

GWAS



Laboratório de Biologia Molecular do Sono
Departamento de Psicobiologia
Universidade Federal de São Paulo

Molecular Core

Associação Fundo de Incentivo à Pesquisa
(AFIP)

E-mail: mazzottidr@gmail.com

Variações genéticas em humanos

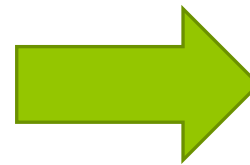
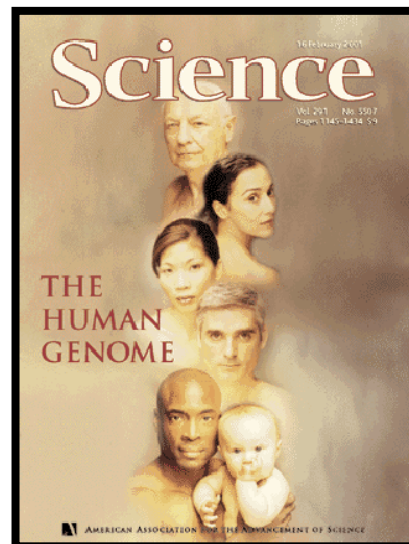


Variações genéticas em humanos

- Fontes de variação genética



Projeto
Genoma
Humano

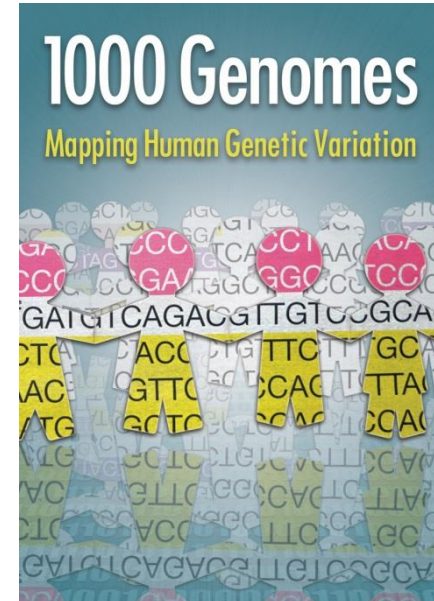


Variações genéticas em humanos

- Projetos para caracterização de variações no genoma humano



Era da genotipagem em larga escala



Era do sequenciamento de nova geração

Variações genéticas em humanos

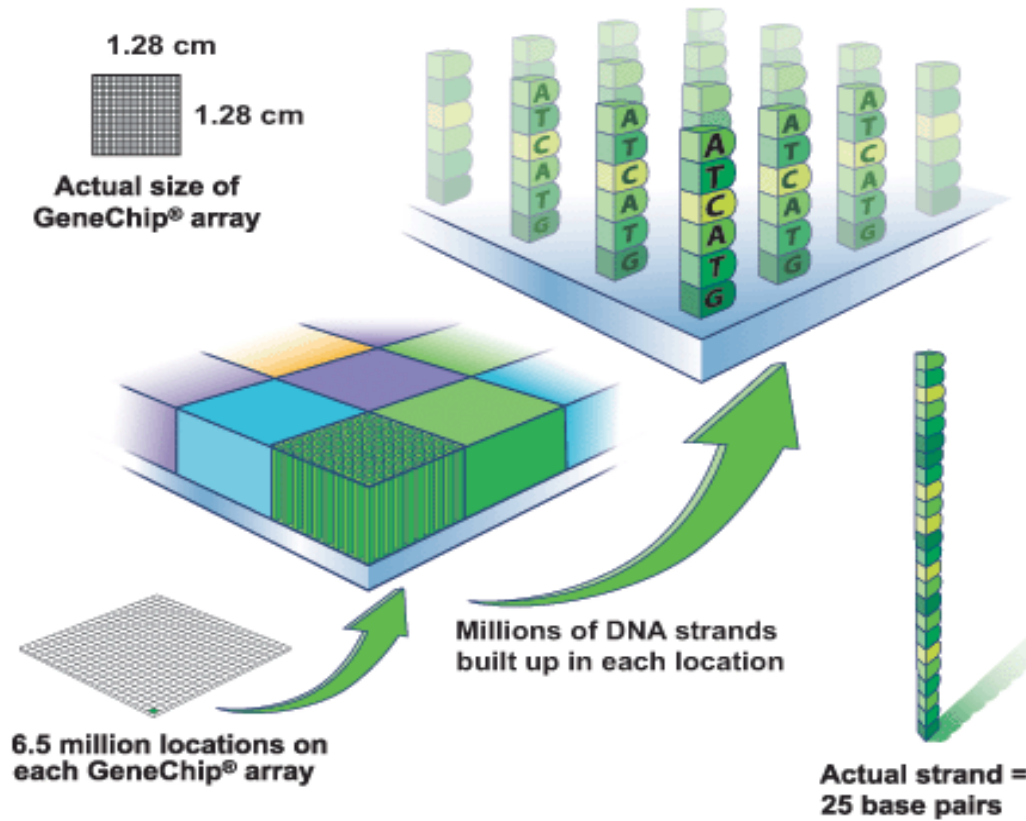
- ◉ Quais são as fontes de variação genética?
 - ◉ *Single nucleotide polymorphisms* (SNPs)
 - ◉ *Variable Number of Tandem Repeats* (VNTRs)
 - ◉ Microssatélites (STRs)
 - ◉ Variações estruturais
 - ◉ Inserções, deleções, inversões, duplicações, translocações
 - ◉ *Copy Number Variations* (CNVs)

Variações genéticas em humanos

- Como medir essas variações?
 - Genes candidatos (poucos de cada vez)
 - PCR-RFLP (SNPs)
 - *TaqMan Real Time PCR* (SNPs e CNVs)
 - MLPA (SNPs e CNVs)
 - ...
 - Abordagem “Genome-Wide” (muitos de uma vez!)
 - **DNA microarrays** (SNPs e CNVs)

