Module 2: ABCs Of Genetics

As genetics is the core of our services, it is important for each practitioner to become familiar with the basics of genetics. In this module we will introduce you to the fundamentals of genetics and variations that lie behind the genetic tests we perform. You will be able to learn about the following five topics:

1. Intro to basics of genetics
2. Genetic variations
3. Inheritance & Environmental factors
4. DNA testing
5. From DNA sample to results.

Topic 2.1: Intro to Basics of Genetics

DNA as a blueprint of life

We are sure that most of you have heard of deoxyribonucleic acid (DNA) – a molecule present literally in every living being as a significant building block of life. Moreover, it is also known as a blueprint of life, as it stores all information for building our bodies.

It is present in every human cell, and animals, plants, and even microorganisms like bacteria possess DNA.

Just saying that DNA contains all necessary instructions for life is not enough. Therefore, we need to go a bit deeper into molecular aspects to explain the structure, location, and how it is used as a blueprint, dictating the development and growth of the organism via the **central dogma of molecular biology** - the flow of genetic information from DNA to protein.

Deoxyribonucleic acid is a polymeric double-stranded macromolecule, composed of smaller biological building blocks called **nucleotides** [4]. Each nucleotide comprises three components – a nitrogen base, a phosphate group, and sugar called **deoxyribose**. The sugar and the phosphate constitute what is called a backbone of a DNA strand (Figure 1) [5].

