



Génomique, promesses et réalités

**Bernard E. Bihain
Nicholas Shork
Lydie Bougueleret
Marta Blumenfed
Ilya Chumakov
Frances Yen
Daniel Cohen**

B.E. Bihain, N. Shork, F. Yen: Genset Corporation, 875 Prospect Street, Suite 206, La Jolla, California 92037, États-Unis. L. Bougueleret, M. Blumenfed, I. Chumakov, D. Cohen: Genset SA, 24, rue Royale, 75008 Paris, France.

► Les études génétiques de liaison ont permis la découverte de nombreux gènes de maladies monogéniques. Avec les études dites d'association, le défi est clair: identifier les gènes impliqués dans les maladies multigéniques, comprendre comment les produits de ces gènes interagissent entre eux et avec l'environnement pour créer des conditions pathologiques. Les retombées thérapeutiques sont sans limites, ne serait-ce que d'un point de vue prédictif. Grâce au programme Génome, nous assistons à la naissance d'une nouvelle discipline, la génomique fonctionnelle, et au renouveau de la physiologie qui devra établir des modèles physiopathologiques et prédictifs de ces maladies humaines. ◀

Au mieux dans 18 mois, au pire dans 3 ans, tous les biologistes disposeront d'un accès direct à la séquence des 3 000 à 4 000 mégabases (Mb) qui composent le génome humain. Pour répondre à l'attente de l'opinion, les séquences brutes sont rendues publiques en 4 phases et en vrac [1]. La phase I fournit des séquences préliminaires sans organisation. L'ordonnancement a lieu en phase II et se poursuit en phase III. En phase IV, la séquence d'un seul *contig* est assemblée et chaque base est validée par une triple lecture. En septembre 1999, le programme génome avait produit 313 Mb phase I; 34 Mb phase II; 30 Mb phase III; 420 Mb phase IV. L'objectif actuel est de fournir, durant le premier trimestre 2000, 1 000 Mb en phases II, III, IV et 1 800 Mb en phase I. Les données s'accumulent à une vitesse telle qu'il est probable que les systèmes d'analyse informatique utilisés dans la plupart des laboratoires se révéleront insuffisants. Cela justifie la mise en place de centres spécialisés de traitement de l'information. L'exploitation des résultats reste toutefois difficile en raison de hauts niveaux de contamination par des séquences bactériennes qui proviennent des vecteurs, de séquençages de faible qualité et d'assemblages imprécis. Toutes ces imperfections seront peu à peu rectifiées et il est raisonnable de prévoir

qu'en 2005, un chercheur disposant d'un élément de séquence protéique déterminera en quelques minutes, grâce à un ordinateur, la séquence du gène codant, et identifiera les différents gènes et pseudogènes homologues. Des logiciels reconstitueront *in silico* le cadre de lecture et déduiront la séquence des ARN messagers à partir d'assemblages d'EST (*expressed sequence tag*) qui, en parallèle du programme génome, s'accumulent dans les banques de données. L'importance du programme génome pour la recherche biologique est telle qu'il se doit de réussir.

Influence du programme Génome sur la recherche génétique

Au cours des trente dernières années, la génétique s'est attaquée avec succès à l'élucidation des causes des maladies dites monogéniques. La découverte des gènes responsables de la mucoviscidose [2-4], de l'hypercholestérolémie familiale (*m/s 1985*, n°7, p.388) [5], de la maladie de Huntington [6], de l'hémochromatose [7] fournissent quelques exemples d'importantes avancées médicales, même si toutes n'ont pas encore abouti à l'introduction de traitements curateurs. Ces découvertes sont, dans la plupart des cas, le fruit du clonage positionnel. Ces études de liaison reposent sur l'identification de marqueurs génomiques qui,

en dépit des réarrangements méiotiques, restent associés au trait phénotypique parce qu'ils sont localisés dans la même région du génome. La puissance de ces études est déterminée essentiellement par le nombre de méioses informatives qui surviennent entre les membres d'une même famille affectée par la maladie. Dans le meilleur des cas, les régions ainsi définies s'étendent sur quelques millions de paires de bases, elles peuvent donc contenir plusieurs centaines de gènes. Le gène responsable de la maladie et la (ou les) mutation(s) qui entraîne(nt) son(s) dysfonctionnement(s) sont ensuite identifiés par éliminations successives. Le processus est long et fastidieux mais il bénéficiera énormément du programme Génome.

