

Les maladies génétiques "complexes"

Collège National des Enseignants et Praticiens de Génétique Médicale

Dr Francis Vasseur

Maître de Conférences en Génétique Praticien Hospitalier

Clinique de Santé Publique - Pôle S3P

Parc Eurasanté

CHRU de Lille

Date de création du document 2010-2011

Table des matières

I	Présentation.....	4
II	La maladie génétique "complexe" est plus complexe que prévu.....	6
III	Quand évoquer une maladie génétique "complexe" ?.....	7
	III.1 La maladie présente un déterminisme génétique.....	7
	III.1.1 La notion d'agrégation familiale.....	7
	III.1.2 La concordance des jumeaux.....	7
	III.1.3 L'existence d'ethnies à risque.....	8
	III.1.4 L'existence de modèles animaux.....	8
	III.2 Le mode de transmission ne répondant à aucun des schémas de transmission de maladie monogénique ou mitochondriale, la maladie est dite "complexe"	9
IV	Maladie génétique "compexe" et phénotype quantitatif continu.....	10
V	Où s'arrête la variabilité interindividuelle et où commence la maladie multifactorielle, la maladie génétique "complexe" ?.....	11
	V.1 Définition arbitraire et mathématique des états "normaux" et "pathologiques"	11
	V.2 Les caractères multifactoriels avec effet de seuil.....	12
	V.3 Définition de la limite pathologique par la survenue de comorbidités.....	12
	V.4 La notion de seuil et l'agrégation familiale.....	14
VI	Quelques exemples de maladies génétiques "complexes"	15
VII	L'identification des facteurs génétiques multiplesimpliqués dans une maladie génétique "complexe"	15
	VII.1 La stratégie gène candidat.....	15
	VII.2 L'approche génome entier (Genome Wide Scan).....	17
	VII.3 L'approche "Genome Wide Association Study" (GWAS).....	20
	VII.4 Le "Whole Exome Sequencing"	23
VIII	En finir avec l'héritabilité manquante ?.....	24

IX Conclusions.....	26
X Annexes.....	28

I PRÉSENTATION

On qualifie de maladies génétiques complexes, les maladies qui possèdent un déterminisme manifestement génétique mais dont la transmission ne correspond à aucun des modes mendéliens classiques de transmission, ni à un mode de transmission de type mitochondrial.

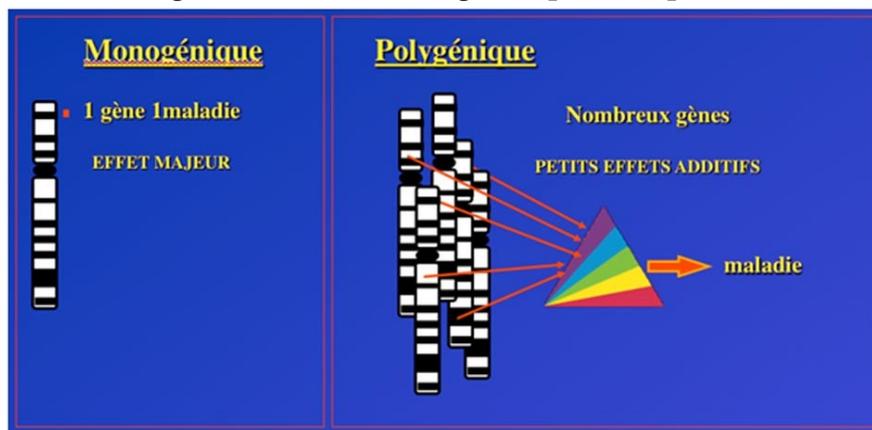
En effet si l'on peut résumer la maladie génétique monogénique à un modèle "simple" (par opposition aux maladies génétiques "complexes"):

l'altération au niveau d'un gène est nécessaire et suffisante pour s'accompagner de l'apparition de la maladie génétique (exception faite des mutations à pénétrance non complète),

dans les maladies génétiques complexes il s'agit d'un déterminisme génétique polygénique:

de nombreuses altérations aux niveaux de nombreux gènes sont les déterminants multiples de la maladie. On parle aussi de maladie "multifactorielle".

Figure 1 : Les maladies génétiques complexes



Chaque effet individuel de chacune des altérations génétiques est à lui seul insuffisant pour induire à la maladie. C'est la présence simultanée de ces altérations génétiques chez un même patient qui conduit à l'apparition de la maladie.

Dans ce contexte de maladie multifactorielle on attribue aussi une part non négligeable à l'"environnement" dans le déterminisme de la maladie génétique "complexe".

L'environnement étant à prendre au sens très large: ce qui environne ou a environné le patient ainsi que certains aspects du mode de vie du patient. La partie génétique du déterminisme de la maladie peut être considérée comme une susceptibilité, une prédisposition génétique à la maladie.

Figure 2 : Maladie génétiques complexes "Multifactorielle"



Il est possible de relier ces concepts à des notions mathématique et statistique et notamment ce qui est appelé la "part de variance expliquée". Dans une situation idyllique où nous connaissons à la fois tous les éléments génétiques qui constituent la composante polygénique ainsi que tous les déterminants "environnementaux" d'une maladie génétique "complexe" il serait possible de construire un modèle mathématique qui relierait de manière univoque la maladie (comme variable à expliquer) à n variables génétiques et n' variables environnementales (comme variables explicatives).

$$Y (\text{maladie}) = f (X_1\text{gène1}, X_2\text{gène2}, \dots, X_n\text{gène } n)$$
$$(X'_1\text{environnement1}, X'_2\text{environnement2}, \dots, X'_n \text{ environnement } n)$$

Ce modèle ne présentant plus aucune part d'incertitude, on dit qu'il expliquerait 100% de la variance de la maladie. Cette variance de la maladie se décompose en deux "sous variances" : la variance résultant des facteurs "environnementaux" et celle résultant des facteurs génétiques. Cette variance génétique est plus connue sous le terme d'héritabilité. Il existe des méthodes mathématiques indirectes pour estimer la valeur de l'héritabilité d'une maladie génétique "complexe" et ainsi avoir une idée du poids de la génétique dans une maladie donnée. Quant à la variance "environnementale" qui estime le poids de l'environnement dans le déterminisme d'une maladie génétique "complexe" on l'estime généralement par différence:

$$\text{Variance environnementale} = \text{Variance totale} - \text{Héritabilité.}$$

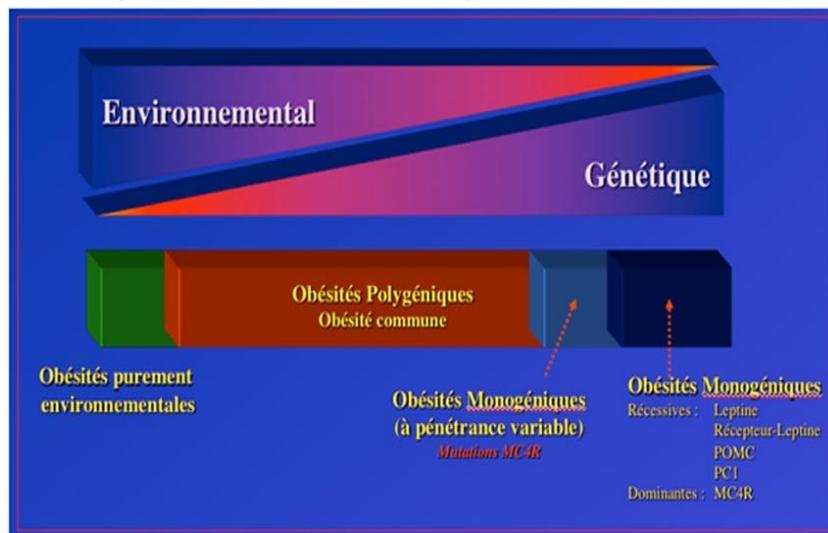
Il apparaît donc bien que derrière le terme "environnemental" soient regroupés tous les déterminants non génétiques dans une maladie génétique "complexe". Le terme "environnemental" est donc bien à prendre au sens très large: à la fois dans son sens écologique (l'environnement que subit ou a subit le patient) mais aussi dans un sens plus individuel (ce que le patient vit ou a vécu, compte tenu de son comportement présent et passé).

II LA MALADIE GÉNÉTIQUE "COMPLEXE" EST PLUS COMPLEXE QUE PRÉVU

On est amené à penser que certaines maladies génétiques "complexes" présenteraient une part imputable à la génétique (héritabilité) plus importante que d'autres. Ceci est certainement une réalité mais les choses s'avèrent encore plus complexes puisque pour une même entité clinique les parts respectives de "génétique" et d'"environnemental" sont variables d'un patient à un autre.

Cette notion transparaît bien pour l'obésité qui est reconnue comme une entité clinique et une maladie génétique "complexe" dont les déterminants génétiques peuvent être multiples et dont au moins une partie de la composante "environnementale" est identifiable sous la forme du mode de vie (activité physique, sédentarité, stress...), du type d'alimentation et du comportement vis à vis de la nourriture (dans la mesure où ce comportement n'est pas déterminé par des facteurs génétiques). Il apparaît que les contributions génétiques et environnementales dans le déterminisme de l'obésité sont très variables d'un patient à l'autre, allant du "tout génétique" avec des formes monogéniques d'obésité à transmission autosomique récessive (mutations du gène de la leptine *LEP* OMIM164160, du récepteur de la leptine *LEPR* OMIM601007, de la pro-opiomélanocortine *POMC* OMIM176830, de la proprotéine convertase1 *PCSK1* OMIM162150) à transmission autosomique dominante (mutations du gène codant le récepteur 4 aux mélanocortines *MC4R* OMIM155541) aux formes d'obésité purement environnementales en passant par des formes monogéniques à pénétrance variable et bien sûr les obésités polygéniques qui représentent les formes génétiques "complexes" de la maladie.

Figure 3 : Obésité : maladie génétique "complexe"



III QUAND ÉVOQUER UNE MALADIE GÉNÉTIQUE "COMPLEXE" ?

Avant d'avancer la nature génétique "complexe" d'une maladie il faut évidemment au préalable disposer d'arguments solides.

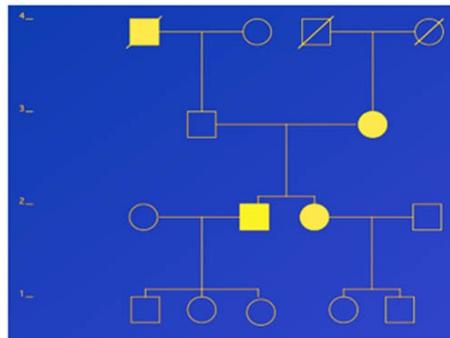
III.1 LA MALADIE PRÉSENTE UN DÉTERMINISME GÉNÉTIQUE

Les arguments qui font penser à un déterminisme génétique sont essentiellement de nature épidémiologique.

