

False Discovery Rate, sensitivity and sample size for microarray studies

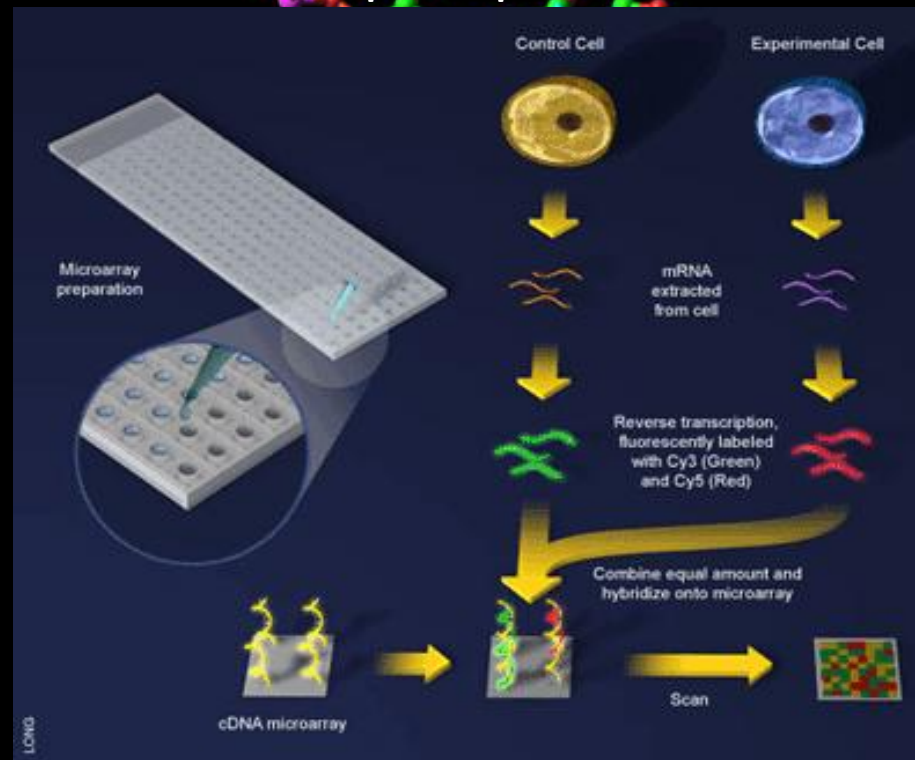
Carlos Paiva – 100378054

Bioinformática 2011

DCC/FCUP 2011/06/25

O que são?

- Conjunto de pontos microscópicos de DNA anexados a uma superfície sólida;
- Superfície sólida composta por vidro especial com tamanho de uma lamina de microscópio óptico.



Vantagens



Rapidez

Baixo custo

Fácil
compreensão

Análise de
grandes
quantidades
de dados

User-friendly

Adaptabilidade



Aplicações

Procura de genes
expressados
diferencialmente

Detecção SNP

Procura de assinatura
molecular

...

Possibilidades Futuras



DNA 'Fingerprints'

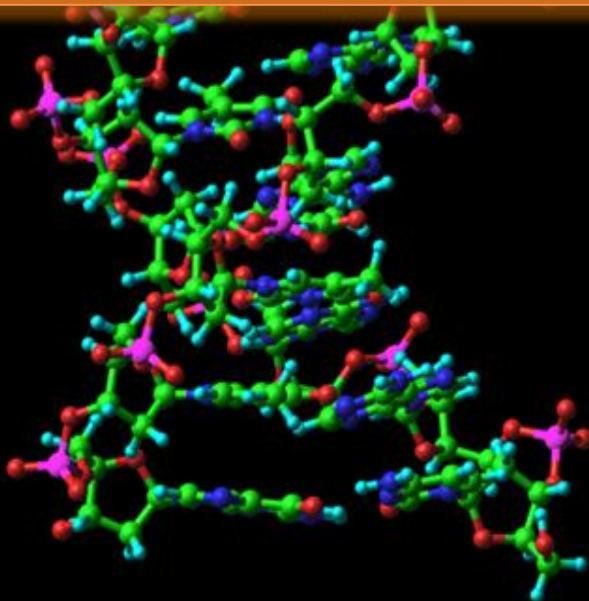
Genome-wide
Studies

Produção de perfis
genéticos



Cuidados

Manter ciente o risco de obtenção de falsos resultados e controlá-los





Valor-p

- Probabilidade de obter uma estatística de valor igual ou mais extrema que a obtida no teste, sabendo que a hipótese nula se verifica. É frequentemente definida como a menor escolha para o nível de significância, tal que a hipótese nula é rejeitada.

- Hipótese nula considerada verdadeira inicialmente;
- Hipótese alternativa é complemento da hipótese nula;

Da opção tomada sobre a hipótese nula podem ocorrer 2 tipos de erros.

- Erros de Tipo I: Rejeitar a hipótese nula, sendo esta verdadeira;
- Erros de Tipo II: Não rejeitar a hipótese nula, sendo esta falsa;

Significância é a probabilidade de cometer erros de Tipo I.

