



Modelos de estudos genéticos para doenças complexas

Prof^a Dr^a Sintia Iole Belangero

Curso de Introdução à Bioinformática Aplicada à Genética

Desenhos de estudos

- Estudos experimentais
- Estudos de gêmeos
- Estudos de adoção
- Estudos de famílias/trios
- Caso x Controle

Metodologias

- Estudos de ligação
- Estudos de associação
- Estudos de varredura genômica ou análise exploratória

Estudos experimentais

- Quando tem interferência do pesquisador (contrário a estudo observacional)
- Transversal
 - Difícil estabelecer relação causa e efeito
- Seguimento
 - Há um interesse ao longo do tempo

Estudos experimentais

- Importante ter uma pergunta
 - “Quem não sabe o que procura, não interpreta o que acha”.
- Definir o método
- Escolher análise estatística adequada
- Definir tamanho amostral

Estudo de gêmeos

- Comparar concordância entre gêmeos MZ e DZ
 - MZ: 100% de compartilhamento genético
 - DZ: média de 50%
- Concordância entre MZ > DZ : > componente genético



Estudo de gêmeos

- Herdabilidade (traços quantitativos)

$$h^2 = \frac{\sigma^2_{DZ} - \sigma^2_{MZ}}{\sigma^2_{DZ}}$$

Onde:

σ^2_{DZ} representa a variância da diferença entre pares DZ.
 σ^2_{MZ} representa a variância da diferença entre pares MZ.

$$h^2 = 0.0-1.0$$

Estudo de gêmeos

- Herdabilidade (traços qualitativos)

$$h^2 = \frac{C_{MZ} - C_{DZ}}{1 - C_{DZ}}$$

Onde:

C_{MZ} representa a taxa de concordância (quantos pares concordam/total de pares) entre pares MZ.

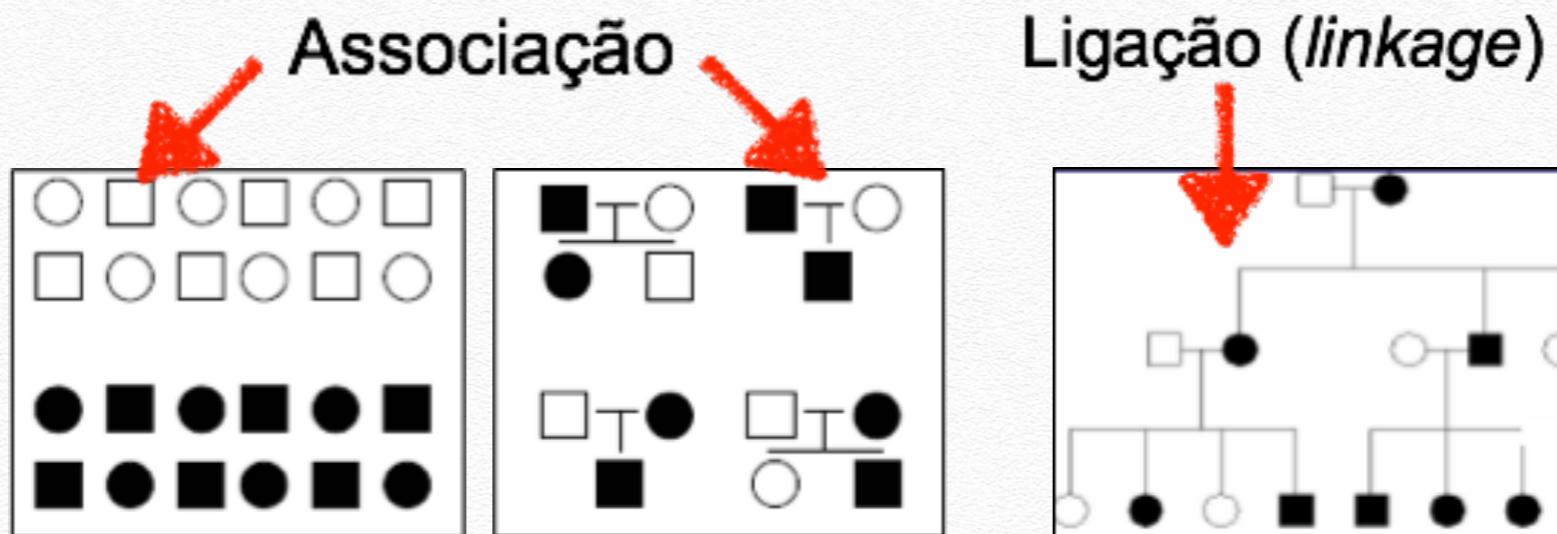
C_{DZ} representa a taxa de concordância entre pares DZ.