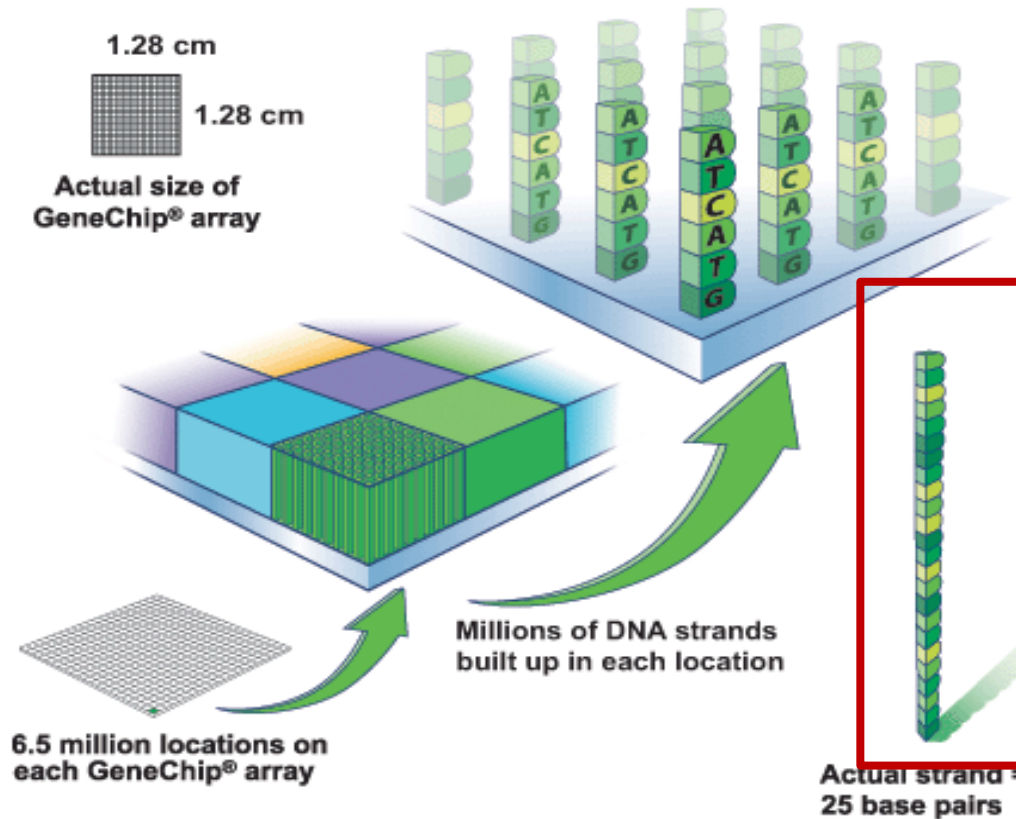
DNA microarrays

From Computer Desktop Encyclopedia
Reproduced with permission.
© 2007 Affymetrix



DNA microarrays

From Computer Desktop Encyclopedia
Reproduced with permission.
© 2007 Affymetrix



Duas sondas,
uma específica
para cada alelo
do SNP

DNA *microarrays*

Sequência Ref. **ATCGGTAGCCAT^YCATGAGTTACTA**

Sonda 1: Alelo A **ATCGGTAGCCAT^TCATGAGTTACTA**

Sonda 2: Alelo B **ATCGGTAGCCAT^CCATGAGTTACTA**

DNA microarrays

Sequência Ref. **ATCGGTAGCCAT****Y**CATGAGTTACTA

Sonda 1: Alelo A **ATCGGTAGCCAT****T**CATGAGTTACTA

Sonda 2: Alelo B **ATCGGTAGCCAT****C**CATGAGTTACTA

Perfect Match



Mismatch



Allele 'A'

Perfect Match



Mismatch

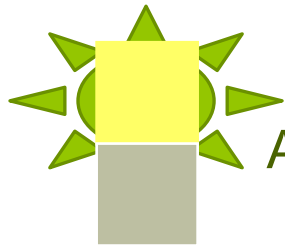


Allele 'B'

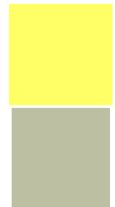


Quatro spots por SNP
(Quarteto)

DNA *microarrays*



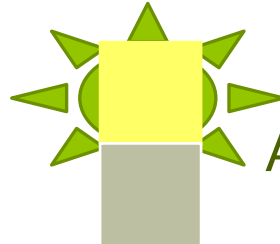
Allele 'A'



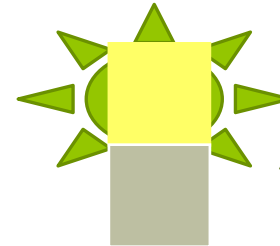
Allele 'B'



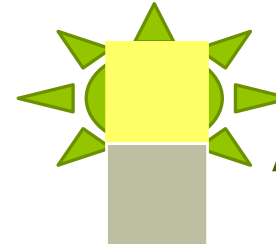
Allele 'A'



Allele 'B'

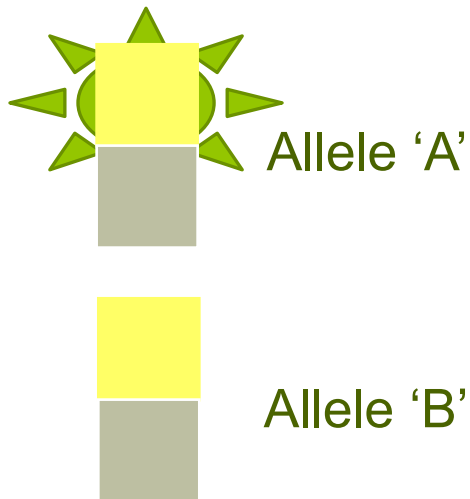


Allele 'A'

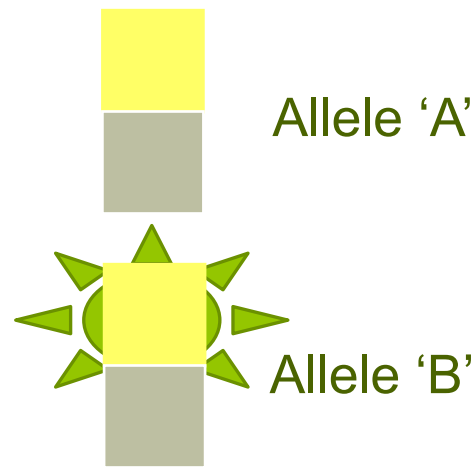


Allele 'B'

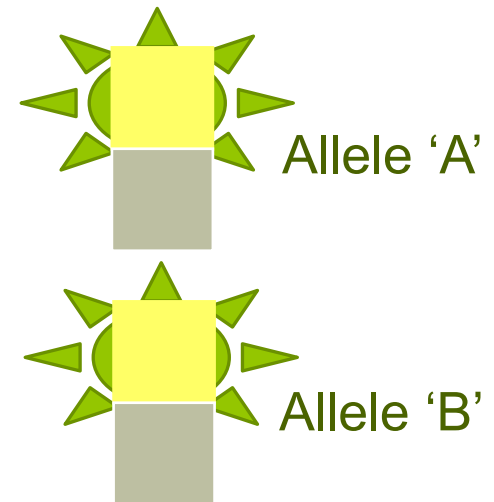
DNA *microarrays*



Homozigoto "AA"



Homozigoto "BB"



Heterozigoto "AB"

DNA *microarrays*

- Genotipagem em larga escala – muitos SNPs genotipados de uma só vez
- Número de SNPs na espécie humana:
 - ~30 milhões catalogados (dbSNP)
 - Dois indivíduos diferem-se por ~3 a 4 milhões de SNPs

DNA *microarrays*

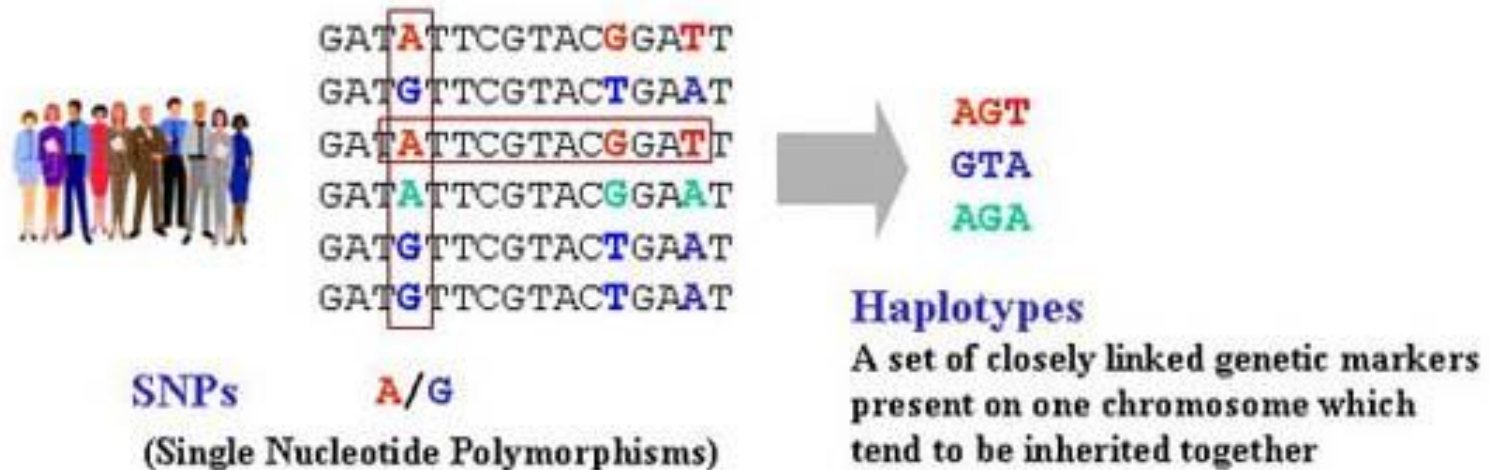
- Genotipagem em larga escala – muitos SNPs genotipados de uma só vez
- Número de SNPs na espécie humana:
 - ~30 milhões catalogados (dbSNP)
 - Dois indivíduos diferem-se por ~3 a 4 milhões de SNPs
- ... **Mas quantos são necessários genotipar?**

