Lecture 32: Gene Variations and Single Nucleotide Polymorphism

Gene Variations

Gene Variations are essentially a change in order of the bases in a DNA sequence. Variations can involve only one or many bases. These variations may be of three different types 1) Polymorphisms- change in one or more nucleotides 2) Deletions- removal of one or more nucleotide and, 3) Insertion- addition of one or more nucleotide. In addition to these changes, a section of DNA on a chromosome may swap places with a part of DNA on another chromosome known as translocation.

The effect of these variations depends upon the region where these variations have occurred. Generally, a majority of gene variations occur in the non-coding region of DNA and thus exhibit no effect at all. Additionally, some variations occur in coding region of DNA and are silent in nature as they do not show any effect. Few variations in coding region do show some characteristic but harmless effect such as difference in heights, color of the iris, shape of the earlobe etc. But, there exist a group of variations in coding and regulatory regions of a gene that has a detrimental effect. These deleterious variations are better known as mutations. These mutations in turn are the reason for many deadly genetic diseases such as Hemophilia, Cancer, Huntington's disease, Cystic fibrosis, Tay Sachs disease, Sickle Cell anemia to name a few. These mutations may be on autosomes {autosome dominant (Huntington's disease) or recessive (Cystic fibrosis)} or on sex chromosomes (X-chromosome linked Hemophilia). Also, some gene variations in coding and regulatory region remain dormant. These variations exert their effect only under certain specific conditions. These changes determine a higher susceptibility of some people to a particular disease and variable response of a drug to different individuals. These dormant variations that do not exert any effect on its own in normal condition but exhibit its effect only under certain environmental/ external conditions are called polymorphisms.

Genetic Variations may occur in any part of the genetic machinery like gene, DNA, chromosome, protein or protein function.

Generation of Genetic variations

Genetic variation is an important part of evolutionary development:

There are three primary sources of genetic variations*:*

Mutations: Mutations are the changes in DNA. A single nucleotide change may cause effective changes in the species. Evolution involves several changes in the genetic make-up.

Sexual changes: Sex may introduce new variations in the community and can cause genetic variations.

Gene flow: When the flow of genes take place in between different populations. It results in genetic variations.

Single Nucleotide Polymorphism (SNP)

SNPs are one of the most common types of genetic variation. A Single Nucleotide Polymorphisms (also referred as SNPs) are the most commonly occurring genetic variations among the populations. They are the small genetic changes that occur within a person's DNA sequence when a single nucleotide, such as G is replaced by any of the three other bases A, C, or T. SNPs occur in both coding and non-coding region of the genome but majority of SNPs are found in the non-coding region and therefore, they do not exhibit any effect. SNPs found within a coding sequence may show silent, harmless, dormant or harmful effect. They are of particular interest because they are more likely to alter the biological function of a protein. The frequency of their occurrence varies from about 1 in 1000 bases to 1 in 100-300 bases. SNPs can help scientists to locate genes that are associated with specific diseases. SNPs play a vital role in disease development when they occur within a gene or in a regulatory region near a gene (Fig. 1)

DNA contains 99.9% of identical sequence in all the individuals with only 0.1% difference. Out of this 0.1% variation, over 80% are single nucleotide polymorphisms (SNPs).

