



Ciência & Saúde Coletiva

ISSN: 1413-8123

cecilia@claves.fiocruz.br

Associação Brasileira de Pós-Graduação em  
Saúde Coletiva  
Brasil

Krieger, Henrique; Feitosa, Mary F.  
O futuro da epidemiologia genética de características complexas  
Ciência & Saúde Coletiva, vol. 7, núm. 1, 2002, pp. 73-83  
Associação Brasileira de Pós-Graduação em Saúde Coletiva  
Rio de Janeiro, Brasil

Disponível em: <http://www.redalyc.org/articulo.oa?id=63070107>

- Como citar este artigo
- Número completo
- Mais artigos
- Home da revista no Redalyc

redalyc.org

Sistema de Informação Científica  
Rede de Revistas Científicas da América Latina, Caribe, Espanha e Portugal  
Projeto acadêmico sem fins lucrativos desenvolvido no âmbito da iniciativa Acesso Aberto

## O futuro da epidemiologia genética de características complexas

### Future of genetic epidemiology in complex traits

Mary F. Feitosa <sup>1</sup>  
Henrique Krieger <sup>2</sup>

**Abstract** Genetic epidemiology has advanced from its early focus on rare mendelian diseases to the genetic dissection of complex traits. With the advent of the complete genome map of humans and other organisms, more than ever genetic epidemiology has an important role in ascertaining the relative importance of genetic and environment causative factors of complex traits. The main methodology strategies (familial resemblance, segregation analysis, association and linkage analysis and meta-analysis) in the study of complex traits are outlined and its advantages and shortcomings are discussed. The importance of sampling and the use of appropriate phenotypes and genetic markers are stressed and an example on the study of BMI (Body Mass Index), showing the role of a major genetic factor located at chromosome 7 illustrates some of the above strategies. It is suggested that in the future, although recognizing that multiplex families will still be the mainstay of linkage studies, new and efficient types of sampling (unrelated controls, for instance) utilizing pooled DNA samples will be universally employed. The recognition of genetic heterogeneity between studies and its interpretation will be one of the prominent features in the forthcoming complex traits studies. **Key words** Familial resemblance, Segregation analysis, Association and linkage analysis, Meta-analysis, Complex traits

**Resumo** A epidemiologia genética evoluiu de um enfoque em estudos sobre doenças mendelianas raras para a análise genética de características complexas. Com o advento de informações sobre a completa seqüência de genes ao longo do genoma humano e de outros organismos, o interesse da epidemiologia genética em desvendar a natureza dos fatores que influenciam essas características se tornou primordial. São apresentados os principais métodos empregados no estudo de doenças complexas bem como suas principais vantagens e desvantagens. Discute-se a importância na determinação da amostra e o uso de fenótipos e marcadores genéticos apropriados. Como exemplo das estratégias citadas tomamos o estudo de índice de massa corporal (BMI) para ilustrar um fator genético principal localizado no cromossomo 7. Em uma discussão sobre tendências no estudo de ligação, embora reconhecendo que famílias e genealogias continuarão sendo o foco principal das amostras, discute-se alguns novos e eficientes tipos de amostragem (como por exemplo, controles não-relacionados) em que amostras de conjunto de DNA serão universalmente empregadas. O reconhecimento da heterogeneidade genética entre estudos e sua interpretação será uma das mais importantes características no futuro das análises de características complexas. **Palavras-chave** Agregação familiar, Análise de segregação, Associação, Análise de ligação, Meta-análise, Características complexas

<sup>1</sup> Division of Biostatistics, Campus Box 8.067, Washington University School of Medicine, 660 S. Euclid. Ave. St. Louis, MO 63110-1093, USA. maryf@wubios.wustl.edu  
<sup>2</sup> Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, Brasil.

A epidemiologia genética progrediu de sua abordagem inicial em estudos sobre doenças mendelianas raras para a análise de características complexas, ou seja, aquelas que são influenciadas por um grande número de fatores, a maioria dos quais de difícil identificação. Doenças cardiovasculares, pressão arterial, doenças infecciosas e certas neoplasias são exemplos de características complexas. Com a aquisição de informações precisas sobre a sequência de genes ao longo do genoma humano e de outros organismos, o interesse da epidemiologia genética em desvendar a natureza dos fatores que influenciam essas características se tornou primordial.

Entretanto, essas características têm-se mostrado muito mais desafiadoras e as esperanças de grandes avanços foram frustradas, principalmente devido ao pequeno efeito de cada uma de suas múltiplas causas e pelo fato de que os métodos planejados para detectar genes com efeitos significativos (genes principais) terem se mostrado insuficientes. Contudo, apesar de os efeitos individuais dos genes serem relativamente pequenos, as interações entre esses genes e outros genes e o meio ambiente podem contribuir significativamente para o desvendamento dessas características. As dificuldades existentes na detecção dessas interações geralmente redundam na impossibilidade de detectar genes desse tipo.

Para a detecção dos efeitos dos genes várias condições precisam ser satisfeitas. Em primeiro lugar, obviamente, é necessário o conhecimento adequado da natureza da característica e da sua biologia, história, distribuição etc. Também fundamentais para esse tipo de estudo são o planejamento do estudo apropriado, amostras robustas e o emprego de metodologia adequada.

Seria muito ambicioso pretender, nesse trabalho, fazer uma completa revisão sobre epidemiologia genética. O intuito do presente trabalho é o de dar uma visão geral da área no desvendamento de fatores atuantes em características complexas com ênfase na importância de um delineamento experimental apropriado, no emprego de metodologia correta e nas nuances de interpretação que essas análises podem propiciar. Algumas reflexões sobre a relevância de estudos genético-epidemiológicos em países em desenvolvimento, nesse começo de milênio, quando o grande enigma do mapeamento do genoma humano está concretizado, serão também apresentados.

## Delineamento experimental

A excelência do delineamento do estudo é de vital importância para o mapeamento de genes associados a doenças complexas (Rao, 2001; Gu & Rao, 2001). Uma abordagem comum para ampliar o poder de um estudo é a utilização de grandes amostras. No entanto, planejamentos simplistas podem conduzir a coleta de amostras enormes que são extremamente difíceis de serem obtidas ou de custo elevadíssimo. Dessa forma, um delineamento criativo é necessário para detectar e ampliar um sinal fraco, porém real.