DNA *microarrays*



Genoma é formado em **blocos de haplótipos!**

DNA microarrays



Desequilíbrio de ligação → medida de quanto dois SNPs estão “ligados” e são herdados juntos

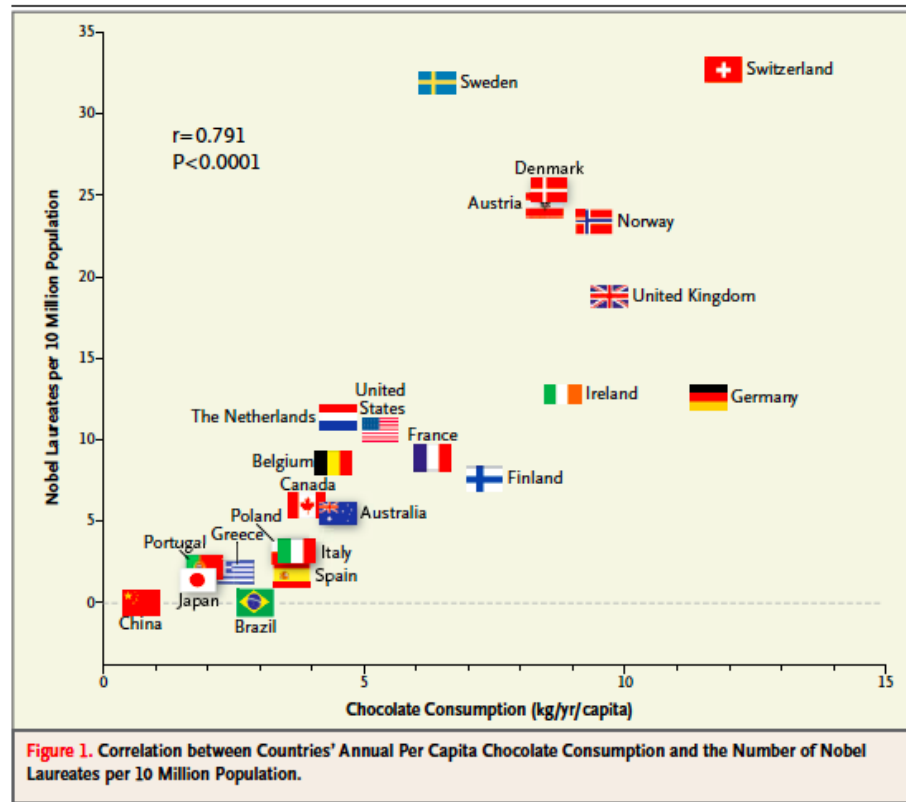
Não preciso genotipar todos os SNPs em um bloco de haplótipos que está em desequilíbrio de ligação

Estudos de associação genética

- Será que um determinado SNP é mais frequente em uma população de casos do que em uma população de controles?
- Verificação da **associação** entre SNP e doença
 - Não é relação causal
 - É sujeita a viés de amostras, estratificação populacional, etc.

Estudos de associação genética

- Cuidado! Associação não implica em causalidade!



Estudos de associação genética

- Cuidado! Associação não implica em causalidade!

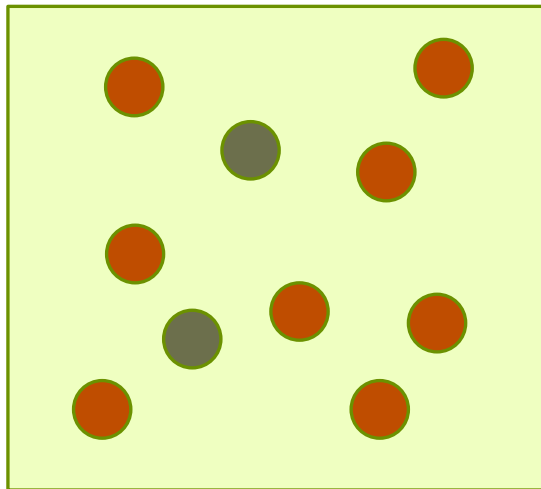


Estudos de associação genética

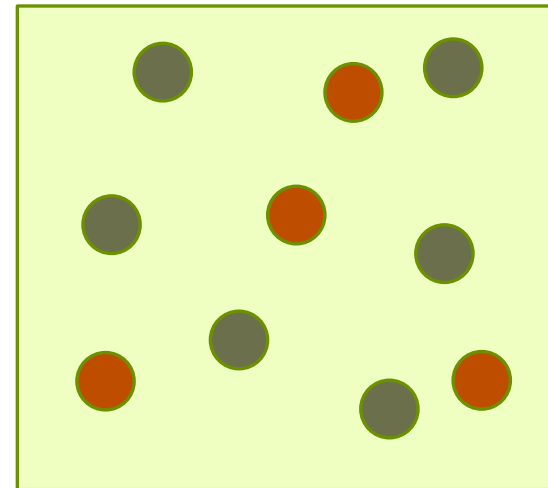
- Exemplo: 10 casos e 10 controles:

CC

TT



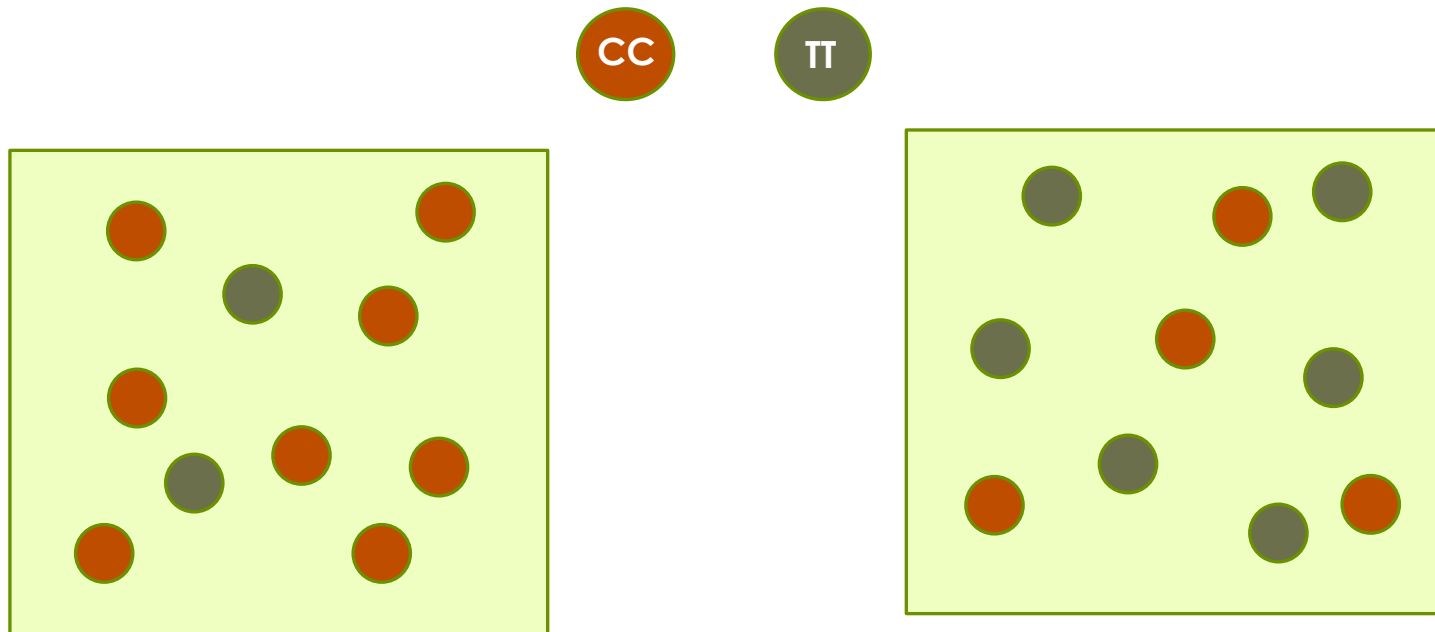
Controles $\rightarrow f(T) = 20\%$



Casos $\rightarrow f(T) = 60\%$

Estudos de associação genética

- Exemplo: 10 casos e 10 controles:



Controles $\rightarrow f(T) = 20\%$

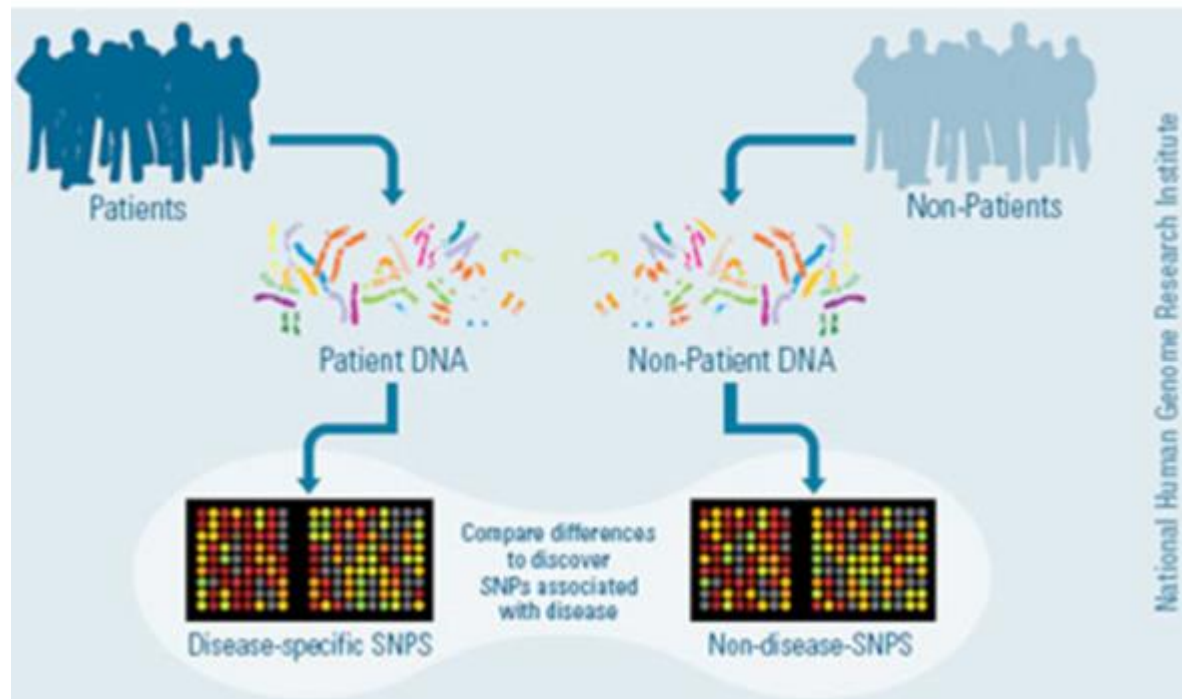
Casos $\rightarrow f(T) = 60\%$

Teste estatístico (qui-quadrado) revelou $p < 0,05$

Alelo T está associado à doença

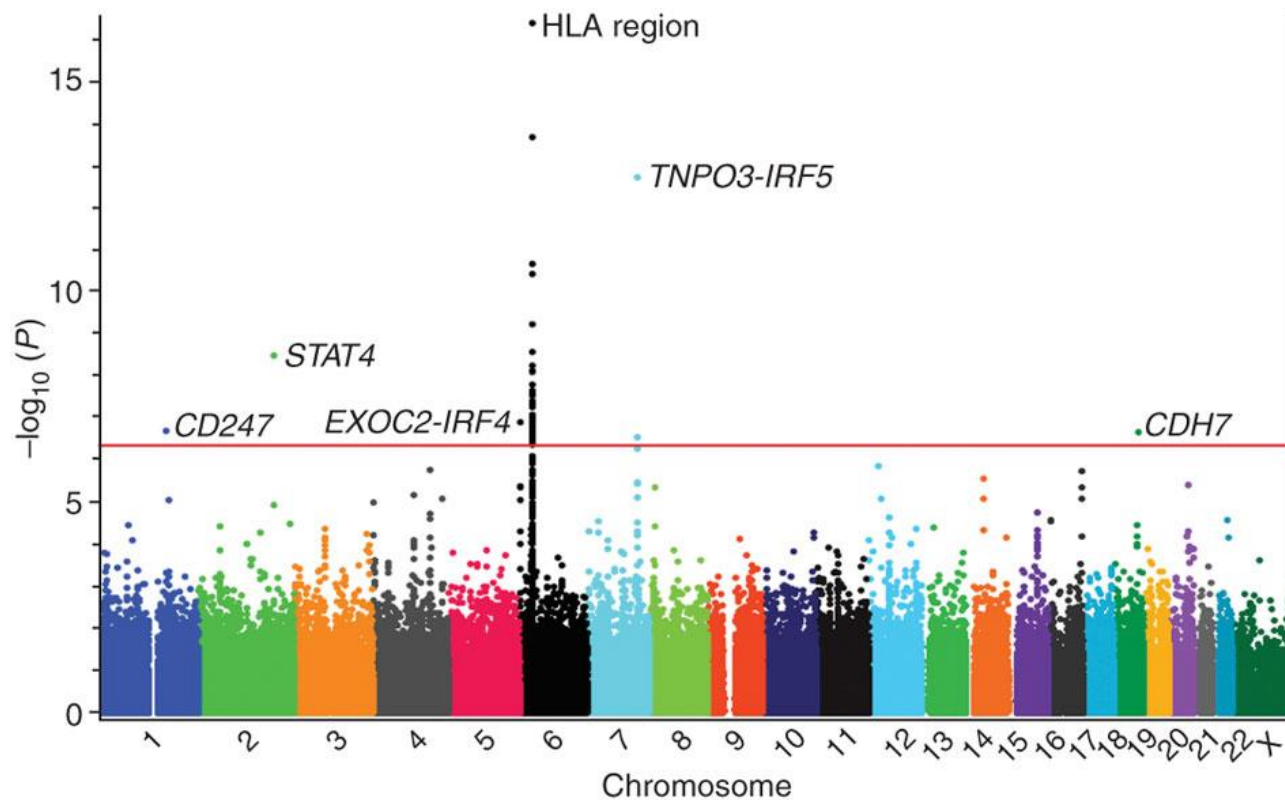
Estudos de associação genética

- Se fizermos isso para 500.000 SNPs → identificar quais SNPs estão associados com a doença!
- Genome-Wide Association Study*

