

Concepts of SNPs, STRs, VNTRs

Single Nucleotide Polymorphisms (SNPs)

Single Nucleotide Polymorphisms (SNPs) are the smallest yet most common type of genetic variation found in humans and many other organisms. They represent a difference in a **single DNA building block (nucleotide)** — adenine (A), thymine (T), cytosine (C), or guanine (G) — at a specific position in the genome.

For example, at a particular site in the DNA sequence, most people might have an adenine (A), while a small fraction might have a guanine (G). This tiny change — just one letter in the genetic code — is called a **single nucleotide polymorphism**.

Although the change seems very small, SNPs are incredibly powerful because they are **abundant, stable, and inherited** across generations. On average, there is **one SNP every 1,000 base pairs** in the human genome, meaning that between any two individuals, there are **millions of single-base differences**.

1. Origin and Stability of SNPs

SNPs originate from **point mutations**, which are small changes in DNA that occur naturally over time during replication or due to environmental factors like radiation or chemical exposure.

If such a mutation occurs in a **germ cell (sperm or egg)** and gets passed on to offspring, it may become a **permanent feature** of that lineage. Over thousands of generations, these small changes accumulate, creating subtle but stable genetic differences among individuals and populations.

SNPs are remarkably **stable** because:

- They involve only one base change, making them less prone to further mutation.
- Once established, they can persist in populations for **thousands or even millions of years**.

Thus, SNPs serve as **molecular fossils**, preserving the record of human evolution and migration.

2. Structure and Distribution

SNPs are distributed **throughout the entire genome**, both in:

- **Coding regions (exons)** that directly determine protein structure, and
- **Non-coding regions (introns, promoters, enhancers, and intergenic DNA)** that control when and how genes are expressed.

Each SNP site typically has **two alternative forms (alleles)** — for example, a cytosine (C) in one version of the genome and a thymine (T) in another. These variations can be **neutral, beneficial**, or sometimes **harmful**, depending on where they occur and how they affect gene function.

3. Types of SNPs

SNPs are categorized based on where they are located in the genome and how they affect gene expression or protein function.

A. Coding SNPs (cSNPs)

These occur in the **exons** — the parts of genes that directly code for proteins.

There are two main kinds:

1. **Synonymous SNPs:**

- Change the DNA sequence but *do not change* the amino acid in the protein.
- This happens because of the redundancy in the **genetic code** (for example, both “GAA” and “GAG” code for the amino acid glutamic acid).
- Although silent, they can still affect how efficiently a protein is produced by influencing the speed of translation.

2. Nonsynonymous SNPs:

- Change the DNA sequence in a way that *does* alter the amino acid sequence of the protein.
- These can be of two types:
 - **Missense mutations**, where one amino acid is replaced by another (e.g., valine replacing glutamic acid in sickle-cell anemia).
 - **Nonsense mutations**, where a change introduces a premature stop codon, resulting in a shortened, usually nonfunctional protein.
- Such SNPs can have major effects on protein structure, stability, or activity, and are often associated with diseases.

B. Non-Coding SNPs

These are found in parts of the genome that do not code for proteins — such as **introns, promoters, enhancers, and intergenic regions**.

Although they don’t alter amino acid sequences, non-coding SNPs can still influence:

- **Gene regulation** (by affecting how transcription factors bind to DNA),
- **mRNA splicing**, and

- **Chromatin structure** (which controls how tightly DNA is packed and how accessible it is for transcription).

For example, a SNP in a gene promoter might increase or decrease the expression of that gene, leading to higher or lower protein levels in different individuals.

C. Regulatory SNPs (rSNPs)

A special category of non-coding SNPs affects **gene regulation** directly.

These SNPs occur in regions where **transcription factors, microRNAs**, or other regulatory molecules bind to DNA or RNA.

By altering these binding sites, rSNPs can change how strongly a gene is turned “on” or “off.”

Such variations are responsible for **differences in height, skin color, immune responses, and even behavior** across human populations.

Evolutionary Significance

From an **anthropological perspective**, SNPs serve as invaluable markers for tracing **human evolution, migration, and population structure**. Because they mutate at a relatively low rate compared to microsatellites (approximately 10^{-8} per base per generation), SNPs provide a **stable record of ancient evolutionary events**.

