix fois sa taille présumée ; l'assemblage des séquences obtenues nécessite un traitement informatique performant, la séquence finale étant ensuite annotée pour localiser les gènes, les séquences répétées, les SNPs (« Single Nucleotide Polymorphism »)... La méthode de Sanger a été l'objet d'une automatisation conduisant à la mise au point d'appareils de plus en plus performants, et à l'origine d'une première version du génome humain en 2001.

Au cours des dernières années est apparue une nouvelle génération de séquenceurs dits à haut débit opérant en parallèle sur un très grand nombre de séquences courtes, maintenant indépendants de la méthode chimique de Sanger et reposant sur de nouvelles technologies physicochimiques. Les capacités de ces nouveaux séquenceurs sont de plus en plus grandes (actuellement 400 millions à 1 milliard de paires de bases par jour) pour un prix de revient de plus en plus faible (actuellement moins d'un dollar par million de paires de base), ce qui permet de penser que le séquençage du génome d'un patient donné sera prochainement un acte courant dans un laboratoire de génétique moléculaire hospitalier. Les coûts peuvent encore être réduits en

ciblent le séquençage sur l'ensemble des exons (« exome »), qui représente environ 1 % du génome et regroupe plus de 90 % des mutations pathogènes [9, 10]. Dès aujourd'hui sont ainsi ouverts des champs nouveaux d'investigation, en particulier l'identification des gènes de maladies très rares : il reste environ 3 500 maladies mendéliennes dont les gènes en cause n'ont pas encore été identifiés. Au lieu d'employer les procédures classiques de clonage positionnel, il suffit maintenant de séquencer un très petit nombre de génomes individuels ; quelques exemples de ce type de stratégie ont été publiés récemment : ataxie dominante et gène TGM6 identifié par séquençage des génomes de quatre patients de la même famille, syndrome de Joubert et gène TMEM216 identifié par séquençage des génomes d'un enfant atteint et de sa mère.... Ce genre de démarche se heurte néanmoins à un certain nombre de difficultés ; c'est ainsi qu'il existe environ vingt mille variations ponctuelles de séquence d'un exome à l'autre, dont quatre mille correspondant à un polymorphisme en acide aminé et une centaine à une perte de fonction potentiellement pathogène (décalage du cadre de lecture, codon stop, anomalies des sites d'épissage....). Il est donc souvent utile d'affiner cette stratégie de capture d'exons/séquençage à haut débit en s'appuyant par exemple sur une liste de gènes candidats ou sur la localisation préalable d'une région candidate (par exemple : identification du gène SPG28 en cause dans une forme de paraplégie spastique par séquençage des exons d'une région d'homozygotie du chromosome 14 cartographiée dans une grande famille marocaine [11]. Le séquençage haut débit offre aujourd'hui une puissance inégalée pour identifier les gènes de maladies héréditaires, mais aussi les facteurs génétiques de prédisposition à de nombreuses pathologies communes ; il est également susceptible de devenir un outil courant de diagnostic en particulier dans les cas d'hétérogénéité génique (plus de 60 gènes responsables d'ataxie cérébelleuse, plus de 50 gènes dans le cas de la maladie de Charcot-Marie-Tooth....) ou d'hétérogénéité allélique (plus de 150 mutations dans le gène de la parkine, quelques centaines dans le gène CFTR....). Toutes ces techniques demandent cependant à être associées à d'énormes moyens de traitement et de stockage informatiques (un seul génome humain représente la capacité de stockage de cinq disques durs d'ordinateur portable...). Enfin ce séquençage haut débit peut générer une quantité d'informations génétiques non liées directement à l'objectif diagnostique initial (mutations hétérozygotes d'autres gènes, allèles de susceptibilité...) et susceptibles de poser des problèmes éthiques, essentiellement en révélant des variants alléliques sans aucune conséquence clinique, mais susceptibles de développer chez les intéressés une anxiété injustifiée.

Un deuxième champ d'application important des techniques de séquençage à haut débit est celui de la génomique tumorale dans le but non seulement d'identifier les différentes anomalies génétiques, mais aussi de faire la distinction entre mutations principales (« drivers ») et mutations secondaires (« pas-

sengers ») avec en arrière plan l'espoir de thérapies ciblées. Le séquençage du génome tumoral présente néanmoins un certain nombre de difficultés (taille et état de conservation de la pièce, ploïdie, hétérogénéité cellulaire clonale et sous-clonale...), en outre les mutations sont somatiques et variées (mutations ponctuelles, délétions, duplications, insertions/délétions, réarrangements...), tous ces éléments rendant indispensable la comparaison génome tumoral/génome constitutionnel du malade. D'un point de vue technique l'analyse de l'ADN tumoral nécessite un séquençage d'un nombre de fragments dont l'ensemble équivaut à au moins trente fois la taille du segment à explorer pour obtenir une précision suffisante : il peut s'agir d'un séquençage du génome entier et dans ce cas de préférence sur des fragments de grande taille, c'est-à-dire d'environ 3Kb, (« jumping libraries ») pour faciliter la détection des réarrangements, d'un séquençage de l'exome ou d'un séquençage après capture de régions particulières du génome. L'instabilité du génome tumoral est démontrée par l'augmentation des anomalies en fonction du stade évolutif et par l'apparition de nouvelles mutations dans les métastases ou dans les xénogreffes (chez la souris NOD/SCID). Les diverses anomalies génomiques observées sont collectées dans un schéma, le « CIRCOS PLOT », qui permet une comparaison aisée des anomalies d'une tumeur à l'autre ou entre tumeurs, métastases et xénogreffes. De telles représentations « CIRCOS PLOT » ont été récemment publiées dans le cas de tumeurs du sein [12], de glioblastomes [13], de cancers du poumon à petites cellules [14] et de mélanomes [15] ; en outre des programmes internationaux (« International Cancer Genome Consortium » ou ICGC, « COSMIC »...) actuellement en cours ont pour objet de mettre en œuvre le même type de recensement sur un grand nombre de tumeurs malignes. Dans le cadre de l'ICGC, qui a pour objectif l'analyse de 50 000 tumeurs, la France via l'INCA est chargée de l'étude des cancers primitifs du foie, du sein et de la prostate.