Estudo de adoção

- Recorrência em família de indivíduos com pais adotados X biológicos
- Maior frequência do caráter em filhos biológicos: > risco genético
- Maior frequência do caráter em filhos adotados: > risco ambiente

Estudos genéticos

- Duas abordagens principais



Estudo de ligação

- Identificação de alelos ou segmentos cromossômicos transmitidos juntamente com a doença de geração em geração (marcador)
- Ocorre devido proximidade física entre o marcador e o causador
- Marcador que está ligado não necessariamente é o responsável pelo fenótipo, mas é herdado junto com o responsável

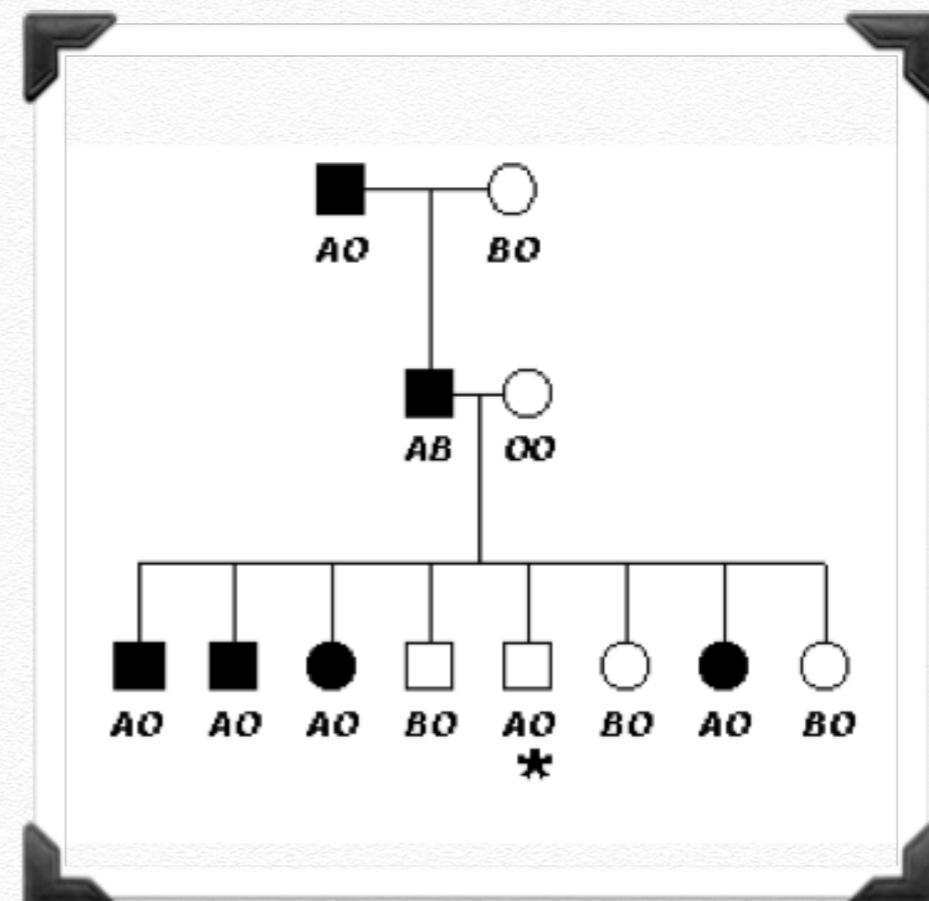
Estudo de ligação

- 2 tipos de estudos de ligação
 - Paramétrico – LOD Scores - precisa de famílias com muitos afetados
 - Não Paramétricos - muitas famílias pequenas

Estudos de Ligação

LOD-Score

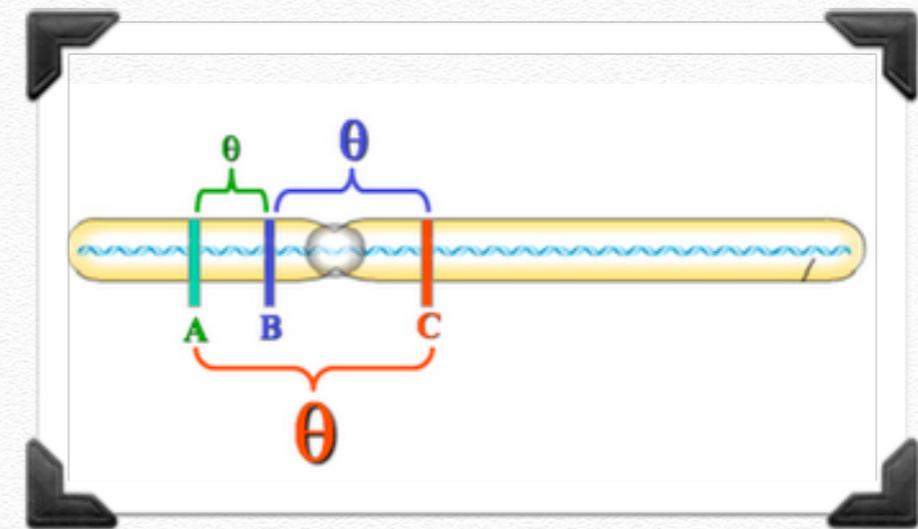
- Taxa de recombinação= 1/9



Estudos de Ligação

LOD-Score

- Depende da recombinação entre o locus e o marcador
- Frequência de recombinação (Θ) – Depende da distância entre 2 loci
 - $\Theta = 0\%$ sem recombinação (há ligação)
 - $\Theta = 1-49\%$ pequena recombinação (há ligação)
 - $\Theta > 50\%$ recombinação (não há ligação)