For example, comparative SNP analysis between ancient and modern genomes has revealed **introgression events** — gene flow from Neanderthals into non-African populations roughly 50,000–60,000 years ago. These findings have reshaped our understanding of human ancestry and show that **modern humans are a mosaic of multiple ancient lineages** rather than a purely separate species.

Short Tandem Repeats (STRs)

Short Tandem Repeats (STRs) — also called **microsatellites** — are short, repeating sequences of DNA that serve as some of the most powerful genetic

markers in modern genetics, anthropology, and forensic science. These sequences are made up of **tiny DNA motifs** that are repeated one after another in a row. The length of each repeat unit is **very short**, usually **2 to 6 base pairs** long. Examples of such repeat motifs include **“CA”, “GATA”, or “AGAT.”**

1. Concept and Structure of STRs

At their simplest, STRs are like small “stutters” in our DNA sequence — a small pattern that repeats several times.

For instance:

- In one person, the sequence **“CA”** may repeat **10 times**,
- While in another person, it may repeat **14 times** at the same position in the genome.

This difference in the number **of repeats** makes each person’s STR pattern unique. That is why STRs are so useful in **DNA fingerprinting, paternity testing, and population studies.**

Each STR location on a chromosome is called an **STR locus**. At each locus, people inherit **two alleles** — one from each parent. These alleles may have **different repeat numbers**, which gives each individual a **distinct genetic signature.**

2. Distribution of STRs in the Genome

STRs are found **throughout the human genome**, spread across both **coding** (gene) regions and **non-coding** regions. However, they are **more common in non-coding and intergenic regions**, where changes in repeat number do not directly affect protein function.

In non-coding DNA, mutations in STRs are usually **neutral**, meaning they do not harm the organism. This allows these repeats to **accumulate freely**, increasing their

diversity. In contrast, STRs found inside genes (in coding regions) tend to be **more stable**, since large changes might disrupt the normal function of proteins.

3. Allelic Diversity

The most important feature of STRs is that **the number of repeat units varies between individuals**. Each variation in repeat number represents a different **allele** of that STR locus.

For example, if one individual has 8 repeats of “GATA” and another has 11 repeats, they possess **different alleles** of that same STR region.

4. Molecular Mechanism of STR Variation

The main reason STRs vary so much among individuals lies in how DNA is copied during cell division. The process that creates this variation is known as **DNA replication slippage**.

a) DNA Replication Slippage

During **DNA replication**, an enzyme called **DNA polymerase** reads the existing DNA strand and builds a new complementary strand.

However, when polymerase encounters repetitive sequences such as STRs (for example, “CACACACACACAC”), it can **lose its place** because the pattern is repetitive and easy to misalign.

If the enzyme **slips backward** on the template strand, it adds **extra repeats** to the new strand (insertion).

If it **slips forward**, it **skips** one or more repeat units (deletion).

Both of these errors change the total number of repeats in the DNA sequence — creating a new allele.

Anthropological Relevance

In anthropology, STRs serve as **powerful tools to study genetic diversity, migration patterns, and population structure**. Because STRs mutate at a relatively high rate (10^{-3} to 10^{-4} per generation), they provide high-resolution insights into **recent evolutionary events**. For example:

- In **population genetics**, STR markers have been used to estimate genetic distances among human groups, reconstruct phylogenies, and infer **bottlenecks, admixture events, and gene flow**.
- In **forensic anthropology**, STRs are central to **individual identification and kinship analysis**, even when DNA is degraded or limited in quantity.
- In **bioarchaeology**, STRs have been applied to study the genetic makeup of ancient populations and establish **biological affinities among skeletal remains**.

Variable Number of Tandem Repeats (VNTRs)

VNTRs are special regions in the DNA where a short sequence of bases — called a **core repeat unit** — is repeated **several times in a row (tandemly)**. The key idea is that the **number of times** this small sequence repeats **differs from one person to another**.

For example, imagine a DNA sequence with a repeat unit “GATA.”

- In one person, this sequence might be repeated **15 times**,
- In another person, it might be repeated **25 times**.

Both individuals have the same type of DNA sequence at that location, but the **length of the region** will differ because of the different **number of repeats**. These length differences can be measured in the laboratory and are what scientists call **polymorphisms** — naturally occurring genetic variations.

