

## **Un nouveau gène du diabète insulino-dépendant sur le chromosome 11q. Le DID : paradigme pour l'étude des maladies multifactorielles**

Les gènes responsables de maladies mendéliennes comme la dystrophie de Duchenne ou la chorée de Huntington ont été identifiés grâce aux méthodes dites de « clonage positionnel ». Ces maladies relativement rares présentent une transmission simple et bien connue, et leur élucidation a pleinement bénéficié des progrès de la cartographie génétique et physique du génome [1, 2]. Leur relative homogénéité, et la présence fréquente de familles multiplex multigénérationnelles a facilité la découverte des gènes responsables, qui repose d'abord sur la mise en évidence d'une liaison génétique dans un petit nombre de familles (c'est-à-dire d'une coségrégation d'un allèle, d'un marqueur polymorphe anonyme ou situé près d'un gène candidat et de la maladie, avec obtention d'un *lod score* > 3) (*m/s* n° 12, vol. 9, p. 1418).

Le problème se pose différemment avec des maladies multifactorielles fréquentes comme le diabète sucré : leur hérédité est indéniable mais leur mode de transmission est inconnu, et leur hétérogénéité clinique et, probablement, génétique est extrême. Les membres d'une famille à haute prévalence de diabète partagent, certes, la plupart de leurs gènes, mais également des facteurs environnementaux qui jouent aussi un rôle important dans l'agrégation familiale de la maladie. Ainsi, la concordance entre jumeaux monozygotes pour le diabète insulino-dépen-

dant (DID) est de 30 à 50 % [3], mais on ne connaît ni le nombre de gènes en cause, ni la nature de l'interaction entre eux ou avec des facteurs de susceptibilité environnementaux (viraux, toxiques ou alimentaires).

Le rôle majeur du *locus* HLA dans le DID a été établi par des études de populations, puis confirmé dans des familles multiplex. La susceptibilité la plus grande est associée aux allèles de classe II des *loci* DR et DQ [4] : près de 90 % des sujets DID appartiennent aux groupes DR3/DR4, mais seulement 40 % de la susceptibilité génétique est expliquée par cette région du chromosome 6 (*m/s* n° 2, vol. 4, p. 118, n° 3, vol. 7, p. 287) [5, 6]. En effet, la concordance pour le DID n'est que de 10 % parmi les frères et sœurs HLA identiques.

Un deuxième gène de susceptibilité a été localisé dans la région du gène de l'insuline [7], sur le chromosome 11p. Une association a d'abord été suggérée entre le VNTR\* 5' adjacent au gène de l'insuline et le DID, mais les études familiales n'ont d'abord pas montré de liaison significative. Les raisons en étaient le relativement faible poids de ce *locus* (environ 10 % du risque génétique), et la très grande fréquence dans la population générale de l'haplotype en déséquilibre de liaison avec le variant morbide. La liaison a pu être finalement établie, en étudiant spécifiquement les familles dont les parents étaient hétérozygotes pour cet haplotype, et en montrant que les paires d'enfants atteints partageaient cet haplotype avec une fréquence de 0,67 (soit significativement plus que les 0,50 attendus [7]). La liaison était plus forte dans les paires DR4<sup>+</sup>, et la distorsion de transmission

était surtout due (du moins dans la population française) au partage non aléatoire des haplotypes des pères, suggérant un effet d'empreinte génomique d'origine paternelle. La combinaison d'études familiales et d'association dans les populations anglaise et française ainsi que le séquençage d'une large région de l'ADN de 19 kb autour du gène de l'insuline a ensuite permis de restreindre la région de susceptibilité à un segment de 4,1 kb, qui comprend la région flanquante en 5' du gène de l'insuline, et s'étend jusqu'au premier promoteur du gène *IGF II* [8]. Cette séquence présente 10 variants associés au risque de DID, chacun d'entre eux étant en étroit déséquilibre de liaison avec les autres. L'analyse simultanée des génotypes possibles chez 600 sujets DID français et 200 témoins vient de permettre de reconnaître quatre de ces variants, dont au moins un serait directement impliqué dans la susceptibilité au DID [9]. Des études fonctionnelles sont en cours, qui devraient aider à déterminer la séquence responsable, et à comprendre le mécanisme physiopathologique en cause dans cette région.