Les puces ADN ou « micro-arrays »

En pathologie, les facteurs génétiques en cause peuvent être classés schématiquement en trois groupes : les facteurs mendéliens, les variants rares à risque relatif élevé [10-12] et les polymorphismes fréquents mais à risque relatif faible [1-5]. L'analyse pan-génomique a pour objet d'identifier ces derniers facteurs génétiques responsables de prédispositions (ou de protection) vis-à-vis des maladies communes ; ces travaux sont effectués à grande échelle (plusieurs milliers de malades et de témoins) par des études GWAS (« Genome Wide Associations Studies ») qui recherchent des déséquilibres de liaison entre le trait pathologique et des SNPs localisés dans une même région du génome, tout en sachant qu'une association n'identifie pas obligatoirement un variant causal, mais très souvent un variant proximal [16]. Ces études utilisent des puces constituées de fragments d'ADN (oligonucléotides) correspondant à des séquences données, déposées sur un support solide selon une disposition

ordonnée (« array »), le fonctionnement de ces puces reposant sur une hybridation par des séquences complémentaires marquées par un fluorochrome. L'évolution technique récente des puces ADN a été considérable : de 2007 à 2010 elles sont passées d'une densité de 370 000 à 2 500 000 SNPs avec des distances les séparant en moyenne sur le génome passant de 4,9 à 0,63 kb. Depuis 2007 une série de SNPs (et donc de loci) associés à la maladie de Crohn [17], à la polyarthrite rhumatoïde [18], au diabète de type I, au diabète de type II et obésité [19], à l'hypertension artérielle [20] mais avec des effets très faibles (risques relatifs à peine supérieur à 1) ont été décrits. Un locus en 9p21 est par ailleurs plus significativement associé aux coronaropathies ischémiques (risque relatif 1,25) et aux anévrysmes abdominaux et intracrâniens (risque relatif 1,7) [21, 22] ; les SNPs correspondants sont localisés dans une région non codante de 58kb située à proximité des gènes suppresseurs de tumeur CDKN2A et 2B (« cyclin dependant kinase inhibitor 2A/2B ») fréquemment délétés dans les processus malins ; chez la souris la délétion de la région homologue diminue le niveau d'expression des gènes CDKN2A/2B et augmente la prolifération et la sénescence cellulaires [23].

Les loci identifiés aujourd'hui par les études GWAS correspondent donc à une augmentation de risque faible (en général inférieur à 1,3). Même associés ils expliquent généralement moins de 10 % de la variabilité d'un trait attribuée aux effets génétiques ; ainsi dans le cas de la surcharge pondérale, l'héritabilité du BMI (« Body Mass Index ») est considérée comme étant de 50 %, le génotypage de 125 000 individus a identifié 32 variants n'expliquant que 1,45 % de la variabilité génétique et les auteurs du travail considèrent par extrapolation que si l'étude de 730 000 sujets devrait conduire à l'identification d'environ 280 loci, ceux-ci n'expliqueraient toujours que 9 % de cette variabilité [24]. Il persiste ainsi une différence considérable entre héritabilité et résultats obtenus par les études GWAS : ce hiatus a été qualifié de « missing heredity ». Plusieurs hypothèses sont avancées pour expliquer cette « héritabilité manquante » : puces ADN ne testant que les SNPs relativement répandus c'est-à-dire dont l'allèle mineur a une fréquence supérieure à 5 % et donc ne prenant pas en cause des variants plus rares mais avec des effets forts, fréquences alléliques différentes d'une population à l'autre, non prise en cause dans ces études des effets non génétiques, non utilisation dans les analyses statistiques des méthodes de la génétique quantitative... Pour répondre à la première explication proposée, un projet (« 1 000 génomes ») se proposant d'identifier des polymorphismes plus rares est en cours (environ 5 millions de SNPs sont nécessaires pour englober 95 % de ces polymorphismes dont l'allèle mineur a une fréquence égale ou supérieure à 1 %) . Il est plus vraisemblable que la résolution de ce problème viendra de l'utilisation des nouvelles technologies de séquençage à haut débit.

