L'ANALYSE DE LIAISON POUR LES MALADIES COMPLEXES: UNE NOUVELLE VIE POUR UNE VIEILLE METHODE

Ahmed Rebaï[1]

1Centre de Biotechnologie de Sfax
B.P. 'K', 3038 Sfax, Tunisie

RESUME

Dans cet article, je présente les principales approches utilisées pour la cartographie de gènes des maladies génétiques complexes par analyse de liaison avec l'aide des marqueurs moléculaires. La première partie se rapporte à la méthode classique du Lod score; après avoir décrit son principe générale, je donne une revue des améliorations apportées à cette méthode, pour tenir compte de facteurs de complications, comme la pénétrance incomplète, l'hétérogénéité génétique ou le mode d'hérédité incertain.
La deuxième partie décrit les trois principales approches de l'analyse de liaison non-paramétriques qui se basent sur des méthodes pour lesquelles on n'a pas besoin de spécifier de modèle génétique de la maladie (mode d'hérédité, fréquences alléliques, ..) et qui semblent a priori mieux adaptées à l'analyse des maladies complexes. Une comparaison de ces trois méthodes à partir des données bibliographiques est également présentée.
Dans la troisième partie et la discussion je m'intéresse à la comparaison entre méthode de Lod score et méthodes non paramétriques en montrant les avantages et les inconvénients de chacune au vu des résultats de recherche les plus récents. J'en déduit que malgré sa mauvaise adaptation apparente et présumée à l'analyse des maladies génétiques complexes, la méthode du LOD score peut concurrencer les méthodes non-paramétriques au niveau de sa puissance et de sa précision et redevient aujourd'hui une approche de choix pour l'analyse de ce type de maladie.
Mots-cles: *Analyse de liaison, maladie complexe, LOD-score, non-paramétrique, modèle génétique*

ABSTRACT

In this paper, I present the main approaches used in gene mapping of
Complex human diseases by linkage analysis based on molecular markers. The First section describes the traditionnal Lod-score analysis and gives a review of different improvements and refinments of this approach, which have been porposed in order to take into account complicating factors such as incomplete penetrance, genetic heterogeneity and unknown mode of inheritance. The second section describes the three main approches of non-parametric linkage analysis, for which there is no need to specify a genetic model (mode of inheritance, allele frequencies, ..). A comparison between these three methods, which may seem to be more appropriate for complex diseases, is given based on the mot recent published studies.In the third section and discussion, I compare the LOD-score and non-parametric methods by stressing the advantages and drawbacks of each approach as found in recent publications. I deduce that, in spite of its apparent and assumed inappropriateness to the analysis of complex diseases, the Lod-score method is still very usefull and could provide,in some circumstances, more power and precision than model-free methods.
Key-words: *Linkage analysis, complex disease, LOD-score, non-parametric, genetic model*

INTRODUCTION

L'analyse de liaison consiste à étudier la co-ségrégation de deux ou plusieurs loci (marqueurs, gènes ou caractères) dans une famille pour estimer les taux de recombinaison entre ces loci (θ) et tester leur liaison génétique ($\theta < 0.5$). Il y a plus de 50 ans, Haldane et Smith (1947) [1] puis Barnard (1949) [2] et Morton (1955) [3] ont construit les bases de la méthode du LOD score en proposant le test du même nom pour tester la liaison entre deux loci. Depuis, cette méthode a connu des améliorations méthodologiques et algorithmiques pour tenir compte des différents facteurs perturbateurs (pénétrance incomplète, hétérogénéité génétique, phénocopie,..) ou mal connus (mode d'hérédité complexe) concrétisées par le développement de logiciels performants rendant ce type d'analyse quasi-automatique (voire Ott, 1999 pour une synthèse) [4].

Le succès de la méthode du LOD score est indéniable pour les maladies monogéniques comme en témoigne le grand nombre de gènes qui ont été identifiés et car-

tographiés grâce à cette méthode. Cependant, certaines limites de l'analyse LOD score sont apparues lorsqu'on a essayé, sans beaucoup de succès, de l'appliquer à des caractères complexes. Un caractère complexe peut être défini comme un caractère pour lequel le mode d'hérédité (par la suite noté MH) est inconnu et qui montre des effets familiaux et génétiques modérés à élevés. Ces maladies correspondent à des désordres communs qui sont soit polygéniques (plusieurs gènes) soit multifactoriels (plusieurs gènes qui interagissent avec le milieu). Les exemples les plus connus sont le diabète, les maladies cardiovasculaires, les désordres psychiatriques et la maladie d'Alzheimer. L'effet des loci multiples (appelé loci de susceptibilité) sur le risque et l'expression du phénotype est généralement expliqué par des modèles génétiques additifs ou multiplicatifs, ce qui rend l'analyse de liaison par LOD score assez délicate.

Pour surmonter certains défauts de l'analyse Lod score et surtout le besoin de spécifier un MH, des méthodes modèle-indépendantes dites abusivement non-paramétriques[5] ont été proposées et largement utilisées. Cependant, de nombreuses études récentes basées sur des simulations extensives tendent à remettre la méthode LOD score sur le devant de la scène en montrant sa puissance, même pour l'analyse des maladies complexes, comparativement aux méthodes alternatives d'analyse de liaison.

L'objectif de cet article est de donner un aperçu sur les principales méthodes d'analyse de liaison en soulignant les forces et les limites de chacune. La comparaison entre ces méthodes, basée sur la bibliographie récente, sera présentée tout au long de l'article et fera l'objet de la discussion. J'espère que la synthèse présentée ici fournira aux utilisateurs non-spécialistes de ces outils d'analyse génétique les éléments nécessaires pour faire face au problème parfois délicat du choix d'une ou plusieurs approches pour analyser leur données.

LA METHODE DU LOD SCORE

L'analyse Lod score est une méthode paramétrique basée sur le principe de vraisemblance pour estimer le taux de recombinaison θ et tester la liaison génétique entre deux loci (analyse deux-points). Les caractères pour ces loci peuvent être qualitatifs (état d'affection, allèles d'un marqueur) ou quantitatifs (maladie, marqueurs) et la méthode peut être étendue à plusieurs loci simultanément (analyse multipoint). Cette méthode a été développé à l'origine pour les caractères ayant un MH et des fréquences alléliques connus.

Quand le mode d'hérédité est inconnu, comme pour les maladies complexes, on peut toujours calculer le Lod score mais dans ce cas celui-ci dépend de plusieurs paramètres que l'on doit spécifier d'avance: le MH (dominant/recessif), le taux de recombinaison, la fréquence des allèles du caractère, les pénétrances pour chaque phénotype de la maladie, les fréquences des allèles des marqueurs. Ces paramètres constituent ce qu'on appelle le modèle génétique. Donc un test paramétrique de liaison est un test non seulement du taux de recombinaison mais de tous les paramètres du modèle génétique.

