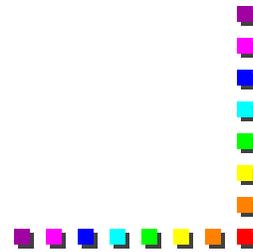


Mapeamento de Genes: Delineamentos com Famílias e Estudos Observacionais

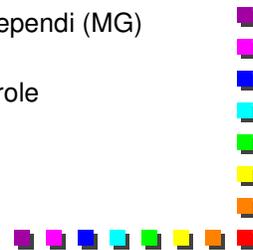
Júlia Maria Pavan Soler
pavan@ime.usp.br

DEST – U.F.P. – Curitiba
Set / 2007



Programa

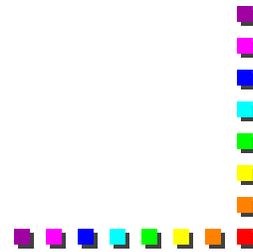
- Aula 1: Conceitos de Genética Quantitativa e Genômica
Mapeamento de Genes
- Aula 2: Delineamentos Familiares – Análise de QTL's
- Aula 3: Aplicação - Projeto Corações de Baependi (MG)
- Aula 4: Estudos Observacionais Caso-Controle
Perspectivas



Aula 1

Genética Quantitativa e Genômica

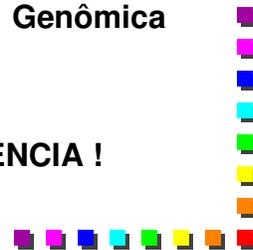
Mapeamento Genético



Tradição da Estatística na Ciência Factual

- Sociometria
 - Psicometria
 - Econometria
 - Geoestatística
 - Atuária
 - **Bioestatística**
 - ...
- Genética Quantitativa
e Genômica

RAMO MULTIDISCIPLINAR da CIÊNCIA !



Formação Multidisciplinar

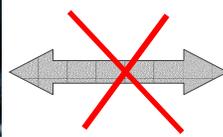
$$p(\psi|y, \theta, \tau^2) = \frac{p(y, \psi|\theta, \tau^2)}{p(y|\theta, \tau^2)} = \frac{f(y|\psi)p(\psi|\theta, \tau^2)}{\int f(y|u)p(u|\theta, \tau^2) du}$$

Tenho interesse nos efeitos de Epistasia, Pleiotropia e Herdabilidades. Qual delineamento usar para garantir Ligação e Desequilíbrio de Ligação entre os QTL's e os Marcadores Moleculares?



Bioestatístico

Comunicação

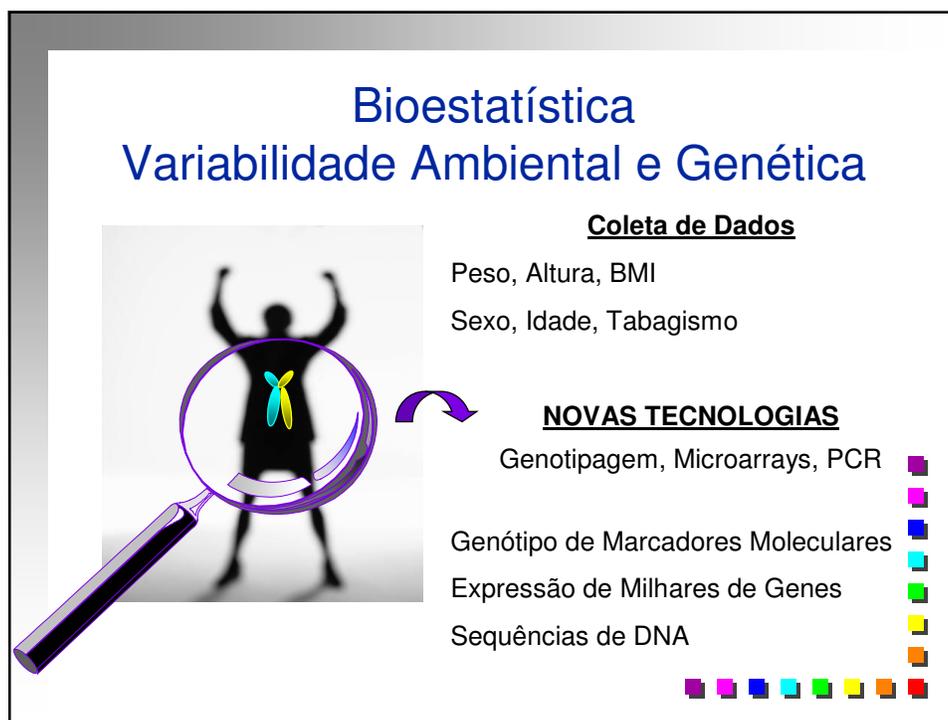
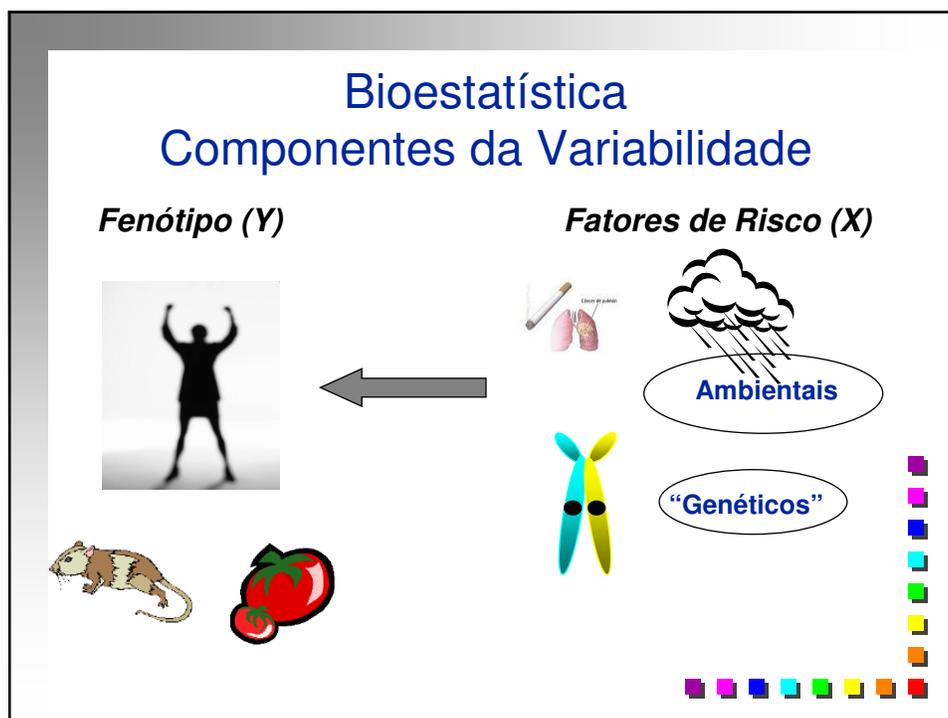