Le concept de « genome disorder » c'est-à-dire de maladie génétique par perte ou gain de gènes, résultant d'une recombinaison non allélique entre deux segments chromosomiques présentant une forte homologie de séquence

(duplicon) a été proposé en 1998 ; il explique les anomalies en cause dans diverses affections : hyperplasies congénitales des surrénales, maladie de Charcot-Marie-Tooth type 1A, neurofibromatose type 1.... La caractérisation de ces réarrangements chromosomiques, entraînant une perte ou un gain de séquence, s'est effectuée par hybridation génomique comparative entre ADN témoin et ADN patient sur puce ADN (« CGH-array »). Les performances de ces puces dépendent de leur taux de recouvrement du génome ; elles se sont progressivement améliorées et dès 2003 des puces 1Mb constituées de 3523 clones BAC ou PAC (densité moyenne d'un clone par mégabase génomique) avaient une résolution supérieure à celle obtenue par un caryotype ; cet outil analytique appliqué chez des patients présentant retard mental et malformations congénitales a démontré la fréquence de ces anomalies (7 délétions et 5 duplications dans un groupe de 50 patients), la fréquence des anomalies de novo (non héritées) (6 délétions sur 7, 1 duplication sur 5 dans ce même groupe) ; il a également conduit à définir de nouveaux syndromes (dont celui correspondant à une microdélétion en 17q21-3) [25] ainsi qu'un nouveau polymorphisme d'inversion ou haplotype H1 caractéristique des populations européennes (26). En 2004 est apparue la notion de CNV (« Copy Number Variant »), c'est-à-dire de variations structurales à grande échelle du génome (un kilobase à plusieurs mégabases), correspondant à des duplications, des délétions, des associations de délétions et de duplications ou des remaniements multialléliques. L'importance quantitative de ces CNV a fait l'objet d'estimations différentes (jusqu'à 12 % du génome) ; l'utilisation de la dernière génération de puces (formées chacune de 2,1 millions d'oligonucléotides), susceptibles d'identifier des CNV d'une taille inférieure à 1 kb, conduit à une évaluation de l'ordre de 4,8 % du génome correspondant à environ 50 000 duplications/délétions. Il est rapidement apparu que chaque génome de sujet phénotypiquement normal présentait tout une série de ces duplications/délétions, qu'il existait des différences significatives suivant les ethnies, compatibles avec une origine africaine de l'humanité et montrant par exemple une adaptation du génome au type d'alimentation : le nombre de copies du gène de l'amylase salivaire est plus élevé dans les populations dont l'alimentation est riche en amidon. D'une façon générale, les fonctions biologiques potentiellement surexprimées par les duplications sont les fonctions de réponse à un stimulus (en particulier immunitaire), de perceptions sensorielles, d'adaptation à l'environnement, de kératinisation et d'adhésion cellulaire. Les CNV sont à l'origine de tableaux pathologiques à transmission dominante (exemple : duplication du locus APP et maladie d'Alzheimer héréditaire à début précoce), récessive (délétion du gène responsable de l'atrophie musculaire spinale....) ou complexe. Certains CNV sont associés à des variations de risque vis-à-vis de certaines maladies : délétion des gènes CFP1 et 3 et diminution du risque de dégénérescence maculaire liée à l'âge, duplication du gène CCL3L1 et augmentation de la susceptibilité au VIH, diminution du nombre de copies du gène β défensine et prédisposition à la maladie de Crohn....Enfin,

dans le cas de maladies multifactorielles (polyarthrite rhumatoïde, diabète de types I et II, sclérose en plaque...) la contribution des CNV est démontrée [27].

Nos 25 000 gènes génèrent environ 300 000 ARN messagers, chaque type cellulaire en exprimant environ 15 000 (transcriptome), mais avec des niveaux quantitatifs très différents (de 10 à 10 000 copies). Comme la protéomique, la « transcriptomique » est un moyen d'analyse globale extrêmement puissant en biologie et en pathologie, en particulier en oncologie, et ceci avec une approche technologique plus simple ; l'étude du transcriptome fait en effet appel à deux types de méthodes : classiquement aux méthodes d'hybridation sur puces, plus récemment au séquençage systématique des ADN complémentaires ou des ARN eux-mêmes. La méthode SAGE a ainsi pour principe d'identifier chacun des cDNA par une étiquette (« tag ») d'une vingtaine de nucléotides ; les étiquettes sont ligaturées pour former de longs concatémères qui sont ensuite clonés puis séquencés : la comparaison des séquences « tag » et des séquences d'ADN génomique permet de déterminer quels sont les gènes exprimés par la cellule ; en outre le nombre de copies de chaque étiquette est proportionnel à l'abondance initiale du messager et cette méthode est donc spécifique et quantitative. Par exemple en physiologie rénale le premier élément à prendre en compte est la grande complexité structurale et la grande diversité cellulaire de l'organe, obligeant pour établir les transcriptomes normaux à des microdissections suivies de tris cellulaires. L'étude transcriptomique des différents segments du néphron caractérise des marqueurs spécifiques de chacune de ces structures et une standardisation à partir de ces gènes de référence permet de corriger la grande variabilité des biopsies rénales. Au cours du syndrome néphrotique l'étude des transcriptomes a identifié un nouveau canal sodium (gène *ACCN1* et protéine *ASic 2*) associé à une protéine chaperonne (gène *HSPA8* et protéine *HSC70*). Par ailleurs la « clusterisation » hiérarchique des gènes exprimés fait apparaître que la souris (couramment utilisée en pathologie expérimentale rénale) n'est en fait sans doute pas un bon modèle de physiopathologie humaine [28].