This variation is the reason VNTRs are sometimes called “**minisatellites.**”

They are similar in structure to **microsatellites** or **Short Tandem Repeats (STRs)**, but differ mainly in size:

- **VNTRs (minisatellites):** Repeat units are typically **10–100 base pairs long.**
- **STRs (microsatellites):** Repeat units are **2–6 base pairs long.**

Thus, VNTRs occupy an intermediate position between the longer repetitive sequences of satellite DNA and the much shorter repeats found in STRs.

2. Molecular Structure of VNTRs

VNTR regions are made up of three main components:

1. **Core Repeat Unit:** A short, specific DNA sequence that is repeated many times in a row.
2. **Flanking Regions:** The non-repetitive DNA sequences on either side of the repeat region. These are important because scientists design **primers** (short DNA pieces used in PCR) that attach to these flanking areas.
3. **Variable Number of Repeats:** The total number of times the core sequence repeats, which varies between individuals and creates allelic diversity.

3. Mechanisms that Cause VNTR Variation

There are two main molecular mechanisms responsible for this variation:

a) Unequal Crossing Over

During **meiosis** — the process that produces sperm and egg cells — pairs of homologous chromosomes line up and exchange segments of DNA.

Sometimes, because of the repetitive nature of VNTRs, these sequences misalign during this process. When this happens, one chromosome may receive **extra repeat units**, while the other chromosome loses some.

This creates two chromosomes with **different numbers of repeats**, introducing new alleles into the population.

Over generations, this process contributes to the diversity observed in human and animal populations.

b) Replication Slippage

During **DNA replication**, the enzyme **DNA polymerase** is responsible for copying the DNA strand. However, when it encounters repetitive sequences like VNTRs, it may “slip” forward or backward by one or more repeat units.

If the enzyme slips backward, it adds **extra repeats**; if it slips forward, it **omits** repeats.

This small error in copying leads to either an **increase** or **decrease** in the number of repeats. Over time, these small changes accumulate, creating many different alleles in the population.

4. Visualization of VNTRs

The analysis of VNTRs involves identifying and comparing the length of DNA fragments containing these repeated sequences. The basic steps are as follows:

a) DNA Extraction

First, DNA is isolated from biological samples such as **blood, saliva, hair roots, skin cells, or bone fragments**. In anthropology, even ancient DNA from archaeological remains can be used, though it is often degraded and must be handled carefully.

b) PCR Amplification

The region containing the VNTR is then amplified using **Polymerase Chain Reaction (PCR)**. Scientists design primers that bind to the **flanking regions** around

the VNTR. This allows the specific region to be copied millions of times in just a few hours, producing enough DNA for analysis.

c) Gel Electrophoresis

After amplification, the DNA fragments are separated using **agarose or polyacrylamide gel electrophoresis**. When an electric current passes through the gel, smaller DNA fragments move faster than larger ones, creating a visible separation based on fragment length.

Since VNTR alleles differ in the number of repeats, each individual produces a **unique pattern of bands** — representing their specific combination of alleles at that locus.

d) Visualization

The separated fragments are then stained or visualized using **dyes, fluorescent markers, or autoradiography**. The final pattern shows the number and size of fragments for each individual.

Each person has **two alleles per VNTR locus** — one inherited from each parent. By comparing band patterns between individuals, scientists can identify relatedness or confirm parentage.

Anthropological Significance

VNTRs bridged the gap between molecular biology and anthropology by providing an accessible and quantifiable measure of **genetic variation at the individual level**. Their applications span across subfields of anthropology, from biological to forensic and medical anthropology.

1. Forensic Identification: VNTR-based **DNA fingerprinting** became the first method used in criminal identification and paternity testing. The historic 1986 case of **Colin Pitchfork** in the UK marked the first time DNA evidence — based on VNTR

analysis — led to a criminal conviction. In anthropology, this method was extended to **identify human remains**, including ancient or disaster-affected populations.

2. Population Genetics: VNTRs provided anthropologists with a molecular tool to study **genetic diversity within and between human populations**. For instance, VNTR variation patterns across populations helped trace **migration routes**, **founder effects**, and **admixture events**. Comparative VNTR data also allowed researchers to infer evolutionary relationships between humans and closely related primates.

3. Genealogical Reconstruction: In field-based anthropological studies, VNTRs have been used to confirm **biological kinship** in societies with complex lineage systems. This enabled the integration of molecular data with **ethnographic kinship models**, offering insights into **social structure, inheritance patterns, and mating practices**.