Estudos multiinstitucionais têm um papel importante na obtenção de grandes amostras coletadas de maneira uniforme pela utilização de protocolos amostrais únicos. Nesse sentido, alguns estudos multiinstitucionais têm proporcionado importantes contribuições científicas em diversas áreas da pesquisa biomédica. Exemplos desse tipo de delineamento são o National Heart, Lung, and Blood Institute Family Heart Study (NHLBI-FHS), estudo multiinstitucional que avalia os determinantes genéticos e não-genéticos que atuam em doenças cardiovasculares (Higgins *et al.*, 1996) e o Heritage Family Study, projeto multiinstitucional planejado para estudar o papel de mecanismos genéticos existentes nos fatores de riscos para doenças cardiovasculares e diabetes (Bouchard *et al.*, 1995), ambos nos EUA. Entre nós, um estudo exemplar, que utiliza desse tipo de abordagem, é o Estudo Colaborativo Latino Americano de Malformações Congênitas (ECLAMC) que é um programa clínico e epidemiológico do tipo caso-controle (Prado, 1992).

Os fatores que determinam o conteúdo informativo de um delineamento, que por sua vez regula o poder de decisão de um estudo, podem ser classificados em algumas categorias: 1) fatores associados à fenotipagem e genotipagem; 2) fatores associados à amostragem; 3) procedimentos analíticos; e 4) argumentos associados a custo-benefício.

Os fenótipos precisam ser claramente definidos. Para evitar que a detecção de um determinado fenótipo seja minimizada, isto é, que a frequência do fenótipo não seja substancialmente subestimada, a característica em estudo deve ser altamente reproduzível. Medidas múltiplas e/ou o uso de informações suplementares (por exemplo, idade do aparecimento da característica, gravidade da doença, história fa-

milial etc.) podem contribuir para o refinamento do critério da escolha do fenótipo e podem ser decisivos na melhoria do poder analítico (Rice *et al.*, 2001).

Apesar da constante melhoria tecnológica dos processos automáticos de genotipagem em grande escala, erros na genotipagem ainda existem. Esse fato, acompanhado de possíveis erros na identificação de indivíduos em uma família, pode produzir ruído suficiente para diminuir o poder analítico de um estudo. Portanto, alta qualidade da genotipagem e grande densidade de marcadores ao longo do genoma são fundamentais para o poder de resolução de um estudo. A escolha do tipo de amostragem (grandes genealogias, famílias nucleares, pares de irmãos, caso-controle, isolados populacionais, mistura de populações) e o tamanho da amostra devem ser cuidadosamente escolhidos a fim de conduzir a uma otimização do quociente sinal/ruído. As características complexas possuem provavelmente mecanismos genéticos distintos em populações diferentes (Krieger & Feitosa 1999) e a escolha da população “apropriada” pode potencializar a qualidade de um estudo (Gu & Rao, 2001).

Os procedimentos analíticos (segregação, ligação, associação, varredura genômica etc.) devem ser escolhidos por ocasião do planejamento da coleta de dados. Dessa forma, a metodologia torna-se parte do delineamento, possibilitando uma antecipação do poder do estudo. Contudo, se os dados já tiverem sido coletados, um esforço deve ser feito no sentido de escolher o método estatístico mais poderoso visando à otimização dos resultados.

Problemas de custo-benefício devem ser enfrentados no sentido de estabelecer um equilíbrio entre poder e sensibilidade (Gu & Rao, 2001). Um orçamento fixo e limitações práticas irão provavelmente determinar o tipo de estudo, independentemente das ambições dos pesquisadores. O balanço entre custo e eficiência deve ser muito bem estudado antes de se começar um projeto, principalmente em países em desenvolvimento, onde as fontes de financiamento são limitadas e, em geral, parcas.

### Procedimentos analíticos

Os métodos analíticos vêm recebendo contínua atenção no estudo de características complexas e sua importância para esse tipo de estudo é medida pela inovação metodológica ob-

servada nos últimos anos. Aceita-se como fato inquestionável a interdependência absoluta entre técnicas moleculares e métodos estatístico-epidemiológicos, na solução de problemas biológicos relacionados a fenótipos mais ou menos distantes de uma correlação absoluta com um determinado genótipo. Aqui, procuraremos dar uma visão simplificada de alguns métodos da epidemiologia genética empregados na atualidade.

### Agregação familiar

A detecção e a avaliação da extensão da agregação familiar são os primeiros passos na análise genética de qualquer característica. É sabido que esta agregação significa o compartilhamento, entre indivíduos da mesma família, de genes e de fatores do ambiente. A existência dessa agregação é evidenciada pela maior semelhança fenotípica entre pares de parentes do que entre dois indivíduos não-aparentados (Elston, 2000, Rice & Borecki, 2001). A magnitude da agregação familiar pode ser medida pela correlação fenotípica entre pares de indivíduos (ex.: pares de irmãos, pai-filho, cônjuges etc.). Os diversos modelos genético-epidemiológicos utilizam essas diversas correlações e a interpretação biológica depende da magnitude e significância dessas correlações.

Vários programas aplicativos encontram-se à disposição do pesquisador para o estudo da agregação familiar. Esses programas baseiam-se fundamentalmente na aderência de modelos a um conjunto de correlações obtido diretamente dos dados, admitindo-se normalidade (no sentido estatístico) multivariada (Hopper & Matthews, 1982). A estimação dos parâmetros do modelo é feita por verossimilhança máxima. Em geral, uma determinada característica é analisada individualmente (análise univariada). Entretanto, alguns estudos podem ser feitos utilizando-se mais de uma característica fenotípica (Coletto *et al.*, 1981). Existem programas altamente eficientes e práticos para trabalhar essas correlações (ex. SEGPAT; Province & Rao, 1995). Em famílias nucleares, o modelo mais simples reconhece a existência de quatro tipos de indivíduos (pais, mães, filhos e filhas), conduzindo a oito tipos de correlações fenotípicas entre indivíduos (pai-mãe, pai-filho, mãe-filho, pai-filho, mãe-filho, filho-filho, filho-filha e filha-filha). No caso de simplificar o modelo, admitindo a inexistência de diferenças

sexuais ou etárias, esse se reduz a três correlações (cônjuges, progenitores-descendentes, irmãos). O modelo bivariado completo, que contempla duas características em cada um dos componentes da família nuclear, trabalha com 34 correlações enquanto que o modelo simplificado, sem diferenças sexuais ou etária, reduz-se a dez correlações diferentes. A validade dessas e outras simplificações ou a existência de determinados fatores causais podem ser cheçadas pelo teste da razão de verossimilhança máxima (RVM) entre o modelo reduzido e um modelo mais geral. As RVM são, assintoticamente, distribuídas como um  $\chi^2$  com um número de graus de liberdade igual a diferença do número de parâmetros estudados nos dois modelos.

