RAMA UNIVERSITY

www.ramauniversity.ac.in

FACULTY OF ENGINEERING &TECHNOLOGY
DEPARTMENT OF BIOTECHNOLOGY

Dr. Simranjit Singh
Assistant Professor
Department of Biotechnology
Rama University, Kanpur

Single Nucleotide Polymorphisms (SNPs), Haplotypes, Linkage Disequilibrium, and the Human Genome

Biological Background

- How can researchers hope to identify and study all the changes that occur in so many different diseases?

- How can they explain why some people respond to treatment and not others?

‘SNP’ is the answer to these questions…

- So what exactly are SNPs?
- How are they involved in so many different aspects of health?

What is SNP ?

- A **SNP** is defined as a single base change in a DNA sequence that occurs in a significant proportion (more than 1 percent) of a large population.

Some Facts

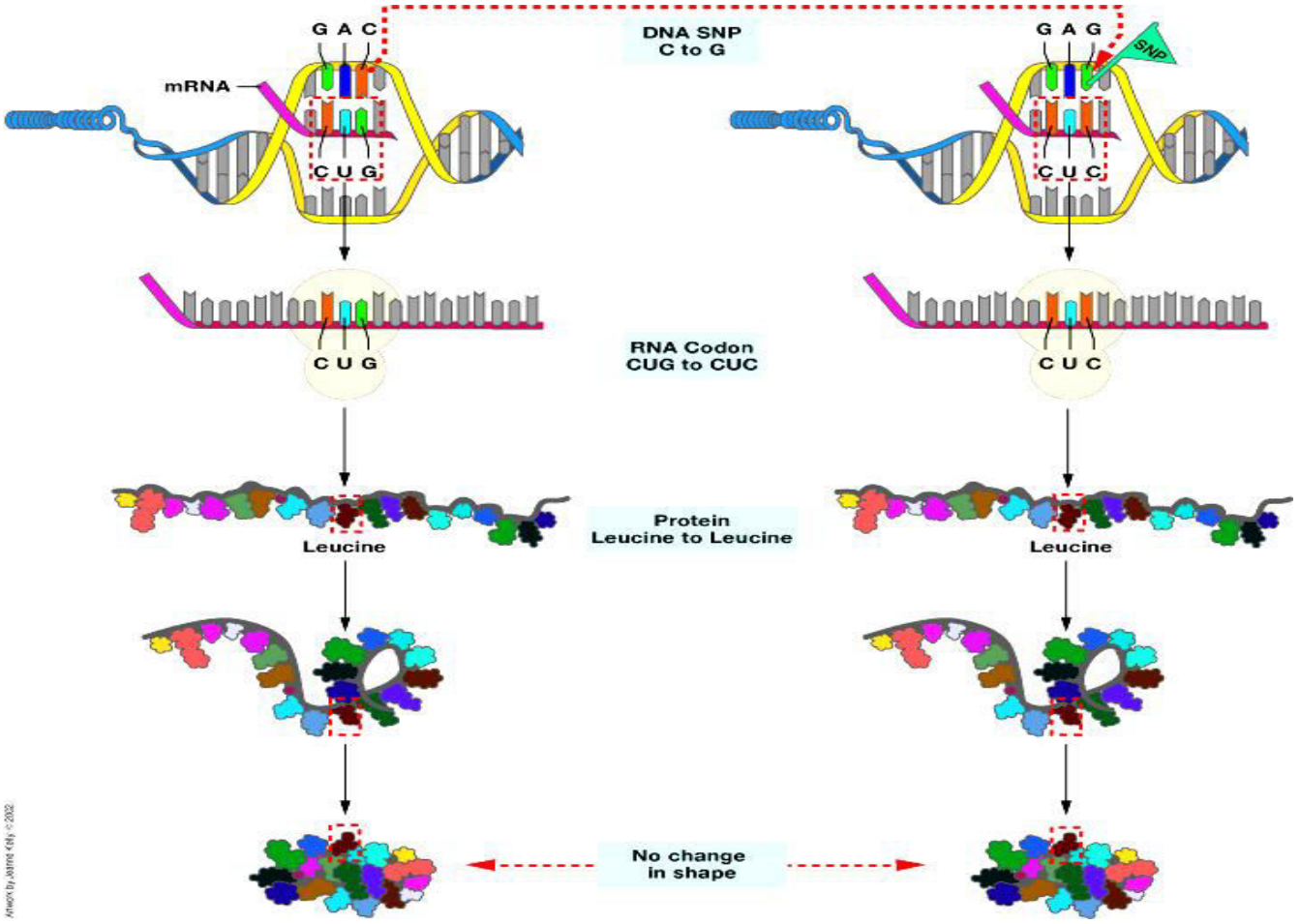
- In human beings, 99.9 percent bases are same.
- Remaining 0.1 percent makes a person unique.
  - Different attributes / characteristics / traits
    - how a person looks,
    - diseases he or she develops.