Geneticista



Dogma da Biologia: Entender Variabilidade



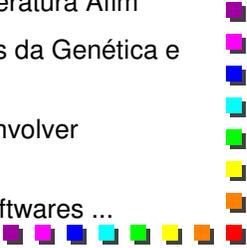


Bioestatística na Era Genômica

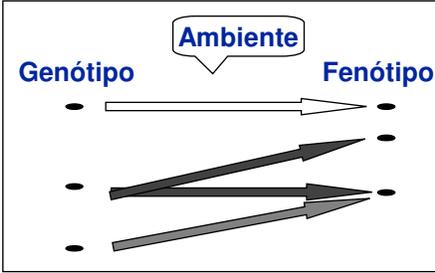


DESAFIOS

- Entender o Problema
- Entender a Coleta dos Dados
- Conhecer os Delineamentos Úteis
- Conhecer a Linguagem da Área
- Conhecer as Literatura Afim
- Entender as Leis da Genética e Genômica
- Adaptar e Desenvolver Metodologias
- Conhecer os Softwares ...



Padrão de Variação Fenotípica



Genótipo

•

•

•

Fenótipo

•

•

•

Modelos Mendelianos
(monogênicos)

Modelos Oligogênicos
ou Poligênicos

Fenótipo (Y) = Genótipo (G) + Ambiente (E)

Efeito de Pleiotropia

$\rho(Y1, Y2)$

Efeito de Epistasia

$G1 * G2$

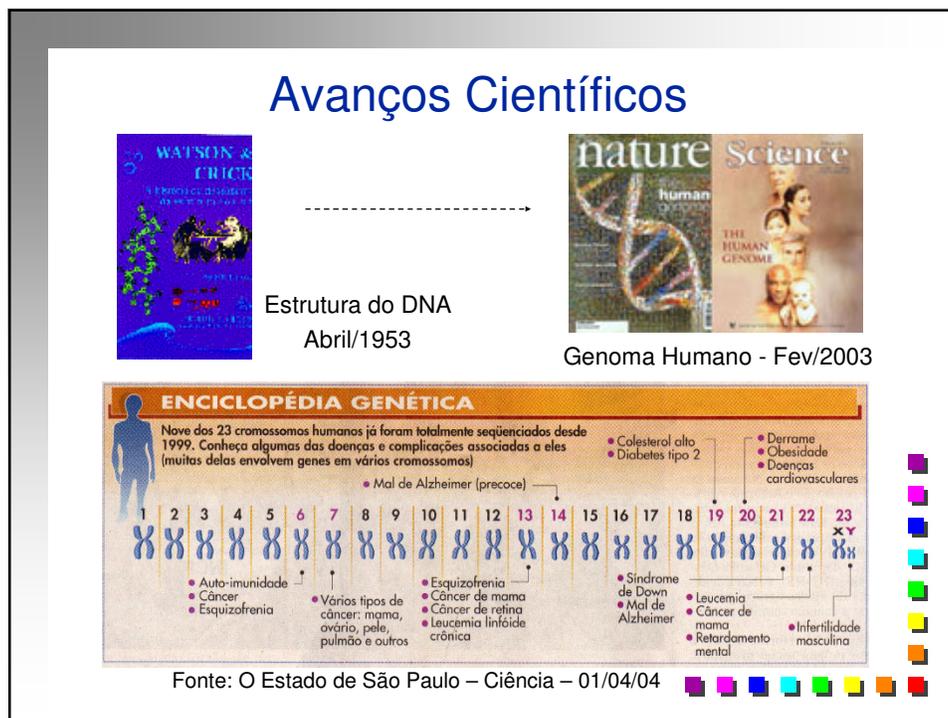
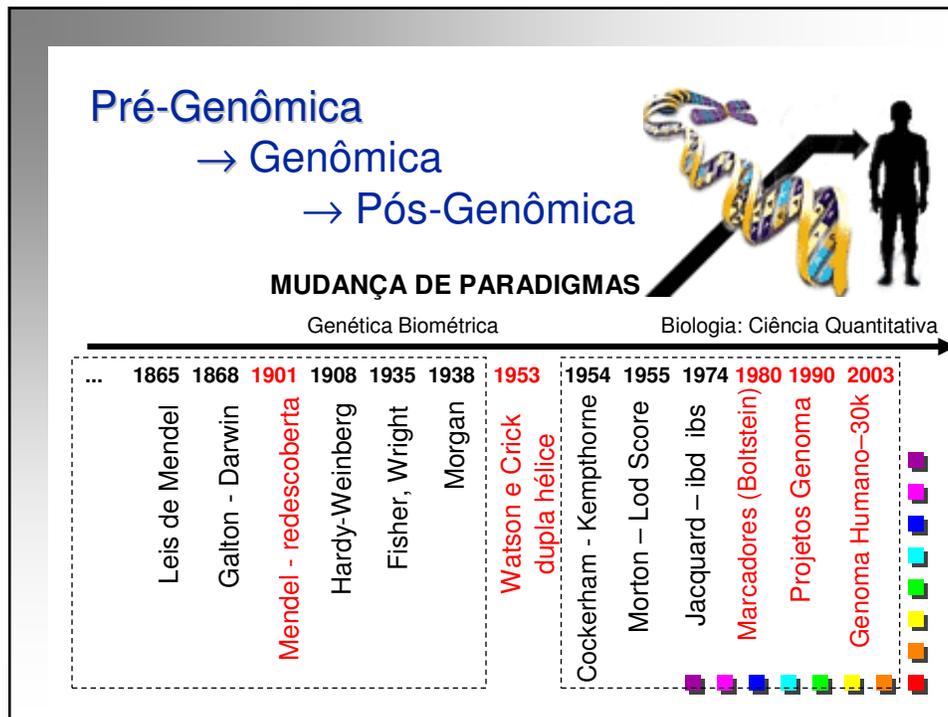
Interação