Convém salientar que os estudos de agregação familiar vêm de longa data, sendo pioneiros os estudos de S. Wright e R. A. Fisher na primeira metade do século passado, em conjunto com numerosas contribuições dos melhoristas, que atestaram pelo caráter clássico desse tipo de estudo e não como sendo devido a modismos tecnológicos.

A correta avaliação da magnitude da contribuição genética à característica fenotípica estudada é fundamental para um delineamento apropriado de estudos que dependem de grandes amostras e técnicas caras de genotipagem.

### Análise de segregação

A análise de segregação é, teoricamente, um método para estudar dados familiares com a finalidade de estabelecer o modo de herança de uma determinada característica, quando o efeito de um gene não pode ser medido diretamente. Mendel, no século XIX, foi o pioneiro nesse tipo de análise, que o conduziu à descoberta das leis básicas da hereditariedade. A história da análise de segregação pode, para fins de um melhor entendimento, ser dividida em duas fases: a clássica e a complexa (Morton *et al.*, 1983). Os métodos clássicos de análise de segregação de fenótipos mendelianos estão fundamentados na estimação e teste de razões de segregação (Rao *et al.*, 1974) que são determinados, em grande parte, pelos efeitos de um único gene.

Modelos mais abrangentes de análise de segregação foram desenvolvidos para abordar os problemas gerados por fenótipos complexos (Elston & Steward, 1971; Morton & MacClean,

1974; Lalouel *et al.*, 1983; Bonney, 1984; Hassstedt, 1982; Blangero & Konigsberg, 1991). Todos esses diferentes tratamentos metodológicos, apesar de diferirem em alguns pormenores, possuem características comuns que podem ser assim resumidas.

Admite-se que o fenótipo seja influenciado pela contribuição independente e aditiva de um gene principal, um componente poligênico/multifatorial e um resíduo ambiental não-transmissível. O efeito do gene principal resulta da segregação, em um único *locus* gênico, de dois alelos (A e a) e os genótipos são distribuídos nas proporções esperadas pelo princípio de Hardy-Weinberg. Os parâmetros desses modelos são a média ( $\mu$ ), a variância ( $\sigma$ ), a frequência genética (q), o grau de dominância do gene (d), a herdabilidade (H) e as taxas de transmissão entre duas gerações ( $\tau$ ).

A análise da segregação tem-se mostrado de grande interesse no estudo de várias características influenciadas por um gene principal ou por um pequeno número de genes (oligogenes) (ex. Blangero *et al.*, 1996; Borecki *et al.*, 1998a; Feitosa *et al.*, 1995, 1996, 1999, 2000a, b, 2002, Olson *et al.*, 2001).

Com as informações de mapas genéticos, modelos de análise de segregação mais reais que incorporem os efeitos de marcadores alojados ou associados (chamados de modelos combinados) serão mais e mais necessários. Certamente, a análise da segregação no futuro próximo utilizará dados sobre marcadores genéticos.

Alguns programas de análise de segregação podem ser obtidos pela internet: POINTER/COMDS: [www.hgmp.mrc.ac.uk/registered/help/comds](http://www.hgmp.mrc.ac.uk/registered/help/comds); PAP (Pedigree Analysis Package): [www.well.ox.ac.uk/docs/indez.html.solar](http://www.well.ox.ac.uk/docs/indez.html.solar); SOLAR (Sequential Oligogenic Linkage Analysis Routines): <http://canag.cit.mih.gov/lserver/solar.htm>, e SAGE (Statistical Analysis for Genetic Epidemiology): <http://darwin.cwru.edu/pub/sage.html>.

### Análise de ligação

Diz-se que há ligação, quando existe uma proximidade mensurável entre genes ao longo do cromossomo. A análise de ligação, em termos de epidemiologia genética, verifica a co-segregação, dentro de uma família, de um gene marcador e um fenótipo para estabelecer se o marcador e a característica estão fisicamente liga-

dos. A análise da ligação tem sido o método principal para mapear doenças mendelianas e também tem exercido um papel importante nas tentativas de mapear aquelas doenças complexas, que supostamente possuem um mecanismo mendeliano, com efeito, significativo sobre a característica.

Independentemente do tipo de dados à disposição do pesquisador (desde pares de irmãos até extensas genealogias), a análise de ligação pode ser realizada por intermédio de técnicas baseadas em modelo (*model-based*) ou independentes de modelos (*model-free*). Deve-se frisar que todas as análises estatísticas admitem um modelo probabilístico, no qual parâmetros são estimados e hipóteses sobre parâmetros podem ser testadas.

Nas análises baseadas em modelo (ABM), também chamadas de *lod-score* ou análise paramétrica, admite-se que todos os aspectos do modelo estatístico, exceto a taxa de recombinação ( $\theta$ ), são conhecidos. Morton (1955) fez uma contribuição fundamental para a genética humana, ao trabalhar o conceito de *lod-score* (*lod*), ou seja, o logaritmo na base 10 de uma razão de verossimilhança (RV). A RV é a probabilidade de se observar um quadro da distribuição de genótipos em uma família, dada a existência da ligação com um valor  $q$  da recombinação versus a mesma probabilidade estimada para segregação independente ( $\theta = 0,5$ ).

Dessa forma, o quociente da verossimilhança ao valor alternativo ( $0 \leq \theta < 0,5$ ) pelo da mesma ao valor  $\theta = 0,5$  possui poder de decisão. Um valor maior que três indica que a hipótese alternativa tem uma verossimilhança mil vezes maior que a hipótese nula com  $\theta = 0,5$ , e é indicativo da existência da ligação entre o marcador e o fenótipo estudado. Valores de *lod-scores* maiores que 1,5 e menores que três sugerem a existência de ligação, enquanto valores iguais ou menores que -2 indicam a inexistência de ligação. Elston & Stewart (1971) produziram um algoritmo para o cálculo de verossimilhança em genealogias complicadas, enquanto Ott (1974) desenvolveu um programa de computador voltado ao usuário para estimar *lod-scores* nas mais variadas genealogias. Os *lod-scores* têm também sido usados com bastante sucesso, no mapeamento de doenças mendelianas em que são consideradas situações complicadas (penetrância incompleta, dominância, fenocópias, aparecimento tardio, entre outras).

