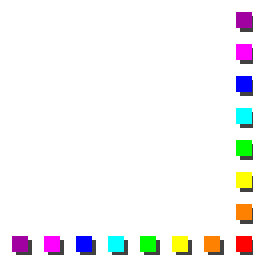


## *Mapeamento de Genes: Delineamentos com Famílias e Estudos Observacionais*

*Júlia Maria Pavan Soler*  
*pavan@ime.usp.br*

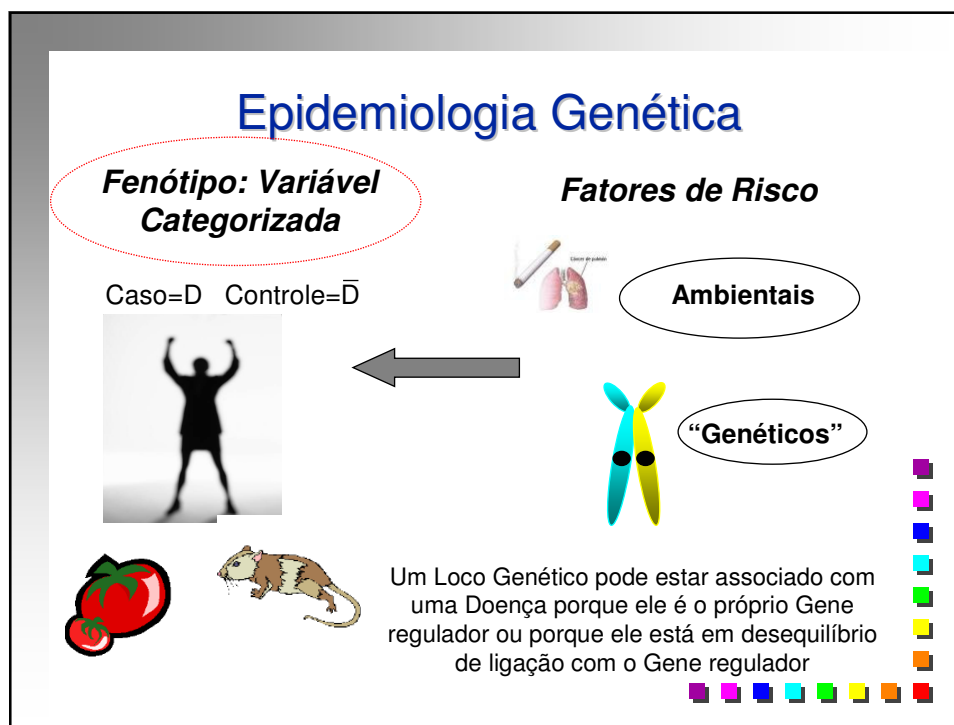
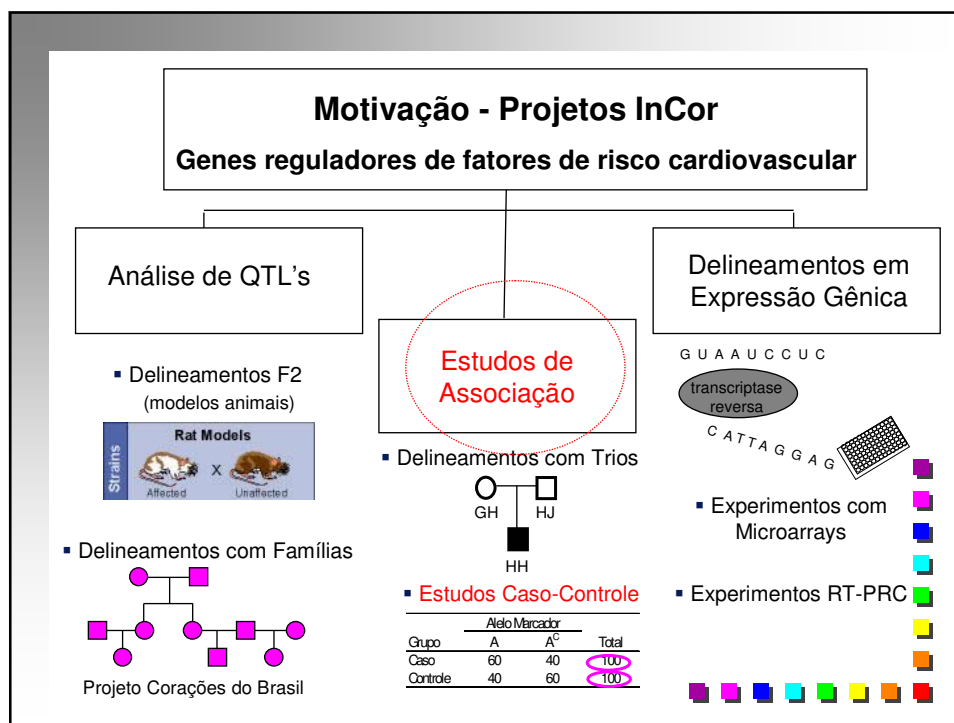
*DEST – U.F.P. – Curitiba*  
*Set / 2007*



