Computational Biology 7/10/19

Optimal - given Scoring Scheme.

Dot matrix + BLAST = misalignment

GWAS → genetic Risk factor identification for diseases.
(AD, diabetes). → SNP identification.

GWAS → 2 groups. (case, control).

H_S minor variant
↳ SNP

H_N major variant

$$OR = \frac{\text{odds of minor}}{\text{odds of major variant}} = \frac{D_S/H_S}{D_N/H_N}$$

if OR > 1 → increase of minor variant
= 1 → no association
< 1 → decrease of minor variant.

Slide 7/45 → see slides for table

$$OR = \frac{D_S/H_S}{D_N/H_N}$$

$$= \frac{2104/2676}{1836/3321} = 1.3871.$$

OR P-value → (χ^2 OR Fisher's exact test)

Fisher's test →

	B_1	B_2	
A_1	$\boxed{a}$	b	$a+b$
A_2	c	d	$c+d$
	$a+c$	$b+d$	n

draw from $\boxed{a}+c$

$$p\text{-value} = \sum_{i=a}^{a+b} \frac{\binom{a+b}{i}\binom{c+d}{a+c-i}}{\binom{n}{a+c}}$$

Fisher's exact test can only be calculated for $\binom{n}{k}$ small numbers & difficult to eliminate in

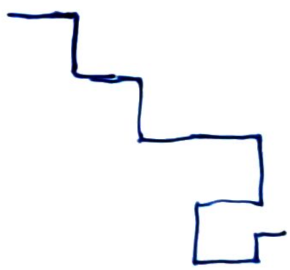
Pearson invented the χ^2-test to account for bigger numbers. Deviance calculated using the hypergeometric distribution

Bonferroni: $\dfrac{\alpha}{\#\ \text{hypothesis}}$

CACTT
GTCC
CTAG

Quantifying variation between sequences.

Markovian process is memoryless.



Substitution rate matrix:

$$Q = \begin{array}{c} T \\ C \\ A \\ G \end{array} \begin{bmatrix} \cdots \\ \cdots \\ \cdots \\ \cdots \end{bmatrix} \begin{array}{l} 0\ (\text{sum}) \\ 0\ (\text{sum}) \\ 0\ (\text{sum}) \\ 0\ (\text{sum}) \end{array}$$

$\tau = k \times \Delta t$

$$P(X > \tau) = (1-\alpha\Delta t)^k = (1-\alpha\Delta t)^{\tau/\Delta t} = \Big(\underbrace{(1-\alpha\Delta t)^{1/\Delta t}}_{\to 0}\Big)^{\tau}$$

$$\lim_{\Delta t \to 0} = e^{-\alpha \tau} \qquad e^{x} = \lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^{n}$$