III.1.1 La notion d'agrégation familiale

La maladie récidive dans des familles et se présente chez de nombreux sujets de la famille. La prévalence de la maladie dans la famille est nettement supérieure à la prévalence observée dans la population générale.

Figure 4 : La notion d'agrégation familiale



En savoir plus: Demenais, F., Martinez, M., and Lathrop, M. (1996). Méthodes statistiques pour identifier les gènes dans les maladies multifactorielles. Ann Instit Pasteur, 7(1):3-12. Feingold, J. (2005). Maladies multifactorielles: un cauchemar pour le généticien. Med Sci, 11(21):927-33.

III.1.2 La concordance des jumeaux

Dans le cas de jumeaux monozygotes, quand l'un des jumeaux présente la maladie, très souvent le second la développe aussi. La concordance des jumeaux monozygotes atteint 80 à 90%. Cette concordance est moindre dans le cas de jumeaux dizygotes, de l'ordre de 16 à 40% ce qui permet de suspecter fortement une composante génétique de la maladie.

B. Newman et al., Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. Diabetologia 30 (1987), pp. 763-768.

III.1.3 L'existence d'ethnies à risque

Dans certaines ethnies la prévalence de maladies complexes est nettement plus élevée. C'est par exemple le cas du diabète de type 2 chez les indiens Pima d'Arizona où près de 50% des sujets présentent un diabète de type 2 et 95% des diabétiques sont obèses

Prevalence of diabetes in Mexican Americans. Relationship to percent of gene pool derived from native American sources. Gardner LI Jr, Stern MP, Haffner SM, Gaskill SP, Hazuda HP, Relethford JH, Eifler CW. Diabetes. 1984 Jan;33(1):86-92.

C'est également le cas des habitants de l'île de Nauru dans l'océan Pacifique où près de 90% de la population souffre d'obésité avec de ce fait une espérance de vie très faible, de 58 ans pour les hommes et 65 ans pour les femmes.

Dans ces deux exemples on admet que ces populations partagent un patrimoine génétique qui leur confère la susceptibilité à la maladie génétique "complexe" en question (diabète de type 2, obésité). On admet également que ces populations n'ont eu que peu d'apport génétique extérieur à même de "diluer" cette susceptibilité génétique. Il s'agit soit d'un isolat culturel dans le cas des indiens Pima d'Arizona, soit d'un isolat géographique dans celui des habitants de l'île de Nauru.

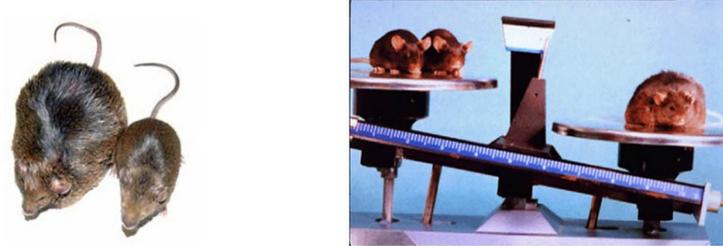
Figure 5 : L'existence d'ethnies à risque



III.1.4 L'existence de modèles animaux

Il existe de nombreux modèles animaux de maladies humaines. Des souches animales présentent spontanément des modèles génétiquement transmissibles de maladies humaines. C'est le cas par exemple des rats Zucker qui sont des modèles animaux d'obésité génétique. C'est d'ailleurs à partir de la souche Zucker que fut cloné et identifié le gène de la leptine. En effet les rats Zucker présentent une mutation (appelée fa, pour fat) du gène codant la leptine. Les rats fa/fa sont obèses et peuvent peser jusqu'à un kilogramme, soit le double de la souche sauvage. TW Kurtz, *The Zucker fatty rat as a genetic model of obesity and hypertension, Hypertension, vol. 13, no 6, 1989, p. 896-901.*

Figure 6 : L'existence de modèles animaux

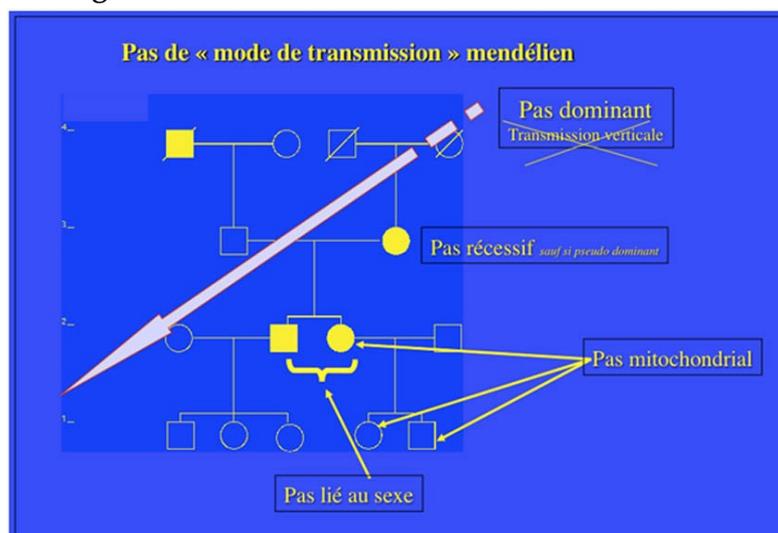


Si le rat Zucker est un modèle animal d'obésité monogénique, il existe des modèles animaux de maladies génétiques "complexes" comme le rat Goto-Kakizaki (GK) qui est un modèle animal de diabète de type 2. Les travaux réalisés chez le rat GK ont permis en 1996 de démontrer que le diabète de type 2 comportait (au moins chez le rat) un déterminisme polygénique. *Galli J et al. Genetic analysis of non-insulin dependent diabetes mellitus in the GK rat. Nat Genet. 1996 Jan;12(1):31-7.* De même la souche de souris KK/Ta représente un modèle animal de diabète de type 2 associé à l'obésité, et il existe des modèles murins d'athérosclérose. *Roberts A, Thompson JS. Inbred mice and their hybrids as an animal model for atherosclerosis research. Adv Exp Med Biol. 1976;67(00):313-327.*

III.2 LE MODE DE TRANSMISSION NE RÉPONDANT À AUCUN DES SCHÉMAS DE TRANSMISSION DE MALADIE MONOGÉNIQUE OU MITOCHONDRIALE, LA MALADIE EST DITE "COMPLEXE"

On constate une forte agrégation familiale mais la transmission ne correspond à aucun grand mode de transmission mendélien (autosomique dominant, autosomique récessif, récessif lié au chromosome X) ni à une transmission mitochondriale.

Figure 7 : Pas de mode de transmission mendélien



En effet la maladie génétique "complexe" étant le résultat des effets combinés de très nombreux variants génétiques, comme chacun d'entre eux considéré individuellement se transmet selon les lois de la génétique mendélienne classique et sur un mode récessif,

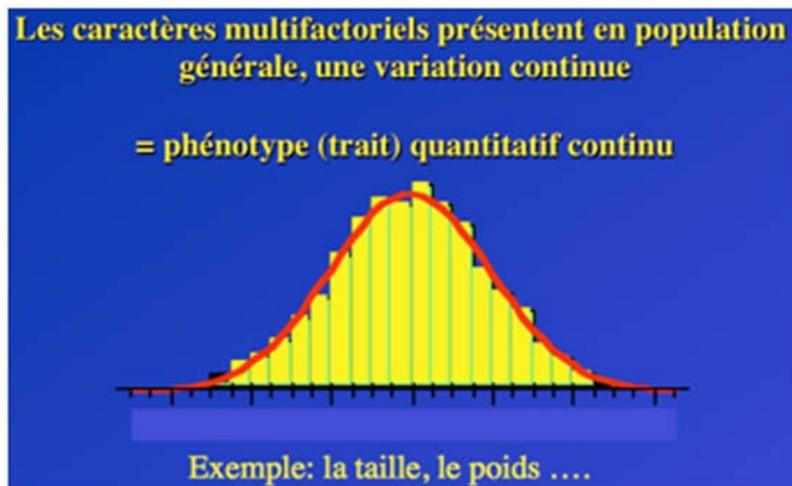
codominant ou dominant, l'hérédité polygénique est une hérédité mendélienne à de multiples loci. Mais globalement la résultante ne correspond à aucun modèle mendélien. Concernant l'hérédité polygénique, on admet par simplification que les effets des différents loci sont additifs ce qui conduit à parler d'hérédité *quantitative*. Il en résulte un phénotype très variable d'un individu à l'autre avec une tendance à obtenir un phénotype qui suit une distribution normale en population générale.

IV MALADIE GÉNÉTIQUE "COMPEXE" ET PHÉNOTYPE QUANTITATIF CONTINU

L'hérédité polygénique des maladies génétiques "complexes" est appelée hérédité multifactorielle, hérédité quantitative. En dehors d'un contexte pathologique les phénotypes sont généralement caractérisés par une variable numérique continue: par exemple la taille des individus, qui peut se définir par une valeur numérique, est un trait (phénotype) polygénique multifactoriel (<http://www.uic.edu/classes/bms/bms655/lesson11.html> : <http://www.uic.edu/classes/bms/bms655/lesson11.html>).

La taille des individus en population générale peut prendre une infinité de valeurs, c'est donc en termes statistiques une *variable aléatoire continue* qui peut se définir par sa moyenne et sa déviation standard (DS) (dans la mesure où sa distribution suit une loi normale). Les situations sont analogues pour d'autres phénotypes comme la couleur de peau et l'indice de masse corporelle.

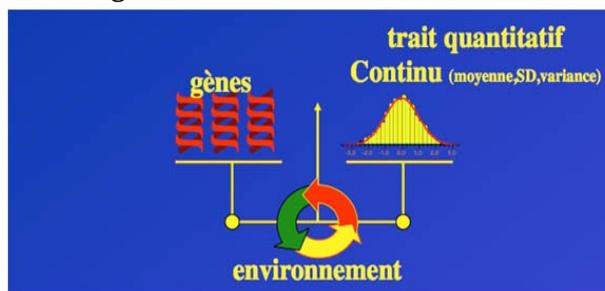
Figure 8 : Phénotype quantitatif continu



V OÙ S'ARRÊTE LA VARIABILITÉ INTERINDIVIDUELLE ET OÙ COMMENCE LA MALADIE MULTIFACTORIELLE, LA MALADIE GÉNÉTIQUE "COMPLEXE" ?

On admet que pour chaque maladie génétique complexe il doit correspondre un phénotype quantitatif continu dont la variation en population devrait permettre de définir les états normaux et pathologiques. Ce phénotype quantitatif étant sous la dépendance d'une hérédité multifactorielle la pathologie est donc bien de nature "complexe", à la fois polygénique et environnementale.

