

## Ilots HTF : le gène annoncé

Des régions particulières de l'ADN riches en séquences CpG et appelées îlots HTF s'avèrent intéressantes à plus d'un titre. Assez fréquentes (il y en aurait environ 30 000 par génome haploïde), elles sont généralement situées dans la région 5' de gènes. Elles ont donc un grand intérêt pour aider à repérer des gènes dans l'ADN, d'autant plus que leur position est facilement décelable, puisqu'elles concentrent les sites de coupure d'enzymes très utilisées pour la cartographie à grande échelle de l'ADN. Généralement non méthylés (sauf dans des cas particuliers importants), les îlots HTF jouent sans doute un rôle dans l'activité des gènes qui leurs sont associés ; ils sont un des éléments d'un mécanisme qui englobe non seulement l'hérédité mendélienne stricte, fondée sur la séquence des nucléotides, mais aussi des modulations épigénétiques, plus subtiles mais non moins importantes.

**Bertrand R. Jordan**

**L**e pratiquant de la génétique inverse a souvent recours aux îlots HTF\* (*Hpa II tiny fragments*) : dans le désert des mégabases d'ADN qu'il vient de cartographier en gel pulsé ou de cloner dans des YACs (*yeast artificial chromosomes*), ces îlots indiquent la présence de gènes car ils leur sont presque toujours associés. Facilement repérables par le cartographe grâce aux nombreux sites de coupure d'enzyme à site rare (comme Not I, Eag I ou d'autres) qu'ils contiennent — à l'opposé du reste de l'ADN où ces sites sont, effectivement, très rares —, ces struc-

tures jouent donc un rôle de balise bien utile lorsqu'on pense que les gènes n'occupent que quelques pour cent de l'ADN, le reste étant constitué de séquences dont la fonction (s'il y en a une) et l'intérêt sont beaucoup moins évidents. Nous nous placerons ici essentiellement dans le cadre de l'étude du génome des mammifères ; notons que les îlots HTF n'existent que chez certains vertébrés ainsi que dans les plantes supérieures.

Qu'est-ce donc qu'un îlot HTF ? Il s'agit en fait de structures qui étaient visibles dans les séquences d'assez nombreux gènes clonés dès l'époque héroïque (1975-1980) avant qu'on n'en reconnaisse réellement l'existence. Dans ma propre équipe, où le premier (pseudo) gène d'histocompatibilité humain avait été cloné puis séquencé en 1980-1981, nous avons remarqué avec François Bregegere (alors à l'Institut Pasteur) une répartition curieuse des dinucléotides

**ADRESSE**

B. Jordan : directeur de recherche au Cnrs, responsable du groupe génétique moléculaire humaine. CIML, Inserm/Cnrs, case 906, 13288 Marseille Cedex 9, France.

\* On dit à peu près indifféremment « îlot HTF » ou « îlot CpG ». J'ai préféré garder la première expression, à laquelle les lecteurs sont sans doute plus habitués.

\*\* La notation CpG signifie 5'-cytosine-phosphate-guanine-3'.

## RÉFÉRENCES

1. Tyrocinsky ML, Max E. CG dinucleotide clusters in MHC genes and in 5'-demethylated genes. *Nucleic Acids Res* 1984 ; 12 : 4385-96.
2. Bird A, Taggart M, Frommer M, Miller OJ, McLeod D. A fraction of the mouse genome that is derived from islands of non-methylated, CpG-rich DNA. *Cell* 1985 ; 40 : 91-9.
3. Ehrlich M, Wang RYH. 5-methylcytosine in eukaryotic DNA. *Science* 1981 ; 212 : 1350-7.
4. Schwartz DC, Cantor CR. Separation of yeast chromosome-sized DNAs by pulsed field gel gradient electrophoresis. *Cell* 1984 ; 37 : 67-75.
5. Brown WRA, Bird AP. Long-range restriction site mapping of mammalian genomic DNA. *Nature* 1986 ; 322 : 477-81.
6. Lindsay S, Bird AP. Use of restriction enzymes to detect potential gene sequences in mammalian DNA. *Nature* 1987 ; 327 : 336-8.
7. Rommens JM, Iannuzzi MC, et al. Identification of the cystic fibrosis gene : chromosome walking and jumping. *Science* 1989 ; 245 : 1059-65.
8. Buiting K, Passarge E, Hosthemke B. Construction of a chromosome 15-specific linking library and identification of potential gene sequences. *Genomics* 1988 ; 3 : 143-9.
9. McDermid HE, Budard ML, Emanuel BS. Towards a long-range map of human chromosomal band 22q11. *Genomics* 1989 ; 5 : 1-8.
10. Antequera F, Boyes J, Bird A. High levels of *de novo* methylation and altered chromatin structure at CpG islands in cell lines. *Cell* 1990 ; 62 : 503-14.
11. Siminovitch L. On the nature of heritable variation in cultured somatic cells. *Cell* 1976 ; 7 : 1-11.
12. Holliday R. The inheritance of epigenetic defects. *Science* 1987 ; 238 : 163-70.
13. Meehan RR, Lewis JD, et al. Identification of a mammalian protein that binds specifically to DNA containing methylated CpGs. *Cell* 1989 ; 58 : 499-507.
14. Antequera F, McLeod D, Bird AP. Specific protection of methylated CpGs in mammalian nuclei. *Cell* 1989 ; 58 : 509-17.