Tendo em vista que a inexistência de um

modelo simples de herança para características complexas e, sobretudo, que a mesma pode depender de co-variáveis (idade de aparecimento, sexo etc.), é muito difícil ou até mesmo impossível obter genealogias multigeracionais informativas sem uma avaliação inicial que possibilite simplificar a análise dessas genealogias. Dessa forma, o método de pares de irmãos (Penrose, 1935) foi revivido para utilizar os padrões de compartilhamento de alelos no estudo de ligação. A atualização tecnológica do método de pares de irmãos foi ampliada, o que permitiu também a utilização de pares de parentes mais afastados, bem como a análise de determinadas genealogias (Weeks & Lange, 1988; Curtis & Sham, 1994; Kruglyak *et al.*, 1996; Almasy & Blangero, 1998).

O conceito de identidade por ascendência (IBD, *Identity by descent*) é de vital importância nos estudos de ligação, pois quantifica a semelhança genética entre pares de indivíduos. Um par de indivíduos aparentados possui um alelo IBD se esse alelo tem origem comum em um ancestral. Dessa forma, um par de irmãos pode possuir zero, um ou dois alelos IBD com as respectivas probabilidades de 1/4, 1/2 e 1/4. Algoritmos para estimar as probabilidades de IBD múltiplos marcadores genéticos constituíram um importante avanço nas pesquisas sobre etiologias complexas (Lander & Green, 1987; Sobel & Lange, 1996; Kruglyak & Lander, 1998).

Ao contrário do método dependente do modelo, os métodos independentes de modelo não dependem de uma especificação *a priori* do tipo de herança da característica. Nesses modelos, não é necessário estabelecer-se, *a priori*, frequências gênicas e penetrâncias, tendo em vista que funções dessas variáveis e a taxa de recombinação ( $\theta$ ) podem ser estimadas simultaneamente.

Para características descontínuas, os métodos independentes do modelo contemplam pares de irmãos afetados ou irmandades com pelo menos dois irmãos afetados. Nos estudos com grupos de dois irmãos afetados, o problema de penetrância incompleta é evitado, uma vez que não entram na amostra tanto os indivíduos não-portadores do mecanismo genético como aqueles que possuem a predisposição genética, porém não tendo manifestado a característica.

Se um determinado marcador e uma característica estiveram ligados geneticamente, as proporções de alelos IBD compartilhado ( $\pi$ )

excedem o valor esperado. Muitas análises estatísticas foram desenvolvidas para testar essa hipótese (Blackwelder & Elston, 1985; Risch, 1990; Kruglyak *et al.*, 1996). Para características contínuas, a semelhança fenotípica pode ser observada diretamente sem a necessidade de escolher irmandades. Sob a hipótese de ligação, irmãos que compartilham uma proporção maior de alelos IBD terão fenótipos mais parecidos, enquanto que sob a hipótese nula (sem ligação) os fenótipos dos pares serão independentes do compartilhamento de alelos, no gene marcador. Várias propostas diferentes foram feitas para estudar esses pares de irmãos e verificar a existência de genes principais influenciando características quantitativas (QTL, *Quantitative Trait Locus*) e o valor de  $\pi$  pode ser estimado através de várias técnicas (Hasegawa & Elston, 1972; Amos, 1994; Almasy & Blangero, 1998; Province *et al.*, 2001). Excelentes revisões recentes sobre ligação e características complexas podem ser encontradas em Olson *et al.* (1999), Elston (2000), e Borecki & Suarez (2001).

### Análise de associação

A análise de ligação pode ser usada em uma ampla procura por todo o genoma pela existência do *locus* responsável pela característica, por intermédio de sua localização cromossômica, enquanto que a análise de associação é mais útil para a confirmação de suspeita da participação de um determinado alelo na manifestação de uma certa característica ou de um alelo existente em um *locus* próximo ao *locus* responsável pela característica. Em outras palavras, a associação pode existir por duas razões: a) por causa do efeito direto do gene em uma característica em estudo ou, b) quando o marcador estiver em desequilíbrio de ligação com o gene principal responsável pela característica estudada. No primeiro caso, o genótipo pode ser determinado imunologicamente, por métodos eletroforéticos ou diretamente no DNA, sendo que seu efeito pode ser medido na característica estudada. Esse tipo foi chamado por Boerwinkle *et al.* (1986) de “genótipo medido”. No segundo caso, genes anônimos ao invés de marcadores funcionais estão presentes, e dessa forma o teste de associação requer a existência de desequilíbrio de ligação entre a característica e o marcador. Quando ocorre uma mutação causadora de uma doença, a mesma tem lugar

em um determinado cromossomo e forma um conjunto haplotípico com os *loci* adjacentes nesse cromossomo. Na geração seguinte, a tendência é que esse alelo mutante ocorra no mesmo haplótipo original, excetuando-se os casos de recombinação. Chama-se desequilíbrio de ligação, a ocorrência, na população, de uma frequência maior de uma determinada combinação entre dois genes, do que a esperada pelo produto de suas frequências individuais. Como visto no exemplo acima, o desequilíbrio da ligação é dependente da taxa de recombinação ( $\theta$ ) e quanto maior a taxa de recombinação, mais rápida será a aproximação do equilíbrio entre os dois *loci*.

Apesar de forças evolutivas outras que a mutação (*e.g.* deriva genética e fluxo gênico) atuarem para alterar o equilíbrio entre *loci*, é interessante notar que quanto mais antigo for o aparecimento da mutação, mais próxima a população vai estar no equilíbrio.

Existem várias maneiras de testar a existência de associação entre um gene marcador e uma determinada característica. O clássico caso-controle, na sua forma mais simples, é uma tabela de contingência 2 x 2. A hipótese nula a ser testada é o risco de um indivíduo apresentar a característica independente do marcador. Alguns fatores podem simular uma associação intrínseca e, na verdade, ela ser devida a fatores alheios à associação propriamente dita. Isso pode acontecer, por exemplo, no estudo de uma população estratificada e equivocadamente olhamos a mesma como sendo uniforme. Nesse caso a associação pode resultar da estratificação e não da biologia do processo. Dessa forma, se pudermos separar os diversos segmentos (estratos) da população e a associação for espúria, verificaremos que não existirá associação dentro de cada segmento, apenas no total da população. Uma das formas de se evitar uma associação devida à heterogeneidade da população é a utilização de tipos de controles familiares. Falk & Rubinstein (1987) propuseram um método para avaliar riscos relativos em estudos que usam controles familiares, embora uma década antes, Krieger & Barbosa (1979) tenham estudado os riscos relativos e as propriedades dependentes das frequências do marcador em pares de irmãos.

Vários testes estatísticos têm sido sugeridos para esse tipo de estudo, como o risco relativo haplotípico (Falk & Rubinstein, 1987; Terwilliger & Ott 1992), controles familiares afetados (Schaid & Summer, 1994) e o Teste de Desequi-

librio de Transmissão – TDT (Spielman *et al.*, 1993).