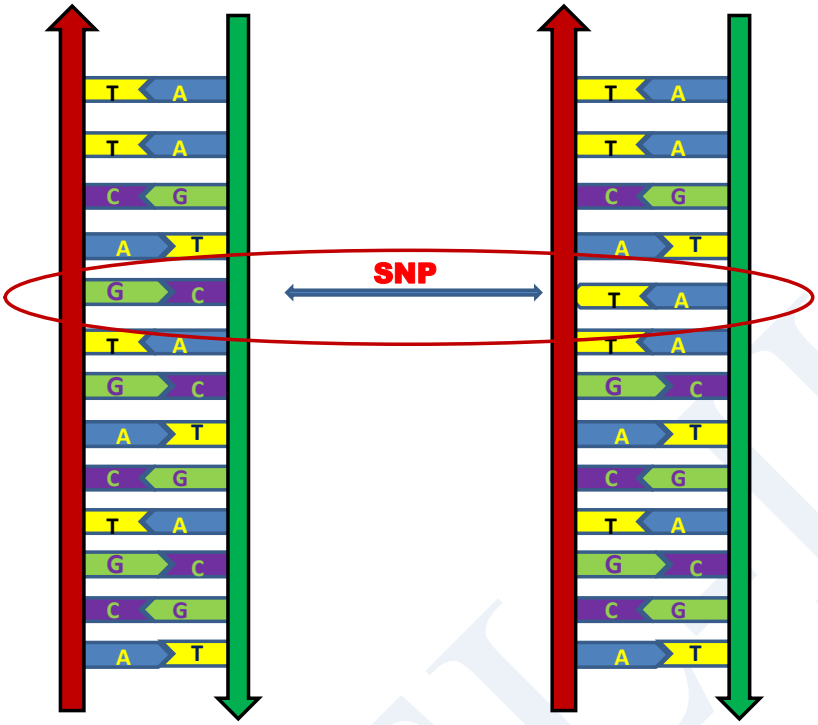

Figure 1: Representation of Single Nucleotide Polymorphism

Although SNPs are commonly occurring throughout the human genome and preferably do not have any side effects. Although scientists believe that SNPs may be proved a cutting edge technology to identify the genes responsible for several deadly diseases such as cancer, diabetes, vascular disease, and some forms of mental illness. Therefore, in April 1999, SNP consortium was developed including ten large pharmaceutical companies and the U.K. Wellcome Trust philanthropy to find and map 300,000 common SNPs to generate a SNP map of human genome as a part of the Human Genome Project.

Once the SNP profiles for specific types of diseases is established, it will be just a click away for physicians to identify susceptibility of a person to a disease by analyzing their DNA samples for specific SNP patterns.

Table 1: Few Gene, Genename, corresponding SNPs and their class.

Gene	Gene Name	SNP	SNP Class
APOE	Apolipoprotein E	rs4966978	NonSyn intron
USF1	Upstream transcription factor 1	Haplotype across USF1	NA
GATA2	GATA binding Protein 2	rs2713604	5'UTR
LTA4H	Leucotreine A4 hydrolase	Haplotype K	NA
PSMA6	Proteasome subunit alpha 6	rs3850641	Intron
GJA4	Gap junction alpha 1	rs1764391	NonSyn intron

Types of SNPs

Non-coding SNPs: They can be 5'-UTR (Untranslated Regions), 3'-UTR, introns, intergenic regions and pseudogenes occurring in the genome. The regulation machinery of these non-coding SNPs includes splicing, transcriptional regulation and translational regulation.

Coding SNPs: These types of SNPs are also called as synonymous SNPs (due to third position variation). Other coding SNPs are replacement SNPs which cause change in amino acid, therefore categorized again in two types: 1) Functional SNPs (acceptable amino acid replacement) and 2) Non-functional SNPs (traits & diseases)

Methods of analysis of SNPs

Analytical methods to discover novel SNPs and detect known SNPs include

Denaturating HPLC and gel electrophoresis: As these techniques separates the samples on the basis of parameters like how long the sample takes to run through the column, therefore the sample can be separated after denaturation of DNA in homozygous and heterozygous bases.

Denaturing high performance liquid chromatography (DHPLC) uses reversed-phase HPLC for identification of SNP and it has differential affinity for single and double stranded DNA. In this process, the DNA molecule is first denatured and then reannealed. These reannealed DNA

fragments have specific melting temperature which determines their time to be retained in the column. Two types of DNA fragments are generated i.e. target DNA containing the SNP polymorphic site and an allele-specific DNA sequence. This second fragment is identical to the target DNA except the diffeernce at the SNP polymorphic site. The fragments are denatured and then allowed to gradually reanneal. After reannealing, the DNA fragments are added to the DHPLC column. In case of normal DNA sequence, perfect matching will take place leading to formation of homoduplex during annealing process, whereas if SNP will be there, it will form heteroduplex DNA.The mismatched heteroduplexes will have a different melting temperature than the homoduplexes and will pass through the column fastly. This generates the difference in separation of normal DNA fragments and SNP bearing DNA sequence. The eluted DNA is detected by UV absorption. DHPLC is easily automated as no labeling or purification of the DNA fragments is needed. The method is also relatively fast and has a high specificity. One major drawback of DHPLC is that the column temperature must be optimized for each target in order to achieve the right degree of denaturation.