Figure 9 : Maladie multifactorielle

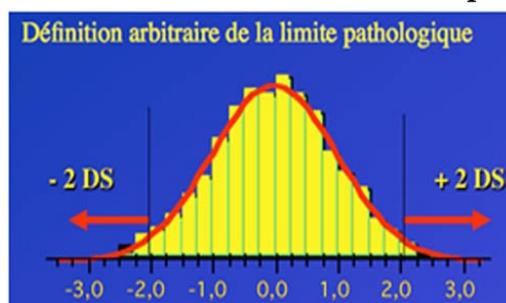


En savoir plus: Plomin R et al.. Common disorders are quantitative traits. *Nature Reviews Genetics* 10, 872 (2009). doi:10.1038/nrg2670.

V.1 DÉFINITION ARBITRAIRE ET MATHÉMATIQUE DES ÉTATS "NORMAUX" ET "PATHOLOGIQUES"

Pour de nombreux caractères multifactoriels il n'existe pas de méthode neutre et objective pour définir la limite entre un état normal et une situation pathologique qui doit en principe déclencher une prise en charge médicale. C'est le cas notamment des paramètres biochimiques comme la triglycémie, la cholestérolémie... Dans ces situations le "pathologique" est alors défini arbitrairement par les valeurs situées en dehors de la fourchette $-2DS/+2DS$ de la distribution du paramètre en population générale.

Figure 10 : Définition arbitraire de la limite pathologique



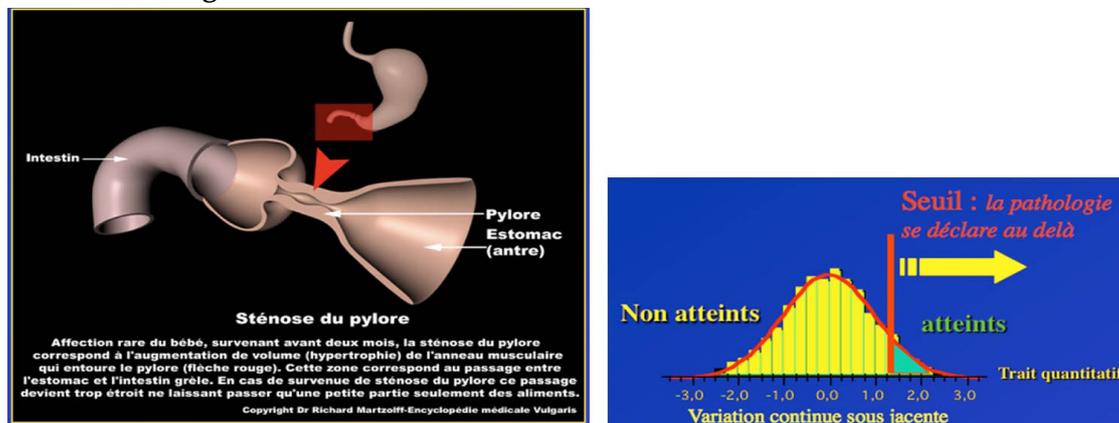
Dans ce contexte, l'intervalle $-2DS/+2DS$ qui correspond aux *valeurs normales* du paramètre, regroupe par définition mathématique et statistique, 95% de la population, le corollaire de

cette définition étant que 5% des sujets sont systématiquement dans la zone dite *pathologique*.

V.2 LES CARACTÈRES MULTIFACTORIELS AVEC EFFET DE SEUIL

Pour certains caractères multifactoriels une situation pathologique évidente se déclare au delà d'un seuil sans qu'il soit besoin de définir des *valeurs normales* en population générale. C'est le cas de la sténose du pylore qui est le rétrécissement au niveau de la jonction estomac - intestin, sténose résultant de l'hypertrophie du muscle qui entoure cette zone. Le développement de ce muscle est sous hérédité multifactorielle et on peut admettre que son épaissement suit une certaine distribution (loi normale ?) en population générale. Le seuil pathologique correspond au maximum d'hypertrophie du muscle et donc au maximum de rétrécissement du pylore compatible avec une fonction physiologique non altérée.

Figure 11 : Les caractères multifactoriels avec effet de seuil



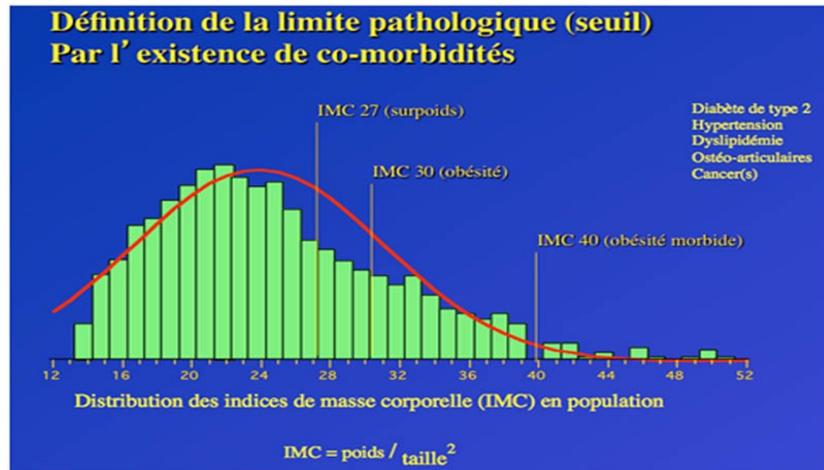
V.3 DÉFINITION DE LA LIMITE PATHOLOGIQUE PAR LA SURVENUE DE COMORBIDITÉS

Pour certaines maladies génétiques "complexes" comme l'obésité ou le diabète de type 2 la valeur en soi du paramètre quantitatif sous-jacent (l'indice de masse corporelle, IMC, la glycémie) ne possède pas de réel seuil physiologique au delà duquel le sujet aurait un ressenti pathologique qui l'amènerait à consulter, ce ne sont pas des caractères multifactoriels à effet de seuil. Pour ces maladies génétiques "complexes" la communauté d'experts médicaux n'a pas opté pour la définition statistique du pathologique sur la base de la déviation standard (DS) mais sur des seuils consensuels au delà desquels l'expérience a montré la survenue de comorbidités ou l'exacerbation des comorbidités.

Pour l'obésité qui peut se définir comme un excès de masse grasse généralement estimée indirectement par la détermination de l'indice de masse corporelle (IMC), un excès de masse grasse ne déclenche pas automatiquement un état pathologique mais les études épidémiologiques ont pu montrer la survenue de comorbidités (diabète de type 2,

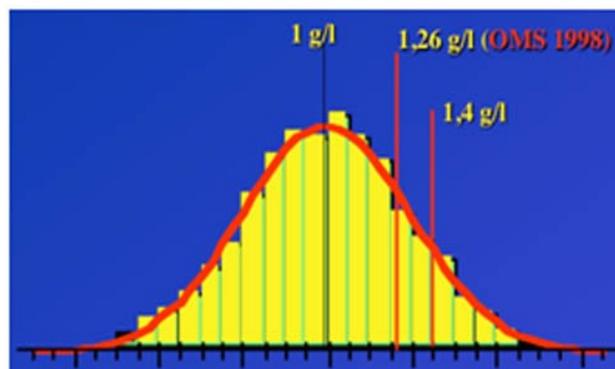
hypertension, dyslipidémie, problèmes ostéo-articulaires, cancers...) au delà d'un IMC de 27 qui est admis comme seuil de *surpoids*. Au delà d'un IMC de 30 la fréquence des comorbidités augmente, et ce seuil de 30 définit l'*obésité*. Au delà d'un IMC de 40 la fréquence des comorbidités devient si grande que le pronostic vital peut être à terme compromis, d'où la définition du seuil de 40 comme celui de l'*obésité morbide*.

Figure 12 : Définition de la limite pathologique



Pour le diabète de type 2, la glycémie à jeun est le paramètre numérique continu dont l'hérédité est multifactorielle. Ce paramètre suit une certaine distribution en population générale. Dans l'absolu on peut considérer qu'une hyperglycémie n'est pas en soi pathologique mais que ce sont les comorbidités et les complications (coronaropathie, neuropathie, néphropathie, atteintes rétinienne...) résultant d'une hyperglycémie chronique qui constituent la pathologie. Dans ce contexte il avait été établi un seuil de glycémie de 1.4g/l au delà duquel le sujet "devenait" un patient diabétique qu'il convenait de traiter afin d'équilibrer sa glycémie pour prévenir les comorbidités. Des études épidémiologiques de grande ampleur ont révélé par la suite que le risque de mortalité coronarienne, d'atteintes rénales et rétinienne était multiplié par deux chez les sujets dont la glycémie à jeun se situait au delà de 1.26g/l. Ces données ont amené en 1998 la communauté médicale à rabaisser le seuil pathologique à 1.26g/l.

Figure 13 : Définition de la limite pathologique



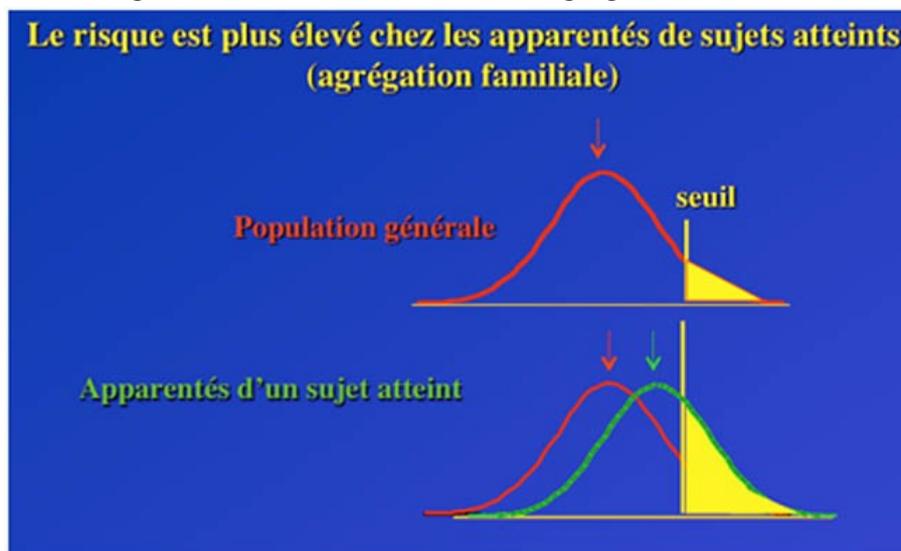
Cet exemple montre toute la difficulté à définir parfois un seuil pathologique dans les maladies génétiques "complexes".