Cancérologie

L'apparition des techniques de biologie moléculaire, notamment des analyses à haut débit a conduit à une nouvelle approche des affections malignes, classiquement effectuée par l'étude des ARN par des méthodes basées sur l'hybridation (puces à ADN) et plus récemment par séquençage génomique massif. Ces études, souvent menées à grande échelle par des programmes internationaux, ont pour objectifs de revoir et d'améliorer la classification histologique des tumeurs, d'identifier leurs gènes « critiques signataires », de définir des critères moléculaires de pronostic et si possible des choix thérapeutiques ciblés. Pour illustrer ce sujet nous avons choisi quatre affections malignes : les cancers colorectaux, les mélanomes, les gliomes et les cancers du sein.

Cancers colorectaux

Les cancers colorectaux sont caractérisés par une instabilité chromosomique en raison d'une perte d'hétérozygotie de gènes comme APC (gène de la polypose adénomateuse familiale), TP53 ou SMAD4, ou par une instabilité des microsatellites du fait de l'inactivation des gènes du système de réparation des misappariements ou MMR (« Mis-Match-Repair »), cette inactivation pouvant être héréditaire (HNPCC pour « Hereditary Nonpolyposis Colon Cancer » ou syndrome de Lynch par mutations de gènes MLH1, MSH2, MSH6) ou acquise par méthylation de leurs îlots CpG. La progression du cancer colorectal est un processus multi-étapes (dysplasie-adénome précoce-adénome intermédiaire-adénome tardif-carcinome), le passage de l'une à l'autre résultant de diverses mutations ou délétions géniques et l'initiation étant déterminée dans 90 % des cas par un dysfonctionnement de la voie Wnt/ β caténine [29]. L'étude des tumeurs par des puces d'expression (cDNA) conduit à les classer en deux groupes correspondant à la stabilité des microsatellites (MSS) ou à leur instabilité (MSI) et à définir deux sous-groupes de MSS (à faible ou forte agressivité) ainsi que deux sous-groupes de MSI (HNPCC et cas sporadiques). Les micro-arrays génomiques montrent pour l'essentiel des gains de matériel dans les MSI (en particulier en 16p13.3, 11q2.1-13.2), tandis que gains et pertes sont sensiblement équilibrés dans les tumeurs MSS. Ces remaniements génomiques entraînent l'apparition prématurée de codons stop responsables d'une inactivation et d'une dégradation prématurée des ARNm correspondants, mais surtout d'un dysfonctionnement d'environ un tiers des gènes en raison de variations dans le nombre de copies (duplications/délétions) (MSI : 702 gènes surexprimés et 362 sous-exprimés, MSS : 318 gènes surexprimés et 867 sous-exprimés). Par ailleurs les cancers MSI ne sont jamais métastatiques et donc de meilleur pronostic. Néanmoins ces résultats de biologie moléculaire n'ont pas encore conduit à des thérapeutiques différenciées [30].

Mélanomes

Les mélanomes cutanés sont des tumeurs malignes fréquentes (8 000 nouveaux cas par an en France) et parmi les plus graves en raison de leur potentiel métastatique élevé et de l'insuffisance des moyens thérapeutiques. La classification anatomo-clinique distingue quatre formes principales : mélanome à extension superficielle, mélanome nodulaire, mélanome lentigo ou de Dubreuilh, mélanome des extrémités ou acral-lentigineux ; à partir du naevus bénin l'évolution passe par les stades de naevus dysplasique puis de mélanome à croissance radiale, à croissance verticale et enfin de mélanome métastatique. 8 à 10 % des mélanomes sont dits familiaux et dus à des gènes de prédisposition majeure : CDKN2A et CDK4 (CDKN2A donne par épissage alternatif deux protéines : p16 qui inhibe le complexe CDK4/cycline D et donc le cycle cellulaire et p14 qui stabilise p53 en sequestrant MDM2 et contrôle l'apoptose). Environ 90 % des mélanomes sont donc dits sporadiques et