CpG\*\* (figure 1) : assez fréquente dans la région 5' du gène, cette séquence était pratiquement absente de la région 3'. La région 5' de ce gène HLA a une composition plus riche en C et en G que la région 3' ; mais cette différence est très insuffisante pour rendre compte de la différence de fréquence de la séquence CpG. Nous avons été, bien sûr, intrigués par ces observations — jamais publiées — et avons évoqué plusieurs hypothèses à son sujet ; mais le mérite d'avoir soulevé le lièvre revient à Tyrocinski et Max [1] qui publièrent en 1984 une étude portant sur 14 gènes de séquence connue, principalement des gènes du complexe majeur d'histocompatibilité de classes I et II. Il en ressortait l'existence fréquente de régions riches en séquences CpG dans les régions 5' de ces gènes [1]. Tyrocinski et Max discutaient ces résultats en les reliant (à juste titre) à la méthylation de l'ADN et au phénomène de *CpG suppression* (voir plus bas) qui en découle sans doute ; en revanche, leurs spéculations sur un éventuel rôle de la richesse en CpG dans l'apparition ou le maintien du polymorphisme dans la région 5' des gènes du MHC n'ont pas été confirmées par la suite.

Le terme d'îlot HTF fut en fait employé pour la première fois par Adian Bird et ses collègues [2]. HTF signifie en fait *Hpa II tiny fragments* et se réfère au fait que cette enzyme (site de reconnaissance CCGG, ne coupant pas si le C est méthylé) produit — après digestion d'ADN de souris — deux classes de fragments : une classe très majoritaire de grands fragments (plus de 2 kb) et une classe minoritaire, environ 1 % du total, de tout petits (*tiny*) fragments de 50 à 500 nucléotides correspondant à des zones où les CpG non méthylés sont fréquents. L'analyse d'un certain nombre de ces fragments permettait aux auteurs de conclure à l'existence dans le génome de la souris d'envi-

pression (voir plus bas) qui en découle sans doute ; en revanche, leurs spéculations sur un éventuel rôle de la richesse en CpG dans l'apparition ou le maintien du polymorphisme dans la région 5' des gènes du MHC n'ont pas été confirmées par la suite.

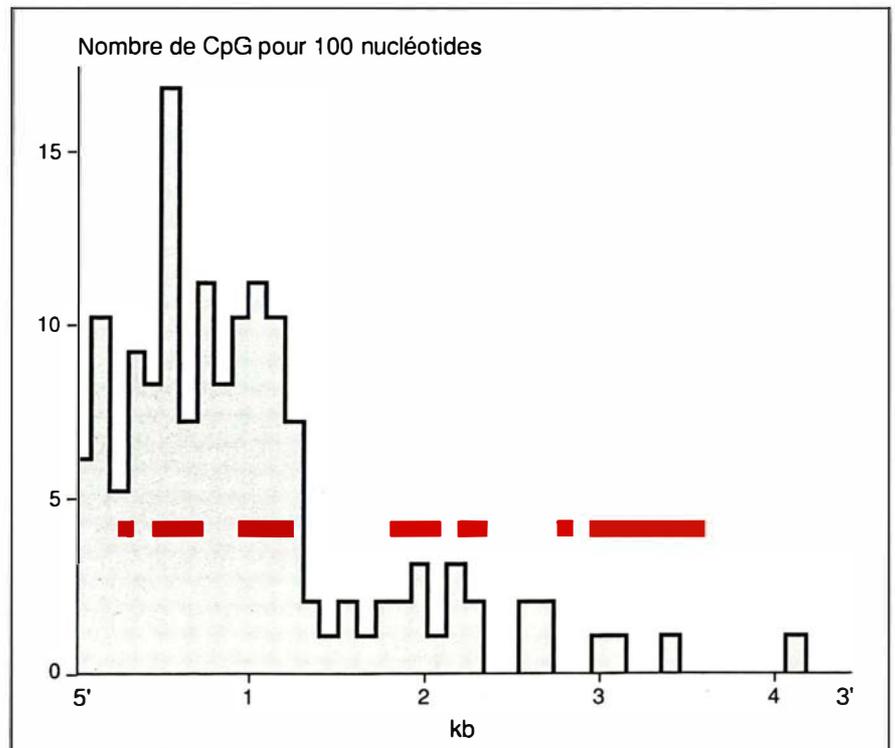


Figure 1. **Fréquence de dinucléotide CpG dans la séquence du gène HLA p12.1.4** [26], déterminée par bloc de cent nucléotides le long du gène dont les exons sont indiqués par les traits rouges.

ron 30 000 îlots HTF, longs de 1 à 2 kb, riches en séquences CpG non méthylées (donc en sites Hpa II) et constitués majoritairement de séquences copie unique (figure 2). Des données préliminaires (recherche de transcrits) leur permettaient de suggérer que la majorité des îlots HTF étaient associés à des gènes, et donc d'entrevoir qu'une bonne partie des gènes (connus ou à découvrir...) contient un îlot HTF puisque l'on estime le nombre total de gènes chez la souris (ou chez l'homme) à 50 000 environ contre 30 000 îlots HTF. Cette histoire de CpG devenait donc très intéressante, et cela à plusieurs titres : tant pour le repérage de gènes que, peut-être, pour la compréhension de leur régulation.

### Méthylation, CpG suppression et expression

Mais à ce point de notre récit, et avant de livrer au lecteur haletant d'impatience la suite de la saga des HTF nous allons procéder — selon les meilleures règles romanesques — à un *flash back* pour resituer quelques notions nécessaires à la bonne compréhension de la suite. Il s'agit essentiellement de la méthylation de l'ADN chez les eucaryotes supérieurs, du phénomène de CpG suppression et du rapport entre méthylation et expression. On sait en effet depuis longtemps que l'ADN est méthylé de façon différente chez les procaryotes et les eucaryotes ; chez les eucaryotes supérieurs (et en particulier les mammifères), la méthylation porte presque exclusivement [3] sur les cytosines qui précèdent une guanine : ce sont donc les C des séquences CpG qui sont méthylés. Cette modification est fréquente puisque la majorité des CpG la portent ; elle n'interfère pas avec l'appariement de l'ADN (en fait, les paires  ${}^m\text{C}=\text{G}$  sont plus stables que les  $\text{C}=\text{G}$ ) mais elle a pour conséquence la formation de thymine en cas de désamination de la méthylcytosine, thymine qui n'est pas toujours éliminée par les systèmes de réparation de l'ADN : les séquences  ${}^m\text{CpG}$  ont donc tendance à se transformer en TpG, une tendance lente mais sensible au niveau de l'évolution. Le