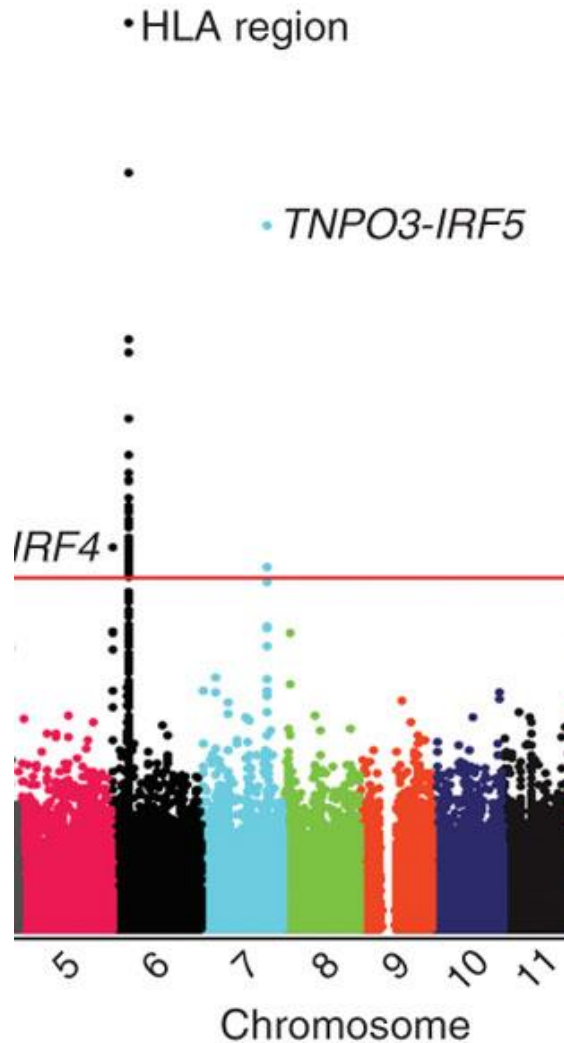


Estudos de associação genética (GWAS)

- Exemplo de visualização → *Manhattan Plot*

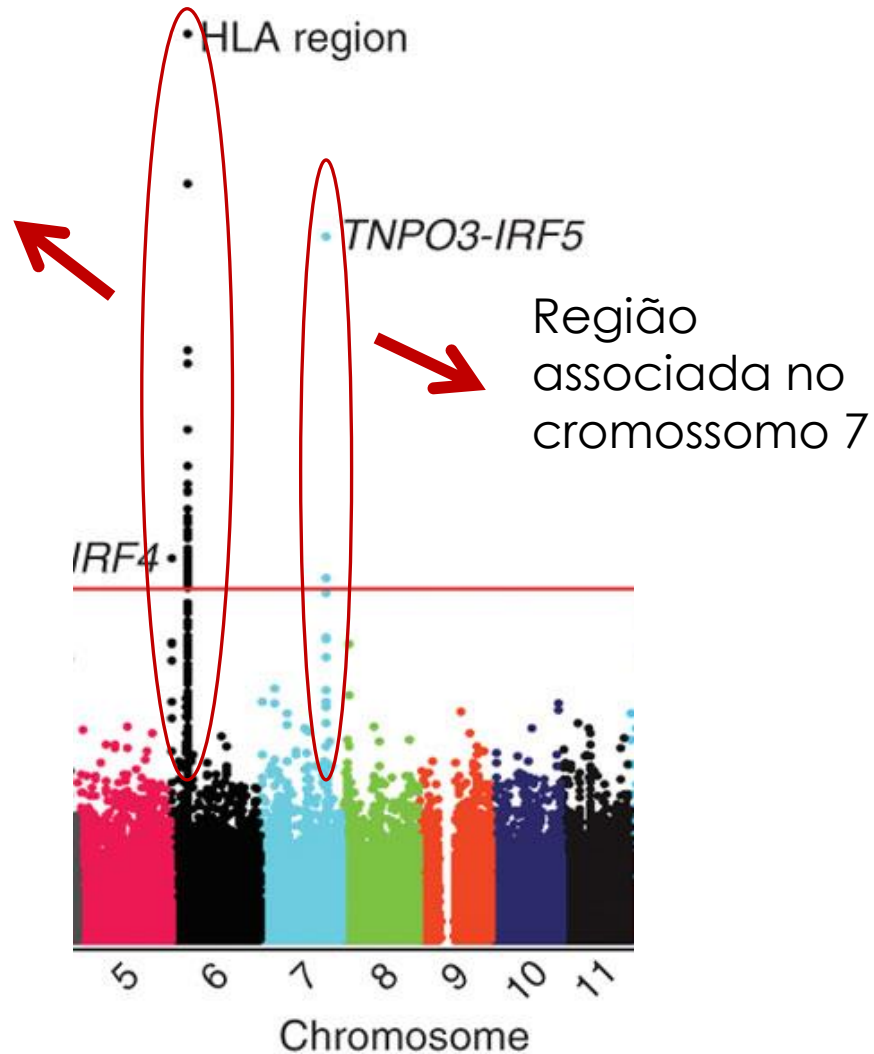


Estudos de associação genética



Estudos de associação genética

Região associada no cromossomo 6



Estudos de associação genética (GWAS)

- **Principais cuidados** quando trabalhamos com estudos de associação genética em larga escala:
 - *Equilíbrio de Hardy-Weinberg*
 - *Estratificação populacional*
 - *Controles de qualidade*
 - *Correção para múltiplos testes*
 - *Replicação*

Estudos de associação genética (GWAS)

- **Principais cuidados** quando trabalhamos com estudos de associação genética em larga escala:
 - *Equilíbrio de Hardy-Weinberg*
 - *Estratificação populacional*
 - *Controles de qualidade*
 - *Correção para múltiplos testes*
 - *Replicação*

Estudos de associação genética (GWAS)

- **Equilíbrio de Hardy-Weinberg (HWE)**

- O princípio de HW diz que:

- 1. A frequência dos alelos e genótipos em uma população permanecem constantes de geração em geração na ausência de influências
- 2. A distribuição da frequência dos genótipos e alelos segue a seguinte distribuição:

$$p^2 + 2pq + q^2 = 1 = p + q = 1$$

Onde:

p = Frequência da alelo selvagem

q = Frequência da alelo alternativo

Estudos de associação genética (GWAS)

- **Equilíbrio de Hardy-Weinberg (HWE)**

- Em HWE, assume-se que:
 - A população é grande
 - Não há fluxo gênico (migração) entre populações
 - Não há mutações
 - Os casamentos são aleatórios
 - Não há efeito da seleção natural

Estudos de associação genética (GWAS)

Equilíbrio de Hardy-Weinberg

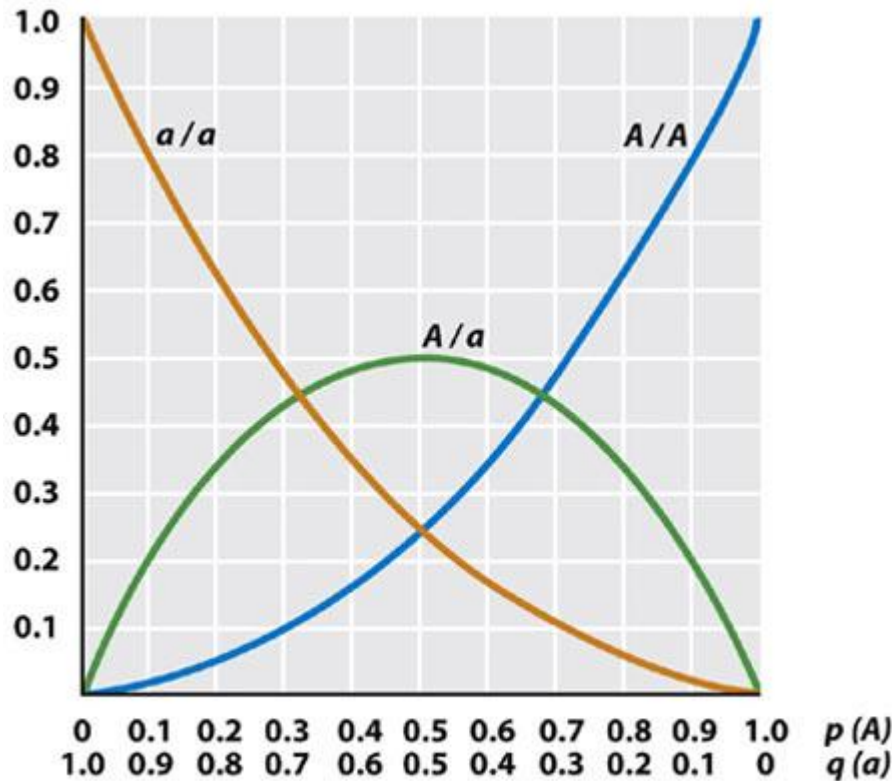
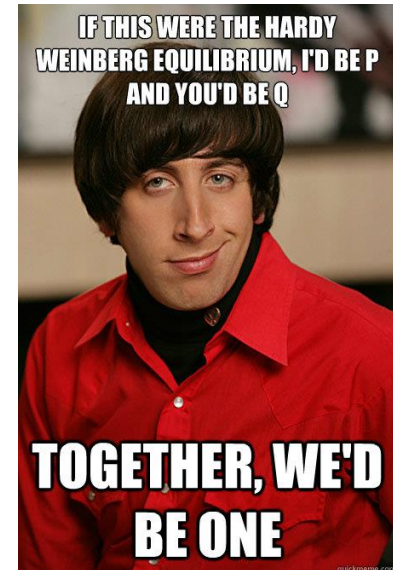