***Use of Nanotechnology*:** SNPs can also be detected using thin film biosensors (Figure below). An optimized method is described as following:

The biosensors optical layer was prepared by coating the surface of silicon wafers with silicon nitride (Si_3N_4). This optical layer was also coated with a layer of T structure aminoalkylpolydimethylsiloxane (TSPS) and poly(Phe- Lys) to facilitate covalent attachment of biomolecules. For detection of SNPs a pair of allele-discriminating oligonucleotide probes (blue) and a single detector oligonucleotide (red) is designed. The allele descriminating oligos have 5'-aldehyde groups, a ''spacer,'' followed by 40 nucleotides complementary to the corresponding SNP target sequence. There is difference only at 3'-terminal nucleotide in these probes and this 3'-terminal nucleotide matches with one of the two SNP genotypes. The detector probe has biotin at its 3' end for detection and phosphate at its 5' end for ligation. The assay involves ligation and hybridization of allele discriminating oligomers attached on the biosensor surface followed by detection by antibiotine-antibody. Briefly, Target DNA hybridization and ligation of allele descriminationg oligonucleotides and biotinylated detector oligonucletide are done

simultaneously during a 20-min incubation in the presence of a thermostable DNA ligase. After that, stringent washing with NaOH is done to remove non-ligated oligomers. Therefore, a perfect match is obtained in case of if no SNP is present in the DNA sequence and *vice versa*. The detector probe gets hybridized to normal DNA sequence and is detected by incubation with an antibiotin IgG-HRP conjugate due to presence of biotine. A purple precipitate is formed which can be detected by eye or a simple digital-imaging system (Fig.2).

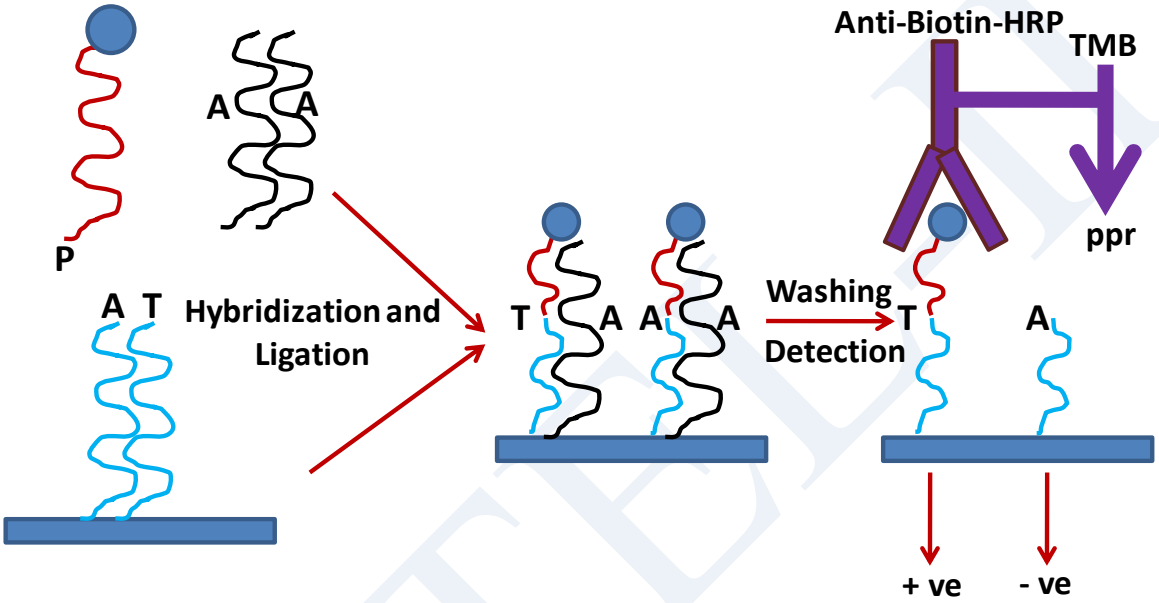

Figure 2: Strategy for SNP detection on thin-film biosensor chips.