## *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-Controlle  
Perspectivas





## Leis de Segregação de Genes

Modelo: 1 Loco Genético



$A_1 \quad A_2$

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

“Equilíbrio de Hardy-Weinberg”

Modelo: Pares de Locos

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



$A_1 \quad A_2$   
 $B_1 \quad B_2$

“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

## Estudos de Associação

Associação entre Doença e Fator de Risco

Estudo Prospectivo

Distribuição da Doença de acordo com o Fator de Risco G

Fator de Risco	Doença		Total
	D	D <sup>C</sup>	
Gene1	$n_{11}$	$n_{1.} - n_{11}$	$n_{1.}$
Gene2	$n_{21}$	$n_{2.} - n_{21}$	$n_{2.}$



Fator de Risco	Doença		Total
	D	D <sup>C</sup>	
Gene1	$\pi_1$	$1 - \pi_1$	1
Gene2	$\pi_2$	$1 - \pi_2$	1

$$H_0 : \pi_1 = \pi_2 \Leftrightarrow RR = 1 \quad (\notin \text{associação})$$

$$RR = \frac{P(D|G1)}{P(D|G2)} = \frac{\pi_1}{\pi_2}$$

Risco Relativo da Doença  
Razão das Probabilidades de Incidência  
(Penetrância do Gene)

Alternativa ⇒ Estudos Retrospectivos ( Caso – Controle )

## Estudos Caso-Controle (Retrospectivo)

### Associação entre Doença e Fator de Risco Genético

Distribuição do Fator Genético nos grupos Caso e Controle

	Gene		Total	Modelo Produto de Binomiais
	G1	G2		
Caso	$p_1$	$1-p_1$	1	
Controle	$p_2$	$1-p_2$	1	

Medida de Associação: Razão de Produtos Cruzados ( $\theta$ : odds ratio)

$$\theta = \frac{p_1 / (1-p_1)}{p_2 / (1-p_2)} = \frac{p_1(1-p_2)}{p_2(1-p_1)}$$

$$\log \theta = \log ito_{G1} - \log ito_{G2}$$

$H_0: \theta = 1$   $\neq$  associação

0	1	$+\infty$	$\theta$
$-\infty$	0	$+\infty$	$\log \theta$

## Estudos de Associação

### Análise de Associação para Dados de Genótipos

Genótipo do indivíduo: é "invariante" ao longo da vida

Estudos Caso – Controle

	Genótipo			Total
	AA	Aa	aa	
Caso	89	369	342	800
Controle	56	250	266	572

$$\hat{\theta}_{AA} = \frac{89 \cdot 266}{56 \cdot 342} = 1,236$$

$$\hat{\theta}_{Aa} = \frac{369 \cdot 266}{250 \cdot 342} = 1,148$$

Risco da Doença para indivíduos que carregam o genótipo AA é 1,24 vezes maior que o risco para aa

Hipótese de Teste ( $\neq$  associação):

$$H_0 : \theta_{AA} = \theta_{Aa} = 1$$

## Estudos de Associação

Distribuição Genotípica nos grupos Caso e Controle

	Genótipo			Total
	AA	Aa	aa	
<b>Caso</b>	89	369	342	800
<b>Controle</b>	56	250	266	572
	145	619	608	