Gene * Ambiente



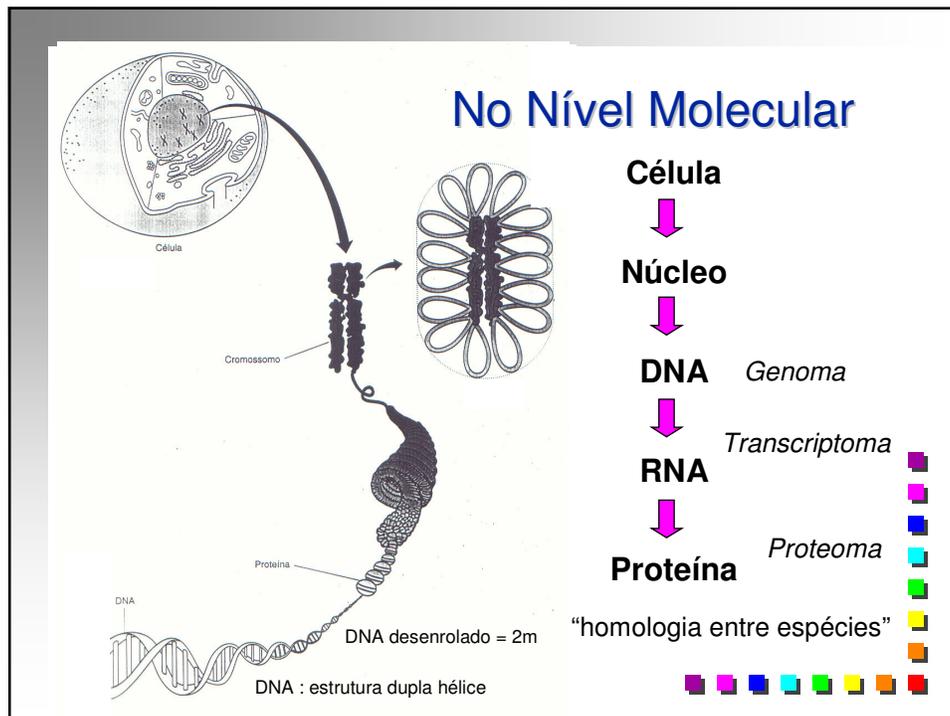
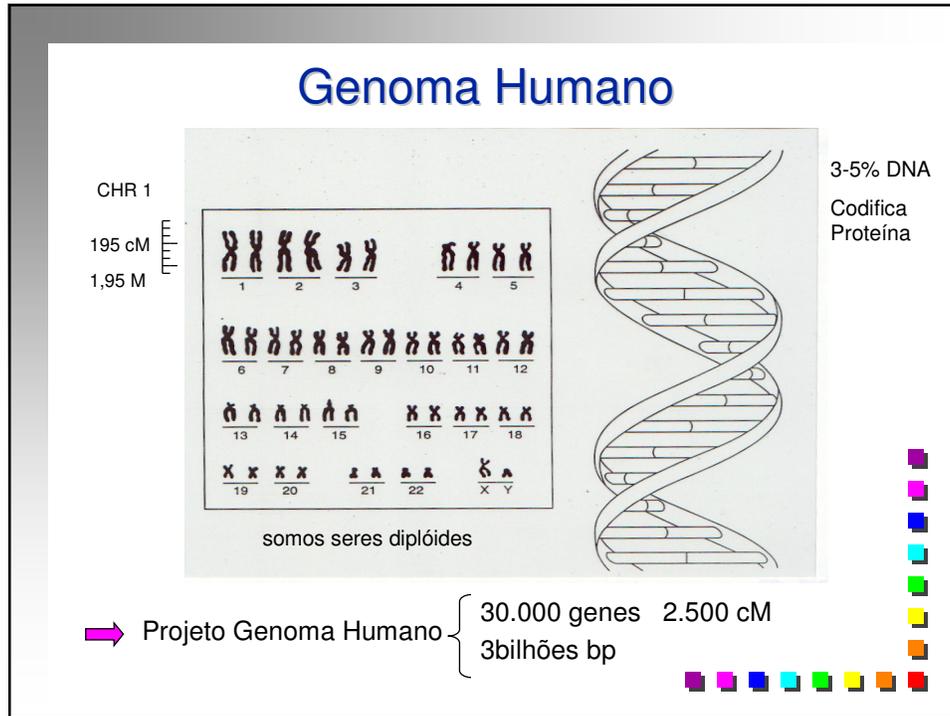
Mapeamento de Genes