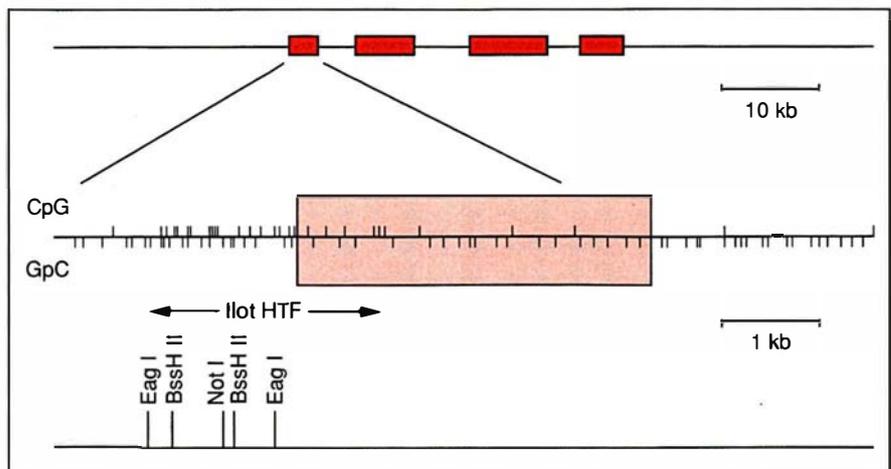


Figure 2. **Portrait robot d'un îlot HTF.** En haut : l'ADN génomique dans la région d'un gène (les rectangles figurent les exons). En bas, agrandissement de la région du premier exon : position des dinucléotides CpG (trait vers le haut) et GpC (trait vers le bas) le long de la séquence ; sites de coupure pour quelques enzymes à site rare.

résultat, on le constate en examinant les séquences de segments d'ADN provenant d'eucaryotes supérieurs : la séquence CpG est rarement rencontrée, cinq ou dix fois moins souvent qu'on ne s'y attendrait d'après la composition en nucléotides de la région considérée. C'est ce phénomène que l'on désigne par *CpG suppression*. Il y a suppression des CpG partout sauf, justement, dans les îlots HTF où la séquence CpG est trouvée à la fréquence attendue d'après la composition de la région. Une autre façon simple et parlante de caractériser une région d'ADN est de comptabiliser le rapport des fréquences des séquences CpG et GpC : ce rapport est voisin de 1 dans les îlots HTF, mais très inférieur (0,1 à 0,2) partout ailleurs, là où règne justement la suppression des CpG (figure 2).

Et l'expression des gènes dans tout cela ? Pour résumer de façon ultralapidement une masse considérable de données, et une masse non moins considérable de spéculations, nous dirons à ce stade que, en général, la méthylation est associée à la répression de l'expression d'un gène. Formulation prudente, car il y a des exceptions ; car la méthylation de certaines bases est sans doute critique

alors que celle d'autres séquences est sans importance ; car enfin le lien de causalité entre méthylation et répression, ou déméthylation et expression n'est pas établi. La tendance générale énoncée ci-dessus est néanmoins claire.

### Les cartographes butent sur les HTF

Revenons donc à nos îlots HTF, et à la façon dont ils apparaissent dans les préoccupations des cartographes de l'ADN. Les premières publications de Schwartz et Cantor [4] sur l'électrophorèse en champ pulsé montrèrent, dès 1984, la possibilité de séparer de très grands fragments d'ADN ; l'intérêt de cette technique pour l'étude du génome humain était évident. Encore fallait-il être capable, après avoir préparé de l'ADN de très grande taille, de le couper spécifiquement en fragments de quelques centaines de kilobases : chose que ne font pas les enzymes habituelles comme Eco RI ou Hind III qui donnent des morceaux de quelques kilobases. On se tourna donc, au fur et à mesure de leur isolement (et de leur commercialisation) vers les enzymes dites à site rare, dont les sites sont peu fréquents sur l'ADN, pour

## RÉFÉRENCES

15. Tazi J, Bird AM. Alternative chromatin structure at CpG islands. *Cell* 1990 ; 60 : 909-20.
16. Holliday R. A different kind of inheritance. *Scientific American* 1989 ; 6 : 40-8.
17. Lorenz K. King Solomon's Ring. New York : Crowell, 1972.
18. Babinet C, Barra J, Renard JP. Le marquage et l'expression différentiels des génomes paternel et maternel : deux conditions nécessaires pour le développement à terme de l'embryon de souris. *médecine/sciences* 1989 ; 5 : 8-15.
19. Swain JL, Stewart TA, Leder P. Parental legacy determines methylation and expression of an autosomal transgene : a molecular mechanism for parental imprinting. *Cell* 1987 ; 50 : 719-27.
20. Schroeder WT, Chao LY, et al. Non-random loss of maternal chromosome 11 alleles in Wilms tumour. *Am J Hum Genet* 1987 ; 40 : 413-20.
21. Junien C, Henry I. Bras court du chromosome 11 : empreinte parentale différentielle, tumorigénèse et pertes d'allèles. *médecine/sciences* 1989 ; 5 : 480-8.
22. Reik W. Genomic imprinting : a possible mechanism for the parental origin effect in Huntington's chorea. *J Med Genet* 1988 ; 25 : 805-8.
23. Hall JG. Genomic imprinting : review and relevance to human diseases. *Am J Hum Genet* 1990 ; 46 : 857-73.
24. Laird CD. Proposed mechanism of inheritance and expression of the fragile-X syndrome of mental retardation. *Genetics* 1987 ; 117 : 587-99.
25. Jordan B. Retard lié à la fragilité du chromosome X : où en est-on en 1989 ? *médecine/sciences* 1989 ; 5 : 450-8.
26. Malissen M, Malissen B, Jordan BR. Exon-intron organization and complete nucleotide sequence of an HLA gene. *Proc Natl Acad Sci USA* 1982 ; 79 : 893-7.
27. Nguyen C, Poustka AM, Djabali M, Roux D, Mattei JF, Jordan BR. Large scale mapping and chromosome jumping in the q27 region of the human X chromosome. *Genomics* 1989 ; 5 : 298-303.
28. Chimini G, Boretto J, Marguet D, Lanau F, Lauquin G, Pontarotti P. Molecular analysis of the human MHC class I region using yeast artificial chromosome clones. *Immunogenetics* 1990 (sous presse).