En savoir plus: D. Chevenne, F. Trivoin, Le diabète sucré : propositions de nouvelles normes de diagnostic et de classification. Annales de Biologie Clinique. Volume 56, Numéro 4, 463-70, Juillet - Août 1998, Pratique quotidienne, http://www.jle.com/fr/revues/bio_rech/abc/e-docs/00/00/C5/9C/article.phtml .

V.4 LA NOTION DE SEUIL ET L'AGRÉGATION FAMILIALE

Les sujets apparentés partagent de nombreux facteurs génétiques y compris les facteurs de susceptibilité à une maladie génétique "complexe". Dans une famille qui présente au moins un sujet atteint de maladie génétique "complexe", la fréquence des variants génétiques de susceptibilité à la maladie est en général plus élevée que dans la population générale. Les sujets de cette famille auront donc une probabilité plus élevée de présenter ces variants génétiques. Comme ces variants génétiques déterminent la valeur du trait quantitatif sous jacent à la pathologie, les apparentés présenteront des valeurs augmentées du trait quantitatif, au prorata du nombre d'allèles de susceptibilité. Dans ces familles la distribution du paramètre quantitatif sera déplacée vers les valeurs élevées et de ce fait comparé à la population générale, une plus grande proportion de sujets seront au delà du seuil pathologique.

Figure 14 : La notion de seuil et l'agrégation familiale



VI QUELQUES EXEMPLES DE MALADIES GÉNÉTIQUES "COMPLEXES"

Il existe un très grand nombre de phénotypes quantitatifs à déterminisme polygénique multifactoriel, chacun pouvant être à priori associé à une pathologie: la fente labiale et palatine, la luxation congénitale de la hanche, certains cancers, les cardiopathies congénitales, les anomalies de fermeture du tube neural, certaines épilepsies, le glaucome, l'hypertension, les cardiopathies ischémiques, la maladie de Crohn, l'ostéoporose, des maladies mentales comme la schizophrénie et la psychose maniaco-dépressive, toutes les pathologies qui reposent sur l'anomalie d'un paramètre biochimique (HDL LDL cholestérol, glycémie...) ... On peut penser que les maladies multifactorielles constituent la grande majorité des pathologies en dehors des maladies infectieuses. Néanmoins comme une certaine variabilité inter individuelle de la susceptibilité aux infections est à déterminisme polygénique, même si la cause de la pathologie reste malgré tout l'agent infectieux, la présentation, la sévérité et l'évolution de la maladie infectieuse peut être considéré à déterminisme polygénique et multifactoriel "complexe".

VII L'IDENTIFICATION DES FACTEURS GÉNÉTIQUES MULTIPLESIMPLIQUÉS DANS UNE MALADIE GÉNÉTIQUE "COMPLEXE"

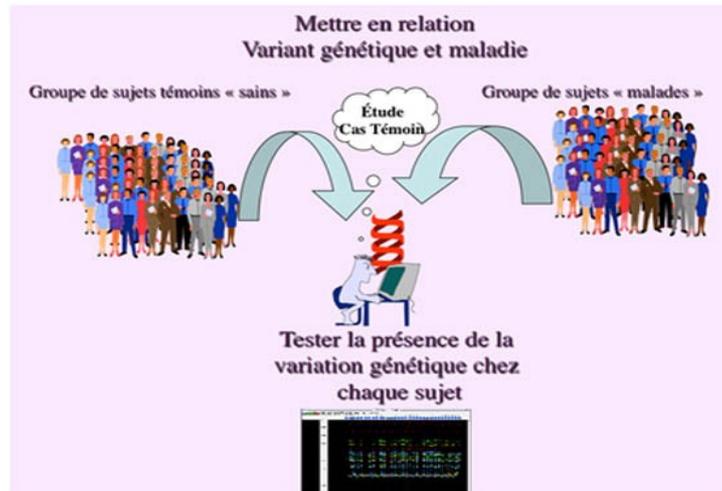
La recherche et la caractérisation des déterminants génétiques impliqués dans les maladies génétiques "complexes" a été initiée dans la décennie 1990 par des travaux qui ont concerné entre autres le diabète de type 2. Compte tenu de la multiplicité des loci potentiellement impliqués, les méthodes employées dans les maladies monogéniques ont dû être adaptées à l'étude de l'hérédité polygénique. On retiendra de cette époque, deux méthodes majeures: la stratégie gène candidat et l'approche génome entier (Genome Wide Scan).

VII.1 LA STRATÉGIE GÈNE CANDIDAT

Cette stratégie consiste à rechercher et analyser les variants génétiques dans des gènes dont la fonction pourrait jouer un rôle dans la pathologie en question. Elle repose sur une bonne connaissance de la physiologie et de la fonction du gène mis en examen et de ses interactions avec les autres partenaires impliqués dans la même voie métabolique qui pourrait être incriminée dans la pathologie. Par exemple le gène *GLUT4* (OMIM138190) codant un transporteur de glucose a été gène candidat dans un contexte de diabète de type 2. Les gènes impliqués dans la réponse immunitaire et l'inflammation ont été candidats dans le cadre du psoriasis. Une fois le ou les variants génétiques identifiés (en général par séquençage direct de l'ADN de sujets atteints et de sujets non atteints) il reste à conclure sur chacun quant à son rôle (ou son absence de rôle) dans la maladie génétique "complexe". Comme dans la maladie "complexe" chaque variant génétique ne possède qu'un effet

minime sur la variation du trait phénotypique et donc sur la survenue ou non de la maladie, cette phase nécessite une approche épidémiogénétique sur des groupes importants de patients atteints et de sujets témoins indemnes de la maladie (études cas-témoins).

Figure 15 : La stratégie gène candidat



La question posée étant : *le variant génétique est-il plus fréquent chez les sujets présentant la maladie que chez les témoins?*

La réponse à cette question ne peut être objectivée que par un test statistique de comparaison de fréquences au risque 5%, dans notre exemple, un test du χ^2 . Si la réponse est oui, on est en droit de conclure (au risque 5%) que:

- le variant en question est l'un des déterminants génétiques de la maladie
- ou il est en déséquilibre de liaison génétique avec l'un des déterminants génétiques de la maladie.

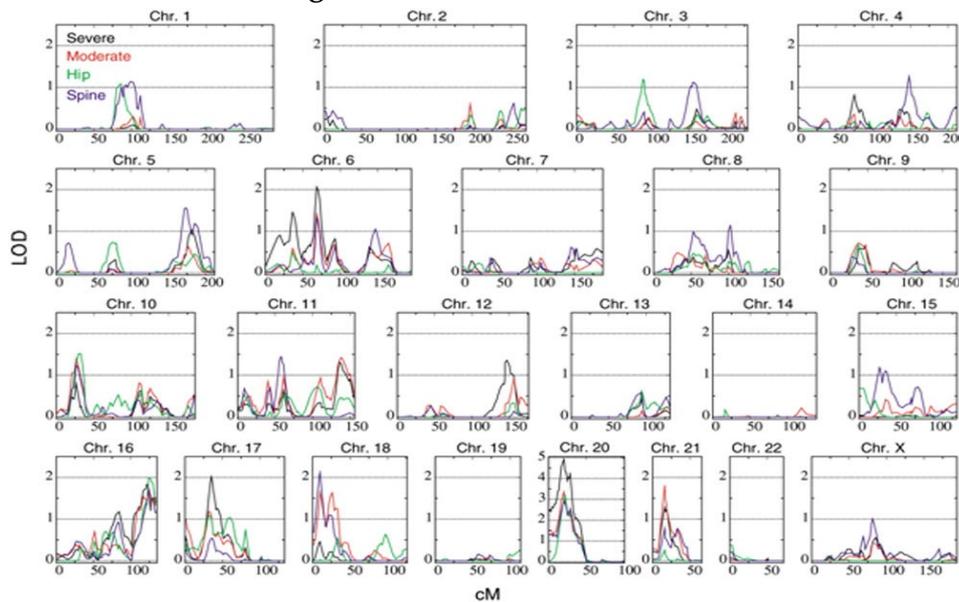
N'ayant aucune information objective et biologique quant au retentissement du variant génétique sur la fonction du gène et sur la maladie, nous ne pouvons à ce stade que constater qu'il existe une association significative au sens statistique entre la présence du variant génétique et la présence de la maladie. C'est pourquoi ces analyses sont appelées **analyses d'association**. Ces analyses d'association peuvent aussi se réaliser par des études familiales: tests de déséquilibre de transmission (Transmission disequilibrium test TDT). On objective par un test statistique la réponse à la question : *le variant génétique est-il transmis d'un parent hétérozygote à un enfant atteint, plus souvent que ne le voudrait le hasard ?*

Ces approches gène candidat qui ont mis en évidence de nombreux variants génétiques impliqués dans des maladies génétiques "complexes", présentent le défaut de ne pouvoir cibler et analyser que les gènes connus dont la fonction est évidemment connue et qui présente un rapport a priori évident avec la pathologie.

VII.2 L'APPROCHE GÉNOME ENTIER (GENOME WIDE SCAN)

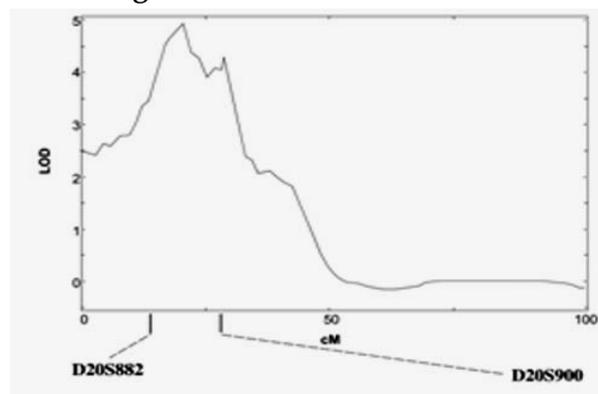
L'approche génome entier consistait à génotyper chez les sujets d'un grand nombre de familles (présentant des patients atteints) un maximum (± 1000) de marqueurs microsatellites très polymorphes, recouvrant la totalité du génome. C'est une méthode de génétique inverse et sans à priori contrairement à l'approche gène candidat. L'étude est basée sur l'analyse de la liaison génétique entre chaque marqueur microsatellite et la maladie "complexe". Par analogie avec le calcul du Lodscore employé dans les maladies monogénique et qui quantifie la liaison génétique entre un locus et la maladie, la méthode Genome Wide Scan quantifie la **liaison génétique** entre chaque marqueur microsatellite et la maladie "complexe". Dans l'exemple ci dessous extrait de Styrkarsdottir U et al. (*Linkage of osteoporosis to chromosome 20p12 and association to BMP2. PLoS Biol. 2003 Dec;1(3):E69. Epub 2003 Nov 3*), il a été étudié 1100 marqueurs microsatellites en relation avec l'ostéoporose.