O TDT tem recebido atenção particular, pois se baseia tanto em informação de ligação como no desequilíbrio, que são subjacentes à associação. A análise molecular do marcador é importante para esse método, uma vez que visa a uma análise de alta sensibilidade de marcador. Tomam-se amostras de pessoas com uma determinada característica (casos) e estuda-se o marcador neles e em seus progenitores. Para o teste ser informativo, pelo menos um dos progenitores tem que ser heterozigoto para o *locus* marcador. O TDT usa um teste  $\chi^2$  de Mc Nemar (1947) para verificar a hipótese nula em que o alelo tido como associado à característica é transmitido em 50% das vezes pelo progenitor heterozigoto.

O método TDT foi modificado para lidar com situações mais complicadas, como marcadores multialélicos (Sham & Curtis, 1995; Bickel & Clérget-Darpoux, 1995; Rice *et al.*, 1995), quando apenas um progenitor é testado (Sun *et al.*, 1999) ou quando ambos os progenitores não podem ser testados, como por exemplo, características de aparecimento tardio (Spielman & Ewens, 1998; Hovarth & Laird, 1998; Boehnke & Langefeld, 1998). Nesse caso, a utilização de irmãos não-afetados fornece informação a respeito dos alelos que não é passada aos irmãos afetados.

Recentemente, foram também desenvolvidos métodos que abordam o problema de características quantitativas, utilizando modelos de componentes de variância por intermédio da similaridade familiar residual (Boerwinkle *et al.*, 1987; Almasy & Blangero, 1998; Province *et al.*, 2001). O método TDT original foi também ampliado para ser utilizado em características contínuas (Allison, 1997), em situações mais complexas como irmandades com mais de dois indivíduos, alelos múltiplos e em casos que não se referem a hipóteses paramétricas sobre a distribuição da característica (Rabinowitz, 1997). Alternativas que permitem uma abordagem através de regressão em dados genealógicos (George *et al.*, 1999) e vários graus de desequilíbrio de ligação entre o marcador e o *locus* responsável pela característica (Xiong, 1998) também foram propostas.

Em resumo, como o desequilíbrio de ligação deve persistir na maioria das populações humanas para genes ligados com uma taxa de recombinação menor que 1%, só em situações peculiares é que se tem a sorte de escolher um

marcador que preencha esse requisito. Por outro lado, estudos de associação ainda que sabidamente com excesso de resultados falsos positivos, podem fornecer informações importantes, uma vez que a “história” de recombinação é, de certa forma, utilizada nas análises, tendo em vista que o TDT resolve de maneira satisfatória os problemas causados por falsos positivos devido ao fluxo gênico.

Considerando o acima exposto, sugerimos que a procura de genes responsáveis por características complexas seja feita em primeiro plano com um estudo de ligação, com varredura genômica não muito sensível com a finalidade de se identificar as regiões cromossômicas potencialmente candidatas para alojar o gene condicionador das características. Após essa identificação deve se utilizar uma análise mais fina nas regiões promissoras usando densos mapas de polimorfismos de um único nucleotídeo (SNP).

Com essa abordagem, a utilização da informação sobre desequilíbrio de ligação na presença de uma associação positiva pode ser considerada um instrumento para mapeamento fino (Olson *et al.*, 1999; Borecki & Suarez, 2001 e Baron, 2001).

## Meta-análise

Meta-análise pode ser entendida como sendo a utilização de toda a informação existente sobre a possível identificação de um gene principal oriundo de vários estudos. Sabidamente, características complexas são identificadas pela existência de interações entre genes e ambiente, portanto o efeito de um gene em particular deve ser relativamente pequeno. Assim, a meta-análise se propõe a juntar sinais relativamente fracos de estudos particulares, em um conjunto mais robusto de evidências sobre efeitos genéticos, bem como permitindo um arcabouço quantitativo para modelar a variabilidade entre estudos. A aplicação desse tipo de análise, utilizando várias abordagens estatísticas (Olkin, 1995), tem sido relativamente frequente nos últimos anos (Li & Rao, 1996; Rice, 1998).

Diferentes estudos de características complexas variam entre si por numerosas razões: definições fenotípicas diferentes; averiguação e escolha tanto dos marcadores genéticos como da metodologia estatística. Dessa forma, diferentes conclusões a respeito de ligação podem ser devidas tanto ao delineamento do estudo,



como pela presença de diferentes fatores genéticos.

Contudo, se um esforço for despendido em rever todos os estudos e os artefatos forem cuidadosamente identificados, a meta-análise dos estudos disponíveis pode resultar em um conjunto poderoso para a detecção de genes com pequenos efeitos individuais.

É de especial interesse o estudo de meta-análise de resultado de ligação levada adiante por Gu *et al.* (1998), apresentando métodos para agregar resultados em uma análise de dados de pares de irmãos, usando a proporção de alelos que compartilham IBD no *locus* marcador e do efeito comum do possível *locus*. Efeitos aleatórios foram incluídos no modelo para caracterizar a variabilidade entre amostras, enquanto que outras estimativas (por quadrados mínimos, no caso) foram usadas para avaliar os prováveis efeitos do gene principal. Vários resultados suplementares puderam ser obtidos incorporando variáveis específicas em análises com co-variáveis que pudessem explicar a variabilidade entre estudos (Gu *et al.*, 1999, 2001).

Uma indicação de importância desse tipo de abordagem pode ser avaliada pelo interesse de firmas independentes em se associar a esse esforço. Existe uma firma dedicada à varredura genômica (Genoma Search Meta-Analysis/GSMA) que permite a agregação de novos dados a estruturas já existentes para integrar os novos dados a bancos de dados já atuantes (Wise *et al.*, 1999). Pormenores sobre meta-análise podem ser vistos em um recente trabalho de Gu *et al.*, (2001).

### Exemplos à guisa de ilustração

Alguns estudos evidenciaram a existência de um possível *locus* influenciando diversos fenótipos complexos (Comuzzie *et al.*, 2001; Baron, 2001) e nessa revisão, tomamos como exemplo o fenótipo Índice de Massa Corporal (BMI – Body Mass Index, que é o clássico índice gerado pelo quociente do peso corporal pelo quadrado da altura). O BMI é a influência de vários fatores causais, uma vez que representa a quantidade de massa corporal, a composição corporal e proporções, sendo certamente efeito de vários processos metabólicos, de efeitos hormonais, de comportamento e de interações entre esses fatores etiológicos de amplo espectro.