des raisons statistiques dans le cas de Not I — dont le site couvre huit nucléotides au lieu de six habituels —, ou parce que bien qu'hexanucléotidiques ils comportent une voire deux séquences CpG : ces séquences étant rares dans l'ADN, les sites correspondants le sont aussi. A partir de 1985-1986, de nombreux laboratoires commencèrent donc à utiliser cette méthode pour établir des cartes de restriction sur l'ADN génomique humain (ou murin). Expériences qui, dans leur principe, suivaient un schéma classique : coupure de l'ADN, séparation des fragments sur gel, transfert et hybridation avec des sondes de la région étudiée. Simple-ment, là où les méthodes classiques établissent la carte sur une ou plusieurs dizaines de kilobases, l'utilisation conjointe d'enzymes à site rare et de séparations en champ pulsé devait permettre de cartographier des centaines ou des milliers de kilobases. Dans ce type d'étude, on procède souvent par simples et doubles digestions : si l'action d'une enzyme donne un fragment A (révélé par une sonde donnée), l'action d'une

enzyme de spécificité différente un fragment B (révélé par la même sonde), l'action conjointe des deux enzymes produit un nouveau fragment C plus petit que A et B et dont la taille donne une information sur la position respective des sites de coupure (figure 3). Nous nous livrions, comme beaucoup, à de telles études dès 1986 dans mon équipe et constatations, à notre grande surprise, qu'en général les fragments A et B avaient la même taille, et que la double digestion n'avait pas d'effet apparent. Bien entendu dans ce cas, l'on incrimine d'abord l'enzyme, le tampon d'incubation ou l'expérimentateur ; mais ces résultats devinrent brusquement compréhensibles à la lecture d'un article d'Adrian Bird (encore lui, on lui doit d'ailleurs la plus grande partie de ce que nous savons sur les îlots HTF), que nous ne fûmes certainement pas les seuls à lire avec attention [5]. Cet article faisait la liaison entre sites de coupure des enzymes à site rare et îlots HTF, liaison évidente (après coup...) puisque richesse en CpG et coupure par des enzymes comme Not I (site

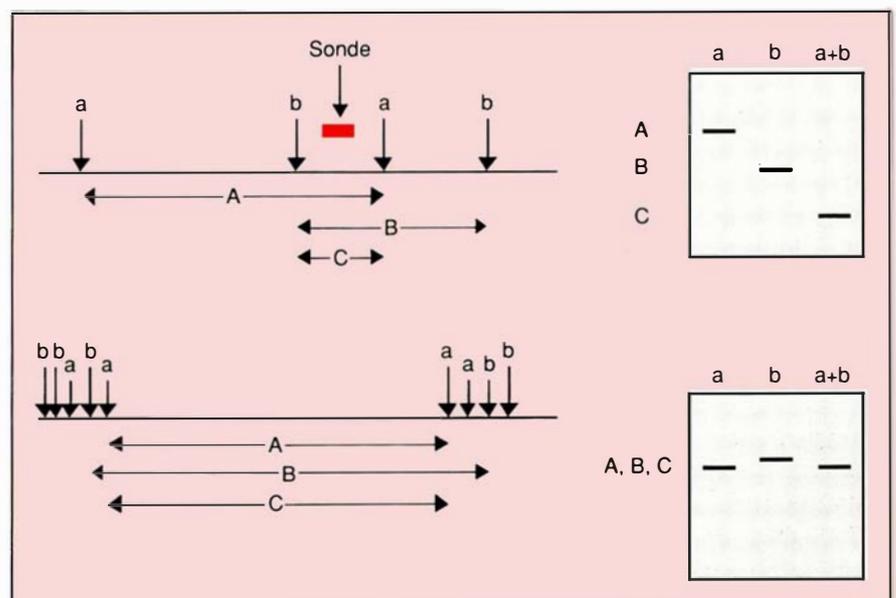


Figure 3. **Cartographie par simple et double coupure (enzymes a et b).** En haut : le cas habituel et les fragments observés par électrophorèse. En bas : la situation en cartographie à grande échelle avec des sites concentrés dans les îlots HTF. Les fragments A (enzyme a), B (enzyme b) et C (a + b) ont des tailles très voisines et donnent peu d'informations pour établir une carte.