Figure 16 : Genome Wide Scan



Les résultats montraient clairement que de nombreux loci présentaient une liaison génétique au moins "suggestive" avec l'ostéoporose, en accord avec la nature polygénique de la maladie. Le maximum de liaison génétique était observé sur le chromosome 20.

Figure 17 : Genome Wide Scan



Une vue détaillée des résultats concernant le chromosome 20 montre que la zone de liaison génétique couvrait presque la moitié du chromosome 20: 50cM soit 50 Mégabases d'ADN (50 000 kb). Les investigations ont été restreintes au pic de liaison génétique compris entre les marqueurs D20S882 et D20S900, la zone à explorer à la recherche d'un (ou plusieurs) variant génétique en rapport avec la maladie restait de 1 726 485 bp (1.7Mb) et même si cela ne représentait plus que 3% du chromosome 20 et 0.05% du génome, cette zone à explorer était encore très grande à l'échelle du biologiste moléculaire puisque compte tenu de la densité des gènes sur le chromosome 20 elle était censée contenir en théorie statistiquement 21 gènes (*Deloukas P. et al. The DNA sequence and comparative analysis of human chromosome 20. Nature. 2001-27;414(6866):865-71*). Il n'est pas rare que la stratégie gène candidat soit utilisée en seconde intention pour restreindre le nombre de gènes dont l'investigation sera être poussée (stratégie dite gène candidat positionnel). Dans l'exemple présenté la région de 1.7 Mb contenait en fait six gène répertoriés et connus en 2003 (*BMP2, CHGB, LOC51605, C20orf154, C20orf155, et C20orf42*). Compte tenu de son rôle dans la formation de l'os et la différenciation des ostéoblastes le gène *BMP2 (Bone Morphogenetic Protein 2 OMIM112261)* est apparu le meilleur gène candidat. Des investigations plus fines ont permis de mettre en évidence une mutation Ser37Ala du gène *BMP2*, fortement associée à l'ostéoporose mais qui n'expliquait que 30% de la liaison génétique sur le chromosome 20.

On perçoit ici les limites de la méthode qui sont (1) la relative grande taille des régions de liaison génétique (qui n'est pas sans rapport avec la distance entre les microsatellites) et donc le grand nombre de gènes à explorer en détail à la recherche de variants génétiques, (2) la part de liaison restant inexpliquée par les variants génétiques associés à la maladie.

Un dernier exemple incontournable concerne l'approche Genome Wide Scan dans la maladie de Crohn, une maladie inflammatoire chronique de l'intestin qui a conduit à la mise en évidence des mutations du gène *NOD2* en cause dans la maladie de Crohn (*Hugot JP et al. Mapping of a susceptibility locus for Crohn's disease on chromosome 16. Nature. 1996 Feb 29;379(6568):821-3*). *distribution: its use in the detection of linkage. Ann Hum Genet. 1978 Jul;42(1):87-94*).

La stratégie genome wide scan repose toujours sur la séquence: analyses de liaison puis clonage positionnel :

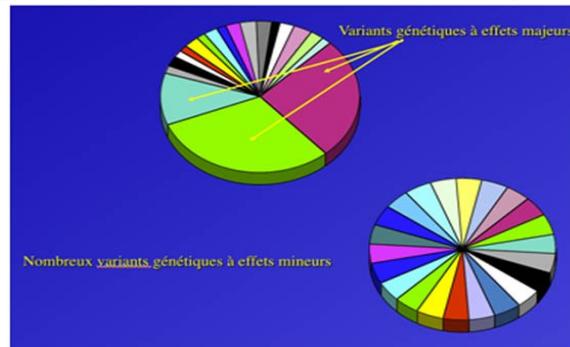
1°-identification de loci de susceptibilité par analyse de liaison génétique dans des familles,

2°-analyse de liaison avec des marqueurs plus rapprochés dans les loci précédemment identifiés ("fine mapping") pour restreindre la taille du locus de susceptibilité,

3°-puis analyse de gènes ou prédictions de gènes candidats positionnels, dont l'implication dans la maladie est validée ou écartée sur la base d'analyses d'association entre des variants génétiques et la maladie.

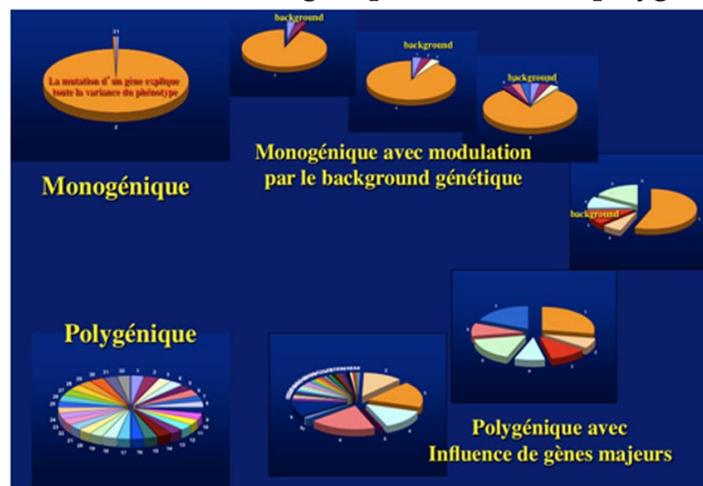
De ces études Genome Wide Scan on peut déjà avancer l'idée que certaines maladies multifactorielles auront une base génétique avec un ou des variants génétiques à effets majeurs (comme dans la maladie de Crohn) alors que pour d'autres maladies multifactorielles il sera question d'une multitude de variants génétiques ayant tous un effet mineur (comme cela semble être le cas pour le diabète de type 2 et l'obésité).

Figure 18 : Variants génétiques à effets mineurs - variants génétiques à effet majeurs



Par extension il faut admettre que la dichotomie sémantique entre maladies monogéniques et maladies polygéniques (multifactorielles) est finalement artificielle car entre la mutation d'un gène qui explique toute la variance du trait phénotypique dans une maladie monogénique et une affection multifactorielle dont la composante génétique résulte d'une myriade de variants génétiques à effets faibles il existe un continuum de situations qui expliquent les atteintes de sévérité variable en fonction du background génétique dans des maladies monogéniques et les maladies polygéniques avec l'effet d'un ou plusieurs gènes majeurs.

Figure 19 : Maladies monogéniques et maladies polygéniques



En effet il est rapporté de nombreux cas de variants génétiques qui modulent la sévérité d'une maladie monogénique comme par exemple dans la mucoviscidose (*Drumm ML et al. Genetic modifiers of lung disease in cystic fibrosis. N Engl J Med 2005 353 pp 1509-1511; Arkwright PD et al. TGF beta 1 genotype and accelerated decline in lung function of patients with cystic fibrosis. Thorax 2000 55 446*).

VII.3 L'APPROCHE "GENOME WIDE ASSOCIATION STUDY" (GWAS)

Le développement des technologies de séquençage et de génotypage à haut débit à partir des années 2000, basées sur les "puces à ADN" (*DNA chips*) a ouvert la voie à l'obtention rapide d'un très grand nombre de génotypes. Il est alors devenu envisageable de génotyper un grand nombre de marqueurs génétiques chez un grand nombre de sujets. Cette opportunité technologique a ouvert la voie au GWAS qui se résume à une étude d'association à très grande échelle: autant d'analyses d'association cas-témoin que de variants génétiques analysés. En effet la méthode consiste à génotyper un maximum de marqueurs génétiques (de 317 000 à 1 000 000 selon le type de puce à ADN) chez un grand nombre de sujets atteints de la maladie génétique "complexe" et un grand nombre de sujets témoins. Les variations d'un nucléotide (SNP, single nucleotide polymorphism) malgré leur nature bi-allélique qui contraste avec le caractère multi allélique des microsatellites, ont été retenues compte tenu de leur très grand nombre sur le génome humain offrant ainsi une très bonne couverture du génome. En effet les SNPs constituent les variants génétiques les plus fréquents. Par ailleurs leur nature bi allélique les rend plus accessibles au génotypage de masse au moyen des puces à ADN que les marqueurs microsatellites qui présentent un très grand nombre d'allèles. La méthode Genome Wide Scan qui s'intéressait à la co-transmission entre des régions du génome (repérées par les marqueurs microsatellites) et le phénotype "malade" ou "non malade", reposait donc sur la liaison génétique entre les marqueurs microsatellites et les loci en rapport avec la maladie "complexe". La méthode GWAS consiste en des études d'association entre un allèle donné d'un SNP et le phénotype "malade" ou "non malade." La méthode GWAS repose donc sur le déséquilibre de liaison génétique qui existe au sein du génome humain entre un allèle d'un marqueur génétique (SNP) et un variant génétique potentiellement impliqué dans la maladie "complexe". Un plus de la méthode est qu'elle permet de cribler le génome avec une haute densité de marqueurs: en effet pour une puce ADN 550K (qui génotype 550 000 SNPs) la densité moyenne est de un marqueur SNP tous les 60 kb.

Figure 20 :Single Nucleotide Polymorphism

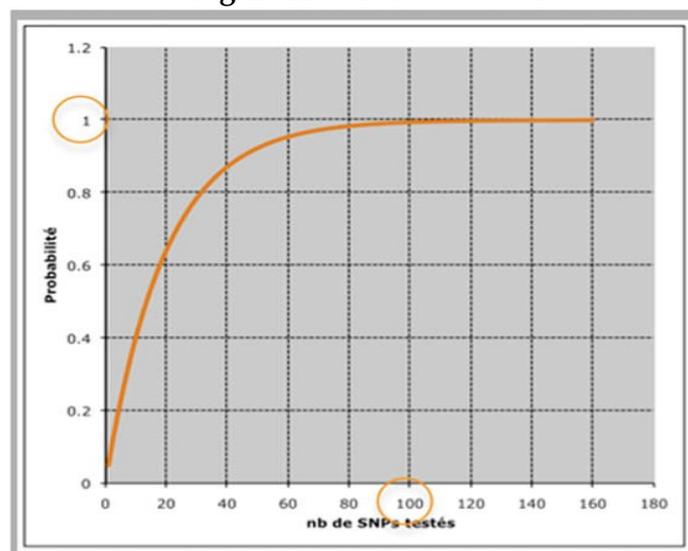


L'association mise à profit dans le GWAS est également fonction (1) de la distance entre le locus du marqueur génétique (SNP) analysé et le locus où réside le variant génétique de susceptibilité à la maladie (ce paramètre sera moins limitant compte tenu de la grande densité de marqueurs : un tous les 60kb pour une couverture de 550 000SNPs), (2) mais surtout sera fonction du déséquilibre de liaison génétique (LD pour « linkage disequilibrium ») qui existe entre un allèle donné du locus du marqueur SNP et l'allèle de susceptibilité au niveau du locus où réside le variant génétique de susceptibilité à la maladie. Comme le LD est variable d'une population et d'une ethnie à une autre, les approches GWAS devront prendre en compte cette spécificité et inclure des sujets homogènes quant au LD (*Salisbury BA et al. SNP and haplotype variation in the human genome. Mutat Res. 2003 May 15;526(1-2):53-61, Evans DM et al. A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations. Am J Hum Genet. 2005 Apr;76(4):681-7. Epub 2005 Feb 17*)

La méthode GWAS pourrait sembler "la" méthode définitive susceptible de repérer tous les déterminants génétiques d'une maladie, compte tenu de sa haute densité de couverture du génome. Néanmoins la méthode souffre d'un certain nombre de limites.