Figure 17-5
Introduction to Genetic Analysis, Ninth Edition
© 2008 W.H. Freeman and Company



Estudos de associação genética (GWAS)

- **Estratificação populacional**



Estudos de associação genética (GWAS)

- Estratificação populacional

ENGLISH:

	Cases	Controls
A	160	160
a	40	40

$$\chi^2 = 0$$

CHINESE:

	Cases	Controls
A	160	40
a	160	40

$$\chi^2 = 0$$

Estudos de associação genética (GWAS)

- Estratificação populacional

COMBINED:

	Cases	Controls
A	320	200
a	200	80

$$\chi^2 = 7.81$$

Estudos de associação genética (GWAS)

- **Estratificação populacional – PROBLEMAS:**

- Associações espúrias
- Aumento de falsos positivos
- Inflação genômica

Estudos de associação genética (GWAS)

- **Estratificação populacional – como contornar?**
 - Marcadores genéticos de ancestralidade (AIMs)
 - PCA / MDS de dados de GWAS

Estudos de associação genética (GWAS)

- **Redução de dimensionalidade**

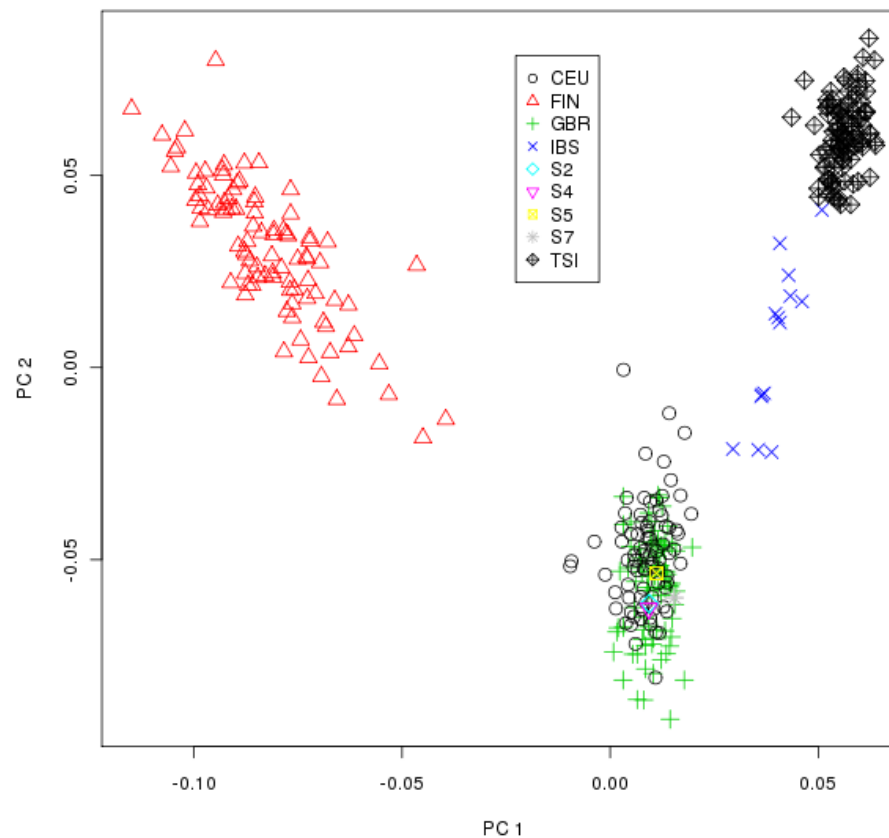
- Técnica matemática/ estatística usada para tentar “resumir” muitas variáveis em poucas variáveis, sem perder informação (mantendo a variação do conjunto)
- Técnicas
 - Escalonamento Multidimensional (MDS)
 - Análise dos Componentes Principais (PCA – caso particular de MDS)

Estudos de associação genética (GWAS)

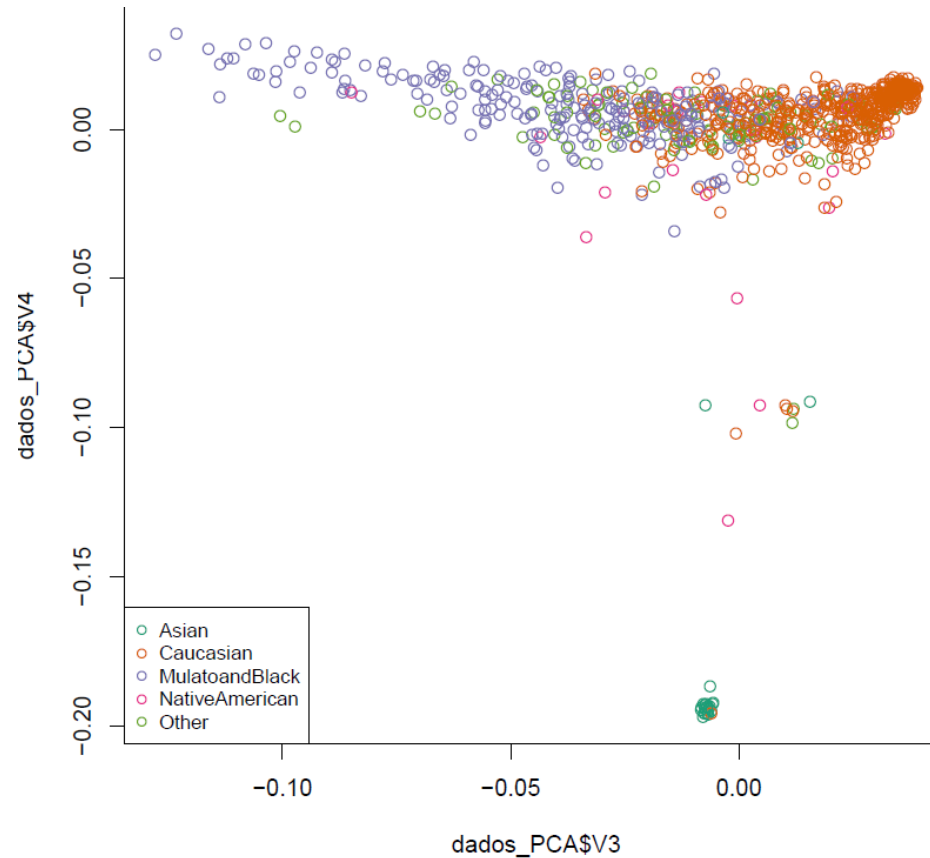
- **PCA / MDS para GWAS**

- Genótipos → Muitas variáveis ($N > 500.000$)
 - 1. Tentar reduzir essas variáveis em poucas, que explicam maior parte da variabilidade de todas (componentes principais)
 - 2. Plotar em um gráfico os primeiros componentes principais (os que explicam maior parte da variabilidade)
 - 3. Identificar grupos (ou *clusters*) que apresentam perfis semelhantes
- **Se estamos derivando variáveis genéticas, o que os componentes principais “explicam” ou “separam” em um GWAS?**