- These variations can be:
  - Harmless (change in phenotype)
  - Harmful (diabetes, cancer, heart disease, Huntington's disease, and hemophilia)
  - Latent (variations found in coding and regulatory regions, are not harmful on their own, and the change in each gene only becomes apparent under certain conditions e.g. susceptibility to lung cancer)

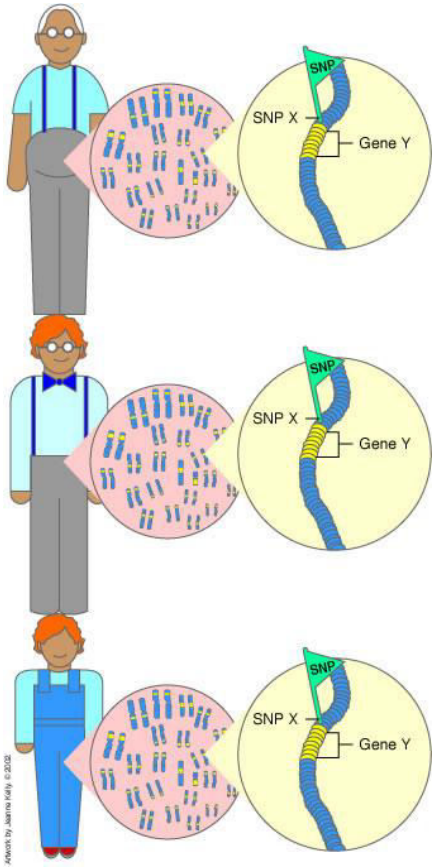
SNP facts

- SNPs are found in
  - coding and (mostly) noncoding regions.
- Occur with a very high frequency
  - about 1 in 1000 bases to 1 in 100 to 300 bases.
- The abundance of SNPs and the ease with which they can be measured make these genetic variations significant.
- SNPs close to particular gene acts as a marker for that gene.
- SNPs in coding regions may alter the protein structure made by that coding region.

SNPs may / may not alter protein structure



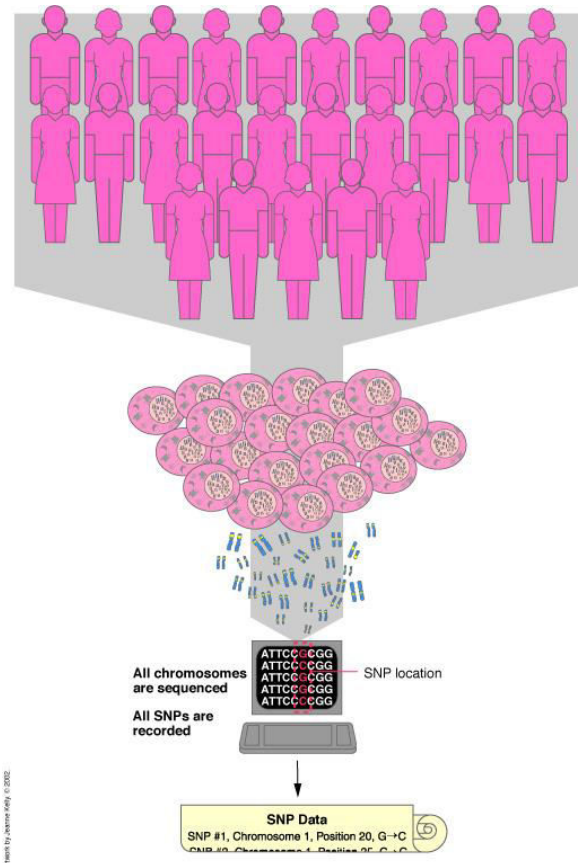
SNPs act as gene markers



SNP maps

- Sequence genomes of a large number of people
- Compare the base sequences to discover SNPs.
- Generate a single map of the human genome containing all possible SNPs => SNP maps