Comme elle se résume à une succession d'études d'association cas-témoin, l'association d'un SNP avec la maladie est testée comme habituellement, au risque statistique 5% (0.05). Une association sera réputée significative quand il n'y a que 5% de chances que cette affirmation soit fausse (ce qui somme toute laisse 95% de chances d'avoir raison). En d'autres termes il y a 5% de chances que l'association soit trouvée positive par hasard. Quand après avoir analysé deux SNPs, on évalue la probabilité que l'un d'entre eux soit associé par hasard, le risque n'est plus de 0.05 mais de 0.1. Il sera de 0.14 pour 3 SNPs....

Figure 21 : Méthode GWAS



On voit qu'à partir d'une centaine de SNPs testés la probabilité que l'un d'entre eux soit associé à la maladie par hasard (calculée par la loi binomiale) est quasiment de 1. Une

méthode classique pour prendre en compte ce problème est de diminuer le seuil de significativité individuel de 0.05 à 0.025 dans le cas de deux SNPs. C'est le principe de la correction de Bonferoni. Mais à combien faut-il baisser le seuil de significativité individuel de chaque SNP pour que globalement une fois les 550 000 SNPs testés, le risque que l'un d'entre eux soit positif par hasard soit encore ≤ 0.05 ? Le calcul montre que si on teste chaque SNP au risque 0.0000001 (10^{-7}), la probabilité pour que l'un d'entre eux soit déclaré associé à la maladie par hasard est voisine de 0.05, ce qui devient statistiquement acceptable.

Il apparaît donc que seuls les SNPs présentant une association très fortement significative ($p \leq 10^{-7}$), pourront être considérés comme étant très vraisemblablement associés à la maladie et donc physiquement proches d'un déterminant génétique de la maladie. Une proximité qui sera fonction de la densité des marqueurs génétiques. La méthode GWAS aura donc tendance à ne pouvoir détecter que les effets génétiques importants. C'est un phénomène que l'on peut présenter comme un "bruit de fond statistique".

Figure 22 : Bruit de fond statistique



Une manière de contourner ce problème est de travailler sur des effectifs pléthoriques de patients et de témoins. En effet pour un même effet biologique ou physiologique testé, si la probabilité associée à la significativité du test statistique est de 0.05 avec 100 patients et 100 témoins elle peut être de 0.001 avec deux fois plus de sujets. C'est l'une des raisons pour laquelle il n'est pas rare que des protocoles GWAS incluent un grand nombre de sujets.

La méthode GWAS reste néanmoins une étude d'association qui implique qu'une détection ne sera possible seulement que si le déterminant génétique est en déséquilibre de liaison génétique avec le marqueur SNP génotypé. Par ailleurs les SNPs inclus dans les puces à ADN par les fabricants, ont été choisis parmi les SNPs « fréquents » en terme de fréquences alléliques pour des raisons de commodité technologique et d'ubiquité d'utilisation. La méthode repose donc implicitement sur le postulat « common disease - common variant » qui s'avère de moins en moins ubiquitaire (Pritchard JK, Cox NJ. *The allelic architecture of human disease genes: common disease-common variant...or not? Hum Mol Genet. 2002 Oct 1;11(20):2417-23*). L'hypothèse « common disease common variant » suppose que les maladies multifactorielles qui sont des maladies fréquentes, reposeraient sur une susceptibilité génétique résultant de variants génétiques fréquents en termes de fréquences

alléliques. La méthode GWAS présente donc la tendance à faire l'impasse sur les variants génétiques rares. Cependant il est manifeste que certains variants génétiques rares jouent un rôle non négligeable de la susceptibilité génétique dans les maladies multifactorielles. Par exemple pour la mutation Y111H du gène ADIPOQ la fréquence allélique du variant n'est que de 0.015 (variant rare) mais cette variation s'accompagne d'un sur-risque de diabète de type 2 avec un OR=7.85 (*Vasseur F et al. Single-nucleotide polymorphism haplotypes in the both proximal promoter and exon 3 of the APM1 gene modulate adipocyte-secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes in French Caucasians. Hum Mol Genet. 2002 Oct 1;11(21):2607-14.*

Néanmoins la méthode GWAS a permis la mise en évidence de nombreux nouveaux variants génétiques de susceptibilité pour toutes les maladies multifactorielles et les divers QTLs analysés. Evidemment la méthode a retrouvé les associations antérieurement rapportées avec les méthodes antérieures comme par exemple les mutations du gène NOD2 dans la maladie de Crohn, les variants du gène APOE dans la maladie d'Alzheimer, les variants du gène PPARG dans le diabète de type 2. Une revue récente de 2010, répertoriait 49 gènes de susceptibilité au diabète de type 2, la plupart mis en évidence par GWAS (*Billings LK, Florez JC. The genetics of type 2 diabetes: what have we learned from GWAS? Ann N Y Acad Sci. 2010 Nov;1212:59-77. doi: 10.1111/j.1749-6632.2010.05838.x*). De même une étude de 2010 rapportait 71 loci de susceptibilité à la maladie de Crohn la plupart détectés au moyen de la méthode GWAS (*Franke et al. Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. Nat Genet. 2010 Dec;42(12):1118-25*). On voit que la méthode a permis une avancée significative dans la mise en évidence des déterminants génétiques en cause dans les maladies multifactorielles.

En savoir plus: *Why do genome-wide scans fail?* (<http://www.genetic-future.com/2008/03/why-do-genome-wide-scans-fail.html>) Lango H Weedon MN *What will whole genome searches for susceptibility genes for common complex disease offer to clinical practice?* *J Intern Med* 2008 vol. 263 (1) pp. 16-27. Manolio TA *Genomewide association studies and assessment of the risk of disease* *N Engl J Med* 2010 vol. 363 (2) pp. 166-76.

VII.4 LE "WHOLE EXOME SEQUENCING"

Malgré les avancées spectaculaires dans la connaissance des facteurs génétiques de susceptibilité aux maladies multifactorielles réalisées entre autres via les GWAS, pour une maladie donnée, l'ensemble des variants identifiés n'explique qu'une faible partie de la variance du phénotype (héritabilité). En moyenne dans les maladies multifactorielles à peine 10% de l'héritabilité est expliquée par les variants génétiques connus. La question est alors « sur quoi reposent les 90% manquants ? » et quelles méthodes d'investigation employer pour caractériser cette part manquante et ainsi expliquer 100% de l'héritabilité?

Une approche récente du problème repose sur une opportunité technologique. Partant du principe que les séquences codantes ne représentent qu'une faible partie du génome mais concentrent 85% des mutations potentiellement responsables des maladies, des méthodes permettant de "capturer" l'ensemble des séquences codantes ("Whole Exome") et d'en déterminer la séquence ("Whole Exome Sequencing") ont été développées (Choi M *et al. Genetic diagnosis by whole exome capture and massively parallel DNA sequencing Proc Natl Acad Sci U S A. 2009 Nov 10;106(45):19096-101. Epub 2009 Oct 27*). Ces méthodes de "Whole Exome Sequencing" (WES) permettent d'identifier en théorie la quasi totalité des variations de séquences qui existent au niveau des séquences codantes entre des sujets atteints de la maladie multifactorielle et des sujets indemnes. On voit que cette approche WES est bien adaptée à la caractérisation des variants rares qui étaient un peu laissés pour compte par la méthode GWAS.

VIII EN FINIR AVEC L'HÉRITABILITÉ MANQUANTE ?

Même si le "Whole Exome Sequencing" doit dans un futur récent révéler de nombreux variants rares qui contribuent à l'héritabilité des maladies multifactorielles, il est vraisemblable qu'il subsistera encore une part d'héritabilité manquante.

Selon Evan Eichler (Eichler EE *et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010 Jun;11(6):446-50*) l'héritabilité manquante résiderait dans les variants de « grande taille » (comparés aux SNPs) comme les délétions, duplications, inversions (copy number variations CNVs) qui sont communes dans les populations humaines et non accessibles par les approches classiques de génétique (GWAS, séquençage...).

En savoir plus: O'Donovan et al. Phenotypic variations on the theme of CNVs. Nat Genet 2008 vol. 40 (12) pp. 1392-3 Wellcome Trust Case Control Consortium Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. Nature 2010 464 pp 713-720

Selon Greg Gibson (Eichler EE *et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010 Jun;11(6):446-50*), comme l'héritabilité est aussi la variance totale du phénotype en population à laquelle on déduit la variance environnementale, il est postulé que les individus partagent un environnement commun, ce qui n'est vraisemblablement pas le cas et constitue un postulat réducteur. Gibson suppose l'existence d'un « environnement caché » qui interagirait avec les variants génétiques biaisant ainsi notre estimation de l'héritabilité. De plus les interactions gène-environnement ne sont pas prises en compte dans les modèles de calcul de l'héritabilité et n'interviennent pas non plus dans les analyses des résultats de GWAS.