Restriction fragment length polymorphism (RFLP): Restriction fragment length polymorphism (RFLP) is the simplest and easiest method of SNP detection. RFLP uses a huge number of restriction endonucleases which have affinity for specific restriction sites. Firstly, the genome sample is digested by these restriction endonuclease and the length of frangments is determined by gel assay to determine whether a particular enzyme digests a specified restriction site or not. Thus a failure in cutting the genome sample at expected restriction site indicates that there is a mutation at the point of the restriction site which is rendering it protected from nuclease activity. There are several drawbacks of this technique as well, such as slow nature of gel assays make RFLP, complexity of most eukaryotic genomes, the requirement for specific endonucleases and the uncertaininty of exact mutation to identify in single experiment.

DNA sequencing: The region of the interest is sequenced using Sanger's Dideoxy DNA sequencing method. Then the sequence is read to find out any variation or SNP. We have already studied method for DNA sequencing in our earlier lectures.

Taq-Man Assay: It uses a probe which binds to the SNPs as it is specific. **TaqMan** assay for SNP genotyping uses Taq DNA polymerase's 5'-nuclease activity. The TaqMan assay is performed concurrently with a PCR reaction and the results can be read in real-time as the PCR reaction proceeds. Since the TaqMan assay is based on PCR it is simple and easy and can detect upto seven SNPs in one reaction. However, each SNP requires different probe for identification, therefore it cannot detect closely assembled SNPs the genome.

Mass spectrometry: In this method four different ddNTPs are used i.e. four different molecular weights. Mass spectroscopy is done after running the PCR to find out the exact weight of the product thus SNP can be detected.

There are some other techniques available for detection of SNPs as well, such as Molecular Beacon, Hybridization analysis etc.

Applications of SNPs

There are several applications of SNPs in disease diagnosis and other aspects as follows

Use of SNPs in disease diagnosis: The genome of every individual has unique SNP pattern made up of different genetic variations. Although, most of these SNPs are housekeeping and does not cause any type of disease, yet they may serve as biological markers for identification of a particular disease as they are commonly occurring near to a gene associated with a disease. Occasionally, SNPs may also be responsible for a disease cause due to genetic variation. Therefore, scientists are actively involved in identification of genetic make up and SNP pattern of individuals affected with any disease and prepare a database for further use as reference.

Use of SNPs in public databases: SNPs are present throughout the genome, thus they serve as biological marker i.e. have identifiable physical location that can be easily tracked and used for

constructing a chromosome map and can show a position of a gene. NCBI has prepared a huge **public SNP database** (dbSNP) for comparision of genome of several species. This database can be assessed by any research community to compare and identify the genome sequence of species.

Use of SNPs in drug development: The SNPs can also have importance in action of a particular therapeutic agent. Currently, no method is present to predict the reaction of any individual against a medication i.e a treatment which is proven effective in one patient may not work in others. Even some patients may experience the side effects. As a futuristic tool, SNPs can be used to predict the possible behaviour of an individual for particular therapeutic agent by analyzing his SNP profile, may be referred as "**personalized medicine".** An overview of application of SNPs in Human genome and personalized medicine is given in Fig.3