### Modelo de Regressão Logística

$$F(p) = \log \frac{p}{1-p} = X'\beta$$

Logito	$\mu$	$\beta_1$	$\beta_2$
$\text{Logito}_{AA} = \log(89/56)$	1	1	0
$\text{Logito}_{Aa} = \log(369/250)$	1	0	1
$\text{Logito}_{aa} = \log(342/266)$	1	0	0

$$\left. \begin{aligned} \log \text{ito}_{AA} &= \mu + \beta_1 \\ \log \text{ito}_{Aa} &= \mu + \beta_2 \\ \log \text{ito}_{aa} &= \mu \end{aligned} \right\} \begin{aligned} \log \theta_{AA} &= \log \text{ito}_{AA} - \log \text{ito}_{aa} = \beta_1 \\ \log \theta_{Aa} &= \log \text{ito}_{Aa} - \log \text{ito}_{aa} = \beta_2 \end{aligned}$$

$$H_0: \theta_{AA} = \theta_{Aa} = 1$$

$$\Downarrow$$

$$H_0: \beta_1 = \beta_2 = 0$$

## Estudos de Associação

Distribuição Genotípica nos grupos Caso e Controle

	Genótipo			Total
	AA	Aa	aa	
<b>Caso</b>	89	369	342	800
<b>Controle</b>	56	250	266	572
	145	619	608	

### Modelo de Regressão Logística

$$\left. \begin{aligned} \log \text{ito}_{AA} &= \mu + \beta_1 \\ \log \text{ito}_{Aa} &= \mu + \beta_2 \\ \log \text{ito}_{aa} &= \mu \end{aligned} \right\} \begin{aligned} \log \theta_{AA} &= \log \text{ito}_{AA} - \log \text{ito}_{aa} = \beta_1 \\ \log \theta_{Aa} &= \log \text{ito}_{Aa} - \log \text{ito}_{aa} = \beta_2 \end{aligned}$$

$$H_0: \theta_{AA} = \theta_{Aa} = 1$$

$$\Downarrow$$

$$H_0: \beta_1 = \beta_2 = 0$$

$$LR = -2 \ln \frac{L(\mu)}{L(\mu, \beta_1, \beta_2)} \sim \chi^2_2 \quad \text{Estat. razão de verossimilhanças}$$

$L$ : função de verossimilhança (produto de multinomiais)

$\Rightarrow$  Estatística de Wald, Estatística Score (Teoria de MQG)

## Estudos de Associação

Análises no Nível de Indivíduo (n) e no Nível Cromossômico (2n)

Distribuição Genotípica

	Genótipo			Total
	AA	Aa	aa	
<b>Caso</b>	89	369	342	800
<b>Controle</b>	56	250	266	572

Análise no nível de indivíduos (n)

⇒ Análise de Associação: 2 graus de liberdade

Distribuição Alélica

	Alelos		Total
	A	a	
<b>Caso</b>	547	1053	1600
<b>Controle</b>	362	782	1144

Análise no nível cromossômico (2n)

⇒ Duplica o tamanho amostral

⇒ Análise de Associação: 1 grau de liberdade

Sob Eq. H-W

## Estudos de Associação

Análises no Nível de Indivíduo (n) e no Nível Cromossômico (2n)

Estudos Caso – Controle

Distribuição Alélica nos grupos Caso e Controle

	Alelos		Total
	A	a	
<b>Caso</b>	547	1053	1600
<b>Controle</b>	362	782	1144

Análise no nível cromossômico (2n)

⇒ Análise de Associação: 1 grau de liberdade

$$\hat{\theta} = \frac{547 * 782}{362 * 782} = 1,122$$

Risco da Doença para o alelo A

⇒ Sob Modelo Multiplicativo (e Eq. Hardy-Weinberg)  $\left\{ \begin{array}{l} \text{Risco do Genótipo Aa} = 1,122 \\ \text{Risco do Genótipo AA} = (1,122)^2 = 1,254 \end{array} \right.$

## Estudos de Associação

Distribuição Genotípica nos grupos Caso e Controle

	Genótipo			Total
	AA	Aa	aa	
<b>Caso</b>	89	369	342	800
<b>Controle</b>	56	250	266	572
	145	619	608	