Delineamentos com Famílias e Estudos Observacionais



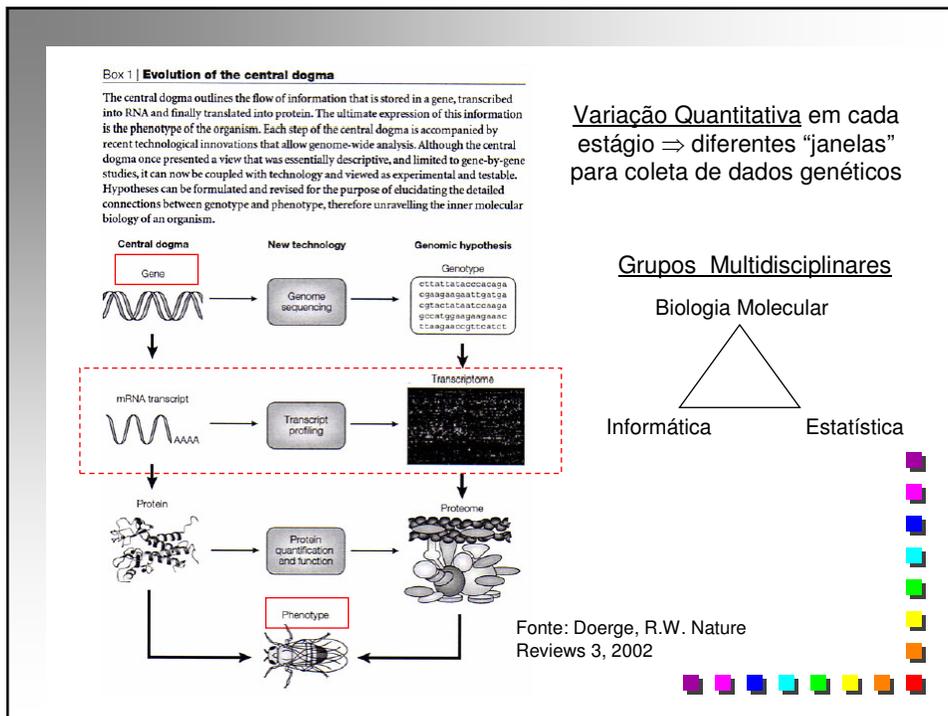
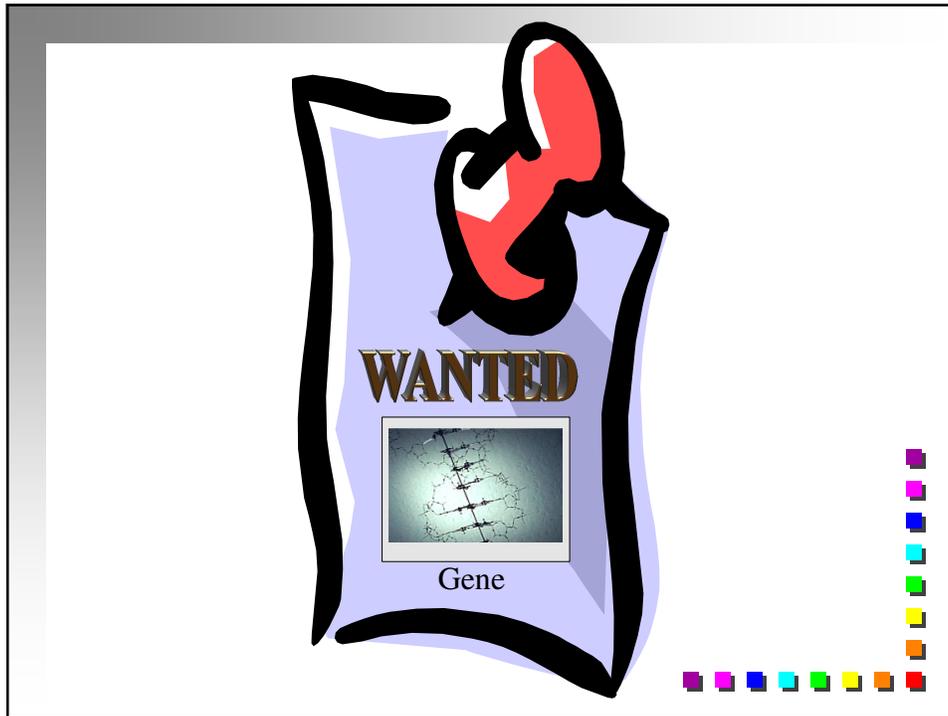
Mapeamento de Genes

Delineamentos com Famílias e Estudos Observacionais



Mapeamento de Genes

Delineamentos com Famílias e Estudos Observacionais



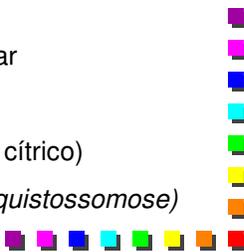
Projetos Genoma Brasil



Logotipo do Projeto Genoma Brasileiro

Construção de Bancos de DNA, RNA e proteínas

- 1997: Rede ONSA (Fapesp)
- 1998-2000: *Xylella fastidiosa* (citros)
- Genoma Câncer
- Genoma da Cana de Açúcar
- Genoma do Eucalipto
- *Xanthomonas citri* (cancro cítrico)
- *Schistosoma mansoni* (esquistossomose)



Mercado de Trabalho para o Estatístico



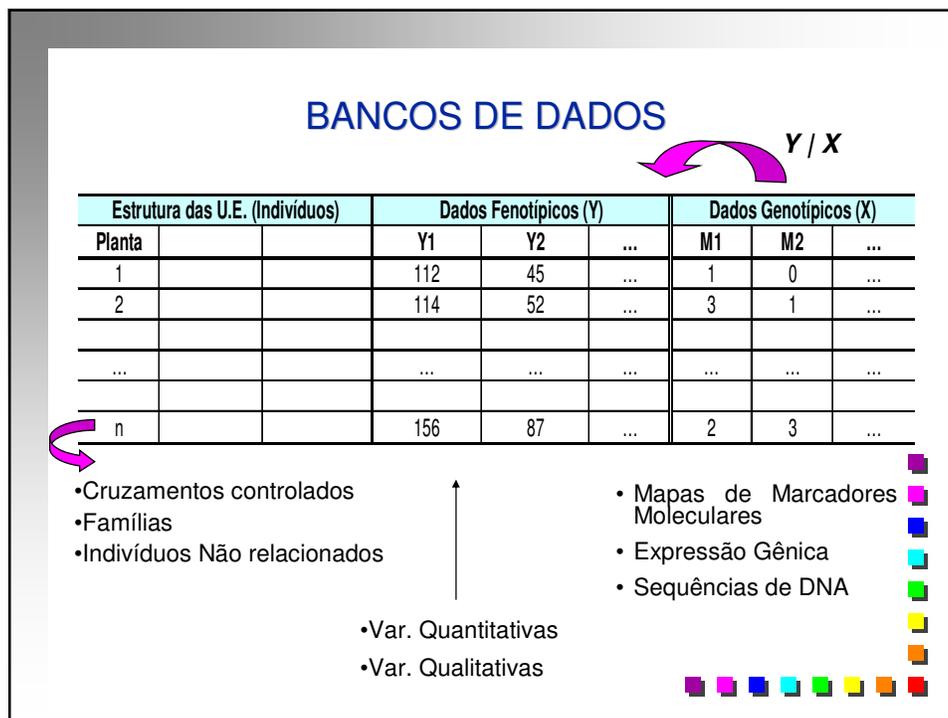
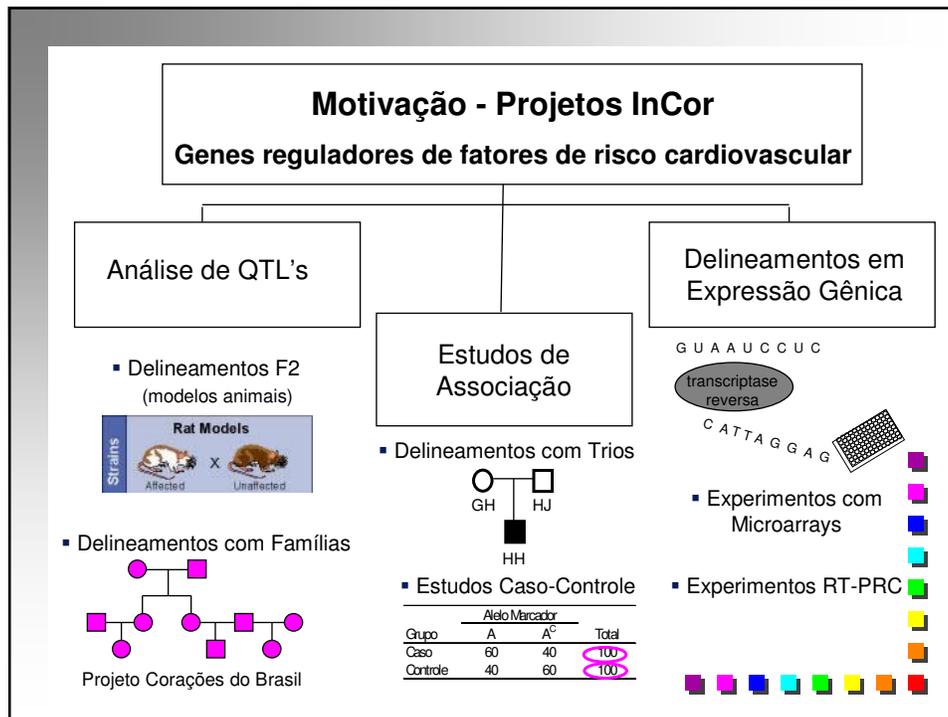
- Empresas de Biotecnologia: Transgênicos (OGM's)
- Indústria Farmacêutica: Farmacogenética
- Laboratórios Clínicos: Testes de Paternidade, doenças congênitas
- Centros de Pesquisas: (Genética Funcional, Células Tronco, Biodiversidade)
- Perícia Técnica