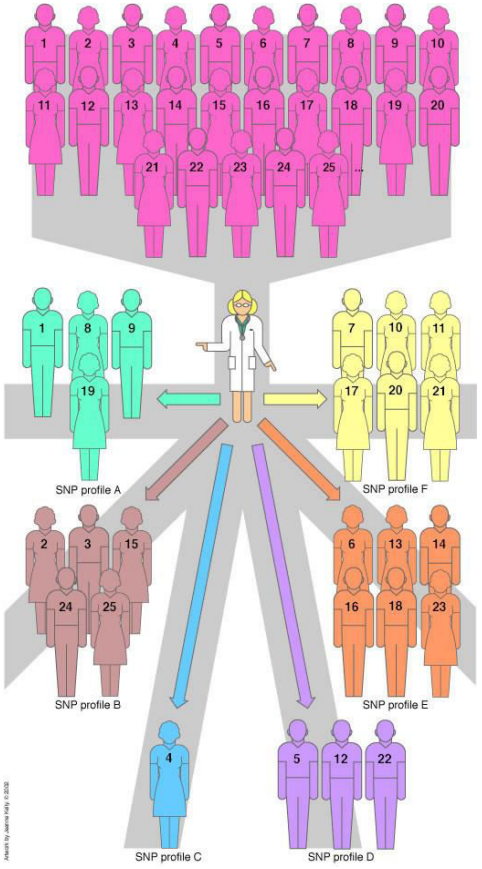
SNP Maps


All chromosomes are sequenced
All SNPs are recorded
SNP location
SNP Data
SNP #1, Chromosome 1, Position 20, G→C

SNP Profiles

- Genome of each individual contains distinct SNP pattern.
- People can be grouped based on the SNP profile.
- SNPs Profiles important for identifying response to Drug Therapy.
- Correlations might emerge between certain SNP profiles and specific responses to treatment.

SNP Profiles



Techniques to detect known Polymorphisms

- Hybridization Techniques
  - Micro arrays
  - Real time PCR
- Enzyme based Techniques
  - Nucleotide extension
  - Cleavage
  - Ligation
  - Reaction product detection and display

Techniques to detect unknown Polymorphisms

- Direct Sequencing
- Microarray
- Cleavage / Ligation
- Electrophoretic mobility assays

Direct Sequencing

- Sanger dideoxysequencing can detect any type of unknown polymorphism and its position, when the majority of DNA contains that polymorphism.
- Misses polymorphisms and mutations when the DNA is heterozygous
- limited utility for analysis of solid tumors or pooled samples of DNA due to low sensitivity
- Once a sample is known to contain a polymorphism in a specific region, direct sequencing is particularly useful for identifying a polymorphism and its specific position.
- Even if the identity of the polymorphism cannot be discerned in the first pass, multiple sequencing attempts have proven quite successful in elucidating sequence and position information.

SIGNIFICANCE OF SNPs

- IN DISEASE DIAGNOSIS
- IN FINDING PREDISPOSITION TO DISEASES
- IN DRUG DISCOVERY & DEVELOPMENT
- IN DRUG RESPONSES
- INVESTIGATION OF MIGRATION PATTERNS

ALL THESE ASPECT WILL HELP TO LOOK FOR MEDICATION & DIAGNOSIS AT INDIVIDUAL LEVEL

SNP Screening

- Two different screening strategies
  - Many SNPs in a few individuals
  - A few SNPs in many individuals
- Different strategies will require different tools
- Important in determining markers for complex genetic states

SNP genotyping methods for detecting genes contributing to susceptibility or resistance to multifactorial diseases, adverse drug reactions:

=> case-control association analysis

case

….GCCGTTGAC….

….GCCATTGAC….