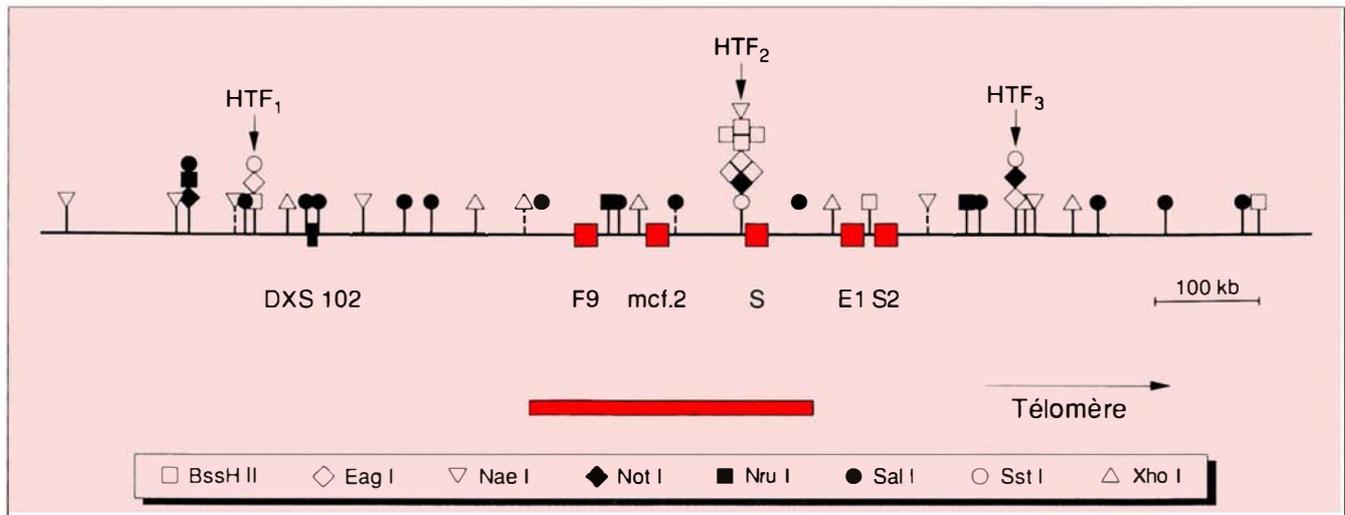


Figure 4. **Un exemple de carte physique** d'une région de 1 500 kilobases autour du gène du facteur de coagulation IX [27]. On remarque l'existence de groupes de sites de coupure qui signalent des îlots HTF (notés HTF1, 2 et 3 sur la carte). Les noms indiqués sous la carte correspondent aux différentes sondes qui ont permis de l'établir : DXS 102, une sonde « anonyme » (c'est-à-dire ne correspondant pas à un gène connu), F9, un segment du gène du facteur de coagulation IX ; mcf.2, une séquence transformante clonée dans cette région ; S1, E1 et S2 trois sondes obtenues par saut le long du chromosome. La barre rouge sous la partie centrale de la carte indique une zone d'environ 300 kb entièrement clonée par marche le long du chromosome. Les différentes références relatives à cette figure se trouvent dans l'article cité [24].

GCGGCCGC), Eag I (CGGCCG) ou Bss H II (GCGCGC) doivent naturellement aller de pair, d'autant plus que ces enzymes ne coupent pas l'ADN lorsqu'il est méthylé... Des calculs fondés sur des hypothèses raisonnables montrent que plus de 70 % des sites existant pour la plupart de ces enzymes sont situés dans les îlots HTF, ce qui explique le résultat obtenu (figures 3 et 4). La cartographie à grande échelle est donc largement gouvernée par l'existence des îlots HTF, dont le franchissement est souvent difficile du fait de l'accumulation des sites en un point de la carte.

### Les îlots HTF balisent les gènes

Nuisibles pour les cartographes donc, dans la mesure où ils gênent l'établissement de cartes physiques sur de longues distances, les îlots HTF sont, en revanche, très utiles dans la recherche de gènes auxquels ils sont assez systématiquement liés. Or, comme nous le disions au début, ce repérage des gènes est une des pha-

ses critiques de la génétique inverse puisque les séquences géniques sont rares et dispersées dans un génome principalement constitué de séquences sans fonction connue et — peut-être — sans intérêt. La recherche des îlots HTF devint donc rapidement une tactique en vogue chez les généticiens moléculaires, selon l'argumentation convaincante développée par Adrian Bird [6]. Elle constitue maintenant une des étapes indispensables dans la « chasse aux gènes » qui devient d'actualité chaque fois qu'une région de quelques centaines de kilobases qui doit contenir le gène responsable d'une maladie donnée est étudiée en détail en champ pulsé ou (mieux) après clonage dans des YAC ou des cosmides. La cartographie repère les îlots (figures 4 et 5, p. 158) ; les séquences avoisinantes peuvent alors être testées par *zoo blot* (hybridation sur l'ADN d'une série d'animaux plus ou moins apparentés à l'homme, les séquences conservées ayant de grandes chances de correspondre à des gènes) et, plus directement, par hybridation sur des banques d'ADNc ou sur de l'ARN mes-

sager (*Northern*). Approche maintenant classique, et qui a de nombreux succès à son actif : elle a, par exemple, joué un rôle important dans l'identification du gène impliqué dans la mucoviscidose [7]. La corrélation gène/îlot HTF a aussi motivé des travaux visant à isoler globalement (différentes stratégies sont possibles, faisant usage de l'existence de sites rares dans ces régions) une série d'îlots HTF provenant d'un chromosome ou d'une région chromosomique donnée : la caractérisation de tels clones a montré qu'en effet ils comportent en général des séquences transcrites [8, 9]. Les clones ainsi obtenus peuvent être répartis au hasard le long de la région chromosomique considérée ; ils n'en sont pas moins particulièrement intéressants puisqu'ils correspondent en général à des séquences exprimées.

La nature des gènes associés aux îlots HTF a été définie par approximations successives. Dans les débuts l'on a pensé, et même écrit, que l'îlot HTF signalait obligatoirement la présence d'un gène à expression ubiquitaire (dit aussi *housekeeping gene*