relèvent d'une étiologie multifactorielle : outre les effets environnementaux avec au premier plan les UV, interviennent également des facteurs de prédisposition génétique : variants de gènes de pigmentation comme MC1R, TYR, MATP, ASIP et TYRP1, de gènes des systèmes de réparation de l'ADN comme ERCC2, XPC, XPF et POLH, de gènes de l'immunité comme FAS et FASLG. La mélanogenèse fait intervenir une voie AMPc dépendante activée par la fixation de l' α MSH (« α mélanocyte stimulating hormone ») à son récepteur MCR1 (« melanocortin 1 receptor ») (dont divers variants comme D84E, R142H, R160W, R151C, D294H... augmentent le risque de mélanome) et relayée par le facteur de transcription MITF. Par ailleurs, divers variants génétiques favorisent également la progression tumorale, ils appartiennent globalement à la voie RAS/MAPK qui contrôle en définitive l'apoptose et le cycle cellulaire ; des mutations de BRAF, NRAS, cKIT, PTEN... augmentent ainsi le risque de mélanome métastatique, ils sont devenus de ce fait des cibles thérapeutiques potentielles. Classiquement, le diagnostic et l'évaluation thérapeutique reposent sur des critères histologiques : indice de Breslow, ulcérations, index mitotique, statut du ganglion sentinelle. En biologie moléculaire, les analyses transcriptomiques rétrospectives sont généralement impossibles, les puces d'expression nécessitant des ARN non dégradés, c'est-à-dire des mélanomes congelés, situation rare car les pièces sont habituellement fixées et paraffinées ; c'est ce qui explique que la plupart des résultats obtenus à ce jour l'ont été sur des lignées de cellules mélaniques : caractérisation d'une expression génique différentielle entre peau normale, naevus, mélanomes primitifs et mélanomes métastatiques ; identification d'un groupe de gènes, contrôlant la prolifération cellulaire et l'apoptose, impliqués dans le passage de la prolifération radiale à la prolifération verticale ; caractérisation d'un ensemble de 254 gènes (cycle cellulaire, réparation de l'ADN, ubiquitinylation..) discriminant pour le pronostic. Les mélanomes présentant une insensibilité importante à la chimiothérapie classique, les efforts pharmacologiques se portent sur la recherche de thérapies ciblées (anti-BRAF, anti-MEK, anti-KIT...) susceptibles d'inhiber la voie RAS/MAPK. Une mutation activatrice de BRAF (V600E) est retrouvée dans 50 % des mélanomes et le traitement spécifique des patients par un inhibiteur sélectif (PLX4032) conduit à une régression complète ou partielle dans la majorité des cas, avec une durée moyenne de réponse de 6,8 mois [31] ; des résistances secondaires à ce produit sont cependant susceptibles d'apparaître en raison d'une mutation de NRAS (Q61K) activant la voie MEK/ERK [32]. D'autres inhibiteurs de BRAF mais aussi de MEK et cKIT sont en cours de développement. Le mélanome représente ainsi un excellent exemple de l'apport de la biologie moléculaire et des technologies d'analyse à haut débit dans la définition de biomarqueurs prédictifs et le choix de thérapies ciblées.

Tumeurs gliales

Les tumeurs cérébrales sont classées en deux grands groupes : les tumeurs neuro-épithéliales (dont les tumeurs gliales, qui représentent 50 % de

l'ensemble tumoral) et non neuro-épithéliales. Les tumeurs gliales ou gliomes se différencient en tumeurs astrocytaires, oligodendrocytaires, oligoastrocytaires ou mixtes, chacune se présentant avec différents grades de malignité (en fonction de la densité cellulaire, des atypies cytonucléaires, des mitoses, d'une prolifération endothéliocapillaire, de nécrose) ; les astrocytomes de grade IV ou glioblastomes représentent environ la moitié de ces tumeurs gliales. Cette classification exclusivement histologique reste imparfaite et subjective comme le prouvent les nombreuses discordances inter et intra-observateurs, ainsi que le nombre important de mauvais résultats obtenus dans les indications pronostiques ou thérapeutiques qui en sont déduites. Ce constat explique l'intérêt potentiel suscité par les techniques d'analyse à haut débit de la biologie moléculaire. L'étude de l'ADN des gliomes par des puces génomiques (BAC ou PAC) caractérise des remaniements, pour l'essentiel des gains de matériel sur les chromosomes 1,7 et 19 ainsi que des délétions sur les chromosomes 11 et 12. Dans les astrocytomes de grade I (ou astrocytomes pilocytiques représentant 5 % des gliomes) une duplication de BRAF (7q34), résultant dans 80 % des cas d'un réarrangement chromosomique, est mis en évidence. Les tumeurs oligodendrogliales de grade III présentent dans deux tiers des cas une codélétion 1p36/19q13 avec translocation 1/19. Les glioblastomes (primitifs ou secondaires à une tumeur de plus faible malignité) ont un très mauvais pronostic (médiane de survie entre 12 et 24 mois) ; leur analyse génomique montre des duplications importantes au niveau des gènes de l'EGFR (chr.7), de MDM4 (chr.1) et de PDGFRA (chr.4) ainsi que des délétions en 9p et 10. Les amplifications d'EGFR et de PDGFRA ont pour conséquence un dysfonctionnement de la voie RTK/RAS/PI3K avec augmentation des capacités cellulaires de prolifération ; en outre des anomalies de la voie p53 (mutations/délétions/amplifications de CDKN2A, MDM2, MDM4, ou TP53) sont retrouvées dans 90 % des cas et dans 75 % des cas des anomalies semblables de la voie Rb (affectant les gènes CDK2A, 2B et 2C, CDK4 ou RB1) ; ces anomalies sont quantitativement croissantes avec le stade évolutif de la tumeur [33]. Au plan du « terrain génétique », un certain nombre de SNPs dans les gènes de réparation de l'ADN et dans les gènes contrôlant le cycle cellulaires sont des facteurs de risque pour les tumeurs gliales ; il faut également noter que des gliomes s'observent au cours de divers syndromes de prédisposition aux cancers (sclérose tubéreuse de Bourneville, neurofibromatose type 1, syndrome de Cowden, syndrome astrocytome-mélanome, syndrome de Li-Fraumeni, syndrome de Turcot B.....). Sur la base de ces anomalies génomiques une classification pronostique des tumeurs gliales en trois groupes (A, B et C) est proposée ; souvent en contradiction avec les résultats de la classification histologique, elle permet de détecter les tumeurs faussement rassurantes et de mieux guider le traitement. Par ailleurs le séquençage à haut débit de 22 gliomes a caractérisé dans 10 % des tumeurs une mutation de l'IDH1 (isocitrate deshydrogénase) responsable d'une réversion de la réaction enzymatique (transformation de l' α cétoglutarate en 2 hydroxyglutarate) et en

définitive entraînant un meilleur pronostic [34]. Enfin il a été également montré que la méthylation et donc l'inactivation du gène MGMT (O-méthyl-guanine-DNA-méthyl-transférase) était associée à une meilleure réponse au traitement par la témozolomide.