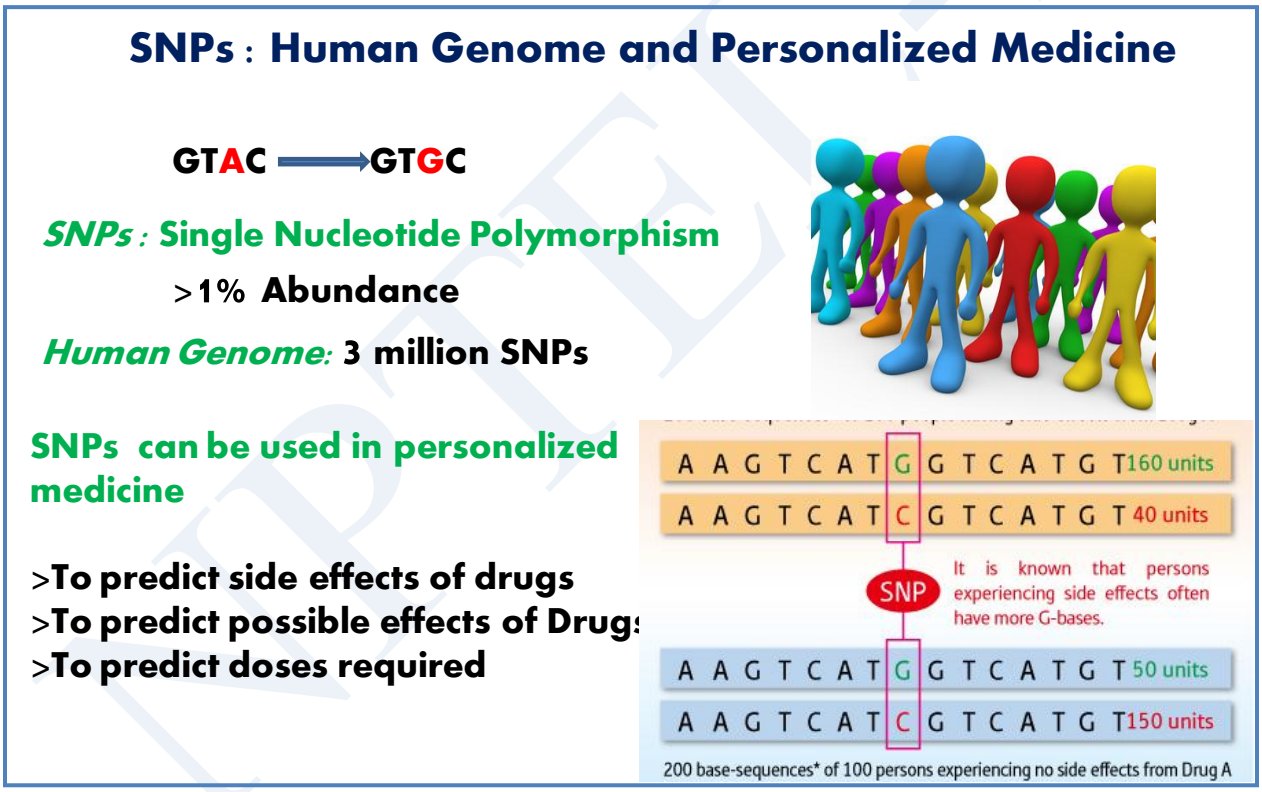

Figure 3: An overview of application of SNPs in Human genome and personalized medicine.

The SNP profiling will be a gem-stone in pharmacogenomics, the genome based drug development. It will lead to development of the right drug, to the right person, in the right dose, at the right time.

SNPs and Human Genome Project

In 1998, during Human Genome Project, several attempts were made to identify the human SNP profile with an aim to develop technologies for rapid, large-scale identification of SNPs and other DNA sequence variations, develop methods for identification of SNPs, identify common variations in coding region of a gene and develop a SNP map of around 100,000 genes.

The use of SNP is not limited to the above applications but there are some other applications as following:

- SNPs are used for identification and forensics
- SNPs are used for mapping and genome-wide association studies of complex diseases
- SNPs are used for immigration & citizenship in the UK
- SNPs are used to predict specific genetic traits

Further reading and recommended article

Single Nucleotide Polymorphism (SNP) Genotyping Techniques—An Overview

Richard M. Twyman
University of York, York, U.K.

INTRODUCTION

Single nucleotide polymorphisms (SNPs) are individual base positions in the genome that show natural variation in a population. They represent the most abundant form of genetic variation in humans, accounting for more than 90% of all differences between unrelated individuals. SNP patterns are likely to influence many human phenotypes; therefore large-scale association studies based on SNP genotyping are expected to help identify genes affecting complex diseases and responses to drugs or environmental chemicals. SNPs these follow the fate of a label either in real time or at the assay end point. Uniquely, mass spectrometry can be used to detect the allele-specific product of a discrimination assay without the need for a label, by distinguishing the masses of DNA molecules containing alternative bases.