Malgré des succès majeurs, les études de liaison manquent de puissance pour s'attaquer avec efficacité à l'identification des gènes impliqués dans des maladies dont la composante génétique est établie mais qui résultent sou-

vent du dysfonctionnement de plusieurs gènes. Sans minimiser l'importance des maladies monogéniques, il est essentiel de garder à l'esprit le fait que les maladies multigéniques ont des conséquences sociales et économiques très lourdes. Le diabète de type II, la schizophrénie, la psychose maniaco-dépressive, la maladie d'Alzheimer, l'hypertension, l'obésité et les maladies auto-immunes cardiovasculaires sont d'origine polygénique [8]. L'influence de chacun des gènes impliqués est limitée, il est donc difficile de distinguer les mécanismes moléculaires précis établissant le lien entre gène et trait pathologique [9]. Ainsi l'isoforme $\epsilon 4$ de l'apoE est statistiquement associée à la maladie d'Alzheimer mais aucun mécanisme moléculaire cohérent ne permet d'expliquer cette observation [10]. Par ailleurs, ces associations statistiques ne sont pas toujours confirmées. Ainsi, l'association entre le gène de l' α_2 -macroglobuline (α_2 MG) et la maladie d'Alzheimer, démontrée dans

une étude [11], n'a pas été retrouvée dans trois études indépendantes [12-14]. Cette controverse n'exclut en rien la possibilité d'un lien moléculaire causal entre α_2 MG et Alzheimer. En effet, le LRP (*LDL receptor-related protein*) – récepteur de l' α_2 MG – est associé à la maladie [9]. Une autre limite des études de liaison est atteinte lorsqu'elles identifient des régions chromosomiques contenant des complexes de gènes codant pour plusieurs protéines qui, toutes, participent au fonctionnement d'une même voie métabolique. La région du chromosome 19q13 offre un tel exemple. Cette région de 10 Mb contient les gènes codant pour l'apo E, l'apo C1, l'apo CII [15], la lipase hormono-sensible [16], et le *lipolysis stimulated receptor* (*m/s 1999, n° 5, p. 763*). D'autres gènes ne participant pas en première analyse au contrôle de la distribution des lipides entre les différents tissus de l'organisme sont également présents dans cette région. Il est par ailleurs intéressant de constater que tous ces gènes contiennent une ou plusieurs copies d'une répétition de 37 paires de bases qui n'est retrouvée dans aucune autre région du génome [17]. Ces liens de localisation chromosomique représentent-ils des observations accidentelles, et donc rares, ou le premier élément de l'ébauche d'une carte génomique fonctionnelle?

Une nouvelle méthodologie génétique est devenue indispensable. Elle repose sur les études dites d'association qui se distinguent des études de liaison par le fait qu'elles s'intéressent à des individus porteurs du trait pathologique mais non apparentés. Le nombre d'événements méiotiques qui séparent ces individus est considérablement plus élevé que le nombre de ceux qui séparent les membres d'une même famille. Le patrimoine génétique des membres d'une famille était identique 3 à 4 générations auparavant. Celui des sujets participant à une étude d'association et appartenant, le plus souvent, à un même groupe ethnique était identique il y a 100 000 ou 300 000 ans. Le nombre de recombinaisons méiotiques s'en trouve considérablement augmenté. Le facteur qui devient limitant dans ce contexte est la densité des marqueurs génomiques. Les *single nucleotide polymorphisms* (SNP) apportent une solu-