1- ANALYSE DEUX-POINTS

Dans cette analyse, la vraisemblance (L) du pedigree est calculée pour une valeur donnée *x* du taux de recombinaison entre le locus de maladie et un locus marqueur θ=*x* et pour θ=0.5 en se basant sur le nombre d'individus recombinants et non-recombinants. Le test du Lod score pour cette position *x* est alors défini comme le log décimal du rapport de la vraisemblance des données sous l'hypothèse de liaison et de non liaison:

$LOD(x)=Z(x)=\log_{10}[L(pedigree/\theta=x)/L(pedigree/\theta=1/2)]$.

Le plus souvent il n'est pas possible de compter les individus recombinants mais du moment où l'on spécifie un modèle génétique, différents algorithmes implémentés dans les logiciels disponibles (voire table 1) peuvent être utilisés pour estime ces recombinaisons et calculer les Lod score sur des pedigrees pouvant avoir des données phénotypiques et génotypiques manquantes (individus décédés ou ADN indisponible) et éventuellement avec des boucles de mariage et/ou de consanguinité. En pratique, les Lod score sont calculés sur chaque famille indépendante et ensuite additionnés pour obtenir un lod score global. Dans le cas d'un très large pedigree avec de nombreuses boucles de consanguinité les méthodes exactes de calcul (implémentées dans LINKAGE/FASTLINK et GENEHUNTER) peuvent devenir impraticables (temps de calcul infini). On peut alors soit découper le pedigree en plusieurs familles quasi-indépendantes soit utiliser des méthodes 'approximatives' du type Monte Carlo comme celle implémentée dans le logiciel BLOCK (tableau 1).

Les Lod scores (notés Z) sont par convention rapportés pour θ=0, 0.001, 0.05, 0.1, 0.2, 0.3, et 0.4. La valeur de q relative au Z maximal (Z_{max}) correspond à l'estimation du maximum de vraisemblance du taux de recombinaison.

A signaler que les Z_{max} ne s'additionnent pas entre familles.

Une valeur seuil de lod score de 3 est généralement utilisée pour déclarer la liaison pour une analyse deux points avec MH connu. Cette valeur critique correspond à une probabilité d'erreur de type I (déclarer vraie une fausse liaison), appelé aussi p-value, de 10^{-4}. Pour une analyse de plusieurs marqueurs (un balayage du génome) on recommande une valeur de 3.3, qui correspond à une p-value de 5.10^{-5}. Une discussion approfondie du seuil approprié selon le type d'analyse souhaité peut être trouvée dans Lander et

Kruglyak (1995)[6], Elston (1997)[5], Kruglyak et Daly (1998)[7], Morton (1998)[8] et Zhao LP et al. (1999)[9]

2 - EFFETS DU MODÈLE GÉNÉTIQUE SUR L'ANALYSE DEUX-POINTS

Les Lod score sont obtenus en utilisant des valeurs données des paramètres. Pour les caractères complexes, où des facteurs comme la pénétrance incomplète, l'hétérogénéité génétique, la phénocopie et le MH peuvent intervenir, le modèle supposé est très vraisemblablement incorrecte, du moins en partie. Ceci peut entrainer une augmentation du taux de faux-positifs (déclarer vraie une liaison qui est fausse) et/ou de faux-négatifs (ne pas déclarer une vraie liaison).

L'impact de la mauvaise spécification des paramètres génétiques sur le Lod score dépend de plusieurs facteurs: le MH, l'étendue de l'erreur et la structure du pedigree. Les études théoriques par simulation tendent à montrer que la puissance du test de liaison (probabilité de déclarer correctement une liaison) dépend, de manière grossière, fortement du degré de dominance (modèle dominant contre récessif), moyennement de la fréquence des allèles aux marqueurs, légèrement de la pénétrance et faiblement de la fréquence des allèles au locus de la maladie. Cependant, l'estimation du taux de recombinaison peut être fortement affectée par une erreur dans chacun de ces paramètres 10.Une étude détaillée de l'influence de ces facteurs peut être trouvée dans Ott (1999, chapitre 13) [4]

Pour ce qui est du MH, la dominance a un impact important sur le Lod score en cas de liaison surtout si un MH dominant est déclaré comme récessif. Les fréquences alléliques étant utilisées uniquement pour reconstruire les génotypes manquants des individus fondateurs (dont les parents ne sont pas présent dans le pedigree) ou pour estimer les probabilités de transmission des allèles, leur effet sur le Lod sera d'autant plus grand qu'il y a d'individus non typés. Quant à la pénétrance, elle n'a pas d'effet sur le LOD (mais uniquement sur l'estimation de θ) du moment où le MH est bien spécifié. Globalement, on peut dire que l'analyse deux points est très robuste par rapport à un faux modèle du moment où la dominance est correctement spécifiée.

3 - L'HÉTÉROGÉNÉITÉ GÉNÉTIQUE

On parle d'hétérogénéité génétique lorsqu'il y a de grands écarts entre les LOD des différentes familles analysées, suggérant que seule une partie des familles montrerait une liaison entre le marqueur et la maladie. Pour la tester, on utilise surtout le 'Admixture test' ou A-test [11] défini par $2\text{Log}[L(\alpha,\theta_1)/L(\alpha=1,\theta_1)]$ où α ($\alpha<1$) est la proportion de familles liées c'est à dire celles ayant un $\theta_1<0.5$ (les $(1-\alpha)$ familles restantes ayant un $\theta_2=0.5$). Ce test suit une loi x^2 à 1 degré de liberté (ddl). Munis des estimations de a et de θ1 on peut calculer pour chaque famille la probabilité conditionnelle qu'elle appartienne au type lié.Le test de liaison en présence d'hétérogénéité (appelé HLOD)peut être défini par $\text{HLOD}(\theta)=Z(\theta)=\log10[L(\alpha,\theta)/L(\alpha=1,\theta=1/2)]$.

Il faut signaler que l'estimation de a change quand θ change, donc ce test peut être interprété comme une mesure de l'évidence de la liaison et/ou d'hétérogénéité. Le seuil de ce test n'est pas facile à déterminer car sa distribution n'est pas régulière (entre un x^2 à 1 ddl et x^2 à 2 ddl). Le seuil approprié se situerait donc quelque part entre 3 et 4 (ce dernier étant un seuil conservateur). Ces tests ainsi que de nombreux autres tests relatifs à l'hétérogénéité peuvent être calculé par le logiciel HOMOG 4.

Falk (1997) [12] a montré par simulation que la prise en compte de l'hétérogénéité (test HLOD) donne une meilleure estimation du taux de recombinaison entre gène de maladie et marqueur mais que la liaison peut être détectée même en utilisant le test LOD classique. Dans une étude récente, Whittemore et Halpern (2001) [13] critiquent l'approche HLOD en montrant qu'elle peut donner des résultats biaisés si le taux de phénocopie est incorrectement spécifié et que l'estimation qu'elle fournie de la proportion de familles liées (α) dépend de la taille et de la structure des familles utilisées et peut donc être très différente d'une étude à une autre (même au sein d'une même population).