⇒ Mapeamento Genético ⇒ Genômica Funcional
⇒ Análises de Risco ⇒ Testes de Paternidade
⇒ Reconhecimento de Padrões



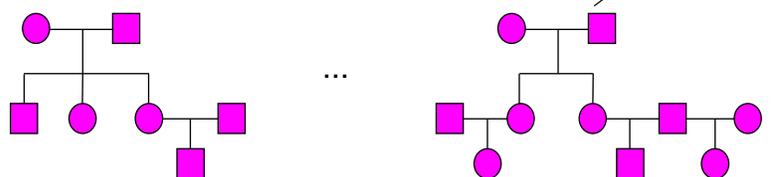
Mapeamento de Genes

Delineamentos com Famílias e Estudos Observacionais



Delineamento com Famílias Projeto Corações de Baependi (MG)

SEGREGAÇÃO HETEROGÊNEA



1.
 - Estrutura Familiar: Indivíduos Relacionados e Não Relacionados
 - Variáveis Fenotípicas de interesse e Covariáveis
 - Variáveis Genóticas: Mapa de Marcadores → 2

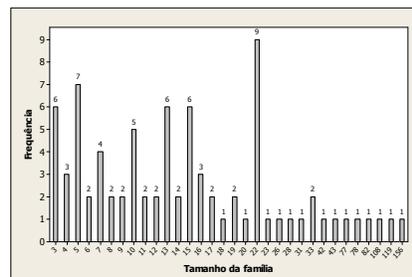
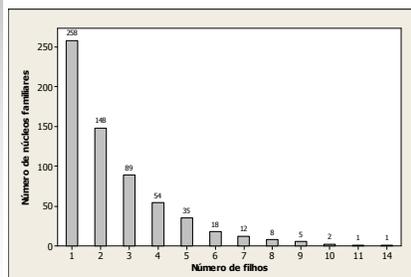
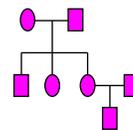
1. O padrão de variação de Fenótipos depende da estrutura familiar ?
2. Se sim, identificar os Genes reguladores.



Projeto Corações de Baependi (MG)

Número de famílias = 119 (⇒81)
 Número de indivíduos = 1666
 Tamanho Médio das Famílias=21 (s=26)
 Total Núcleos Familiares: 631
 Número Médio de Filhos= 3 (s=2)
 ⇒Muitos Fenótipos (Y)

Sexo	Freq	%
Masculino	723	43,4
Feminino	943	56,6
Total	1666	



Estudos Observacionais Caso-Controle População de Vitória (ES) - InCor

Distribuição de Casos e Controles referente
 ao marcador B1BK e cada fenótipo

	B1BK			Total
	AA	AG	GG	
Diabetes				
Caso	50	49	24	123
Controle	564	673	200	1437
Hipertensão				
Caso	210	231	92	533
Controle	406	497	134	1037
Obesidade				
Caso	235	265	87	587
Controle	380	456	139	975

⇒ 9 Marcadores Moleculares

Identificar fatores de risco
 genéticos para diabetes,
 hipertensão e obesidade.

Há efeitos de epistasia?

Há genes pleiotrópicos?

Batista, MJ (2006)
 Dissertação de Mestrado
 IME/USP



A Estatística na Genética Quantitativa e Genômica

Identificar "Genes" (G) envolvidos na regulação de uma característica (Y)

$$Y = G + E$$

Dados Genéticos:

