**Chapter 11**

**Polymorphisms**

Polymorphism

* A DNA polymorphism is a sequence difference compared to a reference standard that is present in at least 1%–2% of a population.
* Polymorphisms can be single bases or thousands of bases.
* Polymorphisms may or may not have phenotypic effects

Polymorphic DNA Sequences

* Polymorphisms are found throughout the genome.
* If the location of a polymorphic sequence is known, it can serve as a landmark or marker for locating other genes or genetics regions.
* Each polymorphic marker has different versions, or alleles.

Types of Polymorphic DNA Sequences

* RFLP: restriction fragment length polymorphisms
* VNTR: variable number tandem repeats (8 to >50 base pairs)
* STR: short tandem repeats (1–8 base pairs)
* SNP: single-nucleotide polymorphisms

Types of Polymorphisms

* Single nucleotide polymorphisms (SNPs)
  + - A **Single Nucleotide Polymorphism**, or SNP (pronounced "snip"), is a small genetic change, or variation, that can occur within a person's DNA sequence.
  + Common SNPs are defined as >1% in at least one population
  + Rare SNPs are hard to identify and validate
  + *But, it is estimated that there are a large number per individual*
  + Some areas more variable than others (HLA)
* Restriction fragment sizes are altered by changes in or between enzyme recognition sites
  + RFLP (often pronounced "rif lip”) is used to follow a particular sequence of DNA as it is passed on to other cells.
  + RFLPs can be used in paternity cases or criminal cases to determine the source of a DNA sample. RFLPs can be used determine the disease status of an individual. RFLPs can be used to measure recombination rates which can lead to a genetic map with the distance between RFLP loci measured in centiMorgans.
  + An RFLP is a sequence of DNA that has a restriction site on each end with a "**target**" sequence in between.
  + A target sequence is any segment of DNA that bind to a **probe** by forming complementary base pairs.
  + A probe is a sequence of single-stranded DNA that has been tagged with radioactivity or an enzyme so that the probe can be detected. When a probe base pairs to its target, the investigator can detect this binding and know where the target sequence is since the probe is detectable.
  + RFLP produces a series of bands when a Southern blot is performed with a particular combination of restriction enzyme and probe sequence
  + RFLP genotypes are inherited.
  + For each locus, one allele is inherited from each parent
  + Let's look at two people and the segments of DNA they carry that contain this RFLP (for clarity, we will only see one of the two stands of DNA). Since Jack and Jill are both diploid organisms, they have two copies of this RFLP. When we examine one copy from Jack and one copy from Jill, we see that they are identical:
    - Jack 1: -GAATTC---(8.2 kb)---**GCATGCATGCATGCATGCAT**---(4.2 kb)---GAATTC-
    - Jill 1: -GAATTC---(8.2 kb)---**GCATGCATGCATGCATGCAT**---(4.2 kb)---GAATTC-
  + When we examine their second copies of this RFLP, we see that they are not identical. Jack 2 lacks an EcoR I restriction site that Jill has 1.2 kb upstream of the target sequence (difference in italics).
    - Jack 2: -GAATTC--(1.8 kb)-*CCCTTT*--(1.2 kb)--**GCATGCATGCATGCATGCAT**--(1.3 kb)-GAATTC-  
      Jill 2: -GAATTC--(1.8 kb)-*GAATTC*--(1.2 kb)--**GCATGCATGCATGCATGCAT**--(1.3 kb)-GAATTC-
    - Therefore, when Jack and Jill have their DNA subject to RFLP analysis, they will have one band in common and one band that does not match the other's in molecular weight:
* Short Tandem Repeat Polymorphisms (STR)
  + STRs are repeats of nucleotide sequences.
    - AAAAAA…: mononucleotide
    - ATATAT…: dinucleotide
    - TAGTAGTAG…: trinucleotide
    - TAGTTAGTTAGT…: tetranucleotide
    - TAGGCTAGGCTAGGC…: pentanucleotide
  + Different alleles contain different numbers of repeats.
    - TTCTTCTTCTTC : four-repeat allele
    - TTCTTCTTCTTCTTC: five-repeat allele
  + STR alleles can be analyzed by fragment size (Southern blot).
  + STR alleles can also be analyzed by amplicon size (PCR).
  + Allelic ladders are standards representing all alleles observed in a population.
  + Multiple loci are genotyped in the same reaction using multiplex PCR.
  + Allelic ladders must not overlap in the same reaction
  + STR genotypes are analyzed using gel or capillary gel electrophoresis.
  + STR genotypes are inherited.
  + One allele is inherited from each parent.
  + Y-STR
    - The Y chromosome is inherited in a block without recombination.
    - STR on the Y chromosome are inherited paternally as a haplotype.
    - Y haplotypes are used for exclusion and paternal lineage analysis.
  + mini-STR
    - Mini-STRs are STRs on smaller amplicons.
    - Recommended for degraded specimens
    - Used to identify remains from mass graves and disaster areas
  + MHC
    - **Major histocompatibility complex** (**MHC**) is a cell surface molecule encoded by a large gene family in all vertebrates. MHC molecules mediate interactions of leukocytes. MHC determines compatibility of donors for organ transplant as well as one's susceptibility to an autoimmune disease via crossreacting immunization. In humans, MHC is also called human leukocyte antigen (HLA).
    - Protein molecules—either of the host's own phenotype or of other biologic entities—are continually synthesized and degraded in a cell. Occurring on the cell surface, each MHC molecule displays a molecular fraction, called an epitope. The presented antigen can be either *self* or *nonself*.
    - The MHC gene family is divided into three subgroups—class I, class II, and class III.
* SNPs
  + Majority are “silent”: No known functional change
  + Alter gene expression/regulation
  + Promoter/enhancer/silencer
  + mRNA stability
  + Small RNAs
  + Alter function of gene product
  + Change sequence of protein