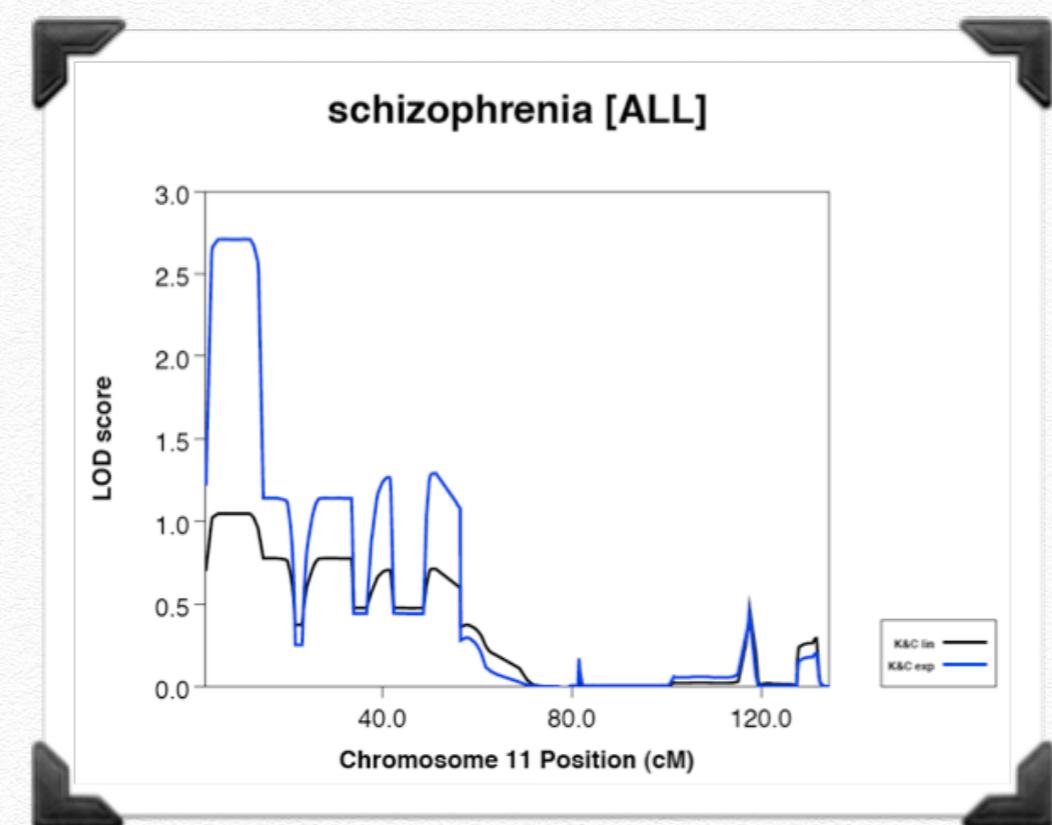
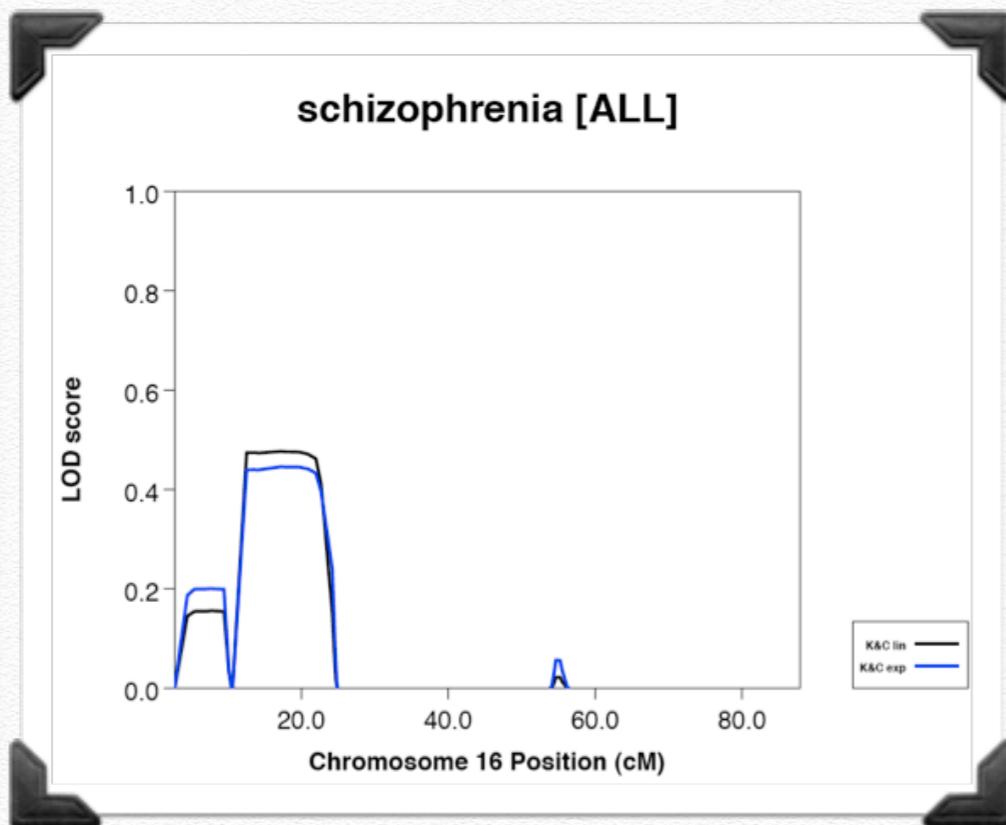
$$\Theta = 1\% \text{ } 1\text{cM} \text{ } 1\text{Mb}$$

Estudo de ligação

$$LOD = Z = \log_{10} \frac{\text{probability of birth sequence with a given linkage value}}{\text{probability of birth sequence with no linkage}}$$
$$= \log_{10} \frac{(1 - \theta)^{NR} \times \theta^R}{\theta^{NR+R}}$$