4 - L'ANALYSE LOD SCORE MULTIPOINT

C'est une extension de l'analyse deux-points, où toute position sur une carte de marqueurs est testée pour la liaison avec la maladie. Le plus souvent le test calculé est le 'Location score', défini comme: $Z(x)=2\text{Log}[L(\theta=x)/L(\theta=\infty)]$, où x est une position donnée sur la carte. Il y a deux avantages à l'analyse multipoint; le premier c'est qu'elle permet de récupérer de l'information génotypique pour un marqueur non informatif à partir de l'haplotype donnant ainsi des résultats qui sont moins sensibles au génotypes manquants ou non informatifs que ceux de l'analyse deux-points. Le second avantage est qu'elle fournit une localisation précise du gène permettant ainsi de mieux définir la région candidate pour des études ultérieures. De ce fait, les résultats de l'analyse Lod score multipoint dépendent de la qualité de la carte génétique (ordre et distances des marqueurs) et toute erreur au niveau de la carte peut avoir des graves conséquences. Göring et Terwilliger (2000c)[14] proposent une approche Lod score qui traite les paramètres des marqueurs (fréquences alléliques, carte génétique,..) comme des paramètres de nuisance mettant ainsi à disposition une méthode plus puissante et plus robuste aux erreurs.

Il convient également d'être attentif à l'interprétation des résultats de l'analyse multipoint en présence d'hétérogénéité (ou en cas de pénétrance incomplète). En effet, pour le test deux-points l'effet de l'hétérogénéité est en grande partie absorbé par une sur-estimation des taux de recombinaison alors qu'en analyse multipoint (la carte étant fixée), elle résulte souvent en une exclusion

incorrecte de la région entre marqueurs, surtout si ceux-ci sont assez éloignés. Ce problème est encore amplifié pour les caractères complexes à cause de la mauvaise spécification du modèle génétique, qui peut entraîner la classification d'individus recombinants en non-recombinants ou l'inverse.

Quand on a une carte de marqueurs très dense avec des intervalles de moins de 2 cM, l'analyse multipoint devient lourde et inutile. Une alternative est alors d'utiliser l'approche polylocus ou pseudo-marqueur qui combine les données de plusieurs marqueurs en un seul superlocus. Les génotypes en ce superlocus sont alloués de manière à rendre chaque méïose informative pour le locus de la maladie [15, 16]. Cette approche est la plus puissante parmi toutes les méthodes modèle-dépendantes [17].

5 - PROBLÈMES ET LIMITES: EXTENSIONS DE L'ANALYSE LOD SCORE

Le problème dans l'utilisation d'une méthode modèle-dépendante dans des situations où le modèle est inconnu (maladies complexes) n'est pas l'augmentation de la probabilité des faux-positifs mais plutôt celle des faux-négatifs (déclarer non significative une liaison qui existe) [10, 18].Lorsqu'on applique l'analyse de Lod score avec un modèle génétique incorrecte on parle de WROD score (Wrong lod score). Cependant, vu la robustesse de l'approche LOD score, sa validité n'est pas forcément mise en doute [4].

La stratégie classique pour sélectionner un modèle génétique pour l'analyse Lod score se base sur les résultats d'une analyse de ségrégation complexe qui fournit des estimations pour les fréquences alléliques, le degré de dominance et les pénétrances. Malgré quelques succès comme pour le cancer familial du sein [19], une telle stratégie n'est généralement pas efficace pour les caractères complexes car elle nécessite la collecte d'un très grand nombre de familles et que l'utilisation des paramètres issus d'une analyse de ségrégation peut entraîner un biais important dans le test Lod [20, 21].

Une façon de contourner les problèmes liés aux maladies complexes est de définir des classes homogènes de phénotypes (*'liability' classes*) selon l'age d'atteinte, la sévérité de la maladie ou d'autres variables liées au risque de la maladie [22]. Toutefois, une telle stratification n'est bénéfique dans une analyse de liaison que si elle a un très grand effet [23]. Une autre façon de réduire l'hétérogénéité des familles est de faire une analyse Lod score en utilisant uniquement les individus affectés (*affecteds-only analysis)* ce qui permet d'éviter les hypothèses concernant les non affectés, mais cette approche n'est pas optimale.

Afin de tenir compte de l'imprécision du modèle génétique dans le cas des maladies complexes, plusieurs extensions de la méthode Lod score ont été proposées. Une approche intéressante consiste à maximiser le Lod score à la fois par rapport à θ et à plusieurs modèles génétiques, ϕ. Cette approche qui défini un test, appelé MOD score ou MMLS (*Maximized Maximum Lod Score*) par $M(x)=\log_{10}[L(\theta=x,\phi)/L(\theta=0.5,\phi)]$ est valide dans certaines conditions (Hodge et Elston, 1994). Elle peut permettre à la fois de tester la liaison et de déterminer le modèle génétique approprié. Cependant, vu la complexité de sa mise en œuvre (choix des paramètres à maximiser, temps de calcul, seuils de signification,..) plusieurs simplifications de l'apporche Mod ont été proposées. La plus intuitive, appelé MMLS-C consiste à effectuer deux analyses LOD score, l'une sous le modèle dominant, l'autre sous le modèle récessif (pour une pénétrance arbitraire de 0.5 et 0 phénocopie par exemple) et à prendre comme statistique de test le maximum des deux LOD score ainsi obtenus [24]. Ces auteurs ont montré par simulation que cette approche appliquée avec un seuil corrigé de 3.3 au lieu de 3 [25] constitue un test puissant et efficace de liaison pour les caractères complexes. Leurs travaux montrent que le facteur critique dans une analyse LOD score est le MH au locus lié et non pour la maladie dans sa globalité [26, 21] ce qui explique la puissance de l'approche proposée (un locus dominant sera détecté sous le modèle dominant et vice-versa).

Parmi les nombreux autres tests du type Mod score qui ont été proposés on peut citer:

- le test MFLOD pour *Model-Free* LOD [27] implémenté dans le programme MFLINK. Dans ce test, le numérateur et le dénominateur du Lod score sont maximisés par rapport à un ensemble réduit de modèles génétiques allant du mendélien dominant à l'effet nul et de l'effet nul au mendélien récessif.
- Les tests MLOD et MALOD pour *Maximized (Admixture)* LOD [28]. Dans le test MLOD, on maximise par rapport à un seul paramètre, la pénétrance de l'hétérozygote *f1*, pour une prévalence donnée, alors que pour le MALOD on maximise par rapport à *f1* et a (proportion de familles liées). Sham et al. (2000) [28] montrent que pour les petits pedigrees, le test MLOD est supérieur au MALOD et MFLOD même en présence d'une forte hétérogénéité.

