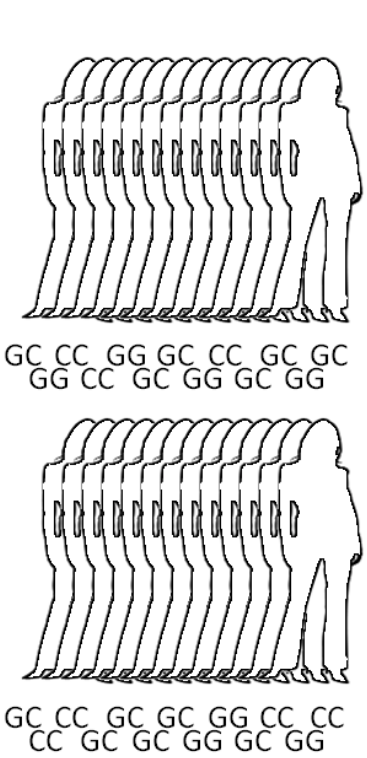
Association

H0: no association ; H1: associated

1. Chi-square (early)

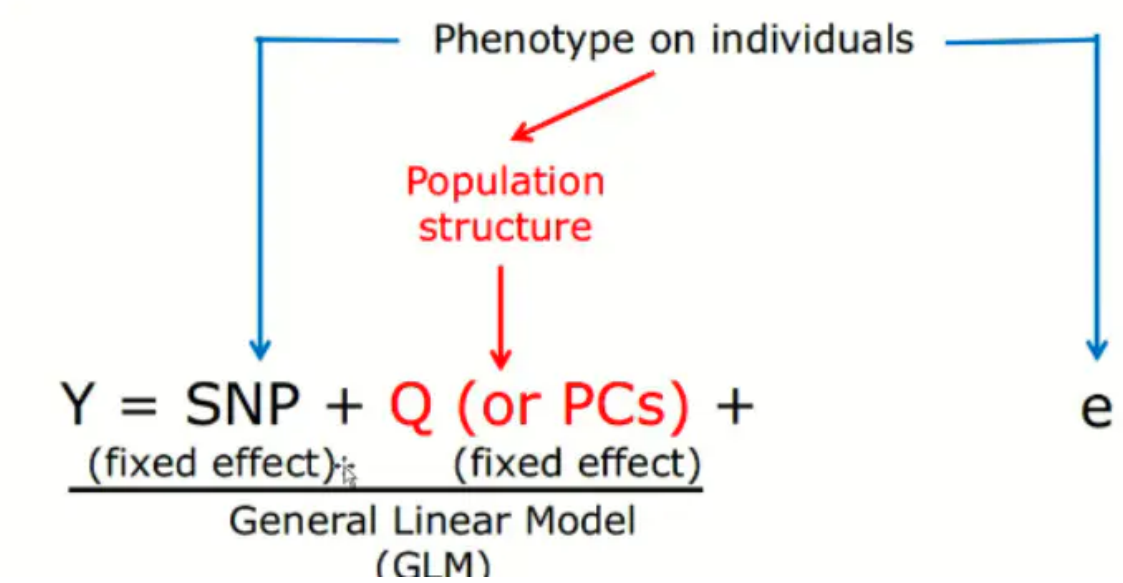
[Example calculation illustrating the methodology of a case-control GWA study]