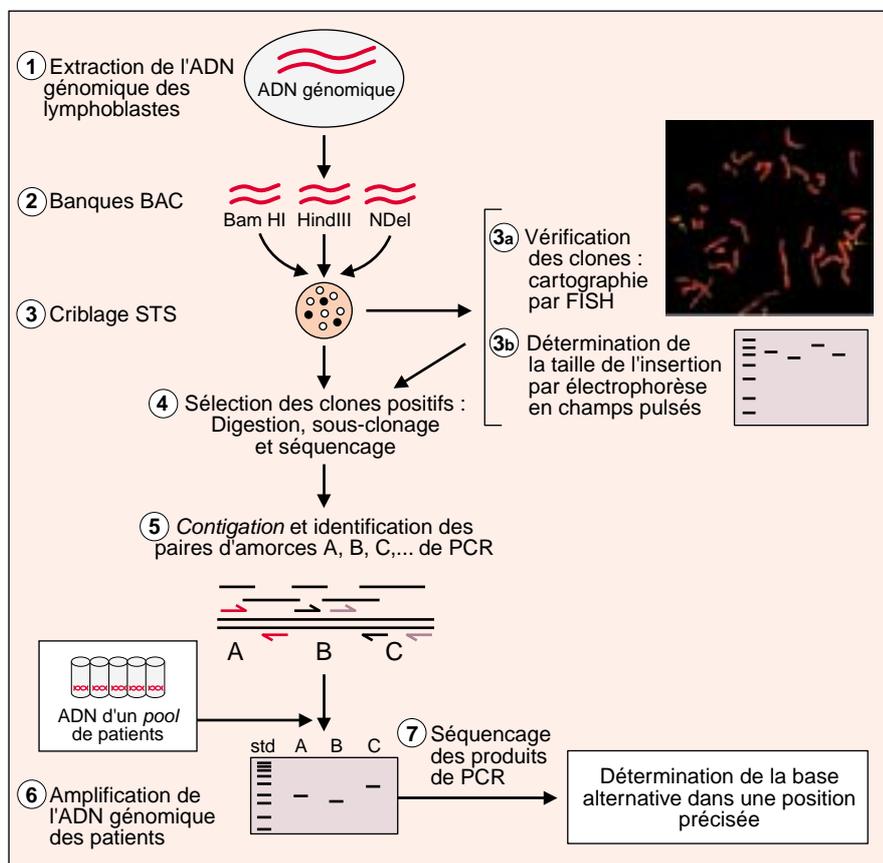


Figure 1. **Détection des SNP.** Les mutations d'une seule paire de base, identifiées à partir d'ADN génomique des lymphocytes circulants, sont très fréquentes et peuvent être utilisées comme marqueurs génétiques.

tion à ce problème (figure 1). Ces mutations d'une seule paire de bases – codantes ou non – sont extrêmement abondantes (10 millions par génome). Distribuées sur l'ensemble du génome, elles sont considérées comme les principales responsables des différences phénotypiques entre individus [8]. Trois cent mille SNP seront identifiés et localisés dans les 24 prochains mois [18]. A ce jour, certains laboratoires réalisent plus de 30 000 génotypages par jour et ce nombre est en augmentation constante (figure 2).

Les études d'association utilisant les SNP permettent de circonscrire le (ou les) gène(s) d'intérêt dans des régions génomiques de moins de 150 000 bases. Ces locus peuvent encore contenir plusieurs gènes candidats; toutefois, le processus d'élimination est simplifié d'un facteur 20 à 30. L'interprétation des résultats d'études d'association pose cependant encore d'importants problèmes d'analyse biostatistique: (1) validation des méthodes d'analyse haplotypique; (2) influence des fréquences alléliques sur les calculs d'haplotype; (3) validation des méthodes de mesure du déséquilibre de liaison entre marqueurs; (4) impact des tests multiples sur l'inférence statistique. La validation au niveau fonctionnel des gènes identifiés par étude d'association sera seule en mesure d'apporter une vali-

ation définitive de cette méthodologie. Dans l'intervalle, le paradoxe d'études génétiques s'intéressant à des sujets sans liens familiaux reste l'objet de controverses.

Les retombées potentielles des études d'association sont à la mesure de la complexité des problèmes biostatistiques qu'elles soulèvent. En effet, elles devraient permettre non seulement d'identifier les différents gènes impliqués dans les maladies polygéniques mais aussi de définir comment les produits de ces gènes interagissent entre eux pour déclencher, dans certaines conditions d'environnement, les processus physiopathologiques. Cette démarche est particulièrement importante lorsqu'elle s'adresse à des voies métaboliques complexes. A titre d'exemple, plus de 30 gènes identifiés contrôlent la répartition des lipides alimentaires entre les différents tissus de l'organisme [19]. Les études d'association rendent possible l'identification des 3 ou 4 gènes-clés qui sont susceptibles d'induire une anomalie de cette répartition et donc un excès de poids. Par ailleurs, il est vraisemblable que les études d'association permettront d'identifier différents groupes de gènes induisant un excès de poids mais dont les conséquences pathologiques divergent. Ainsi, certains individus obèses développent un diabète de type II [20]

tandis que d'autres présentent une augmentation du risque de cancer du côlon [21]. La caractérisation des différents polymorphismes génétiques responsables des maladies polygéniques devrait permettre une classification plus fine de ces entités pathologiques complexes et donner naissance à une véritable médecine prédictive.