L'ANALYSE DE LIAISON MODELE INDEDANTE

Les méthodes d'analyse de liaison modèle-indépendantes (abusivement appelé 'non-paramétriques') se basent sur l'idée que des apparentés qui se ressemblent au niveau du phénotype d'intérêt seront similaires en un locus marqueur, en partageant des allèles identiques par descendance (IBD: *Identity By Descent*), seulement si le phénotype est lié au marqueur. Ces méthodes utilisent pour la plupart la proportion d'allèles IBD aux marqueurs (2 allèles sont dits *ibd* s'ils sont des copies directes d'un même allèle ancêtre). Pour les apparentés où l'état *ibd* ne peut être déterminé sur la

base des données disponibles on peut soit ignorer ces individus, soit utiliser l'identité par état (IBS: *Identity By State*, deux allèles sont identiques par état s'ils sont simplement le même allèle mais pouvant être IBD ou non) ou estimer la probabilité de l'ibd en utilisant la fréquence allélique du marqueur dans la population [29].

1 - MÉTHODE D'ANALYSES DES PAIRES DE GERMAINS AFFECTÉS OU ANALYSE ASP

Malgré son utilisation depuis 1935 par Penrose [30] ce type de familles, n'a connu un grand succès que depuis une dizaine d'années et semble progressivement devenir le matériel de choix pour la cartographie des maladies complexes notamment pour les caractères quantitatifs. Pour les caractères dichotomiques, deux tests simples ont été largement utilisés [31] Ils consistent à comparer les proportions observée et attendue de paires de germains d'un certain type IBD. En effet, deux frères peuvent partager, 0, 1 ou 2 allèles IBD pour un locus donné avec les proportions attendues $1/4$, $1/2$ et $1/4$. Si l'on considère plusieurs paires, on peut calculer la proportion de paires dans chaque catégorie (notés Z_0, Z_1 et Z_2, respectivement). Un premier test appelé 'test de proportion' consiste à comparer Z_2 par rapport à $1/4$. Le second, appelé 'test de moyenne' compare la proportion de paires ayant au moins un allèle IBD (Z_1+2Z_2) par rapport à 1. La comparaison se fait par un test classique de comparaison de moyenne [31].

Quand l'état IBD ne peut être déterminé (génotype des parents indisponibles) on peut utiliser le génotypes d'autres apparentés pour estimer les fréquences des classes d'état *ibd* par des méthodes de maximum de vraisemblance et tester la liaison. L'une de ces méthodes, dite MLS (*Maximum Lod Score*) consiste à comparer la probabilité des données génotypiques observées par rapport à celle attendues sous l'hypothèse de non liaison produisant ainsi l'équivalent d'un LOD score modèle-indépendant [32]:

$MLS=\log10[L(Z0, Z1, Z2)/L(1/4, 1/2, 1/4)]$.

Holmans (1993) [33] a montré que pour être compatible avec l'hypothèse d'équilibre Hardy-Weinberg pour le locus de la maladie la maximisation de la vraisemblance doit se faire sous la condition $2Z_0 \leq Z_1 \leq 1/2$ (appelé triangle génétiquement possible ou plus couramment contrainte du triangle) ce qui permet aussi d'augmenter la puissance du test MLS. Dizier et al. (2000) [34] ont toutefois montré que l'utilisation de cette contrainte peut résulter en une grande perte de puissance en présence d'hétérogénéité c'est à dire si la pénétrance dépend d'un facteur (sévérité de la maladie, présence/absence d'un facteur de risque) qui diffère entre les deux germains d'une paire. Ils proposent alors d'utiliser le test MLS sans la contrainte du triangle et si celui-ci est significatif (liaison) d'utiliser un test (appelé TTS) pour tester l'hétérogénéité en présence de liaison. Différentes études ont montré que le test MLS garde une bonne puissance sous une large gamme d'hypothèses [35, 36].

Une version multipoint de ce test a été également développée (Hauser et al., 1996) et est implémentée dans le logiciel ASPEX (tableau 1, Voir page suivante).

Feingold et Siegmund (1997) [38] on montré qu'un test de moyenne modifié, basé sur la statistique $Z_1+1/4Z_2$ était plus puissant que le test MLS quand le MH est intermédiaire entre dominant et récessif mais moins efficace quand celui-ci est proche des extrêmes.

2 - LA MÉTHODE D'ANALYSE DE MEMBRES DE PEDIGREE AFFECTÉS OU ANALYSE APM

Cette méthode proposée par Weeks et Lange (1988) [39] se base sur l'état IBS entre paires d'apparentés de tout type permettant ainsi de considérer les situations où le génotypage des individus du pedigree est incomplet. Cette méthode d'analyse dépend de trois paramètres: les génotypes des individus affectés, les fréquences alléliques des marqueurs et les relations de parenté entre individus. La statistique de test, *Z*, pour une famille est la somme des Z_{ij} entre toutes les paires (*i*,*j*) d'individus affectés qui est fonction de l'état IBS entre les allèles de *i* et *j* et de la fréquence de ces allèles. Le test final *T* et la somme pondérée des *Z* centrés et réduits pour toutes les familles où la pondération de chaque famille est une fonction linéaire de la racine carrée du nombre d'individus affectés et génotypés dans cette famille. Quand le nombre de familles est élevé (plus de 30), le test *T* suit une distribution normale ce qui permet de calculer les *p-value*s sur la base de cette approximation. La méthode APM possède deux avantages: (1) elle utilise toutes sortes de paires d'apparentés et (2) elle est rapide même pour des pedigrees de structure très complexe avec un grand nombre d'individus non génotypés et des boucles de mariage/consanguinité. Une version multipoint de la méthode a également été développée [40]. Elle permet de mieux utiliser l'information de marqueurs étroitement liés pour définir un haplotype et augmenter ainsi la similarité entre états IBS et IBD.