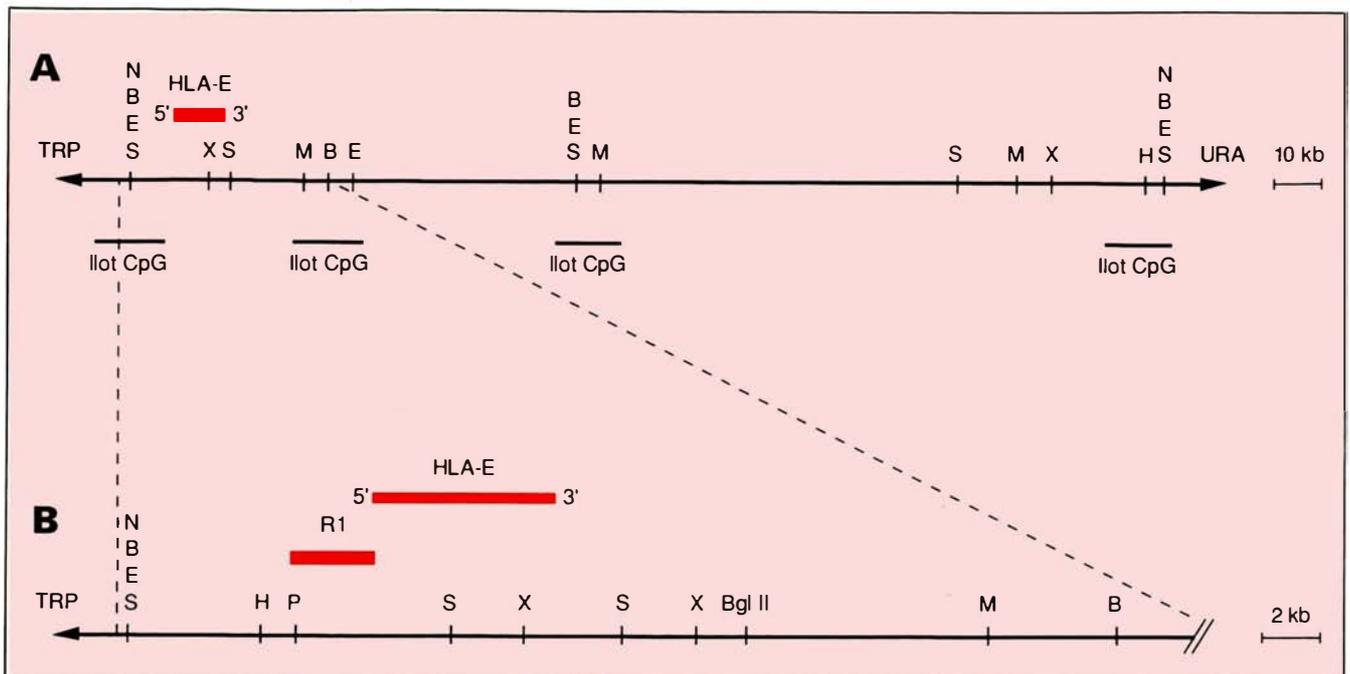


Figure 5. **Recherche de nouveaux gènes via les HTF.** Cartographie et recherche de gènes dans un clone YAC. Le segment d'ADN humain cloné dans la levure et isolé d'une librairie par hybridation avec une sonde HLA a été cartographié à l'aide d'enzymes à site rare (en haut) qui révèlent quatre îlots CpG ou îlot HTF ; la carte détaillée dans la région du premier îlot montre qu'en plus du gène HLA attendu (qui a été caractérisé comme HLA-E), une séquence transcrite appelée R1 est mise en évidence. TRP et URA désignent les deux « bras » du vecteur utilisé pour construire la banque YAC. (Données obtenues par le groupe de P. Pontarotti au centre d'immunologie de Marseille-Luminy [28]).

ou gène de ménage), et que les gènes dont l'expression est spécifique à certains tissus en étaient dépourvus. Ces conclusions ont dû ensuite être nuancées, et un résumé de nos connaissances actuelles (en écartant quelques exceptions...) est présenté dans la figure 6 qui s'applique à la situation *in vivo* chez l'animal (ou l'homme), pas forcément sur les cellules en culture, comme on le verra plus loin. Les gènes de ménage donc, d'expression ubiquitaire, contiennent un îlot HTF en 5' qui recouvre tout ou partie des premiers exons, îlot HTF dans lequel les CpG ne sont jamais méthylés sauf sur celui des deux chromosomes X qui est inactivé chez la femme. Les gènes à expression spécifique, eux, peuvent comporter un HTF qui, lui non plus, n'est jamais méthylé (sauf sur l'X inactif) : il y a donc, dans ce cas, dissociation entre méthylation et arrêt de l'expression puisque dans les tissus où ces gènes ne s'expriment pas, leur îlot HTF reste non méthylé ; si, en revanche, ces gènes ne comportent

pas d'îlot HTF (cas relativement fréquent), leur non-expression est en général corrélée à la méthylation des quelques CpG qu'ils contiennent.

### Un nouveau tournant de l'Histoire

Tout récemment, le groupe d'Adrian Bird a apporté de nouvelles données sur la méthylation d'un nombre important d'îlots HTF dans l'ADN de cellules entretenues en lignée continue et a montré que cette méthylation correspond à l'extinction des gènes correspondants [10]. Ces résultats rejoignent en fait des observations plus anciennes [11] selon lesquelles les cellules en culture présentent un taux apparent de mutation très élevé, puisqu'il est assez facile d'isoler des mutants (comme des mutants thymidine kinase négatifs ou TK<sup>-</sup>), alors qu'avec le taux de mutation attendu *a priori* et en tenant compte du caractère diploïde de ces cellules, on s'attendrait à voir apparaître les mutants à une fréquence

mille fois plus faible... Ce que montrent de façon concluante Antequera *et al.* [10], c'est qu'il y a en fait méthylation d'une fraction importante des îlots HTF dans les lignées cellulaires entretenues au long cours : la moitié environ d'un échantillon d'une trentaine d'îlots pris au hasard, tous non méthylés dans les différents tissus de l'animal, sont méthylés dans des lignées comme les cellules L ou les NIH 3T3. De plus, les îlots méthylés s'avèrent être les mêmes dans les deux lignées : cela suggère très fortement qu'ils correspondent à un jeu de gènes non essentiels à la survie en culture, dont l'extinction a été bien tolérée ou a même été un facteur de sélection positive. Des gènes non essentiels comme ceux de l' $\alpha$ -globine, de Thy-I ou du facteur de fixation du rétinol et qui comportent des îlots HTF sont eux aussi méthylés. En revanche, les îlots appartenant aux quelques gènes de ménage étudiés dans le même article restent, quant à eux, vierges de groupements méthyl (dans les limites de

