18/8 – Association studies II

- Association studies
  - Looks to find traits that correlate with marker alleles in populations
  - Compares unrelated individuals by looking for markers in LD
- Linkage disequilibrium (LD)
  - Non-random association of alleles in the population
  - Alleles at neighbouring loci tend to co-segregate (have stronger LD)
- Linkage Disequilibrium Mapping
  - Population based
  - Look for variant allele in LD with disease
  - If most affected individuals in a population share the same mutant allele, then LD can be used to locate the chromosomal region harbouring the mutant allele
    - Able to see marker alleles in association with disease genes
- Haplotypes
  - Groupings of specific linked SNPs (alleles) on a chromosome segment that stay together and can be tracked through pedigrees
  - Within the same haplotype, all genetic markers (SNPs etc) are inherited together
  - Allows for mapping of particular diseases
    - If a disease gene is produced within a specific haplotype region and passed through generations, it's very likely that people in the population with that disease also have that haplotype
- LD mapping
  - There is correlation between the presence of certain marker allele (or haplotype) and the disease gene allele (LD)
  - The correlation (LD) is based on founder effect:
    - The disease allele was born a long time ago on a certain ancestral chromosome, and majority of disease alleles existing presently predate from that original mutation
- Allelic Association: Three Common Forms
    - Correlation between marker allele and trait
  - Direct Association
    - Marker is directly involved with disease phenotype (within gene)
  - Indirect Association
    - Marker is indirectly involved with disease phenotype (in LD with gene)
  - Spurious association
    - Apparent association due to poor study design
    - Can be caused by population stratification
- Population Stratification
  - Recent admixture of populations
  - Requirements
    - Group differences in allele frequency (unrelated allele differences)
    - Group differences in outcome (diseased/ not diseased individuals)
  - Leads to spurious association
  - In epidemiology, this is a classic matching problem, with genetics as a confounding variable
- Spurious association is caused by two factors in population stratification
  - Different proportions of allele frequencies in subpopulations in case and controls
  - Different allele frequencies of subpopulations
- Population stratification solutions

    - Family-based controls (e.g. TDT - looks at number of parents heterozygous for an allele that pass that allele to affected offspring compared with the nondisease allele - so is really a family association study)
    - Genetic control: extra genotyping – Look for evidence of background population substructure and account for it
- Linkage vs. Association
    - Linkage analyse
        - Looks for relationship between a marker and disease within a family
            - May be different markers for each family
        - Yields broad chromosome regions harbouring many genes
        - Resolution comes from recombination events in families assessed
        - Good in terms of needing few markers
        - Poor in terms of finding specific variants involved
    - Association analyse
        - Looks for relationship between a marker and disease between families
            - Must be same marker in all families
        - Yields fine-scale resolution of genetic variants
        - Resolution comes from ancestral recombination events
        - Good in terms of finding specific variants
        - Poor in terms of needing many markers