Les études génétiques de croisements de la souris NOD (*non obese diabetic*), qui présente un DID spontané, avaient montré qu'au moins 9 autres *loci* participaient au développement de la maladie [10]. Il était donc clair (d'autant que les gènes candidats évidents étaient rares) que seule l'exploration « au hasard » de l'ensemble du génome humain permettrait de dresser l'inventaire des gènes en cause dans le DID. Pour ce faire, il est nécessaire d'étudier un grand nombre de familles présentant plusieurs

\*variable number of tandem repeat, ou minisatellites : régions du génome caractérisées par la répétition en tandem d'une même séquence d'ADN. Le nombre de répétitions varie d'un sujet à l'autre, engendrant un multiallélisme très informatif.

membres atteints dans la fratrie, de manière à rechercher un excès de concordance (> 50 %) dans la transmission génique parmi les paires de germains affectés étudiés (analyse de *sib-pair*). La liaison est affirmée quand un marqueur montre une grande identité par descendance\* dans les familles nucléaires diabétiques. Cette méthode ne nécessite pas de présupposer le nombre de gènes, ni la nature de leurs relations. Le pouvoir résolutif de l'analyse de *sib-pairs* est cependant *a priori* faible, et la cartographie d'exclusion est donc plus aléatoire que dans l'analyse de liaison classique (méthode des *lod scores*) : seule la détermination des génotypes de plusieurs centaines de familles, en utilisant des marqueurs extrêmement polymorphes (> 80 % d'hétérozygotie, [1]), et très proches l'un de l'autre (au plus 10 à 20 cM) donne une probabilité non négligeable de mettre en évidence les *loci* liés au DID. Le développement de techniques semi-automatisées de génotypage ultra-rapide est donc un préalable à la réussite d'un tel projet. C'est en résolvant simultanément les problèmes spécifiques posés par la recherche des gènes de susceptibilité d'une maladie multifactorielle, que deux équipes, l'une française (Inserm U. 358), et l'autre britannique (Université d'Oxford), viennent de localiser un marqueur du DID sur le chromosome 11q, près du *locus FGF3* (Tableau I, [11, 12]). La liaison la plus forte avec ce *locus* est retrouvée dans les fratries HLA DR3 ou DR3/DR4, mais non dans les paires DR4 (Tableau II). D'autres régions potentielles de susceptibilité pourraient exister (du moins dans l'échantillon britannique) sur le chromosome 6q (*locus ESR*, qui est dans la même région que le gène de la superoxyde dismutase 2 pour lequel une association avec le DID a été suggérée), et sur le chromosome 18 (dans la région du groupe sanguin Kidd, dont la liaison avec le DID avait été proposée il y a 10 ans). De manière rassurante, les deux groupes ont retrouvé par l'analyse de paires de germains affectés que la plus forte région liée au DID était HLA (6p). Seul l'échantillon de familles anglaises a montré une liaison avec le gène de l'insuline, ce qui confirme la

difficulté de mise en évidence de ce *locus* par l'analyse de *sib-pairs*. Ces travaux ont d'abord été rendus possibles par la constitution de grandes banques d'ADN en France (constituée par l'association « 200 familles pour vaincre le diabète » grâce à une campagne médiatique), et en Grande-Bretagne et aux USA (grâce à l'aide des associations anglo-saxonnes de

diabétiques). D'autre part, rien n'aurait été possible sans la carte de haute résolution de microsatellites développée récemment au Généthon par l'équipe de Jean Weissenbach [1]. La stratégie utilisée par Cécile Julier [11] a été de caractériser 321 marqueurs polymorphes dans un premier groupe de 43 familles (soit 61 *sib-pairs*), et d'examiner les 188 familles

Tableau I				
ANALYSE DE LIAISON ENTRE LE DID ET LE CHROMOSOME 11q CHEZ 314 PAIRES DE GERMAINS AFFECTÉS				
Chromosome 11	$\theta^*$	Marqueur	Tous diabétiques	Diabétiques DR3
	0,07	AFM211xe1	n.s.	p = 0,0130
	0,01	PYGM	p = 0,0004	p = 0,0003
		AFMa131ye5	p = 0,0124	p = 0,0174
	0,02	FGF3	p = 0,0008	p = 0,0001
	0,06	AFM185ya1	p = 0,0032	p = 0,0005

\* : Fraction de recombinaison, expression de la distance génétique entre le marqueur et le gène.  
n.s. : non significatif.