Estudos de associação genética (GWAS)



Estudos de associação genética (GWAS)

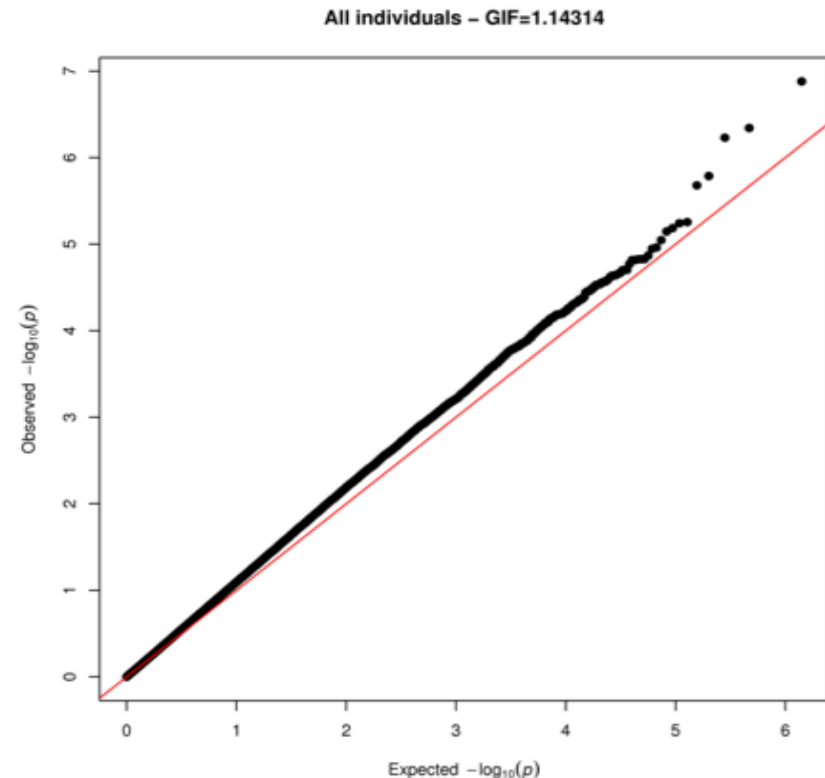


População representativa da cidade de SP (EPISONO)

Estudos de associação genética (GWAS)

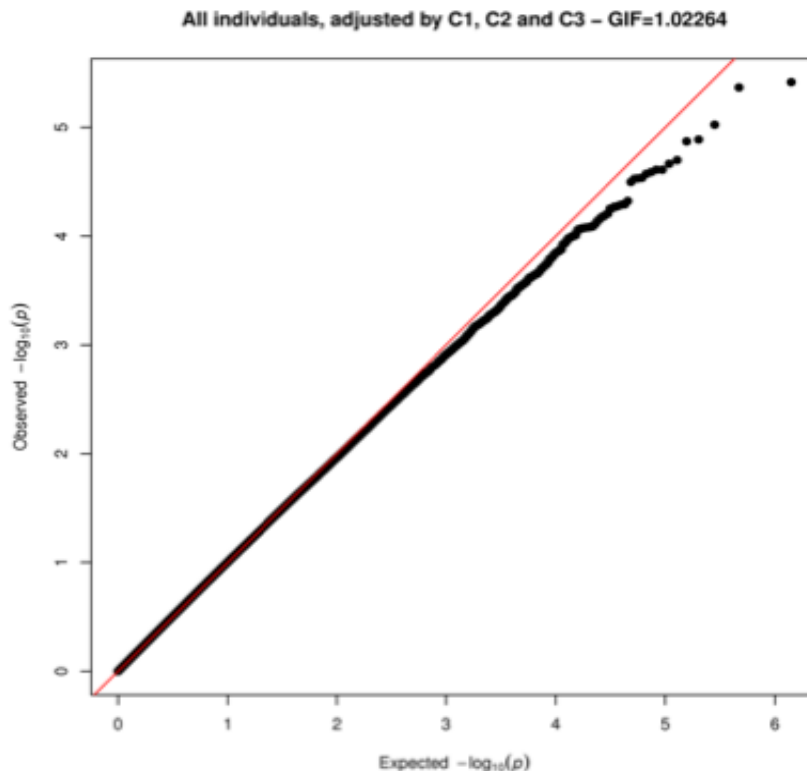
- Estratificação populacional – Inflação genômica:

QQ-plot da
associação sem
ajuste para
ancestralidade



Estudos de associação genética (GWAS)

- Estratificação populacional – Inflação genômica:



QQ-plot da
associação com
ajuste para
ancestralidade

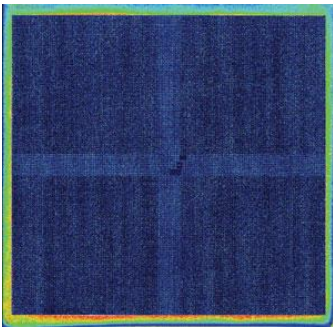
Estudos de associação genética (GWAS)

- **Controles de qualidade**

- Cuidados antes de iniciar os dados, removendo amostras e SNPs que apresentaram baixa qualidade
 - Amostras não genotipadas para mais que 5% dos SNPs
 - Amostras duplicadas ou com alto grau de parentesco
 - SNPs fora do HWE
 - SNPs genotipados em menos que 95% da amostra
 - SNPs raros (*minor allele frequency* < 5% ou 1%)

Estudos de associação genética

Fluxograma de análise



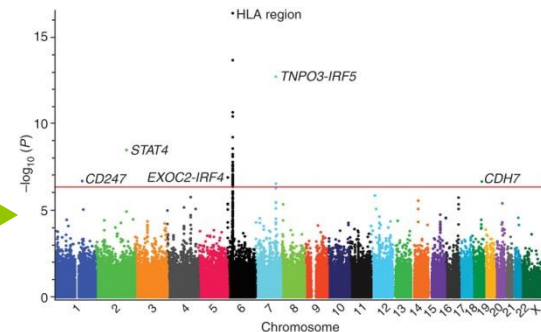
	SNP1	SNP2	SNP3	SNP4	SNP5	...	SNP500.000
Amostra 1	AA	AA	AA	AA	AB	...	AB
Amostra 2	AB	AA	AA	BB	AB	...	BB
Amostra 3	BB	AA	AB	AA	AB	...	AA
Amostra 4	BB	AA	AA	BB	AA	...	AB
Amostra 5	BB	AA	AB	AB	AA	...	AB
...
Amostra 2000	BB	AA	AA	AB	BB	...	AA



- Controle de Qualidade
- Verificar estratificação populacional
- Verificar Inflação Genômica



Teste estatístico apropriado (ex: Regressão Logística)



Estudos de associação genética

- Fluxograma de análise
 - Principais softwares:
 - Análise – PLINK (<http://pngu.mgh.harvard.edu/~purcell/plink/>)
 - Visualização – Ferramentas de visualização do R (<http://www.r-project.org/>)

plink...