L'application de la technologie SNP au domaine de la pharmacologie est également d'intérêt immédiat. L'efficacité et la toxicité d'un traitement sont déterminées pour une grande part par le polymorphisme des gènes cibles, des gènes contrôlant la détoxification et l'élimination des médicaments. La pharmacogénomique a pour objectif de sélectionner les individus bénéficiant au maximum d'un traitement et d'exclure ceux pour lesquels le risque thérapeutique excède les bénéfices attendus. Une retombée particulièrement importante des études de pharmacogénomique réside dans la découverte de nouvelles applications thérapeutiques pour des agents pharmacologiques connus. Ainsi, la mise en évidence de l'implication du gène *COX2* dans la pathogénie du cancer du côlon ouvre un champ d'application thérapeutique potentiel aux inhibiteurs de cette enzyme [22].

Génomique et renouveau de la physiologie

Le programme Génome a créé une nouvelle discipline: la génomique fonctionnelle ou physiologique. Cette discipline ambitionne d'attribuer une fonction, ou, à défaut, un rôle physiopathologique, aux nouveaux gènes identifiés par séquençage direct. Le défi est de taille. En effet, il n'existe à ce jour aucun modèle fonctionnel prédictif. Les études d'homologies de séquences primaires fournissent des informations, utiles dans moins de 30 % des cas. Des méthodes bio-informatiques, reposant sur la recherche d'homologies de structures tridimensionnelles, sont en cours de développement mais n'ont pas encore fait la preuve de leur efficacité en dehors du cadre de programmes pilotes. De plus, chaque niveau d'organisation de l'information présente un problème complexe. Ainsi, un même

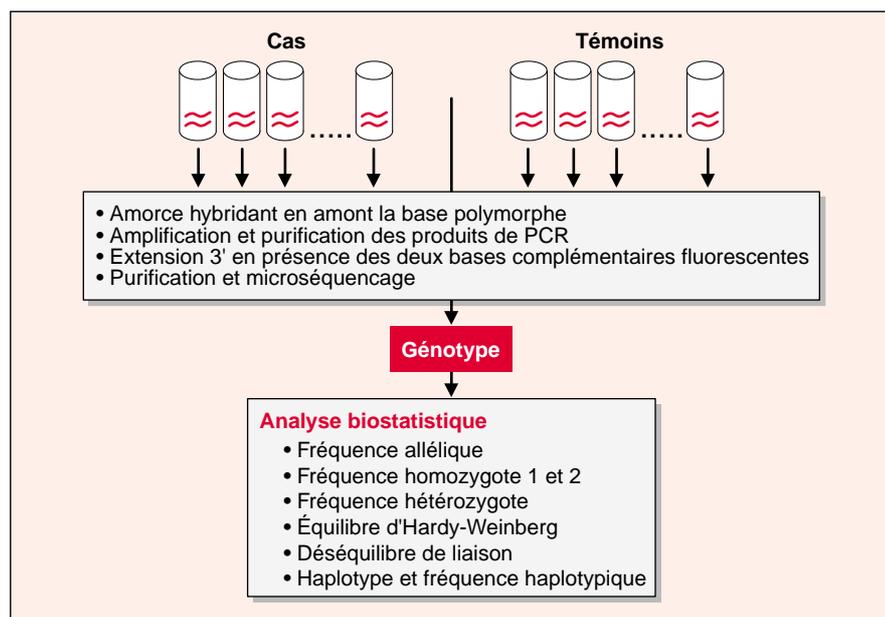


Figure 2. **Génotypage.** Le génotypage obtenu à partir des SNP (single nucleotide polymorphisms) permet de localiser des gènes dans de petites régions génomiques et d'effectuer des études d'association précises.

segment génomique produit dans certains cas des ARN différents. Le même gène produit différents ARN par épissage alternatif. Après traduction, une protéine subit le plus souvent plusieurs types de modifications. Enfin, une même protéine exerce parfois plusieurs fonctions distinctes [23].