- LOD score (Z) se > 2-3 indica ligação

Estudo de ligação

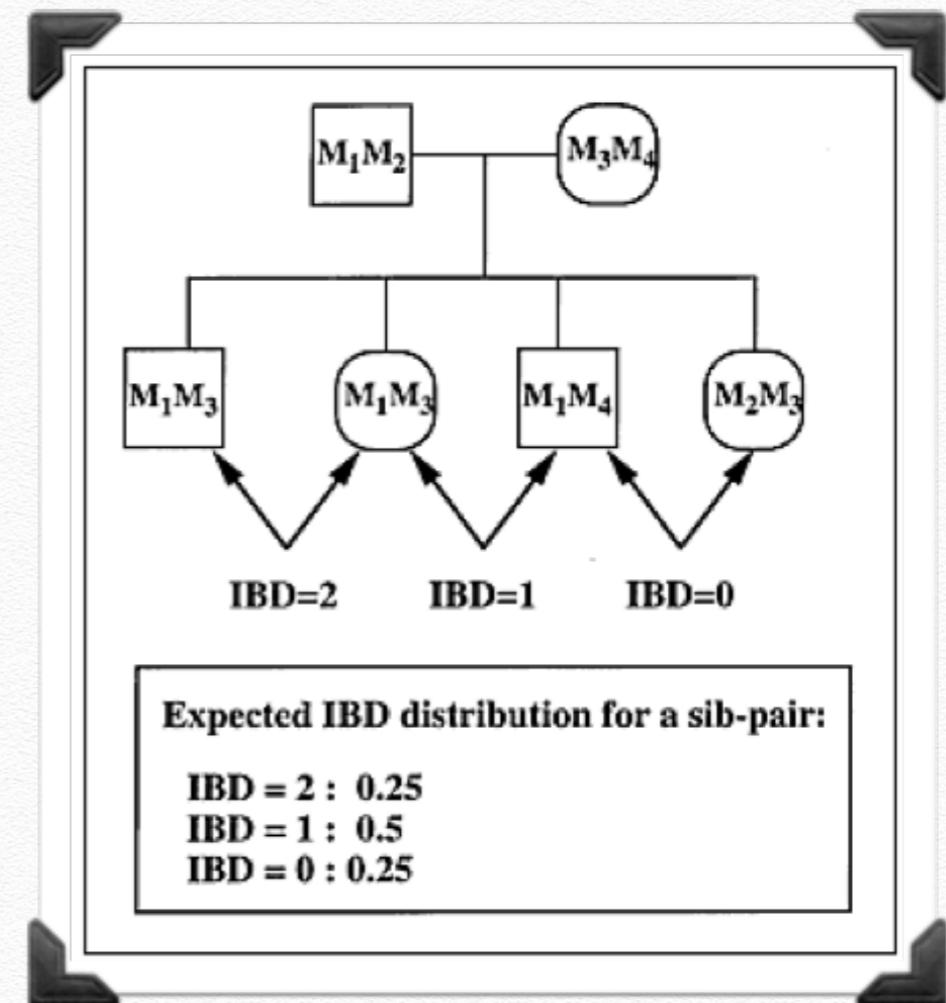


Estudos de ligação

Compartilhamento de genes

- Melhor para doenças complexas (>poder estatístico)
- Irmãos afetados X não afetados
(que compartilham o alelo marcador)
- IBD (Identity By Descendent)
 - Chance de compartilhar 1: 0,5
 - Chance de compartilhar 0 ou 2: 0,25
- Observado x Esperado

Distribuição esperada



Estudos de Associação

- Co-ocorrência estatisticamente significante de alelos e fenótipos
- Pode ser causal, por seleção natural, estratificação populacional ou ao acaso
- 2 tipos principais
 - Populacionais (caso-controle)
 - Baseados em família