Cependant, comme toutes les méthodes qui se basent sur l'IBS, son inconvénient majeur est le mauvais contrôle de l'erreur de type I (probabilité de déclarer une liaison qui n'existe pas) qui entraîne une augmentation du taux de faux positifs, même si celui-ci peut être réduit en utilisant des marqueurs à grand nombre d'allèles et une analyse multipoint. Deux autres inconvénients majeurs sont à signaler: (1) sensibilité aux fréquences alléliques qui doivent alors être estimées sur un échantillon indépendant d'individus non apparentés et (2) l'approximation normale n'est valide que pour un grand nombre de familles sinon il faudra utiliser des méthodes de simulation pour estimer les *p-values*. Ainsi, l'analyse APM (implémentée dans le logiciel du même nom, APM) doit être réservée aux situations où l'on a des pedigrees à structure très complexe et pour lesquels une analyse LOD score n'est pas envisageable et dans le cas où l'on dispose d'un

Tableau I*- Principales méthodes d'analyses de liaison et les logiciels qui les utilisent*

Méthode	Logiciels	Type de familles concernés	S.E.**	Référence
	Méthodes modèle-dépendante			
Lod score deux points uniquement	LIPED	Pedigree t. s.*	Dos, Unix	Ott, 1974
	BLOCK	Pedigree t. s. avec un grand nombre de boucles (<100) de consanguinité et/ou mariage	Dos, Unix	Jensen et Kong, 1995
Lod score Deux-point et multipoint	LINKAGE	Pedigree t.s. avec plus 10 boucles de consanguinité et/ou mariage	Dos, Unix	Lathrop et al., 1984
	FASTLINK	Idem	Idem	Cottingham et al., 1993
	VITESSE	Pedigree t.s. sans boucles	Idem	O'Connel et Weeks, 1995
Lod score multipoint uniquement	GENEHUNTER	Pedigree t.s. avec moins de 12 individus non fondateurs	Unix	Kruglyak et al. 1996
	ALLEGRO	Pedigree t. s. avec moins de 20 non fondateurs	Unix	Gudbjartsen et al., 2000
	HOMOZ	Pedigree t. s. avec boucles de consanguinité	Unix	Kruglyak et al., 1995
	Méthodes modèle-dépendante			
ASP: test de moyenne et/ou MLS multipoint	ASPEX	familles nucléaires	Unix	Hinds et Risch, 1996
	MAPMAKER/SIBS	familles nucléaires	Unix	Kruglyak et Lander, 1995
	SIBPAIR	familles nucléaires	Unix	Terwilliger (1996)
	GENEHUNTER	familles nucléaires	Unix	Kruglyak et al., 1996
	SPLINK	familles nucléaires	Unix	Holmans, 1993; Holmans et Clayton, 1995
APM	APM	Pedigree t. s.	Unix	Weeks et Lange, 1988
NPL	GENEHUNTER	Pedigree t. s.	Unix	Kruglyak et al., 1996
	ALLEGRO	Pedigree t. s.	Unix	Gudbjartsen et al., 2000

*pedigree t.s.: pedigree de toute sorte c'est à dire quelque soit la structure
** S.E. système d'exploitation sous lequel fonctionne le logiciel

grand nombre d'apparentés autres que des germains. Autant que possible, une analyse multipoint est toujours préférable.

3 - L'ANALYSE DE LIAISON NON-PARAMÉRIQUE OU NPL

Cette approche, implémentée dans le logiciel *GENEHUNTER* (tableau 1) permet, à la différence des méthodes ASP ou APM, d'analyser globalement un pedigree (test NPL_{all}) sans le découper en paires d'apparentés (une telle analyse est toutefois possible par le test NPL_{pairs}). Ceci apporte une meilleure précision dans les cas ou plusieurs apparentés d'un même pedigree partagent le même allèle IBD. La méthode NPL est une approche multipoint systématique qui se base sur le calcul de la probabilité IBD en un point donné du chromosome en utilisant tout les marqueurs informatifs de celui-ci. Ces probabilités sont alors utilisées pour estimer le nombre d'allèles partagés IBD dans un groupe d'individus apparentés affectés et ensuite pour en calculer une fonction de score qui mesure le degré de liaison. Le test final est alors une somme pondérée (par $m^{-1/2}$ où m est le nombre d'individus de la famille) des scores centrés et réduits de toutes les familles et suit une distribution normale (pour un nombre de familles supérieur à 30). Cependant, les *p-values* calculées avec cette approximation donnent un test *conservateur*, ce qui peut entraîner une baisse importante de puissance si les données sont incomplètes.

Pour pallier à ce défaut, Kong et Cox (1997) [41] ont proposé une modification du test NPL (implémenté dans le logiciel GENEHUNTER-PLUS) qui permet une meilleure exploitation de l'information disponible et pour laquelle le seuil de signification peut être obtenu de manière exacte. Une seconde approche basée sur le principe de permutation permet également un meilleur contrôle du taux de faux positifs [42].

Un autre inconvénient de la méthode NPL, lié à l'algorithme utilisé [43] et que le temps de calcul, à l'inverse de l'algorithme de Elston-Stewart (utilisé par les programmes LINKAGE/FASTLINK), augmente exponentiellement avec le nombre d'individus (mais linéairement avec le nombre de marqueurs) ce qui limite son utilisation à des petits pedigree (ayant au plus 12 individus non fondateurs). Même si des améliorations au niveau du temps de calcul et du nombre d'individus par famille (ALLEGRO une version rapide de GENEHUNTER permet d'analyser des familles ayant jusqu'à 20 descendants et 15 fondateurs) ont été apportées à l'approche NPL [44, 45], son utilisation pour des pedigrees complexes n'est pas conseillée. Elle reste toutefois un approche de choix pour l'analyse de familles nucléaires ou de petits pedigree ayant au moins une paire d'apparentés affectés surtout si l'on s'intéresse à un grand nombre de marqueurs liés et si le génotypage est incomplet.

4 - COMPARAISON

Davis et Weeks (1997) [17] ont comparé par simulations, dans le cas de familles nucléaires et pour une maladie contrôlée par deux loci, un grand nombre de méthodes modèle-indépendantes. Leur étude confirme le fait que l'approche APM donne un taux de faux positif intolérablement élevé et que le test NPL perd de sa puissance quand l'information disponible diminue. De manière générale, il en ressort que toutes les méthodes basées sur l'IBS sont peu puissantes comparativement à celles basées sur l'IBD même dans les situations les plus favorables pour celles-là (parents non génotypés).

Les méthodes qui utilisent l'information des enfants génotypés non-affectés (comme celles implémentées dans le logiciel SIBPAIR) apportent un gain de puissance de l'ordre de 10% par rapport à celles utilisant les affectés seulement (ASP, APM, NPL,..) ou celles qui traitent les non-affectés comme ayant un phénotype inconnu. Leurs résultats montrent aussi que l'utilisation de toutes les paires de germains d'une fratrie (d'au plus 4 individus) est meilleure (plus puissante sans augmentation du taux de faux positifs) que l'utilisation d'une seule paire de la fratrie ou uniquement les paires indépendantes. En conclusion, il ressort de ce travail que le test de moyenne ASP (implémenté dans les logiciels SIBPAL et SIBPAIR) est le plus puissant pour une large gamme de modèles de maladie et qu'il est robuste pour les données avec des parents non génotypés. Ceci rejoint les résultats de Knapp et al. (1994) [46] qui ont montré que le test de moyenne est uniformément le plus puissant sous mode d'hérédité récessive et pour un allèle de maladie commun.