L'analyse moléculaire des gliomes conduit aujourd'hui à une classification pronostique plus précise ; il est espéré qu'elle mène également à l'établissement d'une carte d'identité génomique de chaque tumeur et en définitive à des traitements personnalisés, ciblés, plus efficaces et moins toxiques.

Cancers du sein

Le cancer du sein est l'affection maligne la plus fréquente chez la femme (46 000 décès en France en 2006) ; 10 à 15 % de ces cancers sont qualifiés d'héréditaires, en raison de mutations à pénétrance élevée des gènes BRCA1, BRCA2, PTEN, RAD51C, TP53, PALB2, BACH. Au plan histologique, les cancers du sein sont classés en deux grands groupes : les cancers *in situ* (canalaires ou lobulaires) et les cancers infiltrants dont 70-75 % de type canalaire et 10-15 % de type lobulaire. La prise en compte du grade de différenciation et de la présence ou non de récepteurs aux oestrogènes (ER) a constitué une première classification pronostique (score de Nottingham). Plus récemment une classification basée sur un profil d'expression génique distingue quatre grands groupes de carcinomes : — « basal-like » ou triple négatif correspondant à l'absence d'expression des gènes ER, PR (« progesterone receptor ») et HER2 (« epidermal growth factor receptor 2 » ou « ErbB2 »). — luminal A qui sont ER positifs et de bas grade histologique. — luminal B également ER+ mais de haut grade. — HER2+ montrant une expression élevée d'ErbB2 et des autres gènes présents dans le fragment d'ADN amplifié par PCR (amplicon) [35]. Les cancers luminal A s'avèrent a priori de bon pronostic, les luminal B et HER2 de mauvais et les triples négatifs de très mauvais pronostic. Afin d'améliorer les critères pronostiques et si possibles les indications thérapeutiques, différents tests d'expression génétique (sur les ARN extraits de la tumeur) centrés sur les gènes de prolifération et de différenciation cellulaires sont actuellement proposés, en particulier MammaPrint qui analyse 70 gènes sur puces ADN et Oncotype DX qui étudie 21 gènes par RT-PCR quantitative. Deux projets internationaux, TAILORx utilisant le kit Oncotype DX et MINDACT le kit MammaPrint, sont en cours, chacun sur plusieurs milliers de tumeurs, afin de préciser l'intérêt médical de ces tests commerciaux [36]. Outre la bonne pertinence du choix des gènes analysés, ces tests se heurtent également, comme dans bien d'autres cancers, à la complexité des tumeurs du sein : hétérogénéités cellulaires et moléculaires, grandes variétés des types histologiques, anomalies chromosomiques.... Dans cette étape l'anatomopathologiste doit jouer un rôle déterminant en réalisant la micro-dissection tissulaire, en précisant le diagnostic histologique, en vérifiant la qualité du prélèvement confié au biologiste moléculaire et en mettant en œuvre différentes approches

in situ, en particulier immunohistochimiques. La mise en œuvre de l'imagerie MALDI sur support solide permet par ailleurs de définir l'arrangement spatial des macromolécules et de générer une représentation 3D du profil protéique tumoral.

Conclusions et recommandations

La spectrométrie de masse est une technique dont les avantages, en particulier sa très grande sensibilité, ne sont sans doute pas suffisamment exploités en milieu hospitalier ; cet état de fait tient aux choix techniques précédemment effectués (dosages radio-immunologiques, dosages immuno-enzymatiques...), à l'importance des investissements matériels à faire, à la nécessité d'un support bioinformatique compétent, mais aussi à la faible culture en physique et biophysique des biologistes médicaux et au cloisonnement traditionnel entre les différentes disciplines biologiques et fondamentales. La technologie dont les développements semblent aujourd'hui les plus prometteurs est celle des séquenceurs, le séquençage d'un ADN génomique de patient devant très rapidement devenir un acte médical courant ; cette évolution nécessitera à nouveau de gros investissements en matériel informatique et le recrutement de bioinformaticiens. Outre le domaine des maladies génétiques, des facteurs génétiques en cause dans les maladies dites multifactorielles, les principales applications du séquençage ou des puces à ADN devraient concerner la cancérologie et la pharmacogénétique. En cancérologie ce seront la définition de signatures moléculaires diagnostiques, pronostiques, et surtout les indications de thérapeutiques ciblées ; quelques exemples de ces indications pharmacologiques sont aujourd'hui bien caractérisés : inhibiteur (imatinib) de l'activité de la protéine kinase BCR/ABL dans la leucémie myéloïde chronique, inhibiteur (PLX 4032) du gène BRAF muté (V600E) dans le mélanome, inhibiteur (gefitinib) du récepteur de l'EGF muté (L858R) dans le cancer bronchopulmonaire....En pharmacogénétique, il s'agira de la détection de mutations dans les gènes impliqués dans le métabolisme des médicaments et susceptibles d'entraîner des accidents thérapeutiques.