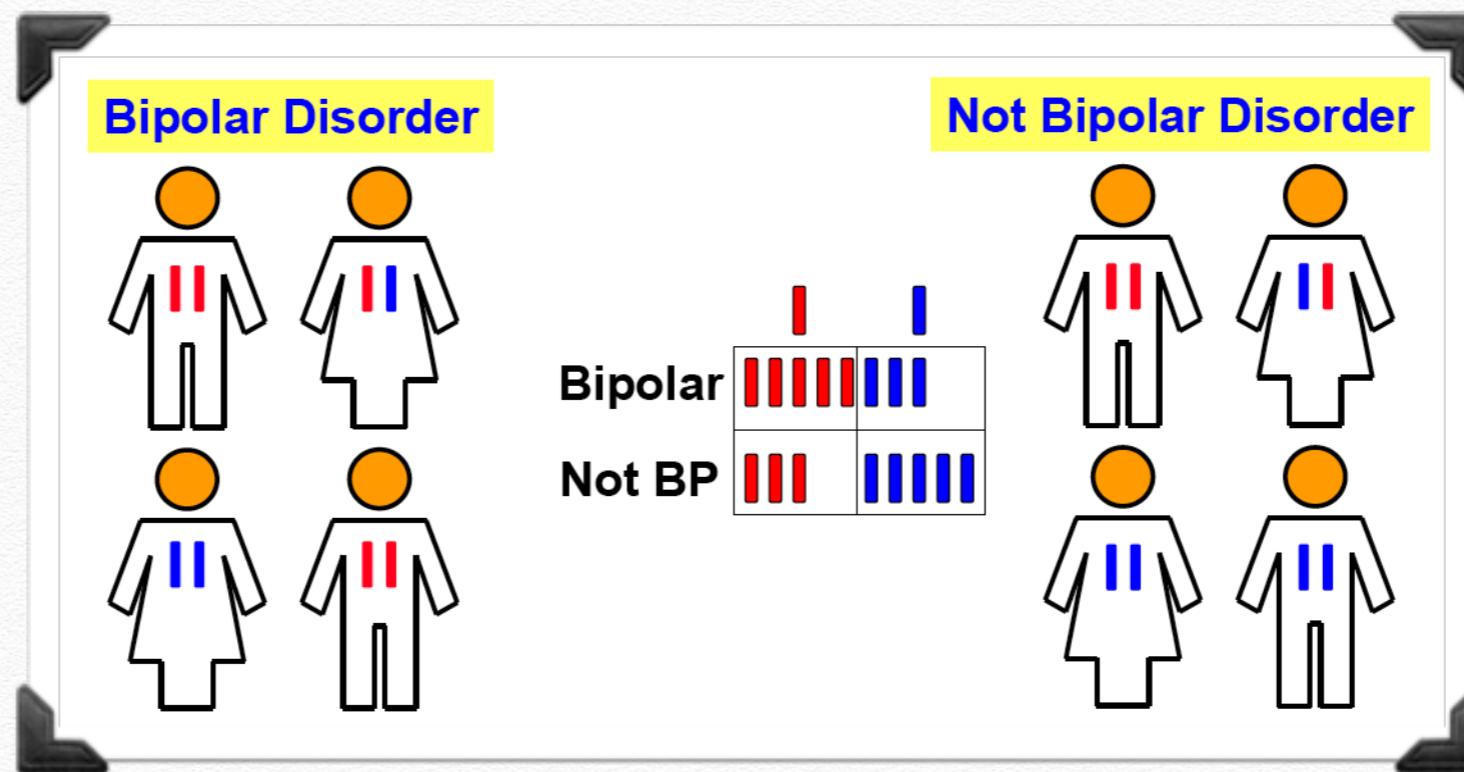
Estudos de Associação

- Populacionais
 - compara frequência entre casos e controles; porém dificuldade de se obter controles fiéis
- Baseados em famílias
 - Família: mais robusto, não afetado pela estratificação populacional

$$\chi^2_c = \sum \frac{(O_i - E_i)^2}{E_i}$$

p<0,05

Estudos de Associação



- comparação de frequências alélica
- populações doentes X não doentes
- Associação entre alelos e fenótipos : puramente estatística

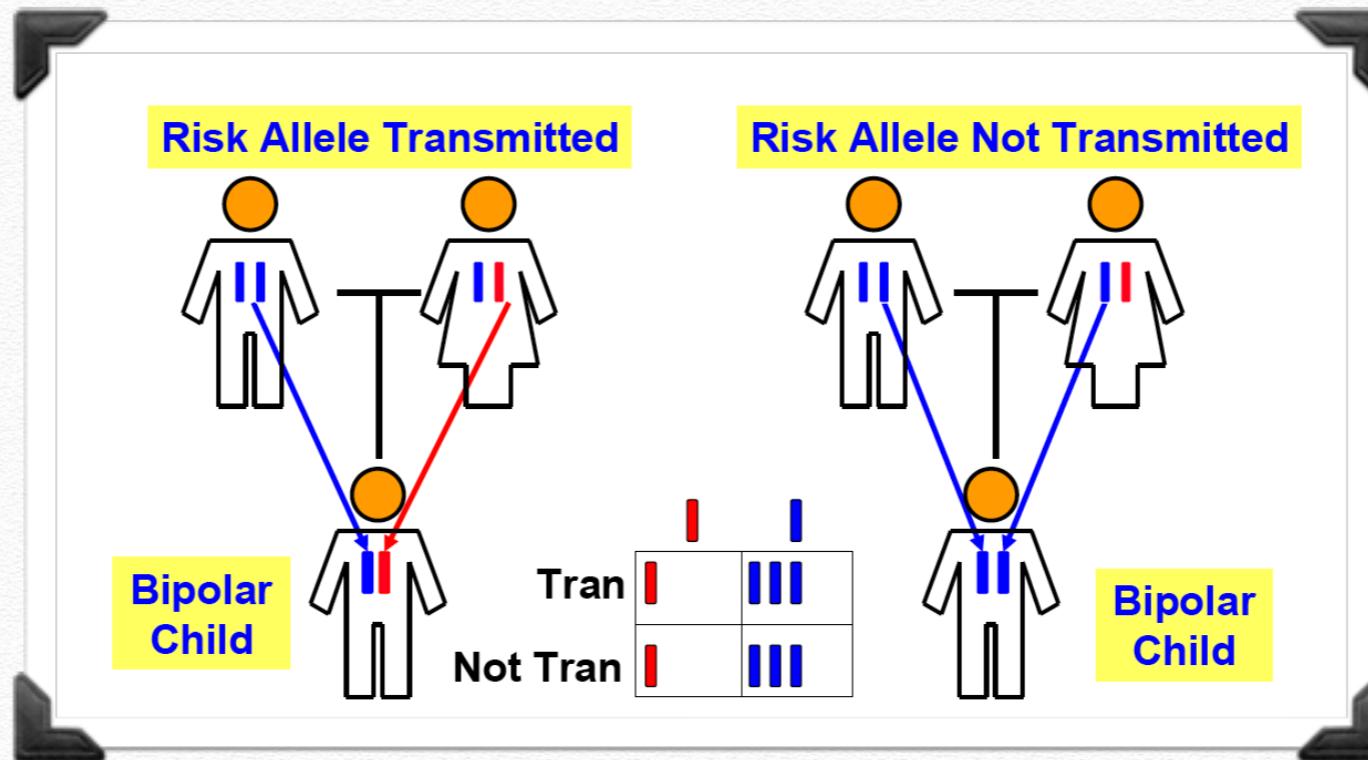
Estudos de Associação

- Desvantagens Caso-controle
- Viés pelas diferenças étnicas
- Necessidade de N amostrais grandes

