Single nucleotide polymorphisms

All individual humans share genome sequences that are approximately 99.9% the same.

0.1% variable region is responsible for the genetic diversity between individuals.

Single nucleotide polymorphisms (SNPs) (pronounced "snips"), where two or more possible nucleotides occur at a specific mapped location in the genome, Most common type of genomic sequence variation among individuals.

e.g. SNP might change the DNA sequence ATGCCTA to ATGCTTA. Individuals may be homozygotes (e.g. T/T or C/C), or heterozygotes with different bases (e.g. T/C) at polymorphic sites.

For a variation to be considered a SNP it must occur in at least 1% of the population.

~7 million common SNPs with a population frequency of at least 5% across the entire human population.

Single Nucleotide Polymorphisms (SNPs)

- **Additional 4 million SNPs exist with an allele frequency of between 1 and 5%.**
- **The least variability occurs on the sex chromosomes (X and Y).**
- **Two of every three SNPs characterized so far involve the replacement of cytosine (C) with thymine (T).**
- **SNPs can occur in both coding and noncoding regions of the genome.**
- **Resequencing certain parts of the genome using DNA from different individuals generates a map of SNPs.**
- **Map is used by researchers to scan the human genome for haplotypes associated with common diseases.**
- **Haplotypes are patterns of sequence variation, i.e. stretches of continuous DNA containing a distinctive set of alleles.**
- **SNPs do not necessarily cause disease, but they can help to determine the likelihood that someone will develop a particular disease.**

Types of SNPs

- **Noncoding SNPs**
  - 5’ UTR
  - 3’ UTR
  - Introns
  - Intergenic Regions
  - Pseudogenes
  - Regulatory
    - Splicing
    - Transcriptional regulation (promoter & TF binding sites)
    - Translational regulation (initiation or termination)
    - Regulatory miRNA target sites
- **Coding SNPs**
  - Synonymous SNPs (third position variation)
  - Replacement SNPs (change Amino acid)
    - Functional SNPs (acceptable amino acid replacement)
    - Non-functional SNPs (traits & diseases)

- **Identifying a polymorphic sequence does not necessarily mean the discovery of a marker for a disease.**
- **So far, there are only a few examples with clear links to disease. Two well characterized examples of SNPs include**
  - **the missense mutation that causes sickle cell anemia, and**
  - **the apolipoprotein E4 allele implicated in susceptibility to late-onset Alzheimer's disease.**
- **Alzheimer's disease is the most common form of dementia among older people. The term "dementia" describes symptoms resulting from changes in parts of the brain that control thought, memory, and language.**
- **Symptoms may include becoming lost in familiar places and not recognizing friends and family, being unable to follow directions, asking the same questions repeatedly, and neglecting personal safety and hygiene.**
- **Dr Alois Alzheimer, a German doctor who, in 1906, described abnormal clumps (amyloid plaques) and tangled bundles of fibers (neurofibrillary tangles) in the brain tissue of a woman who had died of an unusual mental illness.**

- **Plaques and tangles in the brain are considered hallmarks of Alzheimer's disease.**
- **What is understood is that there is probably not one single cause, but several factors that affect each person differently.**
- **The polygenic nature of disorders such as Alzheimer's disease makes genetic testing complicated.**
- **Age is the most important known risk factor. The number of people with the disease doubles every 5 years beyond age 65. Family history is another risk factor.**
- **Most common form of the disease, called late-onset Alzheimer's disease, no obvious family pattern is seen. Symptoms usually begin after age 60, and the risk goes up with age.**
- **The disease progresses slowly, starting with mild memory problems and ending with severe brain damage. On average, Alzheimer's patients live from 8 - 10 years after they are diagnosed. One risk factor of late-onset Alzheimer's is the apolipoprotein E4 allele.**

Apolipoprotein E4 allele: a risk factor for late-onset Alzheimer's disease

- **Apolipoprotein E (ApoE) is a protein involved in lipid metabolism.**
- **The apolipoprotein E (*apoE*) gene contains two SNPs that result in three possible alleles for this gene: *e*2, *e*3, and *e*4.**
- **Each allele differs by one DNA base, and the variable protein product (isoform) of each allele differs by one amino acid.**
- **These single amino acid changes have major effects on protein structure and function. Each individual inherits a maternal and a paternal copy of the *apoE* gene.**
- **The *apoE e*3 allele is the most common allele in all populations studied so far. In this allele, TGC encodes the cysteine at position 112 in the protein, and CGC encodes the arginine at position 158.**
- **In the *e*2 allele, another TGC codon results in a cysteine at position 158 instead. In the *e*4 allele a CGC codon gives rise to an arginine at position 112.**

The three *apoE* alleles determine six genotypes: three homozygotes (*e*4/4, *e*3/3, and *e*2/2) and three heterozygotes (*e*3/4, *e*2/3, and *e*2/4).