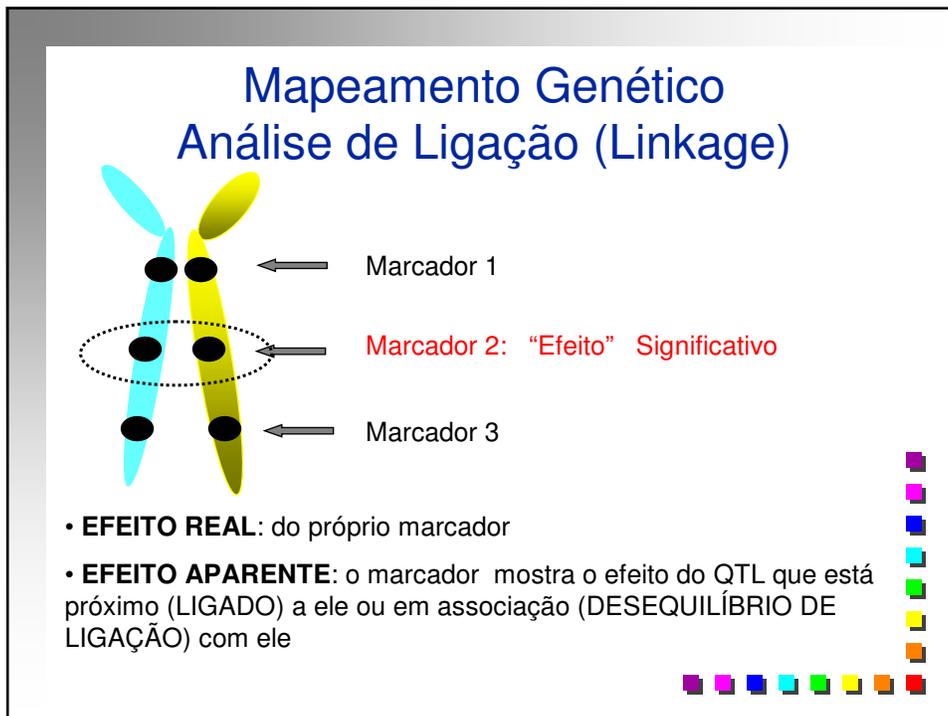
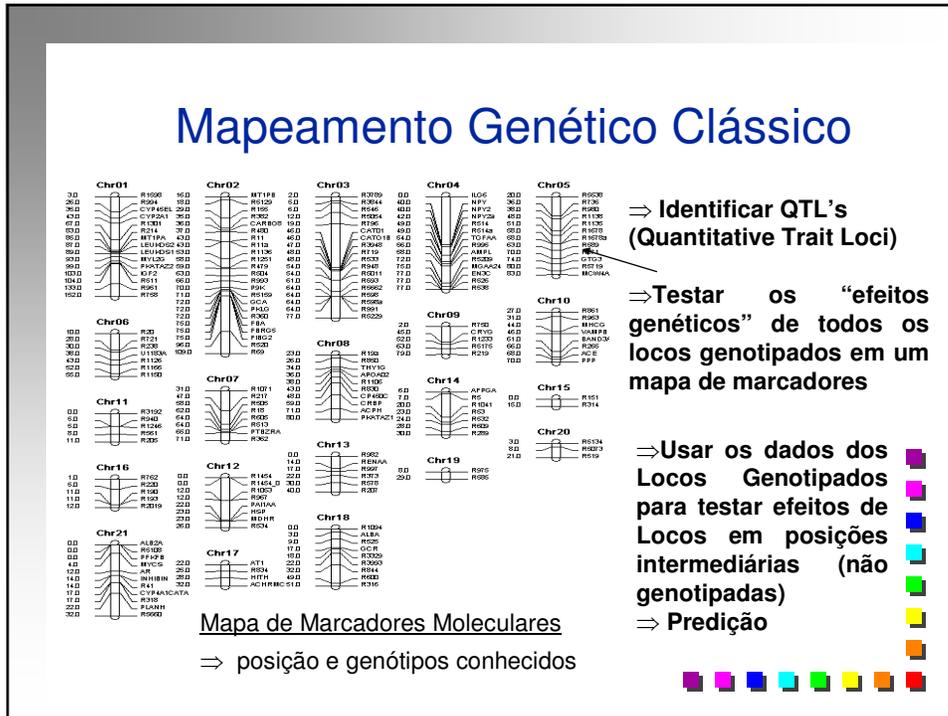
⇒ Mapas Genéticos: informação de Marcadores Moleculares
 distribuídos pelo Genoma (unidade: Morgan ou centi-Morgan)

mapas espaçados → 10cM → 3cM → 1cM → mapas densos

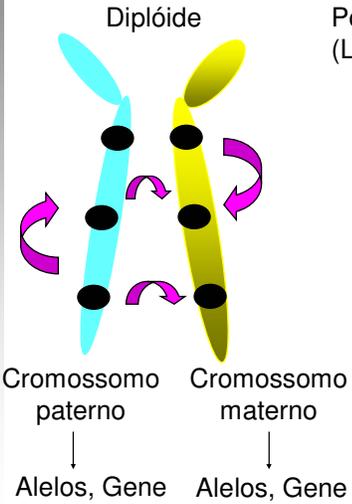
⇒ Mapas Físicos (Genômica): informação de Sequências de DNA
 (ACTGT...), expressão gênica

1cM: aproximadamente 1 milhão de pares de bases de DNA





Como os Genes “Conversam” entre si?



Possíveis Genótipos (Locos Dialélicos):

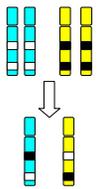
a a	A a	A A
b b	B b	B B
c c	C c	C C

Leis da Genética!

Ligação: Lei de Segregação

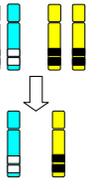
- Ligação entre Locos Genéticos \Rightarrow Lei de distâncias entre locos

Locos não ligados



Segregar material recombinante

Locos ligados



Segregar material não recombinante (íntegro)

\Rightarrow Fração de Recombinação (θ) e Distância (em cM) entre Locos

$\theta = 0$

Locos próximos: “dependência”
na origem dos alelos

\longleftrightarrow

$\theta = 0.5$

Locos distantes: “independência”
na origem dos alelos

\Rightarrow Delineamentos com Famílias (muitas gerações) são úteis

Outras Leis de Segregação de Genes

Modelo: 1 Loco Genético



A_1 A_2

$$P(A_1 A_2) = P(A_1) P(A_2)$$

“Equilíbrio de Hardy-Weinberg”

Modelo: Pares de LOCOS

Além da “ligação”, em termos de eventos de recombinação, temos:



A_1 A_2

B_1 B_2

“Equilíbrio de Ligação”

$$P(A_1 B_1) = P(A_1) P(B_1)$$

⇒ Desequilíbrio de Ligação: associação estatística entre locos

⇒ Estudos de Associação Caso-Controlle exploram esta associação



Efeitos Genéticos

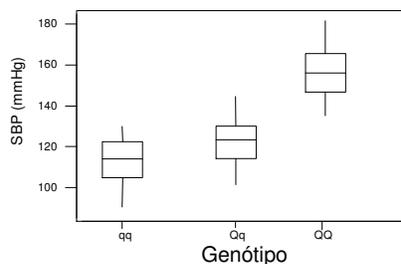


qq Qq QQ

Suposição:

Gene (ou Marcador Molecular)
 Genótipo Conhecido/Observável

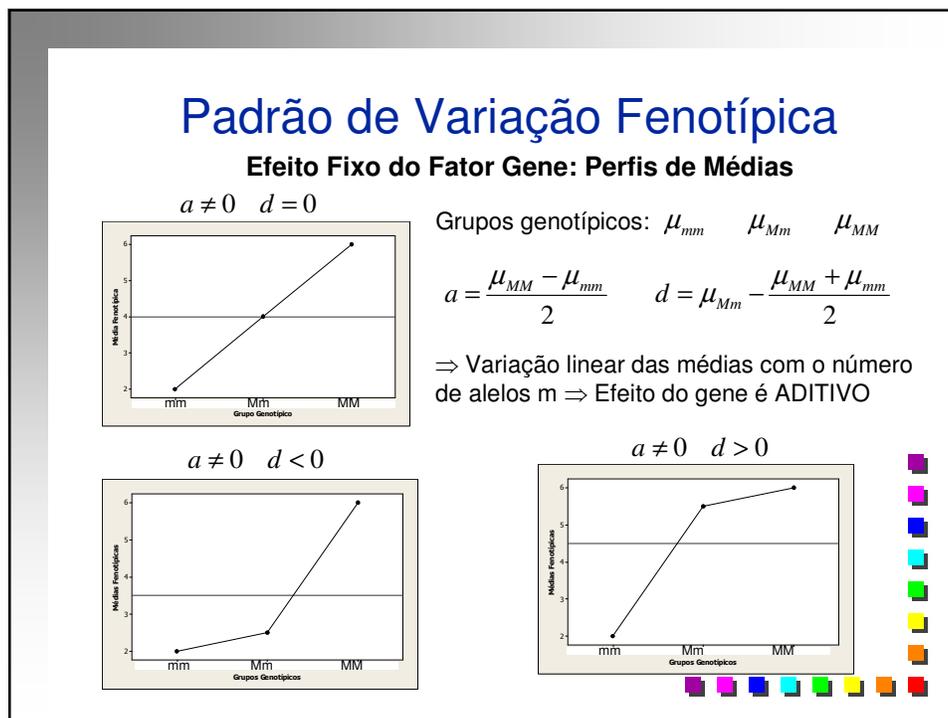
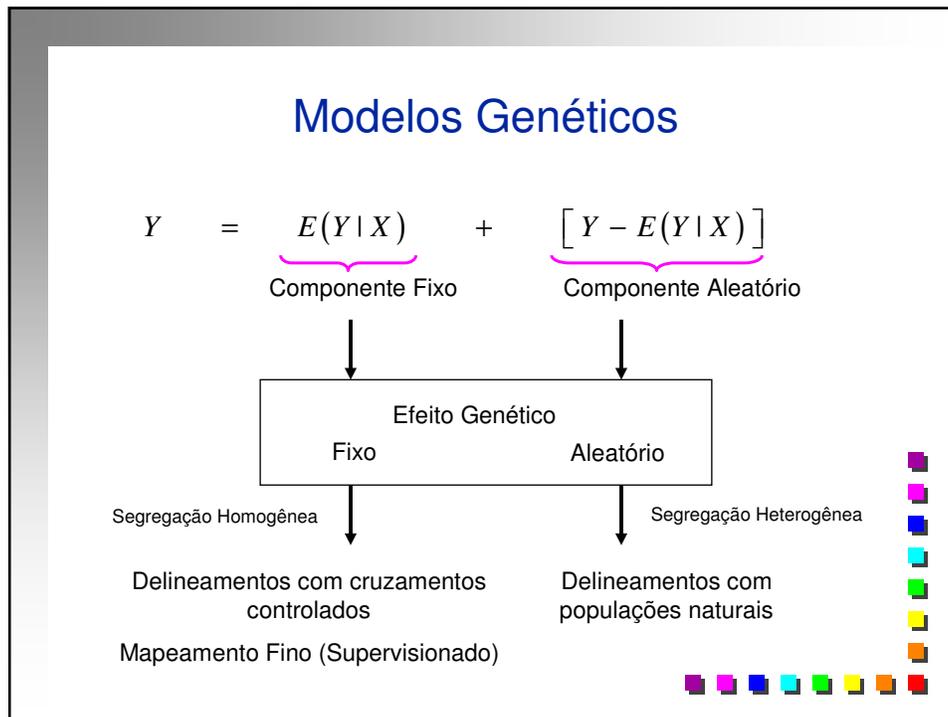
Existe efeito significativo do gene sobre o fenótipo (SBP) ?



Gene: 1 Fator em 3 níveis

⇒ 2 graus de liberdade para estudar o efeito do gene





Modelo Genético Quantitativo

- Modelos com Efeitos Fixos do Gene

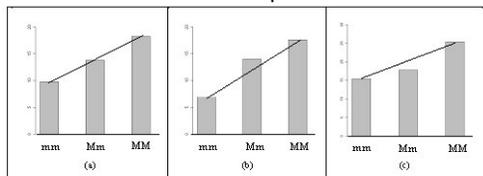
$$y_i = \mu + aX_{ai} + dX_{di} + e_i \Rightarrow E(y_i) = \mu + aX_{ai} + dX_{di}$$

$$X_{ai} = \begin{cases} -1 & \text{para mm} \\ 0 & \text{para Mm} \\ 1 & \text{para MM} \end{cases} \quad X_{di} = \begin{cases} 0 & \text{para mm} \\ 1 & \text{para Mm} \\ 0 & \text{para MM} \end{cases}$$

$$Var(y_i) = Var(e_i) = \sigma_e^2$$

$$H_0 : a = d = 0$$

Efeito do Gene: Contrastes entre Médias Loco com Genótipo Conhecido

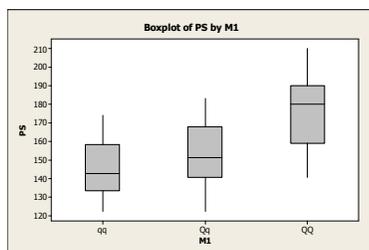


$$a = \frac{\mu_{MM} - \mu_{mm}}{2}$$

$$d = \mu_{Mm} - \frac{\mu_{MM} + \mu_{mm}}{2}$$



PS	M1
150.00	2
190.00	3
135.46	1
148.72	2
178.00	3
...	...
126.49	1
164.54	2
168.64	2
179.58	2



Exemplo

Estimação dos Ef Genéticos

$$\Rightarrow Y = X\beta + e$$

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & -1 & 0 \end{bmatrix} \quad \beta = \begin{bmatrix} \mu \\ a \\ d \end{bmatrix}$$