Teste de desequilíbrio de transmissão (TDT)

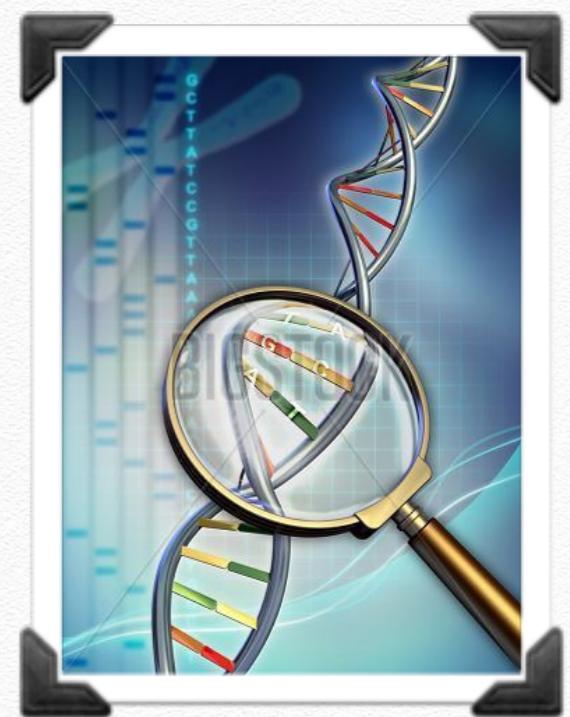
- Associação na presença de ligação
- Baseia-se tanto em princípios de ligação como em desequilíbrio de transmissão
- Determina se a transmissão para o probando é estatisticamente diferente do esperado pelas leis mendelianas
- Os resultados do TDT são dados sob a forma de p (<0,05 são significativos)
- Não precisa de famílias grandes, mas de um grande número de famílias
- Funciona melhor se um dos genitores for heterozigoto para o locus marcador

Teste de desequilíbrio de transmissão (TDT)



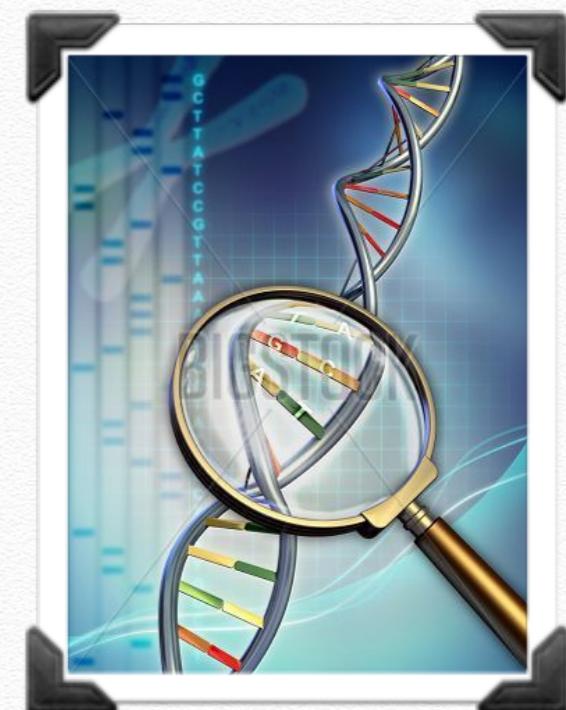
Análise exploratória ou de varredura

- GWAS (Genome Wide Association Studies) - SNP-array e variações
- Expressão gênica em larga escala - Microarray de expressão gênica
- Sequenciamento completo, exoma, RNAseq e variações



Análise exploratória ou de varredura

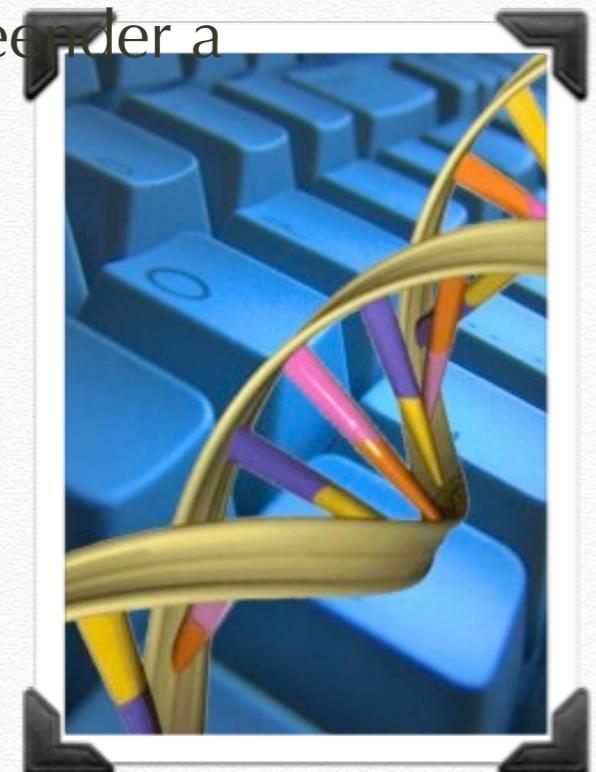
- Busca sem hipótese prévia associada à doença
- Análise exploratória
- Milhares de marcadores simultaneamente
- Sempre acompanhado de correção para múltiplas comparações