Microarrays e valor-p



- Valor-p - variável estatística para testes individuais de hipóteses;
- Microarray junta vários milhares de hipóteses ao mesmo tempo;
- Valor-p volátil com a variação do tamanho da amostra;
- Aplicado a testes múltiplos, o valor-p, é usado para medir a percentagem de falsos positivos de um teste, validando-o com o valor-p < 0.05 ;

FDR, FNR e Sensibilidade



FDR:

- FDR é o valor esperado da proporção de falsos positivos entre os resultados significativos;
- Interpretação útil no estudo de microarrays;

FNR e sensibilidade:

- FNR é o valor esperado da proporção de falsos negativos entre os resultados significativos;
- Sensibilidade é $1 - \text{FNR}$;
- Úteis para não ocorrer perda de demasiados genes expressados diferencialmente;

Factores que determinam as características da FDR para um estudo de microarray



- Proporção dos genes expressados diferencialmente reais;
- Distribuição das diferenças reais;
- Variação nas medidas;
- Tamanho da amostra;

Apenas o tamanho da amostra é controlado pelo investigador.



Resultados

T-statistics:

Em experiências de pequena dimensão onde se espera um grande rácio de genes não-DE, a FDR pode ser bastante elevada, mesmo usando valores críticos altos.

A taxa de falsos positivos expressa em função do valor crítico produz valores inaceitavelmente elevados de FDR.

Tamanho da Amostra



Depende de:

Número de genes
realmente DE

Distribuição dos
genes realmente DE

FDR tolerável

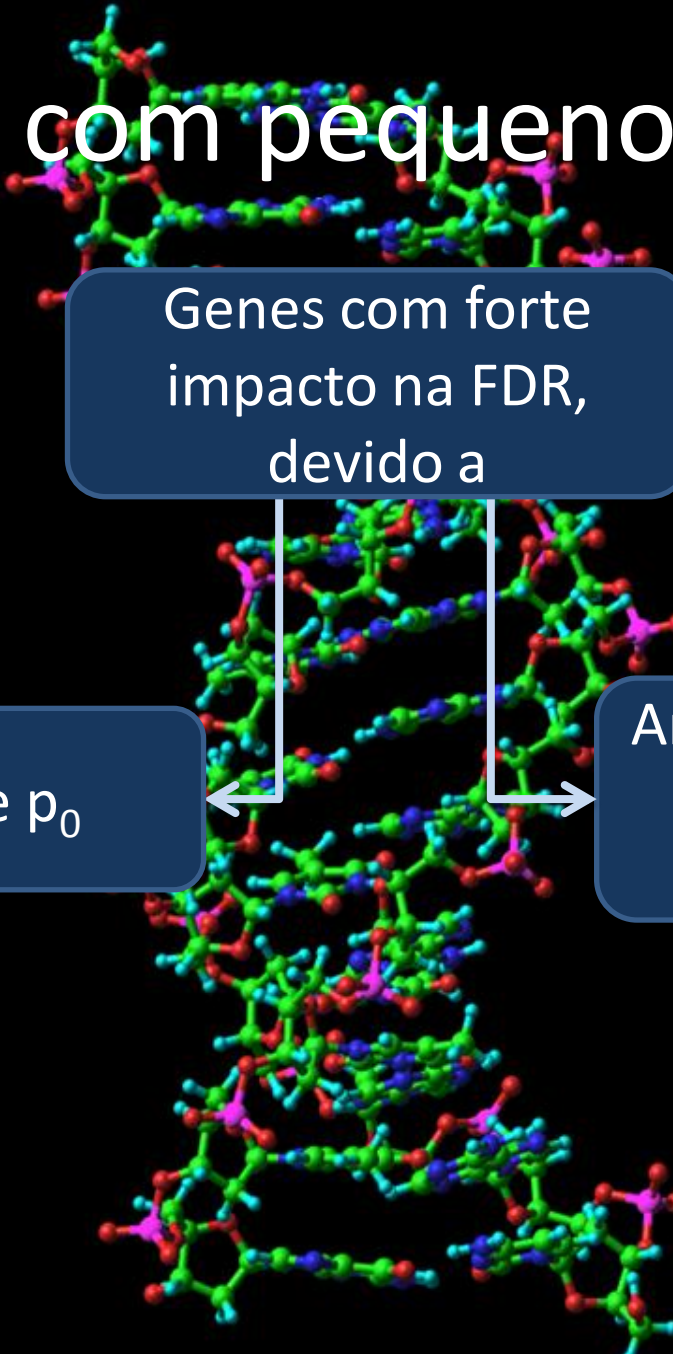
Número de Amostras



Quanto mais próximo de 1 estiver p_0 menos genes significantes poderão ser observados

Amostra pequena ou p_0 próximo de 1 exigem FDR elevado

Genes com pequenos efeitos



Genes com forte
impacto na FDR,
devido a

Aumento de p_0

Ampliar a distribuição
nula de t-statistics
observada

Conclusão

Útil aos cientistas a consciencialização, essencialmente para triagem de dados relevantes ao conceito estudado.