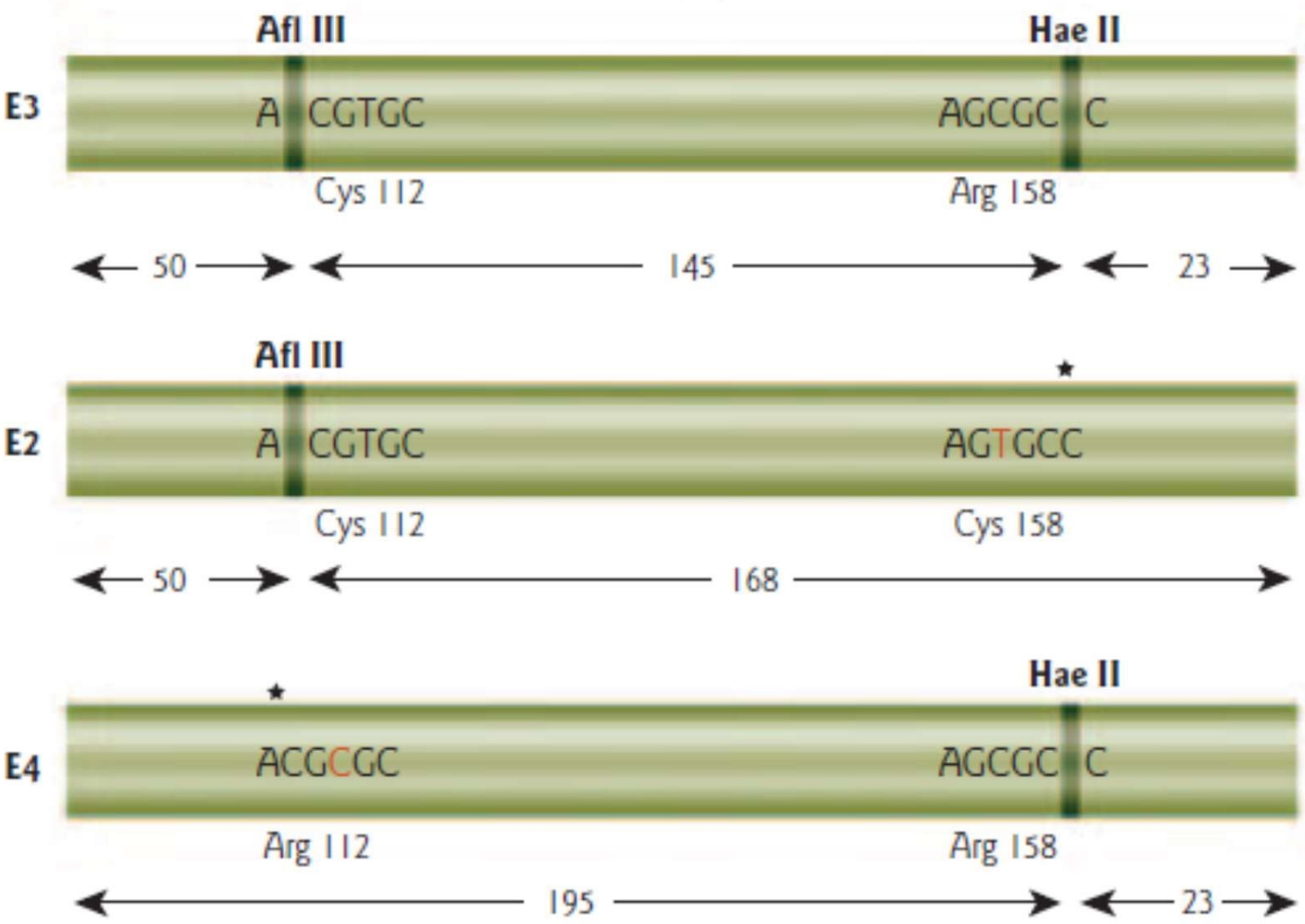
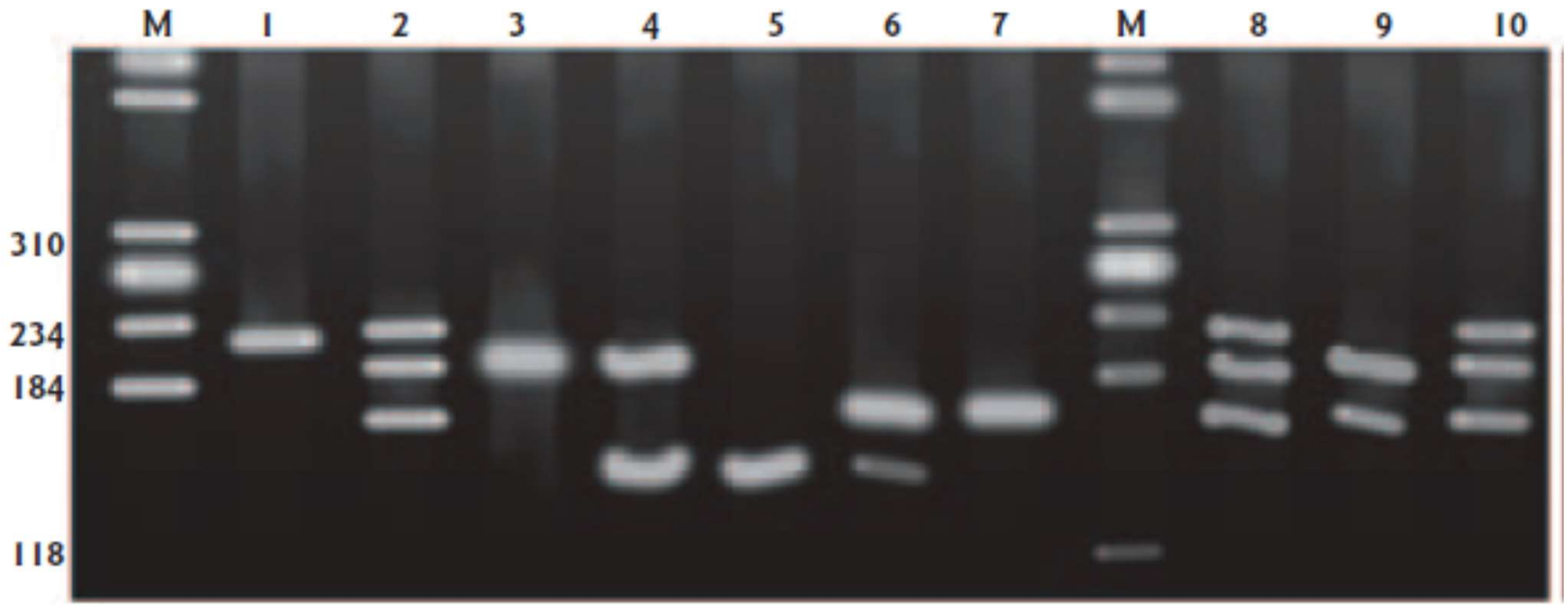


Figure shows an example of *apoE* genotyping by restriction endonuclease digestion of a PCR-amplified fragment.
The restriction endonucleases *Afl*III and *Hae*II recognize the allele-specific nucleotide substitutions at codons 112 and 158, respectively.



The 218 bp amplified product generates 145, 168, and 195 bp fragments that are specific for the *apoE e*3, *e*2, and *e*4 alleles, respectively.

- **Researchers have shown that an individual who inherits at least one *apoE e*4 allele will have a greater chance of getting Alzheimer's disease. Possession of the *e*4 allele may account for 50% of Alzheimer's disease in the USA.**
- **The *e*4 allele appears to increase the rate and extent of amyloid plaque and neurofibrillary tangle formation. In addition, individuals with the *e*4 allele show increased total serum cholesterol and have a greater risk of coronary heart disease.**
- **In contrast, the *e*2 allele is associated with longevity, and individuals with this allele are less likely to develop Alzheimer's. However, it is important to note that SNPs are not absolute indicators of disease development.**
- **Someone who has inherited two *e*4 alleles may never develop Alzheimer's, while another who has inherited two *e*2 alleles may.**
- **The *apoE* gene is just one gene of several that have been linked to Alzheimer's disease.**