Mapping arrays

Main applications of mapping arrays:

- Genotyping for linkage analysis and large-scale association studies.
- Studies of genetic variability in normal populations.
- Study of chromosomal abnormalities in cancer cells.

Definitions:

- **SNP (single nucleotide polymorphism):** a biallelic genetic polymorphism at a specific position in the DNA.
  - A SNP is usually defined to be a locus in which both variants are present in at least 1% of the population.
  - 2/3 of SNPs involve a substitution of cytosine (C) with thymine (T).
  - In humans, SNP’s tend to be separated by around 100-300 bases, so there are roughly $3 \times 10^9/100 \approx 3 \times 10^7$ SNP’s in the human genome. The position of SNP’s is fairly stable over many generations.
  - Most likely, only a small fraction of SNP’s have a phenotypic effect.

- **Copy number:** The number of copies of a chromosome in a cell. The normal copy number for autosomes is two.

- **Polysomy, monosomy, trisomy:** abnormalities in which a chromosome is present more than two times (polysomy), three times (trisomy) or one time (monosomy), rather than at the normal autosomal copy number of two.

- **Amplification:** The presence of additional DNA fragments (beyond the normal complement of chromosomes) that are copies of parts of normal chromosomes. Amplifications consisting of a few copies or hundreds of copies can be present in a cancer cell.

- **Duplication, deletion:** Changes to the structure of a chromosome in which a segment is repeated (duplication) or lost (deletion).

- **Loss of heterozygosity (LOH):** If the normal cell type in an individual at a particular locus is heterozygous, any cell which is found to be homozygous at that locus has undergone LOH. A *copy-neutral LOH* occurs if one copy is deleted and the other copy is duplicated. Duplication or amplification of one copy without change to the other copy may result in an effective LOH, since a 2:1 ratio between the A and B alleles may effectively silence the effect of the B allele.
Biomedical interest in chromosomal alterations:

- **Consequences of chromosomal alterations:** If both copies of a tumor suppressor gene are lost, or if additional copies of an oncogene are present, cancer progression may result.

- **Pharmacogenomics:** There is significant interest in identifying SNP’s that predict the most favorable treatment alternative for a given patient.

- **Identifying disease genes:** A small number of SNP’s may segregate the population into risk groups for a particular disease.

**Technical aspects of the microarray mapping assay:**

- Hybridization of fractionated, labeled genomic DNA to microarrays.

- The mapping arrays contain PM/MM probe pairs for each variant allele at each SNP locus (typically 10 PM/MM pairs for each of the two allelic variants).

- PM-MM differences are used to quantify the specific binding of a given SNP variant. PM-MM differences play the same role in mapping as they do in expression analysis, which is to compensate for cross-hybridization.

- Genotype call: A, B, AB or “nocall”. The ratio of A allele signal to B allele signal should be 1 (heterozygous), 0 (homozygous B), or $\infty$ (homozygous A). If the A and B signal levels are indeterminate, no call is made. This typically happens at 5-15% of positions.

- A LOH call can only be made when paired tumor and normal data are available. In this case, if the tumor sample is AB and the normal sample is either A or B, an LOH call is made.

- For copy number analysis, signals for the A and B alleles are summed and compared to a reference pool of normal controls. The ratio of total signal in the test sample to average total signal in the control pool is a measure of relative copy number (i.e. 1 is normal, 1.5 is a single amplification, 0.5 is a single deletion, etc.).

**Identifying regional defects using SNP array data**

- Amplifications and deletions typically span 0.1Mb up to around 50Mb. Using an array with around 50K SNP’s, a 10MB amplification would cover 250 SNP’s and a 1MB amplification would cover around 25 SNP’s. Thus regional amplifications can be detected by searching for consecutive copy number changes in the same direction. As an alternative, consecutive LOH (or non-informative loci) can be used to identify regional deletions.

- Pooling information from adjacent SNP loci, either by simple averaging or through a model such as a hidden Markov model (HMM), can be used to identify regional gain or loss in a single sample.
Identifying recurrent defects using SNP array data

- In certain types of cancer, certain regions are commonly altered. Typically, anywhere from 10% to 70% of individuals with the disease may experience such alterations. Regions of recurrent alteration are a higher priority for follow-up work than rare alterations.