Análise exploratória ou de varredura

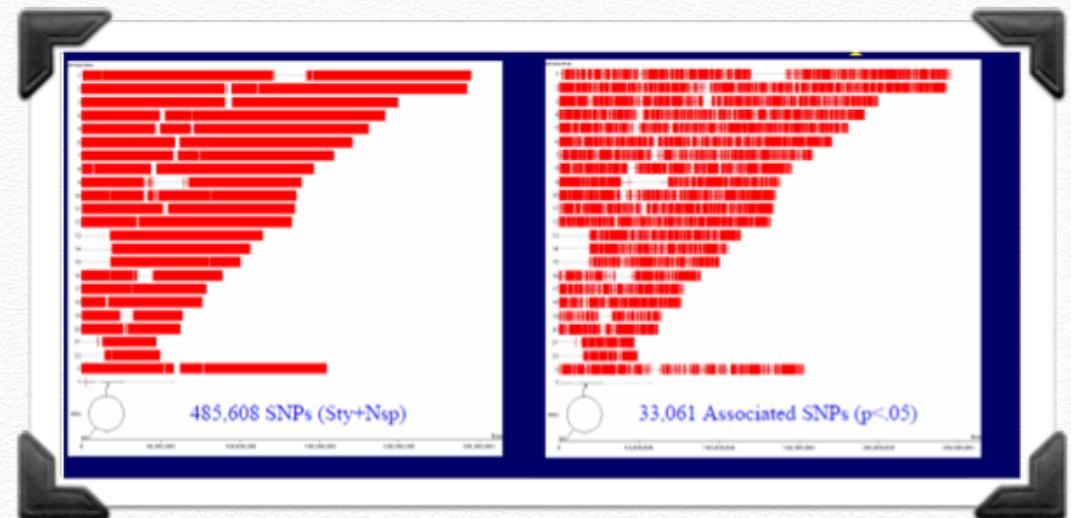
Fusão entre a bioinformática, a bioestatística e a genética molecular

- encontrar regiões ou genes relevantes
- estudar a função biológica desses genes para compreender a fisiopatologia molecular da doença



Limitações da análise exploratória

- GWAS: detecta associação entre amostras de casos e controles
- ~5 % de todos os SNPs têm $p < 0,05$
- Cuidado com os falsos positivos!
- Necessários níveis de significância elevados
- $p < 5 \times 10^{-8}$
- Confirmação com outras metodologias (replicata técnica) e outras populações (replicata biológica)
- Outros estudos que confirmem: reproduzibilidade



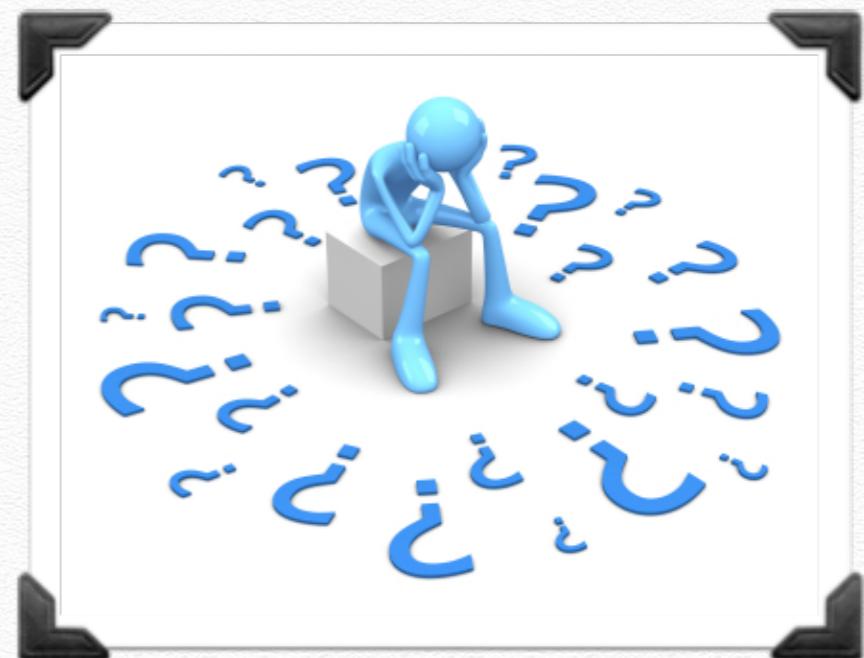
Limitações da análise exploratória

- GWAS
 - não é a resposta final nem a última ferramenta para descobrir variantes e regiões candidatas
 - É preciso muito mais
 - Interações gênicas e redes provavelmente são mais importantes



Como aumentar o poder das análises genéticas??

- Melhorando as populações de estudo
- Melhorando a abordagem
- Melhorando as ferramentas



Como aumentar o poder das análises genéticas??