### Modelo de Regressão Logística Reduzido sob Dados de Genótipo

Logito	$\mu$	$\beta$	
Logito <sub>AA</sub> $\log(89/56)$	1	2	$\log \theta_{AA} = 2\beta \Rightarrow \theta_{AA} = (1,125)^2 = 1,266$
Logito <sub>Aa</sub> $\log(369/250)$	1	1	
Logito <sub>aa</sub> $\log(342/266)$	1	0	$\log \theta_{aa} = \beta \Rightarrow \theta_{aa} = 1,125$

$$H_0: \beta = 0 \quad LR = -2 \ln \frac{L(\mu)}{L(\mu, \beta)} \sim \chi_1^2$$

Dados Genotípicos: testar do efeito linear do # de alelos no risco da doença

⇒ Análise com 1 g.l. : não assume Eq. H-W mas assume modelo multiplicativo



## Estudos de Associação

Análises no Nível de Indivíduo (n) e no Nível Cromossômico (2n)

Distribuição Genotípica

	Genótipo			Total
	AA	Aa	aa	
<b>Caso</b>	89	369	342	800
<b>Controle</b>	56	250	266	572

Distribuição Alélica

	Alelos		Total
	A	a	
<b>Caso</b>	547	1053	1600
<b>Controle</b>	362	782	1144

Dados	Modelo	Suposição		Estimativas	
		Eq. H-W	Mod. Multipl.	$\theta_{Aa}$	$\theta_{AA}$
Genótipos (n)	2 g.l.			1,148	1,236
	1 g.l.		x	1,125	1,266
Alelos (2n)	1 g.l.	x	x	1,122	1,254

Modelo Probabilístico: Produto de Multinomiais (Regressão Logística)

⇒ Modelos Mais Gerais: Produto de Poissons (Modelo Log-Linear)



## Aplicação Estudo de Vitória (ES) - InCor

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

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

Dados de 9  
Marcadores  
Dialélicos.

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



## Estimativas dos Odds Ratios no nível genotípico

Marcador	Hipertensão	valor p	Obesidade	valor p
<b>B1BK</b>	0,7534	0,0765	0,9881	0,9401
	0,6770	<b>0,0130</b>	0,9285	0,6365
<b>NHPS2.5</b>	-	-	-	-
	1,1895	0,4520	0,8961	0,6380
<b>BAR-2(16) Cat</b>	0,7509	0,0787	1,3081	0,0908
	0,7348	<b>0,0198</b>	1,3341	<b>0,0277</b>
<b>BAR-2(27) Cat</b>	1,2212	0,3300	1,4725	0,0556
	1,2061	0,3700	1,1496	0,5003
<b>ECA Cat</b>	1,0835	0,6310	1,0248	0,8795
	1,1026	0,5230	0,9408	0,6796
<b>ANGO Cat</b>	1,4839	<b>0,0108</b>	1,2959	0,0820
	1,2776	0,1045	1,0660	0,6590
<b>ENOS Cat</b>	1,3912	0,1310	0,7768	0,2710
	1,1499	0,2100	0,9961	0,9710
<b>p22 PHOX</b>	0,7879	<b>0,0469</b>	1,0700	0,6990
	0,7532	<b>0,0104</b>	0,9774	0,8370
<b>GNB3</b>	1,2586	0,1250	0,8398	0,2410
	1,0849	0,5120	0,9604	0,7360



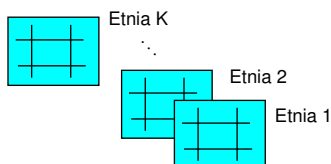


Estimativas dos *Odds Ratios* no nível genotípico e modelo multiplicativo

Marcador	Hipertensão	Valor p	Obesidade	Valor p
B1BK	0,9176	0,2650	1,0104	0,8900
NHPS2,5	1,1413	0,5590	0,8087	0,6920
BAR-2 (16) Cat	0,8407	<b>0,0456</b>	1,1574	0,0816
BAR-2 (27) Cat	1,0436	0,4600	1,2432	<b>0,0086</b>
ECA Cat	1,0307	0,7090	1,0245	0,7800
ANGO Cat	1,2078	<b>0,0122</b>	1,1507	<b>0,0596</b>
ENOS Cat	1,1649	0,0768	0,9998	0,4890
p22 PHOX	0,8068	<b>0,0080</b>	1,0152	0,8480
GNB3	1,1190	0,1920	0,9210	0,2820

Marcador Hipertensão	AIC	
	Genotípico	Genotípico (multiplicativo)
BAR-2 (16) Cat	26,126	26,050*
ANGO Cat	26,676	23,281*
p22 PHOX	26,354	24,123*