control

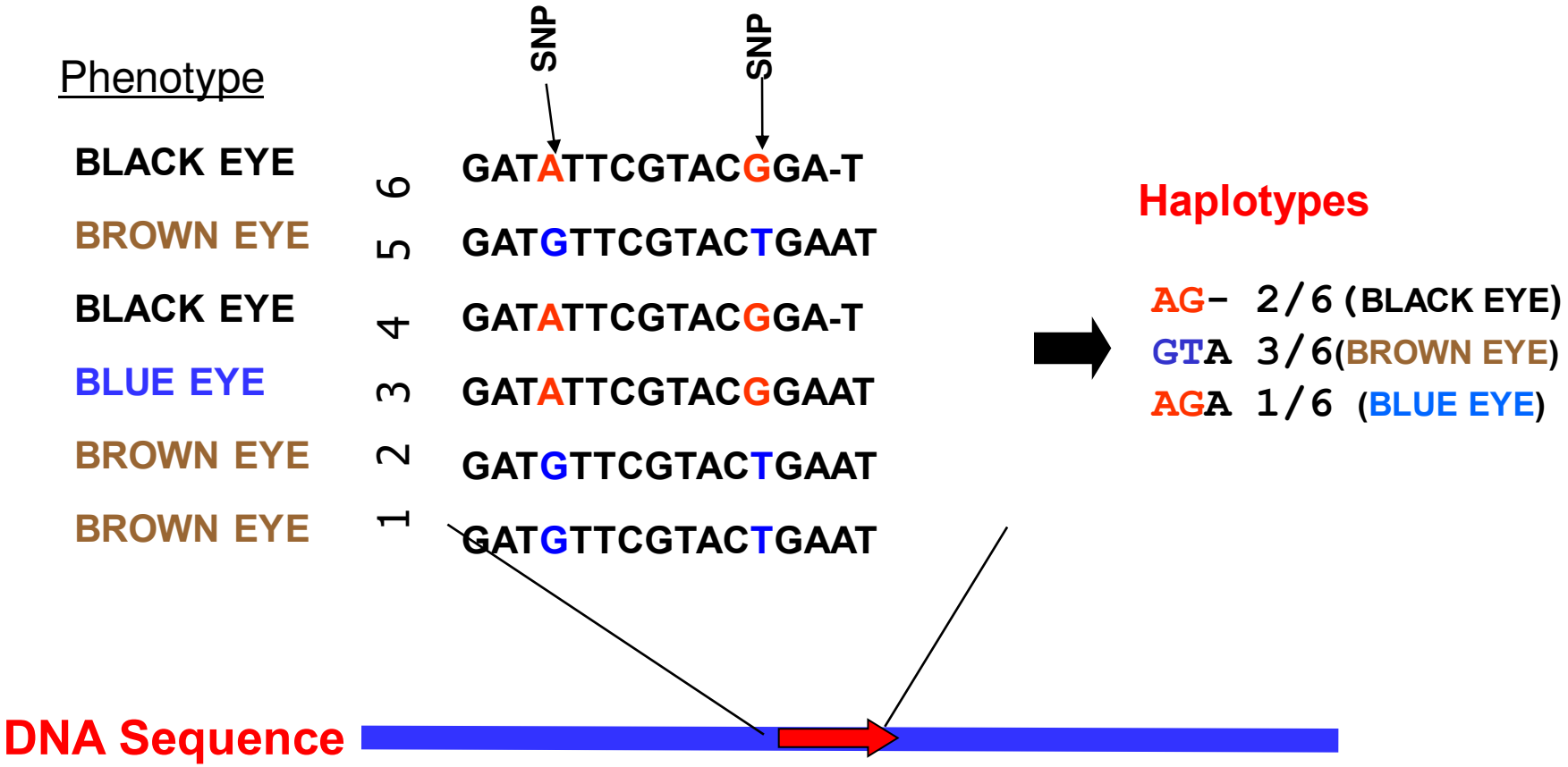
….GCCATTGAC….

….GCCATTGAC….

HAPLOTYPE

A set of closely linked genetic markers present on one chromosome which tend to be inherited together (not easily separable by recombination)

SNP-Haplotype



HAPLOTYPE CORRELATION WITH PHENOTYPE

- The "Haplotype centric" approach combines the information of adjacent SNPs into composite multilocus haplotypes.

- Haplotypes are not only more informative but also capture the regional LD information, which is assumed to be robust and powerful

- Association of haplotype frequencies with the presence of desired phenotypic frequencies in the population will help in utilizing the maximum potential of SNP as a marker.

ADVANTAGES:

1. SNPs ARE THE MOST FREQUENT FORM OF DNA VARIATIONS
2. THEY ARE THE DISEASE CAUSING MUTATIONS IN MANY GENES
3. THEY ARE ABUNDANT & HAVE SLOW MUTATION RATES
4. EASY TO SCORE
5. MAY WORK AS THE NEXT GENERATION OF GENETIC MARKERS

Some important SNP database Resources

1. dbSNP (http://www.ncbi.nlm.nih.gov/SNP/)

   LocusLink (http://www.ncbi.nlm.nih.gov/LocusLink/list.cgi)

2. TSC (http://snp.cshl.org/)

3. SNPper (http://snpper.chip.org/bio/)

4. JSNP (http://snp.ims.u-tokyo.ac.jp/search.html)

5. GeneSNPs (http://www.genome.utah.edu/genesnps/)

6. HGVbase (http://hgvbase.cgb.ki.se/)

7. PolyPhen (http://dove.embl-heidelberg.de/PolyPhen/)

   OMIM (http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM)

8. Human SNP database

   (http://www-genome.wi.mit.edu/snp/human/)