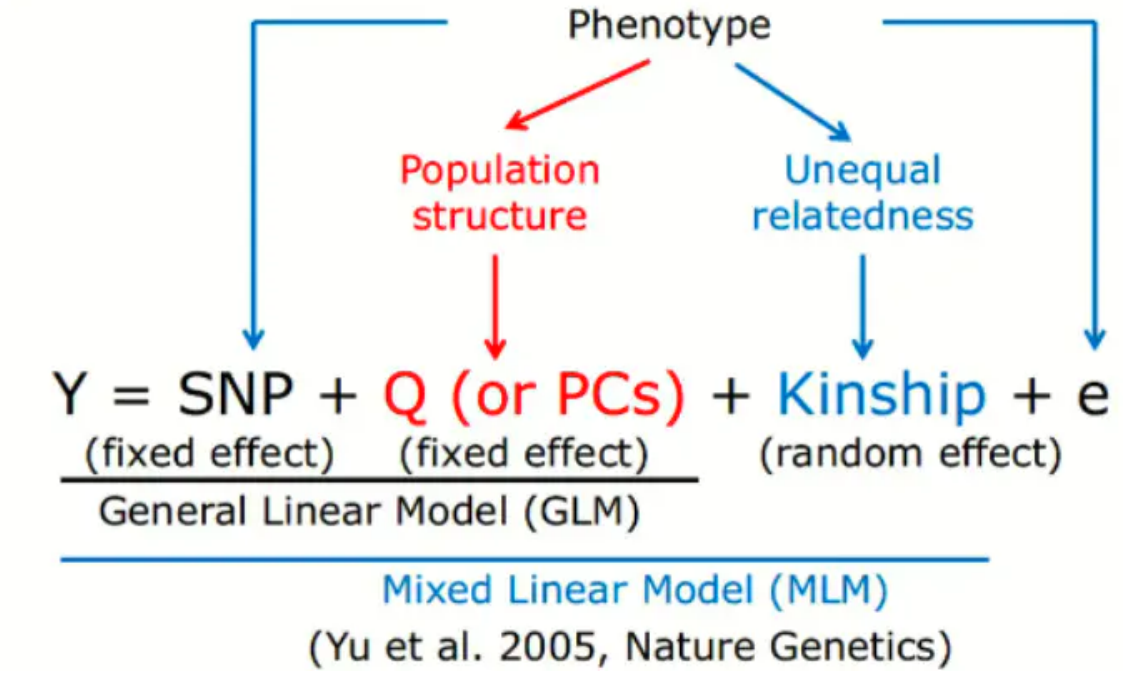
SNP1	*SNP2*	*SNP…*
Cases Count of G: 2104 of 4000 Frequency of G: 52.6%	**Cases** Count of G: 1648 of 4000 Frequency of G: 41.2%	*Repeat for all SNPs*
Controls Count of G: 2676 of 6000 Frequency of G: 44.6%	**Controls** Count of G: 2532 of 6000 Frequency of G: 42.2%	
P-value: $5.0 \cdot 10^{-15}$	**P-value:** 0.33	

2. Linear Regression

a. General linear regression (lr)



b. mixed linear model (MLM)



> The advantages of mixed linear model association (MLMA) include preventing false-positive associations due to population or relatedness structure, and increasing power by applying a correction that is specific to this structure.
>
> Yang, Jian et al. "Advantages and pitfalls in the application of mixed-model association methods." *Nature genetics* vol. 46,2 (2014): 100-6. doi:10.1038/ng.2876

c. logistic regression

$$logit\,(\mu_i) = X_i\alpha + G_i\beta + b_i$$

$$\mu_i = P(y_i = 1 | X_i, G_i, b_i)$$

$$\mathbf{y} = \mathbf{x}_{\text{snp}}\beta_{\text{snp}} + \mathbf{X}_c\beta_c + \mathbf{g} + \mathbf{e}$$

Computational cost of EMMAX, FaST-LMM, GEMMA, GRAMMAR-Gamma and GCTA

For each method we list the computational cost of each step (see main text).

	Building GRM	Variance components	Association statistics
EMMAX	$O(MN^2)$	$O(N^3)$	$O(MN^2)$
FaST-LMM*	$O(MN^2)$	$O(N^3)$	$O(MN^2)$
GEMMA	$O(MN^2)$	$O(N^3)$	$O(MN^2)$
GRAMMAR-Gamma	$O(MN^2)$	$O(N^3)$	$O(MN)$
GCTA	$O(MN^2)$	$O(N^3)$	$O(MN^2)$

*If $M<N$, the computational cost of FaST-LMM can be reduced to $O(M^2N)$.

M~marker , N ~ individual