## Problemas com os Estudos de Associação



**Falsos positivos e Não Reprodutibilidade dos Estudos**

⇒ *Estratificação da População: fator de confundimento (Ex.: raça)*

- Controlar a estratificação via Delineamento Experimental:
  - Pares de Irmãos (afetados e não afetados)
  - Marcadores Moleculares Anônimos
- Fator de Confundimento Conhecido: Uso de análises apropriadas
  - Estatística de Cochran-Mantel-Haenszel
  - Mod. de Regressão Logística com efeito de estratificação

## Problemas com os Testes de Associação

Alelo Marcador	A	a	Total
<b>Caso</b>	60	40	<b>100</b>
<b>Controle</b>	40	60	<b>100</b>

“Paradoxo de Simpson”  
Falsos Positivos!  
Como Combinar Tabelas?

$$\text{Razão de chances} = (60 \cdot 60) / (40 \cdot 40) = 2.25$$

Grupo Étnico I

Alelo Marcador	A	a	Total
<b>Caso</b>	30	30	<b>60</b>
<b>Controle</b>	30	30	<b>60</b>

$$\text{Razão de chances} = 1.00$$

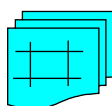
Grupo Étnico II

Alelo Marcador	A	a	Total
<b>Caso</b>	30	10	<b>40</b>
<b>Controle</b>	10	30	<b>40</b>

$$\text{Razão de chances} = 9.00$$

## Estudos de Associação

Estrutura Populacional  $\Rightarrow$  Diferentes Raças



	A	a	Total
Caso	5050	5950	11000
Controle	1095	9005	10100

$$\hat{\theta} = 6,98$$

Etnia 1

	A	a	Total
Caso	50	950	1000
Controle	1000	9000	10000

$$\hat{\theta} = 0,47$$

Etnia 2

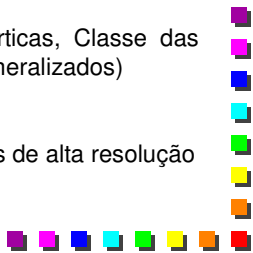
	A	a	Total
Caso	5000	5000	10000
Controle	95	5	100

$$\hat{\theta} = 0,05$$

- Verificar se o Grau de Associação é homogêneo (não conflitante) entre os diferentes estratos.
- Se não for conflitante testar se a Associação é significativa.
- Estimar uma Medida de Associação Comum, se for o caso.

Perspectivas

- Mapeamento Genético Fino (Refinar posições candidatas no nível de sequências de DNA
  - ⇒ Informação IBS, (Análise Combinada de Ligação e Desequilíbrio de Ligação)
  - ⇒ Experimentos com Microarrays (Expressão do Gene)
  - ⇒ Estudos de Associação de Alta Resolução
- Análises Robustas e Correções para Múltiplos Testes:
  - ⇒ Estatísticas de Teste Robustas (Correções)
  - ⇒ Funções de Densidade mais Gerais: Leptocúrticas, Classe das Elípticas, Fenótipos Discretos (Modelos Lineares Generalizados)
- Procedimentos de Busca por Modelos Ótimos
  - ⇒ Como explorar o espaço dos Genótipos Multilocos de alta resolução



Genética Genômica

Banco de Dados

Estrutura das u.a.				Fenótipos				Genótipos Mapa de SNP's				Fenótipos de Expressão					
FAM	INDIV	PAI	MÃE	FENO1	FENO2	...	FENO1	SNP1	SNP2	...	SNP2.882	FENOE1p1	FENOE1p2	...	FENOE1p3.554	REPL1	REPL2
1	1																
1	2																
...	...																
1	14																
2	15																
...	...																
2	28																
...	...																
14																	
14																	
14	~200																

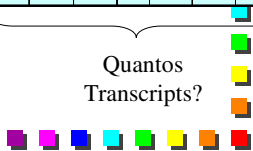
Cruzamentos controlados

Famílias

Indivíduos Não relacionados

Quantos Locos?

Quantos Transcripts?