- Refinamento do fenótipo
- gerar fenótipos mais homogêneos clinicamente

Como aumentar o poder das análises genéticas??



Aumentando número amostral

- Principalmente estudos de associação
- Minimiza todos os viéses inerentes ao desenho do estudo e aumenta a chance de se encontrar variantes alélicas raras com efeito fenotípico da variante.

Como aumentar o poder das análises genéticas??



N em genética depende:

- 1) a frequência alélica das variantes de interesse nos indivíduos afetados ($< \text{freq} > N$)
- 2) a prevalência da doença na população
- 3) o poder estatístico desejado ($> \text{poder} > N$)
- 4) o índice de significância adotado (α) ($< \alpha > N$)
- 5) o tamanho de efeito ($< \text{TE} > N$)

Como aumentar o poder das análises genéticas??



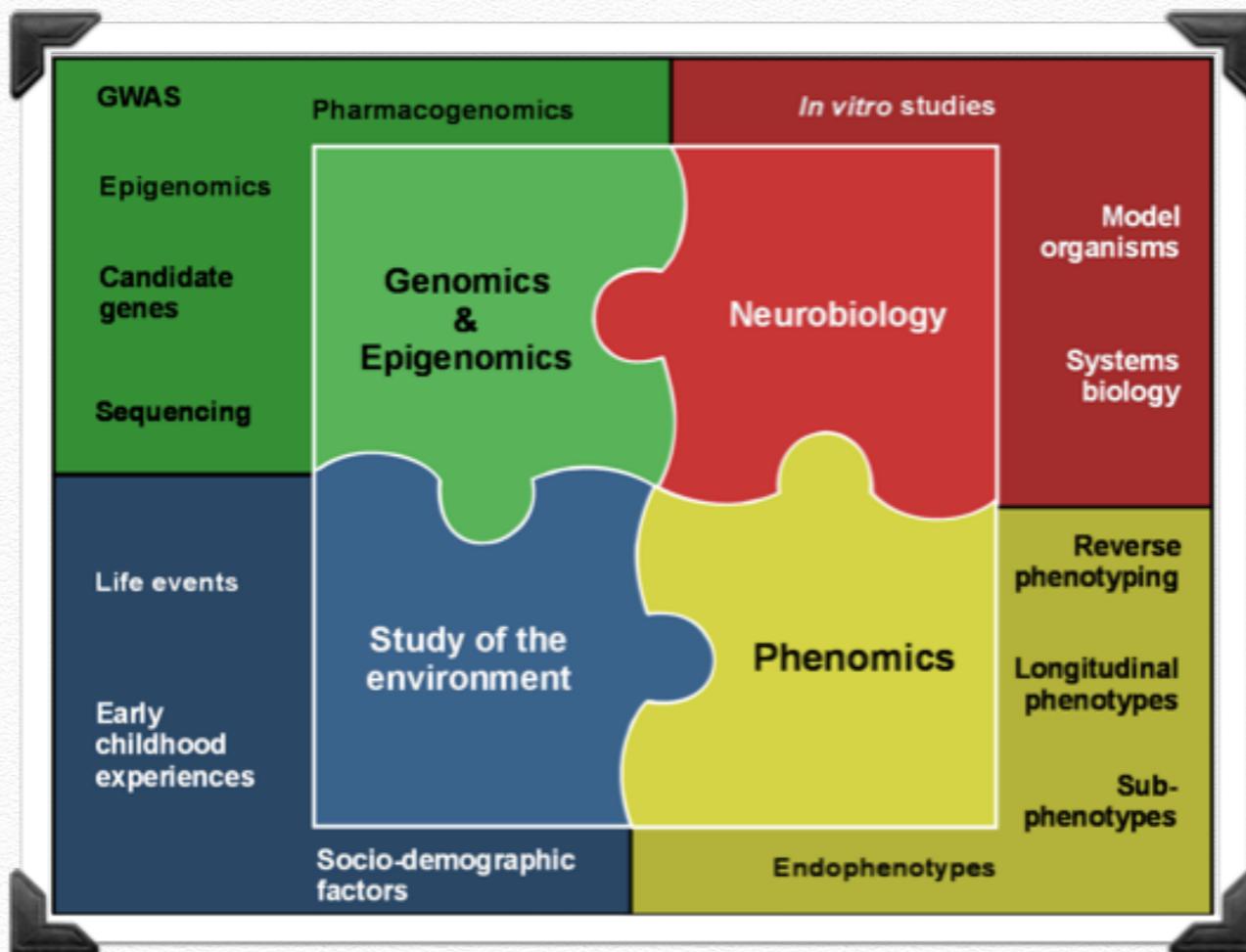
Evitando a estratificação populacional

- a maioria das populações é heterogênea geneticamente, muito mais do que parece à primeira vista
- associações espúrias, já que as associações observadas seriam devido a presença de subpopulações diferentes em uma mesma amostra e não devido a um efeito das variantes genéticas na etiologia da doença
- Para reduzir o viés de uma possível estratificação populacional, é aconselhável utilizar casos e controles com a mesma ancestralidade. Tem sido sugerido que marcadores moleculares podem ser utilizados para avaliar mais acuradamente a ancestralidade dos indivíduos em estudos caso-controle

Como aumentar o poder das análises genéticas??



- Realizando investigações multidisciplinares/multimodais



Fim

Sintia Iole Belangero