A maioria dos estudos estimou a herdabilidade do BMI como estando entre 40-55% (Bouchard *et al.*, 1998; Borecki *et al.*, 1998b e Rice *et al.*, 1999). Contudo, convém salientar que herdabilidades acima de 80% foram descritas em estudos de gêmeos. Apesar do BMI ser uma característica complexa, os estudos de segregação têm sugerido evidência de um ou dois genes controlando a sua variabilidade em diferentes populações/amostras (Province *et al.*, 1990; Hasstedt *et al.*, 1997; Borecki *et al.*, 1998b).

Uma varredura genômica usando componentes de variância baseados em uma análise de ligação multilócica para BMI indicou fortemente a existência de ligação no cromossomo 7q32.3 ( $\text{lod} = 4,9$ ,  $p < 10^{-5}$ ) no estudo familiar NHLBI – Family Heart Study (Feitosa *et al.*, 2002b). Prévios estudos já haviam sugerido evidência de ligação na região cromossômica do *locus* LEP com algumas características de medidas de obesidade (Duggirala *et al.*, 1996; Clément *et al.*, 1996; Roth *et al.*, 1997; Lapsys *et al.*, 1997; Bray *et al.*, 1999). Deve-se também salientar que uma meta-análise baseada em cinco estudos levada a cabo por Allison e Heo (1998) demonstrou forte evidência de ligação e associação do BMI à região LEP ( $p = 1.5 \times 10^{-5}$ ). Não se deve perder de vista, porém, que alguns estudos indicaram associações do BMI com *loci* pertencentes a outros cromossomos (Pérusse *et al.*, 2001).

### O futuro da epidemiologia genética

É claro que o delineamento de estudos ainda será o fator primordial dos estudos futuros e, é claro, também que dados familiares continuarão sendo o melhor arsenal para estudos nessa área.

Contudo, alguma substituição de amostras com controles familiares poderá ser necessária para cobrir lacunas na literatura. Estudos de caso-controle de indivíduos não-relacionados, especialmente para análises de dados com marcadores de DNA, poderão fornecer robustas evidências para estudos de ligação.

A heterogeneidade genética pode ser uma das facetas proeminentes dos estudos sobre características complexas. A amostragem de famílias de procedências étnicas diferentes serão de grande importância para desvendar mecanismos causais e também na identificação de fatores modificadores tanto genéticos como do

## Referências bibliográficas

- Allison DB & Heo M 1998. Meta-analysis of linkage data under worst-case conditions: a demonstration using the human OB region. *Genetics* 148:859-865.
- Allison DB 1997. Transmission-disequilibrium tests for quantitative traits. *Am. J. Hum. Genet.* 60(3):676-90.
- Almasy L & Blangero J 1998. Multipoint quantitative-trait linkage analysis in general pedigrees. *Am. J. Hum. Genet.* 62(5):1.198-1.211.
- Amos CI 1994. Robust variance-components approach for assessing genetic linkage in pedigrees. *Am. J. Hum. Genet.* 54(3):535-543.
- Baron M 2001. The search for complex disease genes: fault by linkage or fault by association? *Mol. Psychiatry* 6(2):143-149.
- Bickeboller H & Clérget-Darpoux F 1995. Statistical properties of the allelic and genotypic transmission/disequilibrium test for multiallelic markers. *Genet. Epidemiol.* 12(6):865-870.
- Blackwelder WC & Elston RC 1985. A comparison of sib-pair linkage tests for disease susceptibility loci. *Genet. Epidemiol.* 2(1):85-97.
- Blangero J & Konigsberg LW 1991. Multivariate segregation analysis using the mixed model. *Genet. Epidemiol.* 8(5):299-316.
- Blangero J, Williams-Blangero S, Mahaney MC, Comuzzie AG, Hixson JE, Samollow PB, Sharp RM, Stern MP & MacCluer JW 1996. Effects of a major gene for apolipoprotein A-I concentration are thyroid hormone dependent in Mexican Americans. *Arterioscler. Thromb. Vasc. Biol.* 16(9):1.177-1.183.
- Boehnke M & Langefeld CD 1998. Genetic association mapping based on discordant sib pairs: the discordant-alleles test. *Am. J. Hum. Genet.* 62(4):950-961.
- Boerwinkle E & Chakraborty R, Sing CF 1986. The use of measured genotype information in the analysis of quantitative phenotypes in man. I. Models and analytical methods. *Ann. Hum. Genet.* 50(Pt 2):181-194.
- Boerwinkle E, Visvikis S, Welsh D, Steinmetz J, Hanash SM, Sing CF 1987. The use of measured genotype information in the analysis of quantitative phenotypes in man. II. The role of the apolipoprotein E polymorphism in determining levels, variability, and co-variability of cholesterol, betalipoprotein, and triglycerides in a sample of unrelated individuals. *Am. J. Med. Genet.* 27(3):567-582.
- Bonney GE 1984. On the statistical determination of major gene mechanisms in continuous human traits: regressive models. *Am. J. Med. Genet.* 18(4):731-749.
- Borecki IB, Blangero J, Rice T, Perusse L, Bouchard C & Rao DC 1998a. Evidence for at least two major loci influencing human fatness. *Am. J. Hum. Genet.* 63(3): 831-838.
- Borecki IB, Higgins M, Schreiner PJ, Arnett DK, Mayer-Davis E, Hunt SC & Province MA 1998b. Evidence for multiple determinants of the body mass index: the National Heart, Lung, and Blood Institute Family Heart Study. *Obes. Res.* 6:107-114.
- Borecki IB & Suarez BK 2001. Linkage and association: basic concepts. *Adv. Genet.* 42:45-66.
- Bouchard C, Leon AS, Rao DC, Skinner JS, Wilmore JH & Gagnon J 1995. The Heritage family study. Aims, design, and measurement protocol. *Med. Sci. Sports Exerc.* 27:721-729.
- Bouchard C, Pérusse L, Rice T & Rao DC 1998. The genetics of human obesity. In *Handbook of Obesity*, GA. Bray, C Bouchard, and WPT James eds. Nova York: Marcel Dekker, Inc, pp. 157-190.
- Bray MS, Boerwinkle E & Hanis CL 1999. Linkage analysis of candidate obesity genes among the Mexican-American population of Starr County, Texas. *Genet. Epidemiol.* 16:397-411.
- Clément K, Garner C, Hager J, Philippi A, LeDuc C, Carey A, Harris TJ, Jury C, Cardon LR, Basdevant A, Demenais F, Guy-Grand B, North M & Froguel P 1996. Indication for linkage of the human OB gene region with extreme obesity. *Diabetes* 45:687-690.
- Coletto GMDD, Krieger H & Magalhães JR 1981. Estimates of the genetical and environmental determinants of serum lipid and lipoprotein concentrations in Brazilian twins. *Hum. Hered.* 31: 232-237.
- Comuzzie AG, Williams JT, Martin LJ & Blangero J 2001. Searching for genes underlying normal variation in human adiposity. *J. Mol. Med.* 79(1):57-70.
- Curtis D & Sham PC 1994. Using risk calculation to implement an extended relative pair analysis. *Ann. Hum. Genet.* 58( Pt 2):151-162.
- Duggirala R, Stern MP, Mitchell BD, Reinhart LJ, Shipman PA, Uresandi OC, Chung WK, Leibel RL, Hales CN, O'Connell P & Blangero J 1996. Quantitative variation in obesity-related traits and insulin precursors linked to the OB gene region on human chromosome 7. *Am. J. Hum. Genet.* 59:694-703.
- Elston RC & Stewart J 1971. A general model for the genetic analysis of pedigree data. *Hum. Hered.* 21(6): 523-542.
- Elston RC 2000. Introduction and overview. Statistical methods in genetic epidemiology. *Stat. Methods Med. Res.* 9(6):527-541.
- Falk CT & Rubinstein P 1987. Haplotype relative risks: an easy reliable way to construct a proper control sample for risk calculations. *Ann. Hum. Genet.* 51 (Pt 3):227-233.
- Feitosa MF, Borecki I, Krieger H, Beiguelman B & Rao DC 1995. The genetic epidemiology of leprosy in a Brazilian population. *Am. J. Hum. Genet.* 56: 1.179-1.185.
- Feitosa MF, Krieger H, Borecki I, Beiguelman B & Rao DC 1996. The genetic epidemiology of Mitsuda reaction in leprosy. *Hum. Hered.* 46:32-35.
- Feitosa MF, Rice T, Nirmala-Reddy A, Reddy PC & Rao DC 1999. Segregation analysis of regional fat distribution in families from Andhra Pradesh, India. *Int. J. Obes Relat. Metab. Disord.* 23(8):874-880.
- Feitosa MF, Rice T, Nirmala A & Rao DC 2000a. Major gene effect on body mass index: the role of energy intake and energy expenditure. *Hum. Biol.* 72(5):781-799.
- Feitosa MF, Borecki I, Hunt SC, Arnett DK, Rao DC & Province M 2000b. Inheritance of the waist-to-hip ratio in the National Heart, Lung, and Blood Institute Family Heart Study. *Obes. Res.* 8(4):294-301.
- Feitosa MF, Rice T, Rosmond R, Borecki IB, An P, Gagnon J, Leon AS, Skinner JS, Wilmore JH, Bouchard C & Rao DC 2002a. A genetic study of cortisol measured before and after endurance training: The Heritage Family Study. *Metabolism*(no prelo).