l'analyse bien sûr). Très clairement donc, une inactivation des gènes, associée à la méthylation des îlots HTF, se produit dans les cellules en culture ; si ces gènes ne sont pas essentiels, la cellule survit et peut même y gagner en vitesse de multiplication ; bien sûr, si le gène est essentiel, la cellule meurt. Cette méthylation des îlots est apparemment aléatoire et irréversible dans les conditions normales de la culture de cellules ; mais, comme on le sait depuis longtemps [12], un traitement à la 5-azacytidine peut réactiver des gènes dont l'expression avait disparu à la suite d'une telle mutation. Le composé en question ayant un effet déméthylant très net (c'est un inhibiteur de certaines méthylases), on comprend son action... et l'on retrouve un lien quasi causal entre déméthylation et expression.

Bien d'autres points de cet article mériteraient d'être discutés, en particulier des résultats un peu paradoxaux, mais finalement cohérents, sur l'accessibilité des îlots HTF dans la chromatine de noyaux intacts. Cela nous entraînerait un peu loin et je préfère renvoyer le lecteur aux trois articles récents du même groupe, qui décrivent l'identification d'une protéine se fixant sur les îlots HTF [13], la protection des CpG dans les noyaux [14] et, plus récemment, les caractéristiques particulières de la chromatine au niveau des îlots HTF [15]. Retenons que la chromatine dans les îlots HTF non méthylés a une conformation ouverte, liée en particulier à l'acétylation des histones H3 et H4 et à l'absence de l'histone H1 ; que cette chromatine est donc accessible, dans les noyaux intacts, aux enzymes de restriction, ce qui fournit un moyen d'isoler assez simplement une fraction active de la chromatine et que, en revanche, les îlots HTF méthylés sont recouverts d'une protéine spécifique (la *methyl CpG binding protein*) et donc à la fois inactifs au point de vue transcriptionnel et protégés contre le clivage par les enzymes de restriction.

### Génétique, épigénétique et imprinting

Je voudrais terminer cet article déjà assez dense en élargissant mon pro-

*ms n° 2, vol. 7, février 91*

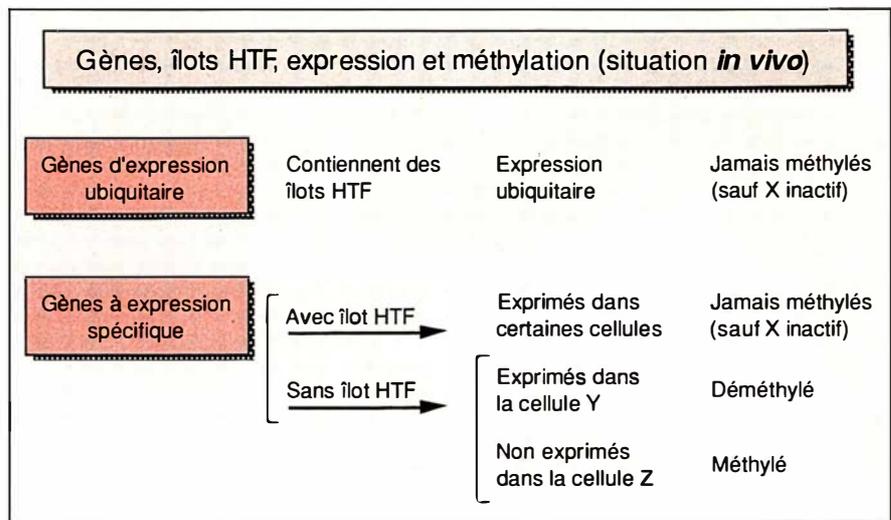


Figure 6. **Résumé schématique de nos connaissances sur les îlots HTF, leur corrélation avec des gènes et leur éventuelle méthylation. Ce tableau est valable *in vivo*, pas pour les lignées de cellules (voir le texte).**

pos vers les mécanismes épigénétiques et les phénomènes d'*imprinting*. Le terme français est « imprégnation », mais il a été fort peu employé pour le phénomène qui nous intéresse et je garderai donc, pour une fois, le mot anglo-saxon. Il est clair, et R. Holliday l'avait fort bien exprimé il y a déjà quelques années [12], que la méthylation de l'ADN peut être un code qui se superpose au premier code, celui de la séquence des nucléotides. La méthylation étant en effet héritée et fidèlement transmise (sauf exception), elle permet effectivement de transmettre et de perpétuer des schémas d'activité de gènes ; son rôle potentiel dans le déroulement du développement peut donc être très important [16].

A un autre niveau et sous un angle un peu différent, le phénomène de l'*imprinting* est lui aussi relié à notre propos. Il ne s'agit pas bien sûr de l'imprégnation au niveau du comportement décrite par Conrad Lorenz [17] mais du fait prouvé, qu'au moins chez les mammifères le matériel génétique ne s'exprime pas de la même façon selon qu'il provient du parent mâle ou femelle [18]. En d'autres termes, il faut pour qu'un embryon se développe qu'il comporte un jeu de chromosomes provenant d'un mâle et un jeu provenant d'une femelle : ce qui réduit à néant les espoirs — et les craintes