**Large and Small Scale Polymorphisms**

* **Copy Number of Polymorphisms**
  + Regional “repeat” of sequence
  + 10s to 100s kb of sequence
  + Estimate of >10% of human genome
  + Multi-copy in many individuals
* **Duplicons**
  + 90-100% similarity for >1 kb
  + 5-10% of genome (5% exons elsewhere)
  + Multi-copy (high N) in all individuals

**Copy Number Variation**

* Across the Genome
* Frequency of CNVs
  + Most are uncommon (<5%)
  + Familial vs Unrelated Studies
* Associated with Disease
  + *OPN1LW Red/green colorblind*
  + *CCL3L1 Reduced HIV Infection*
  + *CYP2A6 Altered nicotine metabolism*
  + *VKORC1 Warfarin metabolism*

**Allogeneic bone marrow transplants are monitored using STR**

* There are two parts to chimerism testing: pretransplant informative analysis and post-transplant engraftment analysis.

**Engraftment Testing Using STR: Informative Analysis**

* STR are scanned to find informative loci (donor alleles differ from recipient alleles).
* There are different degrees of informativity.
* With the most informative loci, recipient bands or peaks do not overlap stutter in donor bands or peaks.
  + Stutter is a technical artifact of the PCR reaction in which a minor product of n-1 repeat units is produced.

**Engraftment Testing Using STR: Engraftment Analysis**

* Using informative loci, peak areas are determined in fluorescence units or from densitometry scans of gel bands.

**Engraftment Analysis of Cellular Subsets**

* Cell subsets (T cells, granulocytes, NK cells, etc.) engraft with different kinetics.
* Analysis of cellular subsets provides a more detailed description of the engrafting cell population.
* Analysis of cellular subsets also increases the sensitivity of the engraftment assay.
* T cells (CD3), NK cells (CD56), granulocytes, myeloid cells (CD13, CD33), myelomonocytic cells (CD14), B cells (CD19), stem cells (CD34)
* Methods
  + Flow cytometric sorting
  + Immunomagnetic cell sorting
  + Immunohistochemistry + XY FISH
* Detection of different levels of engraftment in cellular subsets is split chimerism.

**Single Nucleotide Polymorphisms (SNPs)**

* Single-nucleotide differences between DNA sequences.
* One SNP occurs approximately every 1250 base pairs in human DNA.
* SNPs are detected by sequencing, melt curve analysis, or other methods.
* 99% have no biological effect; 60,000 are within genes.
* SNPs are inherited in blocks or haplotypes.

**Applications of SNP Analysis**

* SNPs can be used for mapping genes, human identification, chimerism analysis, and many other applications.
* The Human Haplotype Mapping (HapMap) Project is aimed at identifying SNP haplotypes throughout the human genome.

**Mitochondrial DNA Polymorphisms**

* Sequence differences in the hypervariable regions (HV) of the mitochondrial genome
* Mitochondria are maternally inherited.
* There are an average of 8.5 base differences in the mitochondrial HV sequences of unrelated individuals.
* All maternal relatives will have the same mitochondrial sequences.
* Mitochondrial typing can be used for legal exclusion of individuals or confirmation of maternal lineage.

**Summary**

* Four types of polymorphisms are used for a variety of purposes in the laboratory: RFLP, VNTR, STR, and SNP.
* Polymorphisms are used for human identification and parentage testing.
* Y-STR haplotypes are paternally inherited; maternal relatives have the same mitochondrial DNA alleles.
* Polymorphisms are used to measure engraftment after allogeneic bone marrow transplants.
* Single-nucleotide polymorphisms are detected by sequencing, melt curve analysis, or other methods.
* SNPs can be used for the same applications as other polymorphisms.
* Mitochondrial DNA typing is performed by sequencing the mitochondrial HV regions.
* Mitochondrial types are maternally inherited.