$$\mu = \frac{\mu_{qq} + \mu_{qq}}{2}$$

Analysis of Variance

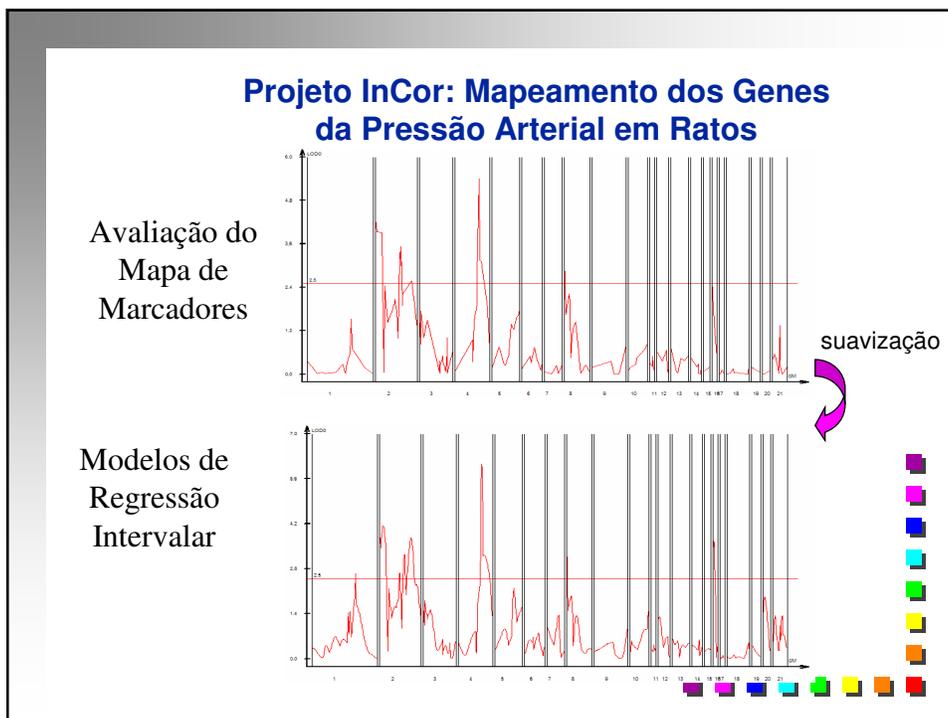
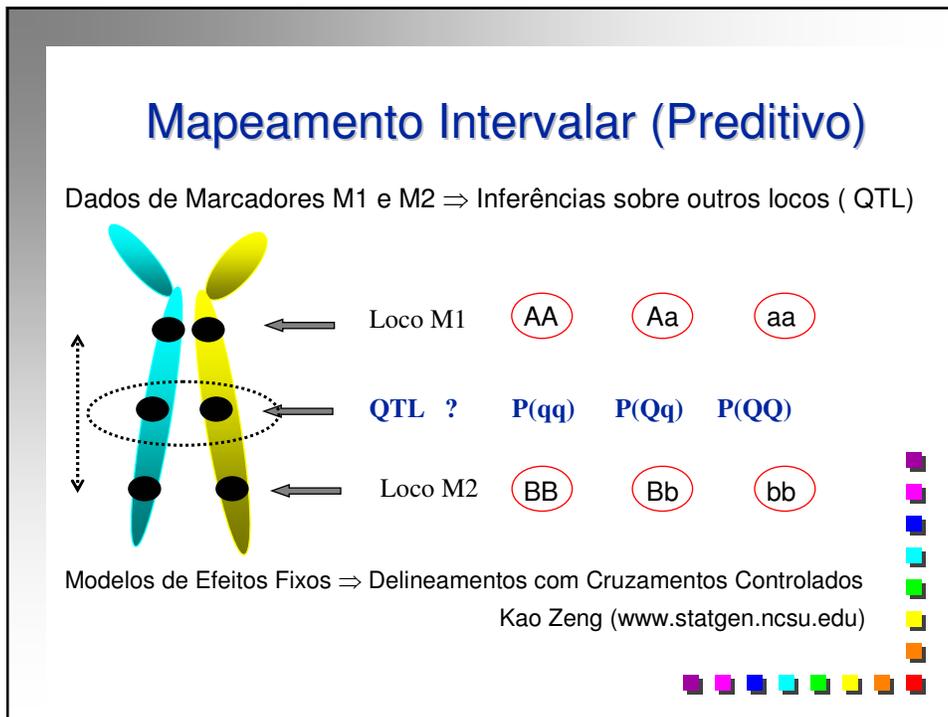
Source	DF	SS	MS	F	p
Regression	2	30149	15074	60.49	0.000
Residual Error	166	41366	249		
Total	168	71515			

$$PS = 161 + 16.3 X_a - 7.09 X_d$$

Predictor	Coef	SE Coef	T	p	
Constant	161.309	1.486	108.54	0.000	$\hat{a} = 16.31$
Adit	16.311	1.486	10.98	0.000	$\hat{d} = -7.089$
Domi	-7.089	2.806	-2.53	0.012	$\hat{\mu} = 161.309$

$$\hat{\mu}_{qq} = 145.00 \quad \hat{\mu}_{Qq} = 154.22 \quad \hat{\mu}_{QQ} = 177.62$$





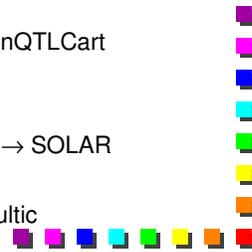
Alguns Recursos da Web

- Genoma Humano, Locus Genéticos, Doenças Genéticas e Tutoriais:
NCBI: <http://www.ncbi.nlm.nih.gov>
NHGRI - <http://www.genoma.gov>
[http://www.ornl.gov/sci/techresources/Human Genome/glossary/](http://www.ornl.gov/sci/techresources/Human_Genome/glossary/)
- Softwares de Análise de dados Genéticos:
<http://www.linkage.rockefeller.edu/soft/> → LINKAGE

<http://statgen.ncsu.edu/qtlcart/WQTLCart.htm> → WinQTLCart

<http://www.sfbr.org/public/software/solar/index.html> → SOLAR

<http://www.r-project.org/> → RQTL, Bioconductor, Multic



Alguns Livros

- Balding, D.J.; Bishop, M. and Cannings, C. (2003). **Handbook of Statistical Genetics**. 2th Ed. Wiley.
- Falconer, D.S. and Mackay, T.F.C. (1996). **Introduction to Quantitative Genetics**. 4th Ed. Prentice Hall.
- Ott, J. (1991). **Analysis of Human Genetic Linkage**. Johns Hopkins Univ. Press.
- Sham, P. (1998). **Statistics in Human Genetics**. Arnold.
- Sorensen, D. & Gianola, D. (2002). **Likelihood, Bayesian, and MCMC methods in Quantitative Genetics**. 2nd printing. Springer-Verlag.
- Speed, T (Ed) (2003). **Statistical of gene expression microarray data**. Chapman & Hall.