En savoir plus: Visscher PM et al. Heritability in the genomics era-concepts and misconceptions. Nat Rev Genet. 2008 Apr;9(4):255-66. Epub 2008 Mar 4

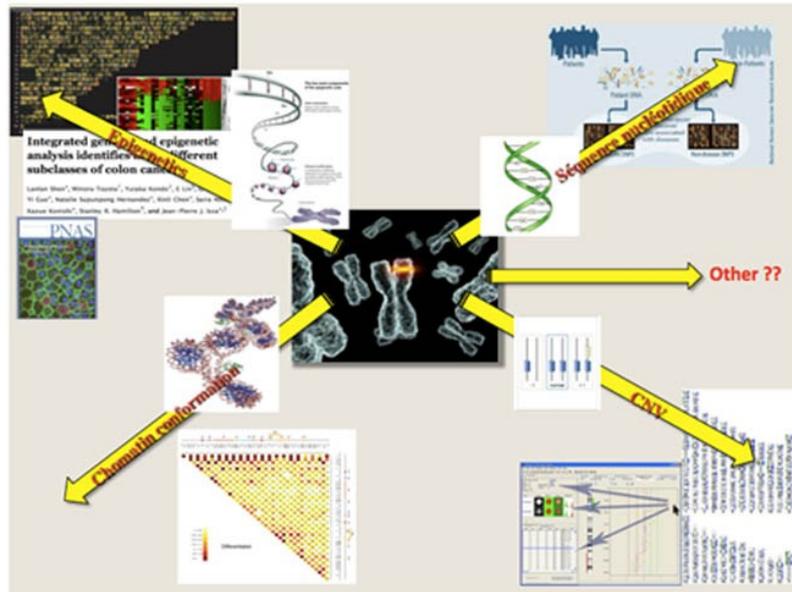
Selon Augustine Kong (*Eichler EE et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010 Jun;11(6):446-50*) comme certains variants impliqués notamment dans les cancers et dans le diabète de type 2 se sont avérés « à risque » uniquement quand ils sont transmis par un parent donné (*Kong A et al. Parental origin of sequence variants associated with complex diseases. Nature. 2009 Dec 17;462(7275):868-74*). En effet des variants « à risque » de diabète de type 2 devenaient « protecteurs » s'ils étaient transmis par l'autre parent. Ces phénomènes permettent de mieux cerner où se cache (une partie de) l'héritabilité manquante. D'une part ces variants sont délicats à mettre en évidence et leur contribution à l'héritabilité est forcément sous estimée si les modèles ne prennent pas en compte l'origine parentale. Un autre exemple concerne des variants génétiques qui augmentent la recombinaison génétique chez les pères et la réduisent chez les mères (*Kong A et al. Sequence variants in the RNF212 gene associate with genome-wide recombination rate. Science. 2008 Mar 7;319(5868):1398-401. Epub 2008 Jan 31*), comme le taux de recombinaison parental affecte la transmission et que la recombinaison peut s'accompagner de mutations il est possible qu'il y ait alors une association qui repose aussi sur la nature du parent transmetteur. Dans le même registre les mécanismes d'empreinte parentale (épigénétique) ne sont pas pris en compte dans les approches GWAS et représentent aussi une part non négligeable de l'héritabilité manquante.

En savoir plus: HT Bjornsson Intra-individual change over time in DNA methylation with familial clustering. JAMA 2008 vol. 299 (24) pp. 2877-83. C Biéumont From genotype to phenotype. What do epigenetics and epigenomics tell us? Heredity 2010 105 pp 1-3. Gertz J et al. Analysis of DNA Methylation in a Three-Generation Family Reveals Widespread Genetic Influence on Epigenetic Regulation. Plos Genet 2011 7 e1002228.

Selon Jason Moore (*Eichler EE et al. Missing heritability and strategies for finding the underlying causes of complex disease. Nat Rev Genet. 2010 Jun;11(6):446-50*) se référant aux micro-RNAs non codants qui sont un mécanisme nouveau de régulation génique, il faut conceptualiser l'héritabilité dans les maladies multifactorielles comme la résultante d'une superposition (juxtaposition ? interaction ?) de différentes couches de complexité. La première serait la variabilité de la structure primaire de l'ADN accessible par les méthodes habituelles (GWAS, séquençage), cependant des variants génétiques peuvent influencer ensuite l'expression de micro-RNAs ajoutant une nouvelle couche de complexité. Sur ces couches peuvent s'adjoindre d'autres couches de complexité comme l'épigénétique, les variations de recombinaison génétique en rapport avec des distorsions de transmission, les variations conformationnelles de la chromatine qui modulent aussi l'expression génétique. Ce concept que des variants génétiques multiples interagissent entre eux à travers plusieurs couches de complexité génomique est à mettre en perspective à ce que Bateson (*Bateson W, Mendel's*

principles of heredity ; Cambridge Univ. Press 1909 ; Bateson P, *William Bateson: a biologist ahead of his time* J Genet. 2002 Aug;81(2):49-58) dénomma en 1909 « epistasis » qui définissait les interactions entre les gènes, terme qui tend à prendre une définition plus large incluant les multiples interactions dans les systèmes biologiques complexes.

Figure 23 : Epistasis



IX CONCLUSIONS

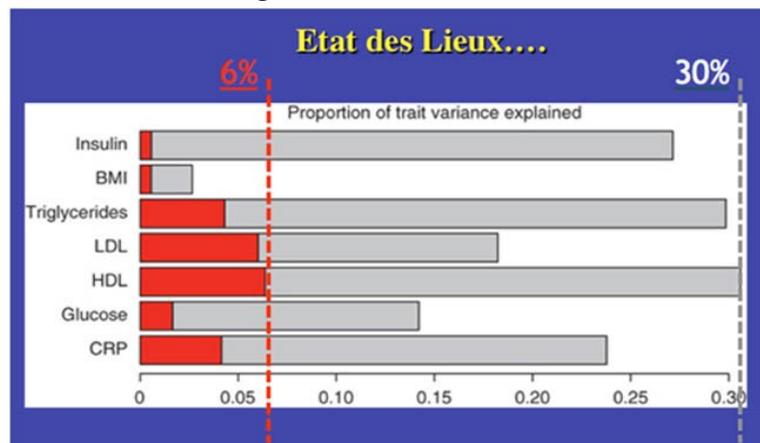
Dans un article de décembre 2008 (The genetic architecture of metabolic traits: a data explosion, (<http://m.wired.com/wiredscience/2008/12/the-genetic-architecture-of-metabolic-traits-a-data-explosion/>), Daniel Mc Arthur réagissait à propos d'articles publiés simultanément dans Nature Genetics concernant l'architecture génétique de traits phénotypiques et métaboliques. Deux papiers concernaient des GWAS sur le taux de glucose sanguin, un troisième rapportait des résultats de GWAS sur le diabète de type 2 et la sécrétion d'insuline; deux articles exploraient la dyslipidémie par GWAS, et un article explorait par GWAS les déterminants génétiques d'une dizaine de traits phénotypiques (triglycéridémie, HDL cholestérol, LDL cholestérol, glycémie, insulinémie, taux plasmatique de CRP, IMC, pression artérielle systolique et pression diastolique).

Prokopenko I et al. Variants in MTNR1B influence fasting glucose levels Nature Genetics 41, 77 - 81 (2008); Bouatia-Naji N et al. A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk Nature Genetics 41, 89 - 94 (2008); Lyssenko V et al. Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion Nature Genetics 41, 82 - 88 (2008); Kathiresan S et al. Common variants at 30 loci contribute to polygenic dyslipidemia Nature Genetics 41, 56 - 65 (2008); Aulchenko YS et al. Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts Nature Genetics 41, 47 - 55 (2008). Sabatti C et al. Genome-wide association analysis of metabolic traits in a

birth cohort from a founder population Nature Genetics 41, 35 - 46 (2008).

Daniel Mc Arthur mettait en relief le décalage entre la masse de variants génétiques révélés par ces études de grande ampleur et donc à priori impliqués dans l'héritabilité des traits, et la petite part de variance génétique expliquée par l'ensemble de ces variants génétiques. Par exemple l'une des études concernant 36610 sujets et la somme des variants détectés n'expliquait que 1.5% de la variance de la glycémie. Il remarquait aussi que deux études concernant la dyslipidémie et impliquant plus de 20000 sujets mettaient en relief un total de 18 loci dont un seul était commun aux deux études. La dernière étude (*Sabatti C et al.*), la plus petite en terme du nombre de sujets (n=4763), avait le grand avantage sur les autres de s'intéresser à une population homogène (toutes les personnes nées en 1966 dans deux des provinces du nord de la Finlande) et les auteurs, en plus de l'étude des déterminants génétiques par GWAS ont aussi évalué un certain nombre d'influences environnementales. Ce travail a l'avantage d'avoir permis de réaliser un état des lieux quant à ce que les données (de l'époque) permettaient de prédire tant du point de la part de variance génétique (héritabilité) que de la part de la variance environnementale. Les barres grises montrent la part de variance expliquée par les connaissances "actuelles" (variance génétique+variance environnementale), la part de variance génétique explicable par les variants génétiques connus est représentée par les barres rouges.

Figure 24 : Etat des lieux



On voit que dans le meilleur des cas, nos connaissances (génétiques et environnementales) actuelles ne permettent d'expliquer que 30% de la variance du HDL cholestérol, dont seulement 6% par les variants génétiques déjà identifiés. Pour les autres traits métaboliques la situation est bien moins avancée surtout en ce qui concerne la variance que la génétique (part d'héritabilité) est capable d'expliquer.

Ces résultats montrent que l'approche des maladies multifactorielles ne peut se dispenser d'inclure les variables "environnementales" dans les modèles d'analyse. Cette nécessité soulève de gros problèmes car contrairement à un génotypage de variant génétique, il est très difficile de recueillir de manière objective les variables environnementales (passées et présentes) d'un sujet donné. Par ailleurs ce recueil risque d'être toujours entaché d'une part

d'incertitude. On voit donc que les données récentes de la littérature pour stimulantes qu'elles soient invitent à une grande modestie quant à notre connaissance des déterminants d'une maladie multifactorielle et néanmoins ouvrent de vastes nouveaux champs d'investigation.

En savoir plus: Tishkoff S et al. Ten years of genetics and genomics: what have we achieved and where are we heading? Nat Rev Genet 2010 vol. 11 (10) pp. 723-33. Cordell HJ Detecting gene-gene interactions that underlie human diseases Nature Reviews Genetics 2009 vol. 10 (6) pp. 392-404. (Sudmant PH Diversity of human copy number variation and multicopy genes. Science 2010 vol. 330 (6004) pp. 641-6. Mc Clellan J Genetic heterogeneity in human disease. Cell 2010 vol. 141 (2) pp. 210-7. Williams A et al. Interchromosomal association and gene regulation in trans. Trends Genet 2010 vol. 26 (4) pp. 188-97. Bogardus C Missing heritability and GWAS utility. Obesity 2009 vol. 17 (2) pp. 209-10.