La voie la plus directe pour l'établissement de la (ou des) fonction(s) d'un nouveau gène repose sur la transgénèse ou la production de souris *knock-out* par recombinaison homologue ou par expression de protéines en doigts de zinc qui inactivent sélectivement l'expression du gène d'intérêt [24]. Ces techniques sont longues même si l'utilisation de vecteurs viraux permet d'envisager la possibilité de gains de temps substantiels. Malgré ces avancées technologiques, de nombreux obstacles subsistent.

1. L'inactivation de gènes vitaux induit souvent une mortalité embryonnaire précoce [25]. La technologie *Cre-lox* permet dans certains cas de contourner l'obstacle (voir l'article de C. Babinet et M. Cohen-Tannoudji, p. 31 de ce numéro) [26].

2. Il reste difficile d'extrapoler à l'homme les données obtenues chez la souris. La découverte de la leptine et de son récepteur représente une avancée majeure pour la compréhension des mécanismes de l'obésité chez le rongeur [27]. Malheureusement, à de très rares exceptions près, ces gènes apparaissent comme parfaitement normaux chez les sujets humains obèses qui, dans une grande majorité des cas, développent un syndrome de résistance à la leptine dont les mécanismes restent inexplicables [28]. Ce problème est particulièrement aigu en ce qui concerne les maladies du système nerveux central. L'inactivation partielle du gène *NMDA a*, toutefois, récemment permis d'induire un syndrome ressemblant à la schizophrénie chez la souris [29]. Il est urgent de développer les méthodes d'analyse – biochimiques, physiologiques et comportementales – des modèles animaux accessibles à la transgénèse.

3. Les coûts d'un programme de transgénèse restent élevés : en moyenne 100 000 à 200 000 Euros par gène. Le budget d'un programme de transgénèse systématique portant sur les quelque 100 000 gènes du

génomique humain ne pourra être mobilisé que sur plusieurs dizaines d'années. Qui plus est, il ne permettra pas à lui seul d'identifier les gènes travaillant en synergie pour induire un processus pathologique.

Dans l'intervalle, il est utile de mettre au point les outils qui permettront de sélectionner rapidement les gènes qui méritent d'emblée une investigation approfondie. La technologie SNP permet d'évaluer rapidement l'effet du polymorphisme d'un gène dans le contexte d'une maladie humaine et de définir ceux qui méritent de façon urgente une analyse approfondie.

Une des principales conséquences du Programme génome est qu'il débarrasse les chercheurs de la nécessité de focaliser leur attention sur un seul gène. Un exemple classique est le cas du récepteur des LDL, dont tous éléments normaux et pathologiques ont été caractérisés en détail [5]. Ce travail représente 20 années d'activité d'un laboratoire extrêmement performant. Rendue incontournable par la complexité des technologies requises pour élucider ces mécanismes moléculaires, le côté réducteur de cette stratégie la rend néanmoins critiquable. En effet, une minorité d'individus hypercholestérolémiques présentent une mutation du récepteur des LDL. A l'inverse, certains sujets dont le récepteur des LDL est déficient ne présentent pas d'hypercholestérolémie franche [30-32]. Clairement, d'autres gènes, aujourd'hui moins connus, participent à la régulation des concentrations plasmatiques de LDL-cholestérol. Même si nous disposons de produits permettant de contrôler les concentrations plasmatiques de cholestérol, identifier ces gènes représente une démarche importante.

La possibilité d'intégrer dans une même analyse un grand nombre de gènes candidats rend essentielle la formulation d'hypothèses physiopathologiques. Cette démarche consiste à intégrer les données de la littérature dans le cadre d'une hypothèse de travail et de définir ainsi les gènes participant en synergie ou en opposition à la régulation d'une même voie métabolique. L'architecture du modèle permet de prévoir les conséquences potentielles d'anomalies de fonctionnement de chacun des éléments qui le composent. Les gènes-clés sont identifiés et de nouvelles stratégies