COMPARAISON DE LA METHODE LOD SCORE AUX METHODES NON-PARAMETRIQUES

Dizier et al. (1996) [20] ont montré, dans le cas de familles nucléaires et pour une maladie à deux loci, que le test ASP était plus puissant que l'analyse Lod score calculé sous un modèle monogénique utilisant les paramètres estimés par analyse de ségrégation. Ce résultat est en accord avec les précédentes études qui avaient montré que l'analyse Lod score utilisée avec un modèle génétique mal spécifié peut devenir très peu puissante au point d'exclure faussement la liaison.

Durner et al. (1999) [21] ont comparé par simulation, pour les même situations, le test Lod score réalisé deux fois, l'une sous modèle dominant et l'autre sous modèle récessif et corrigé pour l'erreur type I [25] et les tests ASP et NPL. Ils ont montré que cette approche LOD score est plus puissante (jusqu'à 5 fois) que les deux autres dans toutes les situations testées et que sa puissance était très souvent proche du test LOD score calculé pour le vrai modèle (ici à deux loci). Dans certaines situations, le test LOD score détecte une vrai liaison que le test NPL ne détecte pas.

Une autre étude récente [47] a confirmé ces résultats en montrant par simulation pour des familles nucléaires

que l'approche MMLS-C (voire § 2.5) était plus puissante que le test NPL pour la grande majorité des situations étudiées, couvrant différentes valeurs de pénétrance, différents MH et différents degrés d'hétérogénéité. De plus, la puissance du test MMLS-C est très proche de celle de l'analyse LOD réalisée avec les vraies valeurs de paramètres génétiques (le vrai modèle génétique) même en présence d'hétérogénéité. Cependant, quand l'information est faible (faible pénétrance) et en présence de forte hétérogénéité les méthodes MMLS-C et NPL sont proches au niveau de leur puissance (qui est alors faible pour les deux).

Dans le cas de petits pedigrees et dans le cadre de maladies complexes, Sham et al. (2000) [28] ont comparé plusieurs test LOD entre eux et avec le test NPL, pour des marqueurs complètement informatifs (cas favorable pour la méthode NPL). Ils ont alors trouvé que pour toutes les situations étudiées, les deux approches LOD et NPL sont équivalentes et ne s'éloignent pas beaucoup en terme de puissance de l'analyse Lod score réalisée sous le vrai modèle génétique (qui est alors l'analyse la plus puissante possible). Ceci confirme le fait que l'analyse Lod score n'est pas fortement affectée par un mode d'hérédité incertain et suggèrent que les méthodes modèle-dépendantes et modèle-indépendantes sont toutes aussi adaptées pour l'analyse des maladies complexes. La méthode NPL modifiée [41] et qui se base sur un modèle statistique à un paramètre constitue un sorte de pont entre ces types d'approches.

DISCUSSION

Il convient de revenir dans cette discussion sur la notion de méthode non-paramétriques ou modèle-indépendant. Kruglyak (1997) [48] défini un test modèle indépendant comme « un test qui est valide sans considération des vrais paramètres génétiques (inconnus) dans le sens qu'il donne un taux de faux positif correct ». Cette définition s'applique non seulement aux méthodes dites classiquement non-paramétriques (ASP, APM, NPL,..) mais aussi à l'analyse LOD score réalisée sous un faux modèle génétique (analyse wrod). Dans la définition courante et d'ailleurs sans intérêt, on attribue le qualificatif de 'non-paramétrique' aux méthodes « pour lesquelles on n'a pas besoin de donner explicitement un modèle génétique et particulièrement un MH », qui elle n'inclus pas l'analyse Lod score. Les méthodes modèle-indépendantes qui ont été proposées comme alternatives à l'analyse Lod score l'avaient été dans le but de fournir des méthodes capables de réaliser une faible perte de puissance tout en retenant une robustesse pour une large gamme d'hypothèses alternatives; de ce fait ces méthodes seraient donc plus adaptées à l'analyse des maladies complexes. Toutefois, l'analyse Lod score, qui sembler a priori exclue de cette catégorie de méthodes, peut atteindre le même objectif sans besoin de tester tous les modèles génétiques (ce qui entrainerait en effet une trop grande perte de puissance) par exemple par une approche du type MMLS-C.

Une autre notion semble être aussi mal comprise; c'est le besoin de spécifier un MH lors d'une analyse de liaison. La plupart des utilisateurs de ces méthodes, pensent que les propriétés statistiques des méthodes non-paramétriques sont indépendantes du MH du caractère étudié. C'est tout le contraire; si en effet le niveau du test (probabilité d'erreur de type I) des méthodes non-paramétriques ne dépend pas du vrai MH, sa puissance en dépend (tout comme pour la méthode Lod score). La ressemblance entre méthodes modèle-indépendantes et Lod score va encore plus loin; Knapp et al. (1994) [46] ont montré que le test de moyenne ASP, qui ne spécifie pas implicitement de MH, avait la même région de rejet qu'un test Lod score sous modèle récessif avec pénétrance complète et donc ces deux tests auront exactement la même puissance quelque soit le MH. Si le vrai MH est dominant les deux tests auront la même perte de puissance. Whittemore (1996) [49] a montré que cette équivalence avec le test Lod score peut être étendue à d'autres méthodes non-paramétriques. De façon générale, on peut trouver pour chaque test non-paramétrique les conditions spécifiques de paramètres pour lesquelles il aura des propriétés statistiques identiques à celle d'une méthode basée sur le maximum de vraisemblance. En particulier, il est théoriquement possible de trouver le ou les ensembles d'hypothèses sur le MH correspondant à chaque analyse non-paramétrique et d'utiliser l'un de ces ensembles d'hypothèses pour faire une analyse LOD score qui fournira alors des résultats statistiques identiques [26].

Dans l'analyse de liaison d'une maladie génétique complexe, toutes les études récentes, tendent à montrer que le paramètre le plus important dans le modèle génétique c'est le MH pour le locus lié au marqueur qui est testé et non pour la maladie dans sa globalité [20, 21, 49]. Comme ce mode d'hérédité peut être approché de manière robuste par une hérédité dominante ou récessive, une analyse du type LOD score réalisée sous les deux modèles et appliquée avec un seuil corrigé constitue un bon compromis. De plus, dans ce type d'analyse, l'effet d'autres loci impliqués dans la maladie et leurs interactions (épistasie, interaction gènes x environnement) peuvent être absorbés en grande partie par l'utilisation d'une pénétrance incomplète [50, 51]. Grâce aux résultats de plusieurs études récentes, les méthodes de type Lod score connaissent un regain d'intérêt dans l'analyse des maladies complexes, après la grande ferveur de ces dernières années pour les méthodes modèle-indépendantes. Il est bien connu que l'analyse Lod score est l'approche la plus puissante si le modèle génétique est correctement spécifié, qu'elle utilise l'information des génotypes et phénotypes de tous les

individus (ce qui n'est pas le cas pour la plupart des méthodes non-paramétriques) et qu'elle permet de tester l'hétérogénéité génétique. Ces propriétés, combinées à sa robustesse par rapport à l'utilisation de faux modèles, la remet à l'ordre du jour.