ALLELE DISCRIMINATION METHODS

Allele-Specific Hybridization

http://www.writescience.com/RMT%20PDFs/Elsevier/Twyman%2005%20EMGP.pdf

Further reading and recommended article

Single-nucleotide polymorphism genotyping on optical thin-film biosensor chips

Xiao-bo Zhong[†], Robert Reynolds[†], Judith R. Kidd[†], Kenneth K. Kidd[†], Robert Jenison[‡§], Richard A. Marlar[¶‖], and David C. Ward[†,††]

[†]Department of Genetics, Yale University School of Medicine, 333 Cedar Street, New Haven, CT 06510; [‡]ThermoBioStar, Inc., Louisville, CO 80027; [¶]Department of Pathology, Oklahoma City Veterans Affairs Medical Center and University of Oklahoma Health Sciences Center, Oklahoma City, OK 73104; and [‖]Denver Veterans Affairs Medical Center, Denver, CO 80220

Contributed by David C. Ward, July 29, 2003

Single-nucleotide polymorphisms (SNPs) constitute the bulk of human genetic variation and provide excellent markers to identify genetic factors contributing to complex disease susceptibility. A rapid, sensitive, and inexpensive assay is important for large-scale SNP scoring. Here we report the development of a multiplex SNP detection system using silicon chips coated to create a thin-film optical biosensor. Allele-discriminating, aldehyde-labeled oligonucleotides are arrayed and covalently attached to a hydrazine-derivatized chip surface. Target sequences (e.g., PCR amplicons) then are hybridized in the presence of a mixture of biotinylated detector probes, one for each SNP, and a thermostable DNA ligase. After a stringent wash (0.01 M NaOH), ligation of biotinylated

accommodate low-, moderate-, or high-throughput needs, takes minimal effort for probe design, and provides rapid SNP scoring with high sensitivity and specificity with little if any instrumentation requirements would be desirable. Here we describe a DNA ligation-based SNP assay conducted on an optical thin-film biosensor chip that meets these criteria. This method will have multiple uses in such areas as determining the correlation between genotype and disease susceptibility (and hence perhaps providing hints to pathogenic mechanisms), genetic epidemiology and population genetics, and diagnosis of infectious diseases (including discrimination between two closely related disease agents with differing virulence).

www.pnas.org/cgi/doi/10.1073/pnas.1934783100 PNAS | September 30, 2003 | vol. 100 | no. 20 | 11559–11564

Further reading and recommended article

High-resolution SNP mapping by denaturing HPLC

Knud Nairz*, Hugo Stocker*, Benno Schindelholz[†], and Ernst Hafen*[‡]

*Zoologisches Institut der Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland; and [†]The Genetics Company, Incorporated, Wagistrasse 27, CH-8952 Schlieren, Switzerland

Edited by Michael S. Levine, University of California, Berkeley, CA, and approved June 13, 2002 (received for review March 7, 2002)

With the availability of complete genome sequences, new rapid and reliable strategies for positional cloning become possible. Single-nucleotide polymorphisms (SNPs) permit the mapping of mutations at a resolution not amenable to classical genetics. Here we describe a SNP mapping procedure that relies on resolving polymorphisms by denaturing HPLC without the necessity of determining the nature of the SNPs. With the example of mapping mutations to the *Drosophila nicastrin* locus, we discuss the benefits of this method, evaluate the frequency of closely linked and potentially misleading second site mutations, and demonstrate the use of denaturing high-performance liquid chromatography to

polymorphisms, and the subsequent development of a detection assay (4). As the exact nature of the polymorphism is irrelevant for mapping purposes, a method permitting the reliable detection of the mere difference in DNA sequence without determining the actual nature of the difference could save the sequencing and assay development steps.

Possibly the most advanced method for mutation detection without sequencing is denaturing HPLC, which permits the resolution of heteroduplex DNA in PCR fragments of up to ≈1,000 bp in length (6–8). The underlying principle of the technique is a slightly altered melting behavior of heteroduplexes