

# GWAS.

9/1

- \* About human genome  $\rightarrow$  3 billion base pairs
- \* De oxyribonucleic Acid
- \* Single Nucleotide Polymorphisms (SNPs)
  - $\rightarrow$  A DNA variation when a single A, T, C, G differ between species or paired chromosomes
- \* Basic Definitions
  - I Locus  $\rightarrow$  A specific location in genome, Could be a single DNA base to gene to bend
  - II Variant  $\rightarrow$  A genetic difference relative to human reference genome.
  - III Polymorphism  $\rightarrow$  A variant common in population (1% Allel, not individual-specific)
  - IV Allele  $\rightarrow$  A genetic variant on a specific locus
  - V Genotype  $\rightarrow$  The combination of allele in an individual

A SNP with two allele, A, a.

P = Percentage of A

$$P + q = 1. \rightarrow \text{Variant}$$

q = Percentage of a

$$P^2 + 2Pq + q^2 = 1. \rightarrow \text{Genotype}$$

↓  
Could used to estimate the  
Genotype derived disease incident rate

### \* Hardy-Weinberg Equilibrium

An ideal population; the observed genotype frequency is identical to expected frequency.

No mutation, no assortative mating, no migration

no natural selection

→ Most population stay in HWE,  $HWE \sim \chi^2$

	AA	Aa	aa	Total
observation	$N_{AA}$	$N_{Aa}$	$N_{aa}$	$N$
Expected	$Np^2$	$Npq$	$Nq^2$	$N$

$$\sum g=AA \left( \frac{(Og - Eg)^2}{Eg} \right) \sim \chi^2$$