Dans la perspective de ces évolutions, l'Académie nationale de médecine recommande :

- 1) Que soit engagée une réflexion officielle sur les problèmes éthiques qui ne manqueront pas d'apparaître en raison des nombreuses informations génétiques résultant du séquençage du génome des patients, comme par exemple la valeur à accorder au caractère prédictif des polymorphismes génétiques ou à la découverte de mutations hétérozygotes récessives. En effet, la mauvaise utilisation de ces nouvelles données génétiques, en dehors du champ médical strict, pourrait connaître des dérives dangereuses, notamment dans le cadre du diagnostic prénatal à partir de l'ADN fœtal.

- 2) Que dans les centres hospitaliers, en particulier universitaires, des plateformes regroupent les moyens analytiques lourds (spectromètres de masse, appareil de résonance magnétique nucléaire, séquenceurs à hauts débits.), cette organisation dépassant le cadre traditionnel des différentes disciplines biologiques.
- 3) Que ces plateformes soient dotées des moyens informatiques indispensables et bénéficient du recrutement d'informaticiens dans le cadre du personnel technique et du personnel enseignant avec dans ce dernier cas la possibilité de recrutement de non médecins.
- 4) Que la détermination des signatures moléculaires en cancérologie devienne systématique et prise en compte dans les nomenclatures d'actes.
- 5) Que se développe la pharmacogénétique en routine hospitalière comme l'a déjà recommandé l'Académie [36] et que soit créée une carte individuelle faisant état des mutations potentiellement dangereuses lors de la prise de certains médicaments.

BIBLIOGRAPHIE

- [1] SOLARY E. — Protéomique clinique en oncologie. *Med. Sci.*, 2007, 23, 1-40.
- [2] SOLASSOL J., BOULLE N., MAUDELONDE T., MANGE A. — Protéomique clinique : vers la détection précoce des cancers ? *Med. Sci.*, 2005, 21,722-728.
- [3] BENNETT M.R., DEVARAJAN P. — Proteomic analysis of acute kidney injury : biomarkers to mechanisms. *Proteomic Clin. Appl.*, 2011, 5, 67-77.
- [4] WANG T.J., LARSON M.G., VASAN R.S. *et al.* — Metabolite profiles and the risk of developing diabetes. *Nat. Med.*, 2011, 17, 448-453.
- [5] BRUNELLE A., LAPREVOTE O. — Lipid imaging with cluster time-of-flight secondary ion mass spectrometry. *Anal. Bioanal. Chem.*, 2009, 393, 31-35.
- [6] LE NAOUR F., GUETTIER C., BRUNELLE A. *et al.* — Imagerie chimique de la stéatose hépatique. *Med Sci.*, 2009, 25, 987-990.
- [7] SHUSHAN B. — A review of clinical diagnostic applications of liquid chromatography-tandem mass spectrometry. *Mass Spectrometry Reviews*, 2010, 29, 930-944.
- [8] ARDAILLOU R., LE GALL J.Y. — Le dépistage néonatal généralisé par des tests d'analyse biologique. *Bull. Acad. Natl. Méd.*, 2006, 190 (8), 1745-59.
- [9] LIFTON R.P. — Individual genomes on the horizon. *N. Engl. J. Med.*, 2010, 362, 1235-36.
- [10] JORDAN B. — Le véritable « génome personnel » ? *Med. Sc.*, 2011, 27, 553-56.
- [11] BOUSLAM N., BENOMAR A., AZZEDINE H. *et al.* — Mapping of a new form of pure autosomal recessive spastic paraplegia. *Ann. Neurol.*, 2005, 57, 567-71.
- [12] DING L., ELLIS M.S., LI S. *et al.* — Genome remodelling in a basal-like breast cancer metastasis and xenograft *Nature*, 2010, 464, 995-1005.
- [13] MARDIS E.R., DING L., DOOLING D.E. *et al.* — Recurring mutations found by sequencing in acute myeloid leukemia genome. *N. Engl. J. Med.*, 2009, 361, 1058-1066.