X ANNEXES

BIBLIOGRAPHIE

- Aulchenko YS et al. : Loci influencing lipid levels and coronary heart disease risk in 16 European population cohorts Nature Genetics 41, 47 - 55 (2008)
- B. Newman et al. : Concordance for type 2 (non-insulin-dependent) diabetes mellitus in male twins. Diabetologia 30 (1987), pp. 763-768.
- Bateson W : Mendel's principles of heredity ; Cambridge Univ. Press 1909 ; Bateson P, William Bateson: a biologist ahead of his time J Genet. 2002 Aug;81(2):49-58
- Billings LK, Florez JC. : The genetics of type 2 diabetes: what have we learned from GWAS? Ann N Y Acad Sci. 2010 Nov;1212:59-77. doi: 10.1111/j.1749-6632.2010.05838.x
- Bogardus C : Missing heritability and GWAS utility. Obesity 2009 vol. 17 (2) pp. 209-10.
- Bouatia-Naji N et al. : A variant near MTNR1B is associated with increased fasting plasma glucose levels and type 2 diabetes risk Nature Genetics 41, 89 - 94 (2008)
- Boutin P et al. : GAD2 on chromosome 10p12 is a candidate gene for human obesity. PLoS Biol. 2003 Dec;1(3):E68. Epub 2003 Nov 3
- Bowden DW et al. : Molecular basis of a linkage peak: exome sequencing and family-based analysis identify a rare genetic variant in the ADIPOQ gene in the IRAS Family Study. Hum Mol Genet. 2010 Oct 15;19(20):4112-20. Epub 2010 Aug 5

- C Biéumont : From genotype to phenotype. What do epigenetics and epigenomics tell us? *Heredity* 2010 105 pp 1-3
- Chen M, Cho J, Zhao H. : Incorporating biological pathways via a Markov random field model in genome-wide association studies. *PLoS Genet.* 2011 Apr;7(4):e1001353. Epub 2011 Apr 7
- Cheyssac C. et al. : EIF4A2 is a positional candidate gene at the 3q27 locus linked to type 2 diabetes in French families. *Diabetes.* 2006 Apr;55(4):1171-6.
- Choi M et al. : Genetic diagnosis by whole exome capture and massively parallel DNA sequencing *Proc Natl Acad Sci U S A.* 2009 Nov 10;106(45):19096-101. Epub 2009 Oct 27
- Cordell HJ : Detecting gene-gene interactions that underlie human diseases *Nature Reviews Genetics* 2009 vol. 10 (6) pp. 392-404
- D. Chevenne, F. Trivin : Le diabète sucré : propositions de nouvelles normes de diagnostic et de classification. *Annales de Biologie Clinique.* Volume 56, Numéro 4, 463-70, Juillet - Août 1998, *Pratique quotidienne,*
- Deloukas P. et al. : The DNA sequence and comparative analysis of human chromosome20. *Nature.* 2001-27;414(6866):865-71
- Delplanque J et al. : Linkage and association studies between the proopiomelanocortin (POMC) gene and obesity in caucasian families. *Diabetologia.* 2000 Dec;43(12):1554-7
- Delplanque J et al. : Mutation screening of the urocortin gene: identification of new single nucleotide polymorphisms and association studies with obesity in French Caucasians. *J Clin Endocrinol Metab.* 2002 Feb;87(2):867-9
- Demenais, F., Martinez, M., and Lathrop, M. (1996) : Méthodes statistiques pour identifier les gènes dans les maladies multifactorielles. *Ann Instit Pasteur,* 7(1):3-12.
- Drumm ML et al. : Genetic modifiers of lung disease in cystic fibrosis. *N Engl J Med* 2005 353 pp 1509-1511; Arkwright PD et al. TGF beta 1 genotype and accelerated decline in lung function of patients with cystic fibrosis. *Thorax* 2000 55 446
- Economou M et al. : Differential effects of NOD2 variants on Crohn's disease risk and phenotype in diverse populations: a metaanalysis. *Am J Gastroenterol.* 2004 Dec;99(12):2393-404
- Eichler EE et al. : Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet.* 2010 Jun;11(6):446-50
- Eichler EE et al. : Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet.* 2010 Jun;11(6):446-50

- Eichler EE et al. : Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet.* 2010 Jun;11(6):446-50
- Eichler EE et al. : Missing heritability and strategies for finding the underlying causes of complex disease. *Nat Rev Genet.* 2010 Jun;11(6):446-50
- Evans DM et al. : A comparison of linkage disequilibrium patterns and estimated population recombination rates across multiple populations. *Am J Hum Genet.* 2005 Apr;76(4):681-7. Epub 2005 Feb 17
- Feingold, J. (2005) : Maladies multifactorielles: un cauchemar pour le généticien. *Med Sci*, 11(21):927-33.
- Fisman PM et al. : A robust method for the detection of linkage in familial disease. *Am J Hum Genet.* 1978 May;30(3):308-21
- Franke et al. : Genome-wide meta-analysis increases to 71 the number of confirmed Crohn's disease susceptibility loci. *Nat Genet.* 2010 Dec;42(12):1118-25
- Galli J et al. : Genetic analysis of non-insulin dependent diabetes mellitus in the GK rat. *Nat Genet.* 1996 Jan;12(1):31-7.
- Gardner LI Jr, Stern MP, Haffner SM, Gaskill SP, Hazuda HP, Relethford JH, Eifler CW : Prevalence of diabetes in Mexican Americans. Relationship to percent of gene pool derived from native American sources
- Gertz J et al. : Analysis of DNA Methylation in a Three-Generation Family Reveals Widespread Genetic Influence on Epigenetic Regulation. *Plos Genet* 2011 7 e1002228.
- Guérardel A et al. : Analysis of sequence variability in the CART gene in relation to obesity in a Caucasian population. *BMC Genet.* 2005 Apr 11;6:19
- Hager J et al. : A genome-wide scan for human obesity genes reveals a major susceptibility locus on chromosome 10. *Nat Genet.* 1998 Nov;20(3):304-8
- HT Bjornsson : Intra-individual change over time in DNA methylation with familial clustering. *JAMA* 2008 vol. 299 (24) pp. 2877-83
- Hugot, J. P. et al. : Mutation screening in the CD19 and CD43 (sialophorin) genes in Crohn Disease patients. *Gastroenterology* 116, A740 1999
- Hugot JP et al. : Mapping of a susceptibility locus for Crohn's disease on chromosome 16. *Nature.* 1996 Feb 29;379(6568):821-3
- Kathiresan S et al. : Common variants at 30 loci contribute to polygenic dyslipidemia *Nature Genetics* 41, 56 - 65 (2008)

- Kong A et al. : Parental origin of sequence variants associated with complex diseases. *Nature*. 2009 Dec 17;462(7275):868-74
- Kong A et al. : Sequence variants in the RNF212 gene associate with genome-wide recombination rate. *Science*. 2008 Mar 7;319(5868):1398-401. Epub 2008 Jan 31
- Lango H Weedon MN : What will whole genome searches for susceptibility genes for common complex disease offer to clinical practice? *J Intern Med* 2008 vol. 263 (1) pp. 16-27. Manolio TA Genomewide association studies and assessment of the risk of disease *N Engl J Med* 2010 vol. 363 (2) pp. 166-76.
- Liu et al. : Genome-wide interaction-based association analysis identified multiple new susceptibility Loci for common diseases., *PLoS Genet*. 2011 Mar;7(3):e1001338. Epub 2011 Mar 17
- Lyssenko V et al. : Common variant in MTNR1B associated with increased risk of type 2 diabetes and impaired early insulin secretion *Nature Genetics* 41, 82 - 88 (2008)
- Mc Clellan J : Genetic heterogeneity in human disease. *Cell* 2010 vol. 141 (2) pp. 210-7.
- O'Donovan et al. : Phenotypic variations on the theme of CNVs. *Nat Genet* 2008 vol. 40 (12) pp. 1392-3 Wellcome Trust Case Control Consortium Genome-wide association study of CNVs in 16,000 cases of eight common diseases and 3,000 shared controls. *Nature* 2010 464 pp 713-720
- Olavesen, M. G. et al. : Analysis of single-nucleotide polymorphisms in the interleukin-4 receptor gene for association with inflammatory bowel disease. *Immunogenetics* 51, 1-7 2000
- Oostenbrug LE et al. : CARD15 in inflammatory bowel disease and Crohn's disease phenotypes: an association study and pooled analysis. *Dig Liver Dis*. 2006 Nov;38(11):834-45. Epub 2006 Aug 21
- Plomin R et al.. : Common disorders are quantitative traits. *Nature Reviews Genetics* 10, 872 (2009). doi:10.1038/nrg2670.
- Pritchard JK, Cox NJ : The allelic architecture of human disease genes: common disease-common variant...or not? *Hum Mol Genet*. 2002 Oct 1;11(20):2417-23
- Prokopenko I et al. : Variants in MTNR1B influence fasting glucose levels *Nature Genetics* 41, 77 - 81 (2008)
- Roberts A, Thompson JS. : Inbred mice and their hybrids as an animal model for atherosclerosis research. *Adv Exp Med Biol*. 1976;67(00):313-327.

- Sabatti C et al. : Genome-wide association analysis of metabolic traits in a birth cohort from a founder population *Nature Genetics* 41, 35 - 46 (2008)
- Salisbury BA et al. : SNP and haplotype variation in the human genome. *Mutat Res.* 2003 May 15;526(1-2):53-61
- Siddiq A. et al. : A synonymous coding polymorphism in the alpha2-Heremans-schmid glycoprotein gene is associated with type 2 diabetes in French Caucasians. *Diabetes.* 2005 Aug;54(8):2477-81.
- Styrkarsdottir U et al. : Linkage of osteoporosis to chromosome 20p12 and association to BMP2. *PLoS Biol.* 2003 Dec;1(3):E69. Epub 2003 Nov 3
- Suarez BK et al : The generalized sib pair IBD distribution: its use in the detection of linkage. *Ann Hum Genet.* 1978 Jul;42(1):87-94
- Sudmant PH : Diversity of human copy number variation and multicopy genes. *Science* 2010 vol. 330 (6004) pp. 641-6.
- Tishkoff S et al. : Ten years of genetics and genomics: what have we achieved and where are we heading? *Nat Rev Genet* 2010 vol. 11 (10) pp. 723-33
- TW Kurtz : The Zucker fatty rat as a genetic model of obesity and hypertension, *Hypertension*, vol. 13, no 6, 1989, p. 896-901.
- Vasseur F. et al. : Single-nucleotide polymorphism haplotypes in the both proximal promoter and exon 3 of the APM1 gene modulate adipocyte-secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes in French Caucasians. *Hum Mol Genet.* 2002 Oct 1;11(21):2607-14
- Vasseur F et al. : Single-nucleotide polymorphism haplotypes in the both proximal promoter and exon 3 of the APM1 gene modulate adipocyte-secreted adiponectin hormone levels and contribute to the genetic risk for type 2 diabetes in French Caucasians. *Hum Mol Genet.* 2002 Oct 1;11(21):2607-14
- Vionnet N et al. : Genomewide search for type 2 diabetes-susceptibility genes in French whites: evidence for a novel susceptibility locus for early-onset diabetes on chromosome 3q27-qter and independent replication of a type 2-diabetes locus on chromosome 1q21-q24. *Am J Hum Genet.* 2000 Dec;67(6):1470-80
- Visscher PM et al. : Heritability in the genomics era-concepts and misconceptions. *Nat Rev Genet.* 2008 Apr;9(4):255-66. Epub 2008 Mar 4
- Williams A et al. : Interchromosomal association and gene regulation in trans. *Trends Genet* 2010 vol. 26 (4) pp. 188-97.