## Modelos de C.V. Multilocos

- Efeito de Epistasia – “Interação entre Genes”

$$Y_f = \mu + X_f \beta + q_{1f} + q_{2f} + q_{12f} + g_f + e_f$$

um novo efeito aleatório associado ao componente de variância devido ao efeito de epistasia

$$\Omega_f = \hat{\Pi}_1 \sigma_{q_1}^2 + \hat{\Pi}_2 \sigma_{q_2}^2 + \underbrace{\hat{\Pi}_1 \oplus \hat{\Pi}_2}_{\text{Produto de Hadamard}} \sigma_{12}^2 + 2\Phi \sigma_g^2 + I_f \sigma_e^2$$

	$j$	$j'$	
	$\pi_{1jj'}$	$\pi_{2jj'}$	
QTL1	$\pi_{1jj'}$	$\pi_{2jj'}$	
QTL2	$\pi_{2jj'}$	$\pi_{1jj'}$	

Produto de Hadamard

$\pi_{12jj'} = \pi_{1jj'} \pi_{2jj'}$  Proporção de Alelos ibd nos dois Locus

Suposição: Locus em Equilíbrio de Ligação

## Estudos de Associação Multilocos

- Análise no Nível Genotípico (Efeito Linear)

$$\text{logit}[P(Y=1|X_m)] = \beta_0 + \beta_m X_m \quad \text{para } m = (1, \dots, M)$$

$$\text{logit}[P(Y=1|X_1, \dots, X_M)] = \beta_0 + \sum_{m=1}^M \beta_m X_m$$

$$\text{logit}[P(Y=1|X_1, \dots, X_M)] = \beta_0 + \sum_{m=1}^M \beta_m X_m + \sum_{m=1}^M \sum_{l=m+1}^M \beta_{m*l} X_{m*l} + \dots$$

- Análise no Nível de Haplótipo

$$\text{logit}[P(Y=1|X_1, \dots, X_H)] = \beta_0 + \sum_{h=2}^H \gamma_h X_h$$

$X_k = 0, 1, 2$  especifica o # de cópias do haplótipo

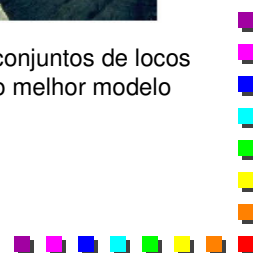
## Seleção de Modelos



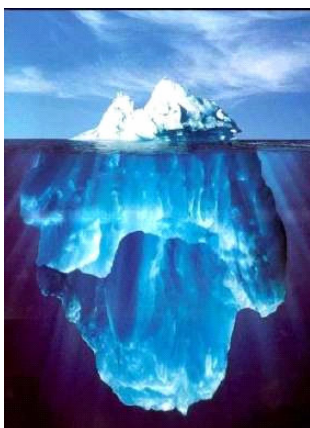
⇒ Problema de alta dimensionalidade: o número de subconjuntos de locos do mapa de marcadores a serem avaliados em busca do melhor modelo cresce exponencialmente

$$\binom{n}{k} = \frac{n!}{k!(n-k)!}$$

n: # de locos no mapa  
k: # de QTL's



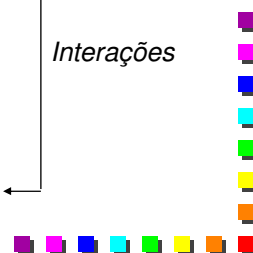
## O estado da arte da Genômica “a ponta do iceberg”



DNA → Genes → Proteínas → Fenótipo

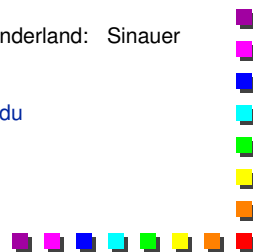
Núcleo  
Célula  
Órgão  
Indivíduo  
População  
biosfera

Interações



## Bibliografia

- Balding, D.J.; Bishop, M. and Cannings, C. (2003). **Handbook of Statistical Genetics**. 2th Ed. Wiley.
- Conti, D.V. & Gauderman, W.J. (2004). SNPs, haplotypes, and model selection in a candidate gene regions: the SIMPlE analysis for multilocus data. *Genet. Epidemiol.* **27**, 429-441.
- 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.
- Terwillinger, J.D. and Ott, J. (1994). **Handbook of Human Genetic Linkage**. Johns Hopkins Univ. Press.
- Weir, B.S. (1996). **Genetic data analysis II**. Sunderland: Sinauer Associates.
- Programas Computacionais: <http://linkage.rockefeller.edu>  
STAT4



Muito Obrigado!

Júlia Maria Pavan Soler  
[pavan@ime.usp.br](mailto:pavan@ime.usp.br)