Récemment, Göring et Terwilliger (2000a,b) [51, 52] ont proposé l'utilisation de taux de recombinaison à valeur complexe ou hypercomplexe dans une approche Lod score 'complexe' multipoint qui est plus robuste que l'analyse Lod score classique par rapport aux erreurs du modèle et équivalente à une analyse deux-points. De plus elle étend l'équivalence entre méthodes modèle dépendantes et certaines modèle indépendantes aux approches multipoints [52]. Cette approche est très prometteuse et constituera très certainement un pont solide entre les deux familles de méthodes.

Toutes les recherches à ce jour nous amènent à penser que les maladies complexes peuvent être considérées comme étant sous le contrôle d'une série de contributions génétiques dont chacune peut être déterminée indépendamment des autres et le problème devient alors d'estimer la contribution de chacun de ces loci dans la détermination de la maladie et éventuellement son interaction avec les autres loci. Quelle méthode sera « meilleure » dépend des données disponibles ainsi que d'autres facteurs liés à la maladie et à la population étudiées. Ce qui devient donc important en premier lieu ce n'est plus la méthode qui est utilisée pour l'analyse (puisque plusieurs peuvent être essayées et leur résultats comparés) mais la collecte de données suffisamment 'propres' et avec une définition précise du phénotype qui permettent d'exploiter toute la puissance des méthodes disponibles.

BIBLIOGRAPHIE

1 - J.B.S. Haldane et Smith C. A. B. (1947) A new estimate of linkage values and calculation of distance between the genes for colour-blindness and hemophilia in man. *Ann Eugen*, **14:** 10 - 31

2 - G. A. Barnard (1949) Statistical inference. *J. R. Stat. Soc*. **B11:** 115 - 135

3 - N. E. Morton (1955) Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* **7:** 277 - 318.

4 - J. Ott (1999) Analysis of *Human Genetic Linkage*. John Hopkins University Press. Third Edition. 382 pp.

5 - R.C. Elston (1997) Akgorithms and inferences: the challenge of multifactorial diseases.

6 - E.S. Lander et L Kruglyak (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage analysis. *Nat. Genet.* **11:** 241 - 247

7 - L. Kruglyak, and M. J. Daly (1998) Linkage thresholds for two-stage genome scans. *Am. J. Hum. Genet.* **62:** 994 - 996

8 - N. E. Morton (1998) Significance Levels in complex inheritance. *Am. J. Hum. Genet.* **62:** 690 - 697

9 - L. P. Zhao, R. Prentice F. Shen et Hsu L. (1999) On the Assessment of statistical significance in disease-gene discovery. *Am. J. Hum. Genet.* **64:** 1739 - 1753

10 - F. Clerget-Darpoux, C. Bonaiti-Pellie et Hochez J. (1986) Effects of misspecifying genetic parameters in lod score analysis. Biometrics **42:** 393 - 399

11 - J. Ott (1983) Linkage analysis and family classification under heterogeneity. *Ann. Hum. Genet.* **47:** 311 - 320

12 - C. T. Falk (1997) Effet of genetic heterogeneity and assortative mating on linkage analysis: a simulation study. *Am. J. Hum. Genet.* **61:** 1169 - 1178

13 - A. S. Whittemore et Halpern J. (2001) Problems in the definition, interpretation and evaluation of genetic heterogeneity. *Am. J. Hum. Genet.* **68:** 457 - 465

14 - H. H. Göring and Terwilliger J. D. (2000) Linkage analysis in the presence of errors III: marker loci and their map as nuisance parameters. *Am. J. Hum. Genet.* **66:** 1298 - 1309

15 - R. L. Trembath Clough, J. L. Rosbotham, A. B. Jones, R. D. R. Camp, A. Frodsham, J. Browne et al. (1997) Identification of a major susceptibility locus on chromosome 6p and evidence for further disease loci revealed by a two-stage geneome search in psoriasis. *Hum. Molec. Genet.* **6:** 813 - 820

16 - H. H. Göring and Terwilliger J. D. (2000) Linkage analysis in the presence of errors IV: joint pseudomarker analysis of linkage and/or linkage disequilibrium on a mixture of pedigrees and singletons when the mode of inheritance cannot be accurately specified. *Am. J. Hum. Genet.* **66:** 1310 - 1327

17 - S. Davis et Weeks D. E. (1997) Comparison of non-parametric statistics for detection of linkage in nuclear families: single maker evaluation. *Am. J. Hum. Genet.* **61:** 1431 - 1444

18 - J. A. Williamson et Amos C. I. (1990) On the asymptotic behavior of the estimate of the recombination fraction under the null hypothesis of no linkage when the model is misspecified. *Genet. Epidemiol.* **7:** 309 - 318

19 - D. F. Easton, D. T. Bishop, D. Ford et Crockford (1993) Genetic linkage analysis in familial breast and ovarian cancer: results from 214 families. The Breast Cancer Linkage Consortium. *Am. J. Hum. Genet.* **52:** 678 - 701

20 - Dizier M. H., M. C. Babron et Clerget-Darpoux F. (1996) Conclusion of lod score analysis for family data generated under two-locus models. *Am.J.Hum.Genet.* **58:** 1338 - 1346

21 - M. Durner, V. J. Vieland and Greenberg D. A. (1999) Further evidence for the increased power of LOD scores compared with non-parametric methods. *Am. J. Hum. Genet.* **64:** 281 - 289

22 - J. D. Terwilliger et Ott J. (1994) Handbook of human genetic linkage. John Hopkins University Press, Baltimore

23 - S. M. Leal et Ott J. (2000) Effect of stratification in the analysis of affected-sib-pair data: benefits and costs. *Am. J. Hum. Genet.* **66:** 567 - 575

24 - D. A. Greenberg, P. Abreu and Hodge S. E. (1998a) The power to detect linkage in complex disease by means of simple LOD-score analyses. *Am. J. Hum. Genet.* **63:** 870 - 879

25 - S. E. Hodge, O. C. Abreu and Greenberg D. A. (1997) Magnitude of type I error when single-locus linkage analysis is maximized over over models: a simulation study. *Am. J. Hum. Genet.* **60:** 217 - 227

26 - D. A. Greenberg., S. E. Hodge, V. J. Vieland and Spence N. A. (1998b) Power, mode of inheritance and type I error in Lod scores and affecteds-only methods: reply to Kruglyak. *Am. J. Hum. Genet.* **62:** 202 - 204

27 - D. Curtis et Sham P. C. (1995) Model-free linkage analysis using likelihoods. Am. J. Hum. Genet. **57:** 703 - 716

28 - P. C. Sham, M. W. Lin, J. H Zhao, and Curtis D. (2000) Power comparison of parametric and nonparametric linkage tests in small pedigrees. *Am. J. Hum. Genet.* **66:** 1661 - 1668