d'intervention thérapeutique sont planifiées. Après une période de succès, illustrés par l'élucidation des grandes voies métaboliques – glycolyse ou cycle de Krebs – le concept de modèle physiopathologique a perdu de son prestige. Relégués au rang de curiosité scientifique, ces travaux sont confinés dans certains journaux très spécialisés (*Perspectives in Biology and Medicine* a un facteur d'impact de 0,6). Cette perte d'influence résulte du fait que la plupart de ces hypothèses étaient jusqu'ici trop complexes pour être vérifiables par une approche moléculaire. La technologie SNP rend possible la vérification d'hypothèses incluant parfois plusieurs dizaines de gènes. Ces hypothèses représentent aujourd'hui les filtres indispensables au tri de la masse d'information produite par le programme génome. L'exploitation des données génomiques impose donc une participation plus active des physiologistes et des investigateurs cliniques qui sont en mesure de créer ces nouveaux outils de travail. Le programme Génome est rendu possible par les avancées technologiques des années 1990. Chercheurs participant à cet effort global, nous avons présenté ici notre vision des retombées de cette nouvelle technologie sur la pratique de la recherche en biologie et en médecine. Nous avons conscience d'avoir passé sous silence des aspects importants de cette nouvelle discipline [33], en particulier les nouvelles techniques d'analyse de l'expression différentielle des gènes. D'autres que nous sont probablement plus qualifiés pour en parler. Nous nous sommes focalisés sur les techniques que nous utilisons quotidiennement. L'impact de ces nouvelles technologies sur notre vie quotidienne sera important. Elles doivent faire l'objet d'un important débat démocratique, lucide et serein. Notre seule mission est d'alimenter ce débat en évaluant de façon critique les nouvelles pratiques rendues possibles par la révolution génomique ■

RÉFÉRENCES

- Collins FS. The human genome project and the future of medicine. *Ann NY Acad Sci* 1999; 882: 42-55.
- Cheng SH, Gregory RJ, Marshall J, et al. Defective intracellular transport and processing of CFTR is the molecular basis of most cystic fibrosis. *Cell* 1990; 63: 827-34.