- [14] PLEASANCE E.D., STEPHENS P.J., O'MEARA S. *et al.* — A small lung cancer genome with complex signatures of tobacco exposure. *Nature*, 2010, 463, 184-190.
- [15] PLEASANCE E.D., CHEETHAM R.K., STEPHENS P.J. *et al.* — A comprehensive catalogue of somatic mutations from a human cancer genome. *Nature*, 2010, 463, 191-196.
- [16] HARDY J., SINGLETON A. — Genomewide association studies and human disease. *N. Engl. J. Med.*, 2009, 360, 1759-1768.
- [17] BARRETT J.C., HANSOUL S., NICOLAE D.L. *et al.* — Genome wide association defines more than 30 distinct susceptibility loci for Crohn's disease. *Nat. Genet.*, 2008, 40, 955-62.
- [18] PLENGE R.M., SEIELSTAD M., PADYUKOV L. *et al.* — TRAF1-C5 as a risk locus for rheumatoid arthritis- a genomewide study. *N. Engl. J. Med.*, 2007, 357, 1199-1208.
- [19] MCCARTHY M.I. — Genomic, type 2 diabetes, and obesity. *N. Engl. J. Med.*, 2010, 363, 2339-2350.
- [20] NEWTON-CHEH C., JOHNSON T., GATEVA V. *et al.* — Genome-wide association study identifies eight loci associated with blood pressure. *Nat. Genet.*, 2009, 41, 666-676.
- [21] HELGADOTTIR A., THORLEIFSSON G., MAGNUSSON K. *et al.* — The same sequence variant on 9p21 associates with myocardial infarction, abdominal aortic aneurysm and intracranial aneurysm. *Nat. Genet.*, 2008, 40, 217-224.
- [22] MCPHERSON R. — Chromosome 9p21 and coronary artery disease. *N. Engl. J. Med.*, 2010, 362, 1736-1737.
- [23] VISEL A., ZHU Y., MAY V. *et al.* — Targeted deletion of the 9p21 non coding coronary artery disease risk interval in mice. *Nature*, 2010, 464, 409-412.
- [24] ALLEN H.L., ESTRADA K., LETTRE G. *et al.* — Hundreds of variants clustered in genomic loci and biological pathways affect human height. *Nature*, 2010, 467, 832-838.
- [25] SHAW-SMITH C., PITTMAN A.M., WILLATT L. *et al.* — Microdeletions encompassing MAPT at chromosome 17q21-3 is associated with developmental delay and learning disability. *Nat. Genet.*, 2006.
- [26] STEFANSON H., HEDGRSON A., THORLEIFSSON G. *et al.* — A common inversion under selection in Europeans. *Nat. Genet.*, 2005, 37.
- [27] CRADDOCK N., HURLES M.E., CARDIN N. *et al.* — Genome-wide association study of CNVs in 16 000 cases of eight common diseases and 3000 shared controls. *Nature*, 2010, 464, 713-719.
- [28] CHEVAL L., PIERRAT F., DOSSAT C. *et al.* — Atlas of gene expression in the mouse kidney : new features of glomerular parietal cells. *Physiol. Genomics*, 2011, 43, 161-73.
- [29] OLIVIER S., MIR A.M., MICHALSKY J.C., LEFEBVRE T. — Signalling and metabolomic predispositions linked to the colorectal cancer. *Med. Sci.*, 2011, 27, 514-520.
- [30] MARKOWITZ S.D., BERTAGNOLLI M.M. — Molecular origins of cancer: molecular basis of colorectal cancer. *N. Engl. J. Med.*, 2009, 361, 2449-2460.
- [31] FLAHERTY K.T., PUZANOV I., KIM K.B. *et al.* — Inhibition of mutated, activated BRAF in metastatic melanoma. *N. Engl. J. Med.*, 2010, 363, 809-819.
- [32] SOLIT D.B., ROSEN N. — Resistance to BRAF inhibition in melanomas. *N. Engl. J. Med.*, 2011, 364, 772-774.
- [33] WEN P.Y., KESARI S. — Malignants gliomas in adults. *N. Engl. J. Med.*, 2008, 359, 492-507.
YAN H., PARSONS D., GENGLIN J. *et al.* — IDH1 and IDH2 mutations in gliomas. *N. Engl. J. Med.*, 2009, 360, 765-773.
- [34] SMEITINK J. — Metabolism, gliomas and IDH1. *N. Engl. J. Med.*, 2010, 1144-1145.

- [35] SOTIRIOU C., PUSZTAI L. — Gene-expression signatures in breast cancer. *N. Engl. J. Med.*, 2009, 360, 790-800.
- [36] BOUREL M., ARDAILLOU R. — Pharmacogénétique et pharmacogénomique. *Bull. Acad. Natl. Med.*, 2006, 190 (1), 9-22.

Experts auditionnés :

Professeur F. Galibert, Académie Nationale de Médecine.
Professeur A. Brice, INSERM U 679, Groupe Pitié-Salpétrière, Paris
Professeur F. Soubrier, membre correspondant de l'Académie Nationale de Médecine
Professeur F. Cambien, INSERM U937, université Paris VI
Professeur F. Dardel, faculté de pharmacie, avenue de l'observatoire, Paris
Professeur F. Sigaux, Institut universitaire d'hématologie- hôpital Saint-Louis Paris
Monsieur R. Redon, INSERM-UMR 915, Institut du thorax, Nantes
Professeur O. Laprevote, EA CNRS 4463, faculté de pharmacie, avenue de l'observatoire
Professeur A. Edelmann, INSERM U845, hôpital Necker, Paris
Docteur A. Doucet, UMR 872, centre des cordeliers, Paris
Docteur S. Mourah, INSERM U940, hôpital Saint-Louis, Paris
Docteur A. Idbah, CRICM-INSERM-UMR975, Pitié-Salpétrière, Paris
Docteur A. Duval, INSERM-UMR938, hôpital Saint-Antoine, Paris
Docteur AV Salomon, INSERM U830, Institut Curie, Paris
Professeur V.Paradis, INSERM U773, hôpital Beaujon, Paris

*
* *

L'Académie saisie dans sa séance du mardi 17 janvier 2012 a adopté le texte de ce rapport à l'unanimité