Whole genome association analysis toolset

[Introduction](#) | [Basics](#) | [Download](#) | [Reference](#) | [Formats](#) | [Data management](#) | [Summary statistics](#)

| [Profiles](#) | [ID helper](#) | [Resources](#) | [Flow chart](#) | [Misc.](#) | [FAQ](#) | [gPLINK](#)



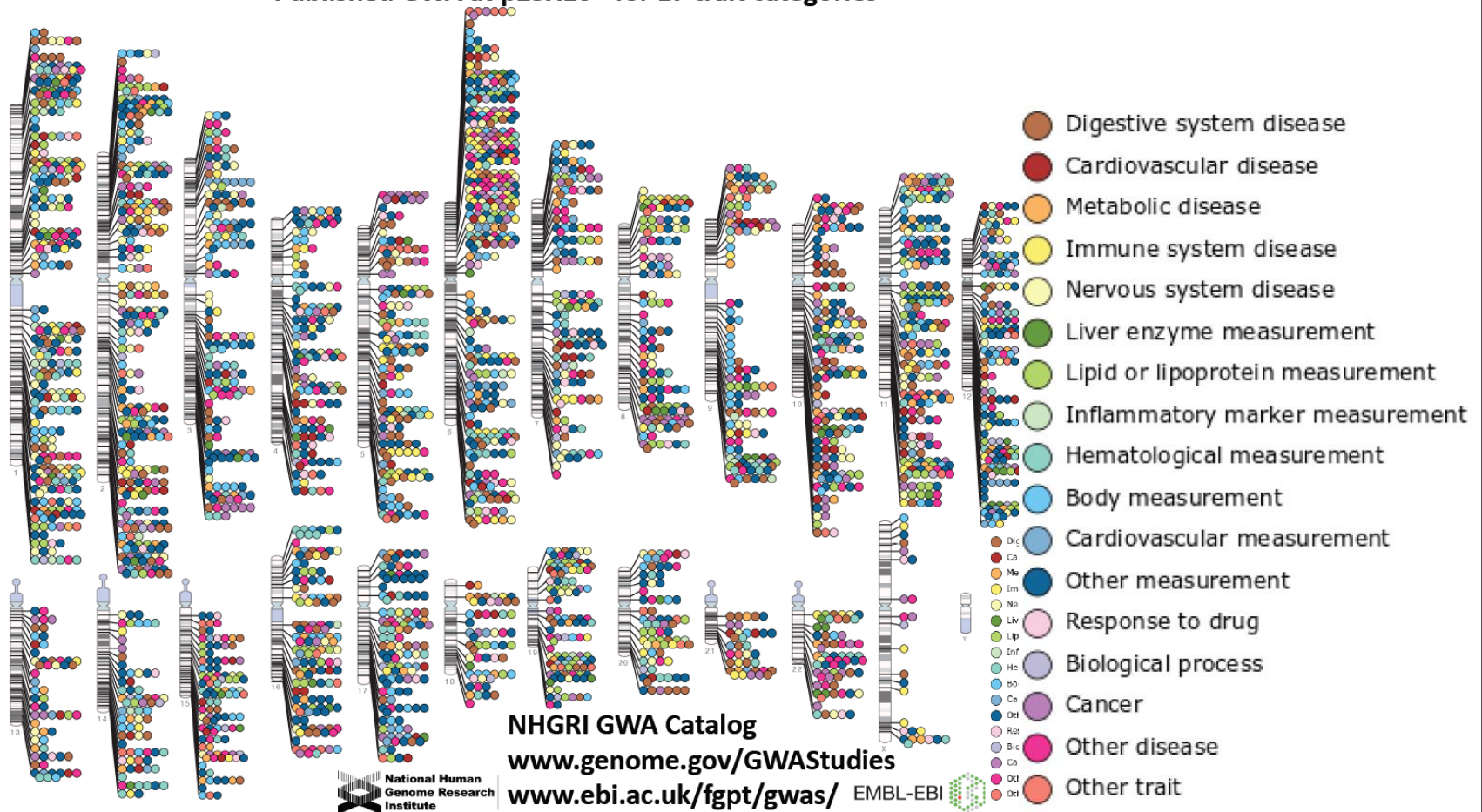
Estudos de associação genética



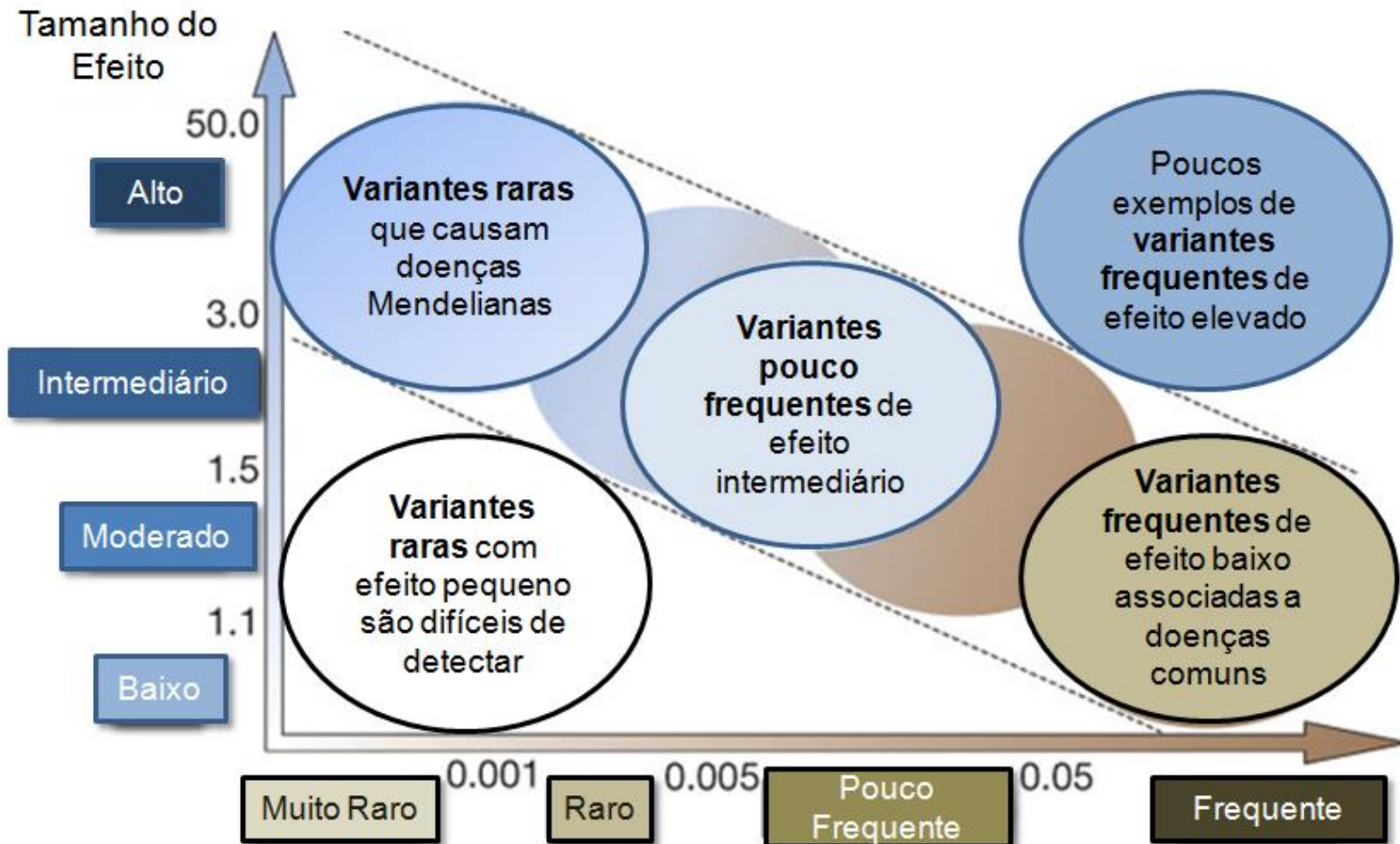
Estudos de associação genética

Published Genome-Wide Associations through 12/2012

Published GWA at $p \leq 5 \times 10^{-8}$ for 17 trait categories



Estudos de associação genética



Manolio et al, 2009

Frequência alélica do polimorfismo

Estudos de associação genética



Manolio et al, 2009