- Feitosa MF, Borecki I, Rich SS, Arnett DK, Rao DC, Phyllis Sholinsky P, Myers RH, Leppert M & Province M 2002b. Quantitative Trait Loci Influencing Body Mass Index Resides on Chromosomes 7 and 13: The National Heart, Lung, and Blood Institute Family Heart Study (FHS). *Am. J. Hum. Genet.* 7:72-82.
- George V, Tiwari HK, Zhu X & Elston RC 1999. A test of transmission/disequilibrium for quantitative traits in pedigree data, by multiple regression. *Am. J. Hum. Genet.* 65(1):236-245.
- Gu C, Province M, Todorov A & Rao DC 1998. Meta-analysis methodology for combining non-parametric sibpair linkage results: genetic homogeneity and identical markers. *Genet. Epidemiol.* 15(6):609-626.
- Gu C, Province M & Rao DC 1999. A meta-analysis approach for pooling sibpair linkage studies with study-specific covariates: mixed effects models. *Genet. Epidemiol.* 17(suppl. 1):599-604.
- Gu C, Province M & Rao DC 2001. Meta-analysis for model-free methods. *Adv. Genet.* 42:255-272.
- Gu C & Rao DC 2001. Optimum study designs. *Adv. Genet.* 42:439-457.
- Haseman JK & Elston RC 1972. The investigation of linkage between a quantitative trait and a marker locus. *Behav. Genet.* 2(1):3-19.
- Hasstedt SJ 1982. A mixed-model likelihood approximation on large pedigrees. *Comput. Biomed. Res.* 15(3): 295-307.
- Hasstedt SJ, Hoffman M, Leppert MF & Elbein SC 1997. Recessive inheritance of obesity in familial non-insulin-dependent diabetes mellitus, and lack of linkage to nine candidate genes. *Am. J. Hum. Genet.* 61: 668-677.
- Higgins M, Province M, Heiss G, Eckfeldt J, Ellison RC, Folsom AR, Rao DC, Sprafka JM & Williams R 1996. NHLBI family heart study: objectives and design. *Am. J. Epidemiol.* 143(12):1.219-1.228.
- Hopper JL & Mathews JD 1982. Extensions to multivariate normal models for pedigree analysis. *Ann. Hum. Genet.* 46 (4):373-383.
- Horvath S & Laird NM 1998. A discordant-sibship test for disequilibrium and linkage: no need for parental data. *Am. J. Hum. Genet.* 63(6):1.886-1.897.
- Krieger H & Barbosa CAA 1979. Smallpox and the ABO system association: a critical review. *Rev. Brasil. Biol.* 39(1):195-199.
- Krieger H & Feitosa MF 1999. Genetic epidemiology of infectious disease. *Ciência e Cultura: Journal of the Brazilian Association for the Advancement of Science* 51:191-198.
- Kruglyak L & Lander ES 1998. Faster multipoint linkage analysis using Fourier transforms. *J. Comput. Biol.* 5(1):1-7.
- Kruglyak L, Daly MJ, Reeve-Daly MP & Lander ES 1996. Parametric and nonparametric linkage analysis: a unified multipoint approach. *Am. J. Hum. Genet.* 58(6):1.347-1.363.
- Lalouel JM, Rao DC, Morton NE & Elston RC 1983. A unified model for complex segregation analysis. *Am. J. Hum. Genet.* 35(5):816-26.
- Lander ES & Green P 1987. Construction of multilocus genetic linkage maps in humans. *Proc. Natl. Acad. Sci. USA.* 84(8):2.363-2.367.
- Lapsys NM, Furler SM, Moore KR, Nguyen TV, Herzog H, Howard G, Samaras K, Carey DG, Morrison NA, Eisman JA & Chisholm DJ 1997. Relationship of a novel polymorphic marker near the human obese (OB) gene to fat mass in healthy women. *Obes. Res.* 5:430-433.
- Li Z & Rao DC 1996. Random effects model for meta-analysis of multiple quantitative sibpair linkage studies. *Genet. Epidemiol.* 13(4):377-383.
- McNemar Q 1947. Note on the sampling error of the difference between correlated proportions or percentages. *Psychometrika* 12:153-157.
- Morton NE & MacLean CJ 1974. Analysis of family resemblance. 3. Complex segregation of quantitative traits. *Am. J. Hum. Genet.* 26(4):489-503.
- Morton NE 1955. Sequential tests for the detection of linkage. *Am. J. Hum. Genet.* 7:277-318.
- Morton NE, Rao DC & Lalouel JM 1983. *Methods in genetic epidemiology.* (Contributions to genetic epidemiology and biostatistics; v. 4). Ed. Basel, Nova York: Karger. 261 pp.
- Olkin I 1995. Meta-analysis: reconciling the results of independent studies. *Stat. Med.* 14(5-7):457-472.
- Olson JM, Witte JS, Elston RC 1999. Genetic mapping of complex traits. *Stat. Med.* 18(21):2.961-2.981.
- Olson JE, Atwood LD, Grabrick DM, Vachon CM, Sellers TA 2001. Evidence for a major gene influence on abdominal fat distribution: the Minnesota Breast Cancer Family Study. *Genet. Epidemiol.* 20(4):458-478.
- Ott J 1974. Estimation of the recombination fraction in human pedigrees: efficient computation of the likelihood for human linkage studies. *Am. J. Hum. Genet.* 26(5):588-597.
- Penrose LS 1935. The detection of autosomal linkage in data which consist of pairs of brothers and sisters of unspecified parentage. *Ann. Eugen. (London)* 6:133-138.
- Pérusse L, Chagnon YC, Weisnagel SJ, Rankinen T, Snyder E, Sands J & Bouchard C 2001. The human obesity gene map: the 2000 update. *Obes. Res.* 9(2):135-169.
- Prado RAR 1992. *Curriculum vitae del estudio colaborativo latinoamericano de malformaciones congénitas (ECLAMC), 1967-1992.* Universidade Federal do Rio de Janeiro, Rio de Janeiro. 188 pp.
- Province MA, Arnqvist P, Keller J, Higgins M & Rao DC 1990. Strong evidence for a major gene for obesity in the large unselected total community healthy study of Tecumseh. *Am. J. Hum. Genet.* 47(suppl.):A143.
- Province MA & Rao DC 1995. A general purpose model and a computer program for combined segregation and path analysis (SEGPATH): automatically creating computer programs from symbolic language model specifications. *Genet. Epidemiol.* 12:203-219.
- Province MA, Rice T, Borecki IB, Gu C & Rao DC 2001. A multivariate and multilocus variance components approach using structural relationships to assess quantitative trait linkage via SEGPATH. *Genet. Epidemiol.* (no prelo).
- Rabinowitz D 1997. A transmission disequilibrium test for quantitative trait loci. *Hum. Hered.* 47(6):342-350.
- Rao DC, Morton NE & Yee S 1974. Analysis of family resemblance. II. A linear model for familial correlation. *Am. J. Hum. Genet.* 26(3):331-59.
- Rao DC 2001. Genetic dissection of complex traits: an