Tableau II					
LIAISON DU DID AU LOCUS FGF3 SUR LE CHROMOSOME 11 DANS 251 PAIRES DE GERMAINS					
Tous DID		Paires DR3 (1)		Paires DR4 (1)	
IBD = 1	IBD = 0	IBD = 1	IBD = 0	IBD = 1	IBD = 0
173	258	95	46	98	77
p < 0,01		p < 0,0001		n.s.	

(1) Les deux membres de la paire doivent satisfaire au critère retenu.

(2) IBD = Identity by descent (voir encadré).

n.s. : non significatif.

**Identité par descendance (*identity by descent*, IBD) :** quant une paire de germains partagent un même allèle d'un locus donné, cet allèle est considéré « identique par descendance » lorsqu'il représente le même allèle hérité d'un même ancêtre. Dans le cas où on ne peut déterminer l'origine parentale de l'allèle partagé par les germains affectés, on parle d'identité par état (*identity by state*, IBS). L'analyse de liaison génétique par étude de paires de germains affectés est théoriquement plus puissante quand on peut rechercher un excès de transmission d'allèles partagés par IBD, plutôt que par IBS. Toutefois, la différence de puissance est faible dans le cas de marqueurs extrêmement polymorphes (hétérozygotie supérieure à 80 %), comme c'est le cas pour les marqueurs de la carte génétique du Généthon. De plus, dans le cas où un ou les deux parents sont décédés, il existe des méthodes d'analyse de paires de germains qui reconstituent les haplotypes parentaux par calcul de probabilités, à partir des génotypes des germains affectés et non affectés.

restantes pour tous les marqueurs montrant une indication de liaison ( $p < 0,05$ ). Le groupe de John Todd [12] a d'abord étudié 289 marqueurs dans 96 *sib-pairs* anglaises, puis a cherché à reproduire les résultats positifs dans deux autres groupes de familles anglaises ( $n = 102$ ) et américaines ( $n = 84$ ). Ainsi 40 marqueurs (12 %) appartenant à 15 régions génétiques ont donné une probabilité significative de liaison à  $p < 0,05$  dans les 43 familles françaises. La caractérisation des autres familles a restreint la possibilité de liaison à 17 marqueurs, dont la plupart étaient situés dans la région 11q (mais aussi 1q et 8p). La région *FGF3*, est en fin de compte celle qui donne la probabilité la plus élevée de liaison ( $p < 0,001$ ), mais il n'est pas exclu que certains des autres *loci* potentiellement intéressants soient aussi liés au DID. Une analyse plus fine de ces régions, dans les mêmes groupes de familles et dans d'autres populations, permettra certainement de répondre à cette question.

L'objectif clinique de ces travaux est notamment de faciliter le dépistage des personnes à risque de DID, dans un but de prévention ou de traitement précoce de cette maladie. Pour cela, il reste bien entendu à identifier le(s) gène(s) en cause (et en premier lieu le gène situé sur le chromosome 11q), et pour ce faire, l'agrandissement des banques d'ADN à de nouvelles familles est nécessaire. Une fois de plus, le diabète insulino-dépendant peut être considéré comme un modèle pour l'étude des traits complexes. La méthodologie développée pour l'étude du gène de l'insuline (*transmission disequilibrium test*, TDT), qui permet de mettre en évidence une distorsion dans la transmission à des patients d'un haplotype ou d'un variant [14], peut être utilisée dans d'autres pathologies. Ainsi, la transmission d'un mutant d'un gène candidat qui paraît associé à la maladie dans des études cas/contrôles, pourra aisément être testée dans des familles multiplex porteuses du variant [15]. Au-delà du DID, l'exploration du génome entier, réservée jusqu'ici aux seules maladies monogéniques, est probablement également réalisable dans les maladies multifactorielles [13]. Cette stratégie