RÉFÉRENCES

3. White MB, Amos J, Hsu JM, Gerrard B, Finn P, Dean M. A frame-shift mutation in the cystic fibrosis gene. *Nature* 1990; 344: 665-7.
4. Dean M, White MB, Amos J, *et al.* Multiple mutations in highly conserved residues are found in mildly affected cystic fibrosis patients. *Cell* 1990; 61: 863-70.
5. Goldstein JL, Hobbs HH, Brown MS. Familial hypercholesterolemia. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. *The molecular basis of inherited disease*. New York: McGraw-Hill, 1995: 1981-2030.
6. Goldberg YP, Rommens JM, Andrew SE, *et al.* Identification of an Alu retrotransposition event in close proximity to a strong candidate gene for Huntington's disease. *Nature* 1993; 362: 370-3.
7. Feder JN, Gnirke A, Thomas W, *et al.* A novel MHC class I-like gene is mutated in patients with hereditary haemochromatosis. *Nat Genet* 1996; 13: 399-408.
8. Collins FS. Shattuck lecture-medical and societal consequences of the human genome project. *N Engl J Med* 1999; 341: 28-37.
9. Saunders AM, Strittmatter WJ, Schmechel D, *et al.* Association of apolipoprotein E allele epsilon 4 with late-onset familial and sporadic Alzheimer's disease. *Neurology* 1993; 43: 1467-72.
10. Tanzi RE. A genetic dichotomy model for the inheritance of Alzheimer's disease and common age-related disorders. *J Clin Invest* 1999; 104: 1175-9.
11. Blacker D, Wilcox MA, Laird NM, *et al.* Alpha-2 macroglobulin is genetically associated with Alzheimer disease. *Nat Genet* 1998; 19: 357-60.
12. Rudrasingham V, Wavrant-De Vrieze F, Lambert JC, *et al.* Alpha-2 macroglobulin gene and Alzheimer disease. *Nat Genet* 1999; 22: 17-9.
13. Dow DJ, Lindsey N, Cairns NJ, *et al.* Alpha-2 macroglobulin polymorphism and Alzheimer disease risk in the UK. *Nat Genet* 1999; 22: 16-7.
14. Rogaeva EA, Premkumar S, Grubber J, *et al.* An alpha-2-macroglobulin insertion-deletion polymorphism in Alzheimer disease. *Nat Genet* 1999; 22: 19-22.
15. Shachter NS, Zhu Y, Walsh A, *et al.* Localization of a liver-specific enhancer in the apolipoprotein E/C-I/C-II gene locus. *J Lipid Res* 1993; 34: 1699-707.
16. Levitt RC, Liu Z, Nouri N, Meyers DA, Brandriff B, Mohrenweiser HM. Mapping of the gene for hormone sensitive lipase (LIPE) to chromosome 19q13.1→q13.2. *Cytogenet Cell Genet* 1995; 69: 211-4.
17. Das HK, Jackson CL, Miller DA, Leff T, Breslow JL. The human apolipoprotein C-II gene sequence contains a novel chromosome 19-specific minisatellite in its third intron. *J Biol Chem* 1987; 262: 4787-93.
18. Hacia JG, Fan JB, Ryder O, *et al.* Determination of ancestral alleles for human single-nucleotide polymorphisms using high-density oligonucleotide arrays. *Nat Genet* 1999; 22: 164-7.
19. Havel R, Kane JP. Structure and metabolism of plasma lipoproteins. In: Scriver CR, Beaudet AL, Sly WS, Valle D, eds. *The molecular basis of inherited disease*. New York: McGraw-Hill, 1995; 1841-51.
20. Weyer C, Bogardus C, Mott DM, Pratley RE. The natural history of insulin secretory dysfunction and insulin resistance in the pathogenesis of type 2 diabetes mellitus. *J Clin Invest* 1999; 104: 787-94.
21. Kim YI. Diet, lifestyle, and colorectal cancer: is hyperinsulinemia the missing link? *Nutr Rev* 1998; 56: 275-9.
22. Sheehan KM, Sheahan K, O'Donoghue DP, *et al.* The relationship between cyclooxygenase-2 expression and colorectal cancer. *JAMA* 1999; 282: 1254-7.
23. Krieger M, Herz J. Structures and functions of multiligand lipoprotein receptors: macrophage scavenger receptors and LDL receptor-related protein (LRP). *Annu Rev Biochem* 1994; 63: 601-37.
24. Liu Q, Segal DJ, Ghiara JB, Barbas CF III. Design of polydactyl zinc-finger proteins for unique addressing within complex genomes. *Proc Natl Acad Sci USA* 1997; 94: 5525-30.
25. Herz J, Clouthier DE, Hammer RE. LDL receptor-related protein internalizes and degrades uPA-PAI-1 complexes and is essential for embryo implantation. *Cell* 1992; 71: 411-21.
26. Rohlmann A, Gotthardt M, Hammer RE, Herz J. Inducible inactivation of hepatic LRP gene by cre-mediated recombination confirms role of LRP in clearance of chylomicron remnants. *J Clin Invest* 1998; 101: 689-95.
27. Friedman JM, Halaas JL. Leptin and the regulation of body weight in mammals. *Nature* 1998; 395: 763-70.
28. Caro JF, Sinha MK, Kolaczynski JW, Zhang PL, Considine RV. Leptin: the tale of an obesity gene. *Diabetes* 1996; 45: 1455-62.
29. Mohn AR, Gainetdinov RR, Caron MG, Koller BH. Mice with reduced NMDA receptor expression display behaviors related to schizophrenia. *Cell* 1999; 98: 427-36.
30. Hobbs HH, Leitersdorf E, Leffert CC, Cryer DR, Brown MS, Goldstein JL. Evidence for a dominant gene that suppresses hypercholesterolemia in a family with defective low density lipoprotein receptors. *J Clin Invest* 1989; 84: 656-64.
31. Varret M, Rabès JP, Boileau C. L'hypercholestérolémie familiale 25 ans après. I. Défauts du récepteur des LDL. *Med Sci* 1997; 13: 1399-408.
32. Rabès JP, Varret M, Boileau C. L'hypercholestérolémie familiale 25 ans après. II. Formes non liées au récepteur des LDL. *Med Sci* 1997; 13: 1409-18.
33. Shimkets RA, Lowe DG, Tai JT, *et al.* Gene expression analysis by transcript profiling coupled to a gene database query. *Nat Biotechnol* 1999; 17: 798-803.

TIRÉS À PART

D. Cohen.

MS2000

Summary

Genomics: promises and reality

Linkage studies have contributed to identify genes implicated in monogenic diseases. The polygenic traits of numerous diseases have been more difficult to elucidate. It is now potentially possible to identify genetic determinants and functionally important polymorphisms in genes that govern multigenic disease. Indeed, the human genome project, coupled with DNA array technology, high throughput screening systems and advanced bioinformatics will permit rapid elucidation of complex genetic components of human health and disease. New disease models should provide the physiological tools for a new type of medicine.