29 - R. C. Elston (1998) Methods in linkage analysis and the assumptions underlying them. *Am. J. Hum. Genet.* **63:** 931 - 934

30 - J. H. Edwards (1998) Penrose and sib-pairs. *Ann. Hum. Genet.* **62:** 365 - 377

31 - W. C. Blackwelder et Elston R. C. (1985) A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet Epidemiol,* **2:** 85 - 97

32 - N. Risch (1990) Linkage strategies for genetically complex traits. II. The power of affected relative pairs. *Am. J. Hum. Genet.* **46:** 229 - 241

33 - P. Holmans (1993) Asymptotic properties of affected sib-pair linkage analysis. *Am. J. Hum. Genet.* **52:** 362 - 374

34 - M. H. Dizier, H. Quesneveille, B. Prun, H. Selinger-Leneman et Clerget-Darpoux F. (2000) The triangle test statistic (TTS): a test of genetic homogeneity using departure from the traingle constraints in IBD distribution among affected sib-pairs. *Ann. Hum. Genet.*, **64:** 433 - 442

35 - L. Kruglyak, M. J. Daly, M. P. Reeve-Daly et Lander E.S. (1996) Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* **58:** 1347 - 1136

36 - J. Teng et Siegmund D. (1997) Combininig information within and between pedigrees for mapping complex traits. Am. J. Hum. Genet. **60:** 979 - 992

37 - E. R. Hauser, M. Boehnke, S. W. Guo et Risch N. (1996) Affected sib-pair interval mapping and exclusion for complex genetic traits: sampling considerations. *Genet. Epidemiol.* **13:** 117 - 137

38 - E. Feingold et Siegmund D. (1997) Strategies for mapping heterozygous recessive traits by allele-sharing methods. *Am. J. Hum. Genet.* **60:** 965 - 978

39 - D. E. Weeks et Lange K. (1988) The affected pedigree member mehod of linkage analysis. *Am. J. Hum. Genet.* **42:** 315 - 326

40 - D. E. Weeks et Lange K. (1992) A multilocus extension of the affected pedigree member mehod of linkage analysis. *Am. J. Hum. Genet.* **50:** 859 - 868

41 - A. Kong and Cox N. J. (1997) Allele-sharing models: Lod scores and acurate linkage tests. *Am. J. Hum. Genet.* **61:** 1179 - 1188

42 - H. Zhao, K. R. Merikangas et Kidd K. K. (1999) On a randomization procedure in linkage analysis. *Am. J. Hum. Genet.* **65:** 1449 - 1456

43 - E. S. Lander et Green P. (1987) Construction of multilocus genetic maps in humans. Proc. *Natl. Acad. Sci.* USA **84:** 2363 - 2367

44 - L. Kruglyak et Lander E. S. (1998) Faster multipoint linkage analysis using Fourrier transforms. *J.Comput.Biol.* **5:** 1 - 7

45 - D. F. Gudbjartsen, K. Jonasson, M. L. Frigge et Kong A. (2000) Allegro: a new program for multipoint linkage analysis. *Nat. Genet.* **25:** 12 - 13

46 - M. Knapp, S. A. Seuchter et Baur M. P. (1994) Linkage analysis in nuclar families. II. Relationship between affected sib-pair tests and lod score analysis. *Hum. Hered.* **44:** 44 - 51

47 - P. C. Abreu, D. A. Greenberg and Hodge S. E. (1999) Direct power comparisons between simple LOD scores and NPL scores for linkage analysis in complex diseases. *Am. J. Hum. Genet.* **65:** 847 - 857

48 - L. Kruglyak (1997) Nonparametric linkage tests are model free. *Am. J. Hum. Genet.* **61:** 254 - 255

49 - A. S. Whittemore (1996) Genome scanning for linkage: an overview. *Am. J. Hum. Genet.* **59:** 704 - 716

50 - V. J. Vieland, D. A. Greenberg and Hodge S. E. (1992) Adequacy of single locus appoximations for linkage analyses of oligogenic traits. *Genet Epidemiol.* **9:** 45 - 59

51 - V. J. Vieland, D. A. Greenberg and Hodge S. E. (1993) Adequacy of single locus appoximations for linkage analyses of oligogenic traits: extension to multigenerational pedigree sructures. *Hum Hered.* **43:** 329 - 336

52 - H. H. Göring and Terwilliger J. D. (2000) Linkage analysis in the presence of errors I: complex valued recombination fractions and complex phenotypes. *Am. J. Hum. Genet.* **66:** 1095 - 1106

53 - H. H. Göring and Terwilliger J. D. (2000) Linkage analysis in the presence of errors II: marker-locus genotyping errors modeled with hypercomplex recombination fractions. *Am. J. Hum. Genet.* **66:** 1107 - 111

54 - R. W. Cottingham, R. M. Indury and Schäffer A. A. (1993) Faster sequential genetic linkage computations. *Am. J. Hum. Genet.* **53:** 252 - 263

55 - C. S. Jensen and Kong A. (1999) Blocking Gibbs sampling for linkage analysis in large pedigrees with many loops. *Am. J. Hum. Genet.* **65:** 885 - 901

56 - D. Hinds and Risc N. (1996) The ASPEX package: affected sib-pair mapping.ftp://lahmed.stanford.edu/pub/aspex.

57 - S. E. Hodge et R. C. Elston (1994) Lods, wrods and

mods: the interpretation of lod scores calculeted under different mdels. *Genet. Epidemiol.* **11:** 329 - 342
58 - P. Holmans and Clayton D. (1995) Efficiency of typing unaffacted relatives in an affected sib-pair linkage study with single-locus and multiple tightly linkae markers. *Am. J. Hum. Genet.* **57:** 1221 - 1232
59 - L. Kruglyak, M. J. Daly et Lander E. S. (1995) Rapid multipoint linkage analysis of recessive trait in nuclear families. *Am. J. Hum. Gene*t. **56:** 519 - 527
60 - L. Kruglyak and Lander E. S. (1995) Complete multipoint sib-pair analysis of qualitative and quantitative traits. *Am. J. Hum. Genet.* **57:** 439 - 454
61 - G. M. Lathrop, J. M. Lalouel, C. Julier et Ott J. (1984) Strategies for multilocus linkage analysis in humans. *Proc. Natl. Acad. Sci.*, **81:** 3443 - 3446
62 - J. R. O'Connell and Weeks D. E. (1995) The VITESSE algorithm for rapid exact multilocus linkage analysis via genotype set-recoding and fuzzy inheritance. *Nature Genetics*, **11:** 402 - 408
63 - J. Ott (1974) estimation of recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am. J. Hum. Genet.* **26:** 588 - 597
64 - J. D. Terwilliger (1996) Program SIBPAIR: sib-pair analysis on nuclear families.
ftp://linkage.cpmc.columbia.edu