pourra donc être utilisée dans des maladies aussi diverses que le diabète non insulino-dépendant, l'asthme, l'hypertension artérielle ou la polyarthrite rhumatoïde. Certes, dans ces affections à début souvent tardif, ou entraînant une mortalité précoce, et extrêmement hétérogènes, les conditions optimales de mise en évidence d'une identité par descendance ne sont pas toujours réunies : une carte génétique de plus haute résolution (5 à 10 cM) la caractérisation d'un plus grand nombre de *sib-pairs* ( $> 600$ ) mais aussi des membres non affectés des fratries, et l'utilisation systématique des nouvelles méthodes de génotypage automatisées par marqueurs fluorescents à l'aide de séquenceurs, et un effort de phénotypage précis des familles permettant la détermination de sous-groupes homogènes, devraient faciliter l'obtention de résultats positifs, comme dans le DID ■

## RÉFÉRENCES

- Weissenbach J, Gyapay G, Dib C, Vignal A, Morissette J, Millasseau P, Vaysseix G, Lathrop M. A second generation linkage map of the human genome. *Nature* 1992; 359: 794-801.
- Cohen D, Chumakov I, Weissenbach J. A first generation physical map of the human genome. *Nature* 1993; 366: 698-701.
- Deschamps I, Khalil I. Le rôle du système HLA dans la génétique du diabète de type 1. *Diabète Metab* 1992; 18: 253-63.
- Vague Ph. Multidisciplinarité et progrès dans le diabète. *médecine/sciences* 1994; 10: 791-4.
- Boitard C, Bach J. Pathogénie du diabète insulino-dépendant, maladie polygénique d'origine auto-immune. *médecine/sciences* 1991; 7: 226-34.
- Khalil I, Lepage V, Degos L, Galibert F, Deschamps I, Hors J. Diabète insulino-dépendant et HLA-DQ: un modèle moléculaire de susceptibilité. *médecine/sciences* 1991; 7: 235-9.
- Julier C, Hyer RN, Davies J, Merlin F, Soularue P, et al. The insulin-IGF2 region encodes a gene implicated in HLA-DR4 dependent diabetes susceptibility. *Nature* 1991; 354: 155-9.
- Lucassen AM, Julier C, Beressi JP et al. Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR. *Nature Genet* 1993; 4: 305-10.
- Julier C, Lucassen A, Villedieu P, Delepine M, Czernichow, Danze PM, et al. Multiple DNA variant association analysis: application to the insulin gene region in type 1 diabetes. *Am J Hum Genet* 1994 (sous presse).
- Ghosh S, Palmer SM, Rodrigues NR, Cordell HJ, Hearn CM, Cornall RJ et al. Polygenic control of auto-immune diabetes in non obese diabetic mice. *Nature Genet* 1994; 4: 404-9.
- Hashimoto L, Habita C, Beressi JP, Delepine M, Besse C, Cambon, et al. Genetic mapping of a susceptibility locus for insulin-dependent diabetes mellitus on chromosome 11q. *Nature* 1994; 371: 161-4.
- Davies JL, Kawaguchi Y, Benneet ST, Copeman JB, Cordell HJ, Pritchard LE, et al. A genome-wide search for human type 1 diabetes susceptibility genes. *Nature* 1994; 371: 130-136.
- Goodfellow PN, Schmitt K. From the simple to the complex. *Nature* 1994; 371: 104-5.
- Spielmann RS, McGinnis RE, Ewens WJ. Transmission Test for Linkage Disequilibrium: The Insulin Gene Region and Insulin-Dependent Diabetes Mellitus (IDDM.. *Am J Hum Genet* 1993; 52: 506-16.
- Hager J, Vaisse C, Vionnet N, Philippi A, Poller W, Cambien F, Passa Ph, et al. A missense mutation in the glucagon receptor gene is linked to familial NIDDM. *Diabetologia* 1994; 37: A16 (abstr).

## Philippe Froguel

Centre d'Étude du Polymorphisme Humain (CEPH), 27, rue Juliette-Dodu, 75010 Paris, France, service d'Endocrinologie et Inserm U.358, Hôpital Saint-Louis, Paris, France.

## TIRÉS A PART

P. Froguel.