— de clonage d'êtres humains par injection de noyaux somatiques dans des ovules préalablement énucléés : ce thème, qui a fait, l'objet de quelques romans correspond à une expérience qui réussit chez la grenouille (où l'*imprinting* n'existe pas) mais pas chez la souris et encore moins chez l'homme. En fait, les ovules humains qui par accident reçoivent deux jeux de chromosomes mâles donnent des moles hydatiformes, une sorte de tumeur ne contenant que des tissus désorganisés de type placentaire, alors que s'ils ont deux jeux de chromosomes femelles, l'on a un tératome formé exclusivement de tissus embryonnaires. L'on observe de même, à l'occasion de la construction de souris transgéniques, des phénomènes inattendus : il arrive assez souvent que l'expression du transgène dépende du sexe du parent qui l'a transmis. Par exemple, le transgène hérité d'un mâle sera correctement exprimé, mais la descendance d'une souris transgénique femelle ne présentera aucune expression alors qu'elle porte le transgène. Dans ce cas, l'extinction (provisoire) du transgène est clairement corrélée à sa méthylation [19]. Chez l'homme aussi, l'on commence à connaître des segments géniques dont l'expression dépend du sexe du parent ; un exemple qui vient immédiatement à

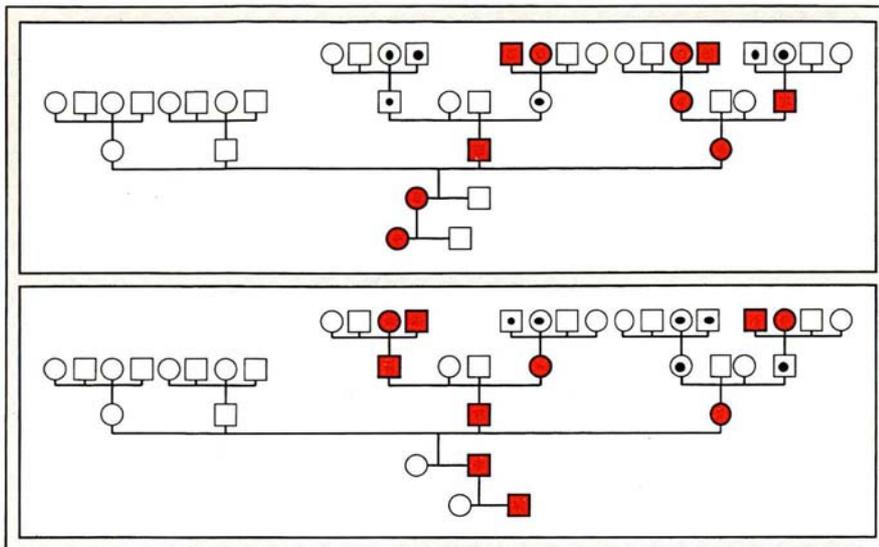


Figure 7. **Imprinting et transmission d'une maladie.** On montre ici deux arbres généalogiques imaginaires dans lesquels on suit la transmission d'une maladie génétique dont l'expression est régie par l'imprinting. Les individus atteints sont représentés en rouge, les individus porteurs du gène muté mais non atteints sont signalés par un point noir. Le gène est supposé autosomique et les partenaires ne sont pas représentés à partir de la génération III ; quatre enfants sont figurés pour chaque fratrie. En haut l'allèle est transmis de façon mendélienne mais la maladie n'est exprimée chez un individu que s'il a reçu cet allèle de son père (symbole rouge) ; s'il le reçoit de sa mère, il est porteur mais non atteint (point noir). En bas, le cas inverse où la maladie ne s'exprime que si l'allèle muté est transmis par la mère.

l'esprit est celui de la tumeur de Wilms : les délétions partielles du chromosome 11 rencontrées dans les tumeurs de Wilms sporadiques intéressent pratiquement toujours le chromosome d'origine maternelle [20, 21], comme si seule la région du 11 d'origine maternelle pouvait jouer un rôle d'anti-oncogène, sa délétion permettant donc le développement d'une tumeur que le 11 d'origine paternelle ne pourra pas empêcher. De même, l'imprinting a été proposé comme explication au fait que les cas de chorée de Huntington précoces correspondent tous à une transmission paternelle [22]. On peut multiplier les exemples, et l'on trouvera une excellente discussion de l'imprinting et des maladies génétiques dans une récente revue de Judith Hall [23]. Il y est très justement souligné qu'une maladie liée à l'imprinting paternel ou maternel d'un gène va donner un arbre généalogique apparemment aberrant où la transmission de la maladie se fera de façon non mendélienne

(figure 7) : une réinterprétation de données sur l'homme ou la souris mises de côté car semblant inutilisables donnerait sans doute quelques surprises. N'oublions pas non plus qu'un des modèles les plus séduisants avancés pour expliquer le mode de transmission bizarre du retard mental lié à la fragilité du chromosome X est celui de Charles Laird qui fait appel à l'imprinting [24, 25].

Il y a là tout un domaine fascinant où l'on voit la méthylation introduire un deuxième niveau de déterminisme génétique — ou plutôt épigénétique — qui module, corrige et, dans certains cas, met en échec le premier niveau, celui de la séquence des nucléotides. Les îlots HTF auxquels nous avons consacré l'essentiel de cette synthèse sont certainement un élément important de ces mécanismes ; et la poursuite de leur étude va sans doute apporter de nouveaux éclairages sur la façon dont notre patrimoine génétique est transmis, contrôlé et utilisé ■

## Summary

### HTF islands : the announced gene

The existence of HTF islands was recognized in the early eighties, and their knowledge has made significant progress recently. HTF island are short (1-2 kb) stretches of mammalian genomic DNA in which « CpG suppression » does not occur, so that CpG frequency is much higher than in bulk DNA. There are approximately 30 000 HTF islands per haploid genome ; these structures contain most of the sites for « rare cutter » enzymes used in large-scale mapping, and they are generally associated with housekeeping or, occasionally, tissue-specific genes. Thus they provide extremely useful landmarks for the gene-search strategies which are set up in the second phase of the reverse genetics approach. Methylation of HTF islands, although generally excluded, does occur in specific instances which are of great interest as they lead into the fields of imprinting mechanisms and of epigenetic inheritance.

## Remerciements

Je remercie le Dr Judith Hall [21] de m'avoir autorisé à utiliser une de ses illustrations (figure 7 de cet article).

## TIRÉS A PART

B. R. Jordan.