- overview. *Adv. Genet.* 42:13-34.
- Rice JP, Neuman RJ, Hoshaw SL, Daw EW & Gu C 1995. TDT with covariates and genomic screens with mod scores: their behavior on simulated data. *Genet. Epidemiol.* 12(6):659-664.
- Rice JP 1998. The role of meta-analysis in linkage studies of complex traits. *Am. J. Med. Genet.* 74:112-114.
- Rice JP, Saccone NL & Rasmussen E 2001. Definition of the phenotype. *Adv. Genet.* 42:69-76.
- Rice T, Perusse L, Bouchard C & Rao DC 1999. Familial aggregation of body mass index and subcutaneous fat measures in the longitudinal Quebec family study. *Genet. Epidemiol.* 16(3):316-34.
- Rice T & Borecki IB 2001. Familial resemblance and heritability. *Adv. Genet.* 42:35-44.
- Risch N 1990. Linkage strategies for genetically complex traits. I. Multilocus models. *Am. J. Hum. Genet.* 46(2):222-228.
- Roth H, Hinney A, Ziegler A, Barth N, Gerber G, Stein K, Bromel T, Mayer H, Siegfried W, Schafer H, Remschmidt H, Grzeschik KH & Hebebrand J 1997. Further support for linkage of extreme obesity to the obese gene in a study group of obese children and adolescents. *Exp. Clin. Endocrinol. Diabetes* 105:341-344.
- Schaid DJ & Sommer SS 1994. Comparison of statistics for candidate-gene association studies using cases and parents. *Am. J. Hum. Genet.* 55(2):402-409.
- Sham PC & Curtis D 1995. An extended transmission/disequilibrium test (TDT) for multi-allele marker loci. *Ann. Hum. Genet.* 59 (Pt 3):323-336.
- Sobel E & Lange K 1996. Descent graphs in pedigree analysis: applications to haplotyping, location scores, and marker-sharing statistics. *Am. J. Hum. Genet.* 58(6):1.323-1.337.
- Spielman RS, McGinnis RE & Ewens WJ 1993. Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). *Am. J. Hum. Genet.* 52(3):506-516.
- Spielman RS & Ewens WJ 1998. A sibship test for linkage in the presence of association: the sib transmission/disequilibrium test. *Am. J. Hum. Genet.* 62(2):450-458.
- Sun F, Flanders WD, Yang Q & Khoury MJ 1999. Transmission disequilibrium test (TDT) when only one parent is available: the 1-TDT. *Am. J. Epidemiol.* 150(1):97-104.
- Terwilliger JD & Ott J 1992. A haplotype-based 'haplotype relative risk' approach to detecting allelic associations. *Hum. Hered.* 42(6):337-346.
- Weeks DE & Lange K 1988. The affected-pedigree-member method of linkage analysis. *Am. J. Hum. Genet.* 42(2):315-326.
- Wise LH, Lanchbury JS & Lewis CM 1999. Meta-analysis of genome searches. *Ann. Hum. Genet.* 63(Pt 3):263-272.
- Xiong MM, Krushkal J & Boerwinkle E 1998. TDT statistics for mapping quantitative trait loci. *Ann. Hum. Genet.* 62 (Pt 5):431-452.