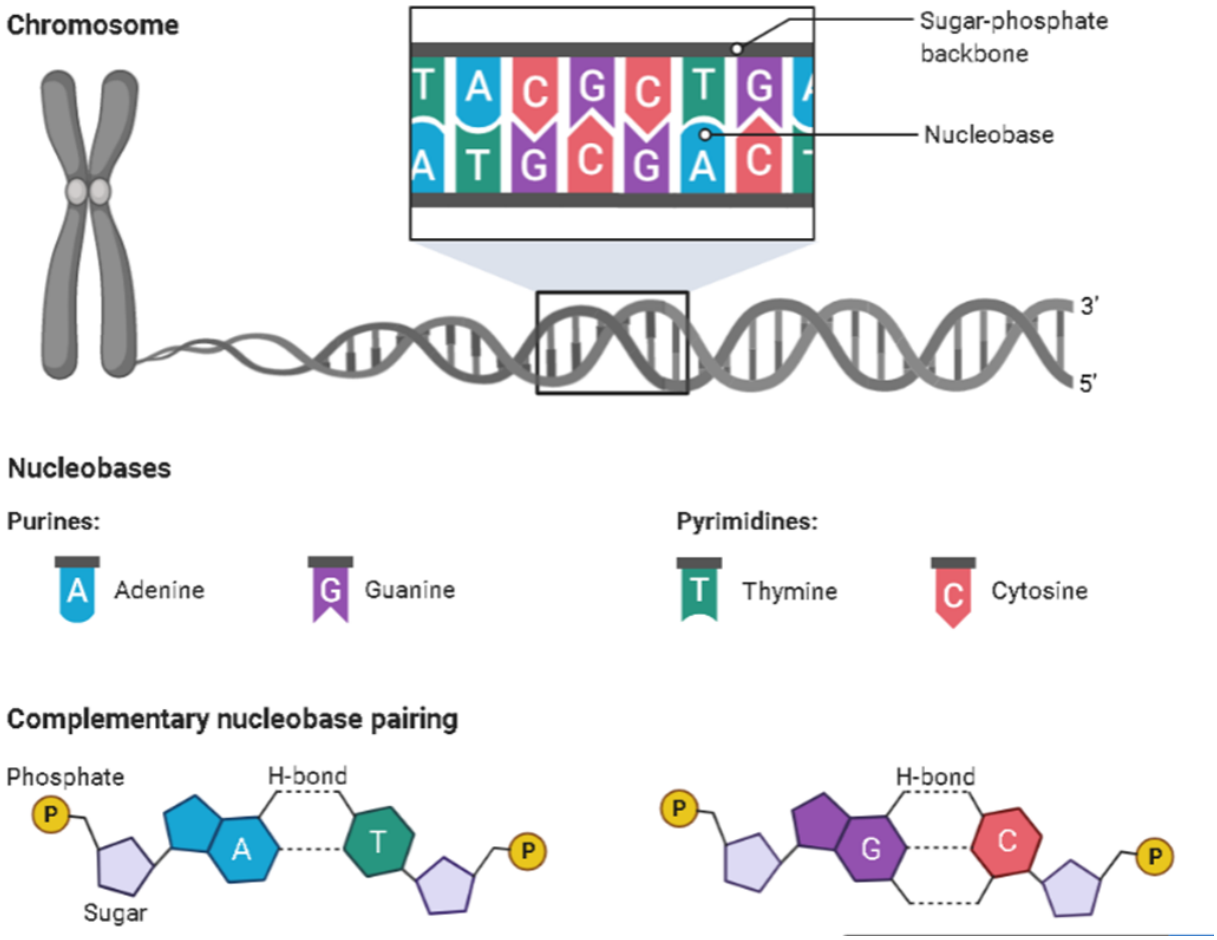

Figure 1: Structure of DNA

The nitrogen base is perhaps the most important as it provides an identity to the nucleotide. Each nucleotide contains either of the following four bases: adenine (A), guanine (G), cytosine (C), or thymine (T) [5].

Within the DNA molecule, adenine always pairs with thymine and cytosine with guanine. Two strands run in opposite directions and are linear in human cells. In contrast, DNA in bacterial cells is circular [4].

DNA is primarily known for its hereditary nature, as it contains our genes and all genetic information. However, it plays an essential role in other crucial life processes.

Cell location

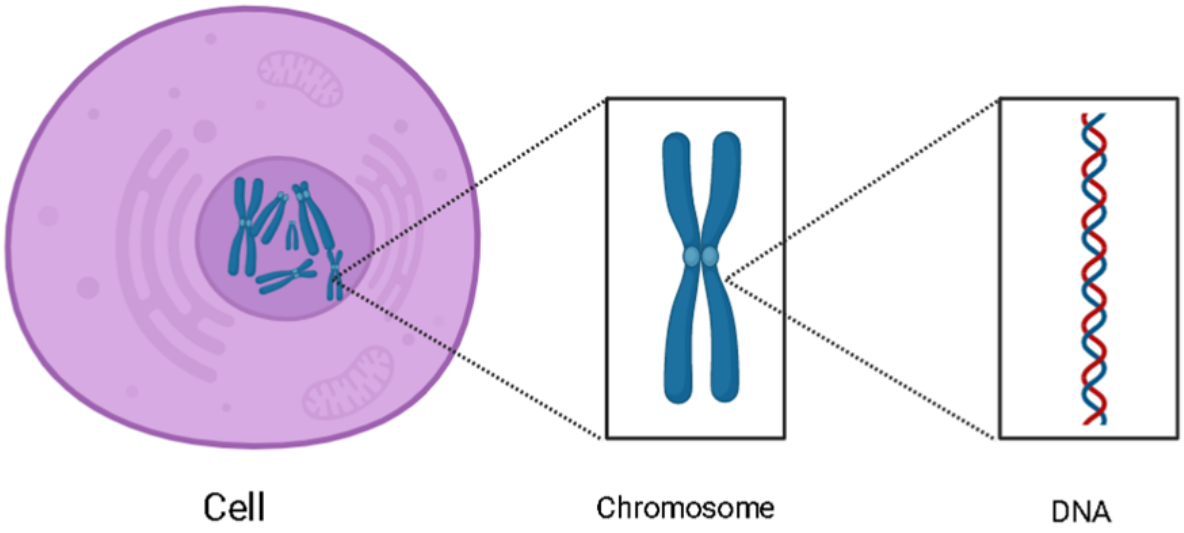

Figure 2: Location of DNA in the cell

As is the case with all eukaryotic cells, DNA is located in the nucleus. It is tightly packed into several structural forms, the last one being the most prominent and well-known– **chromosomes**. Within chromosomes, there are sections of DNA called **genes** that contain information that encodes proteins. There are also stretches of DNA sequences between genes. These are called **intergenic regions** and are a subset of noncoding DNA that mostly have a regulatory function, which means they control gene expression. In other words, each gene contains its coding part - **exons** and non-coding DNA - known as **introns**. Introns are removed during RNA processing before the translation of DNA information to protein.

Each human individual contains a complete set of DNA, called the **genome**. The human genome contains 3 billion bases, around 30,000 genes, packed into 23 pairs of chromosomes, where each pair is inherited from either parent [5].

Each individual has 22 pairs of autosomal chromosomes and 1 pair of heterosomes (sex chromosomes) that can be either XY for males or XX for females (Figure 3). Visual

representation of individual chromosomes, where homologous chromosomes are arranged and divided into groups by decreasing size is called karyogram.

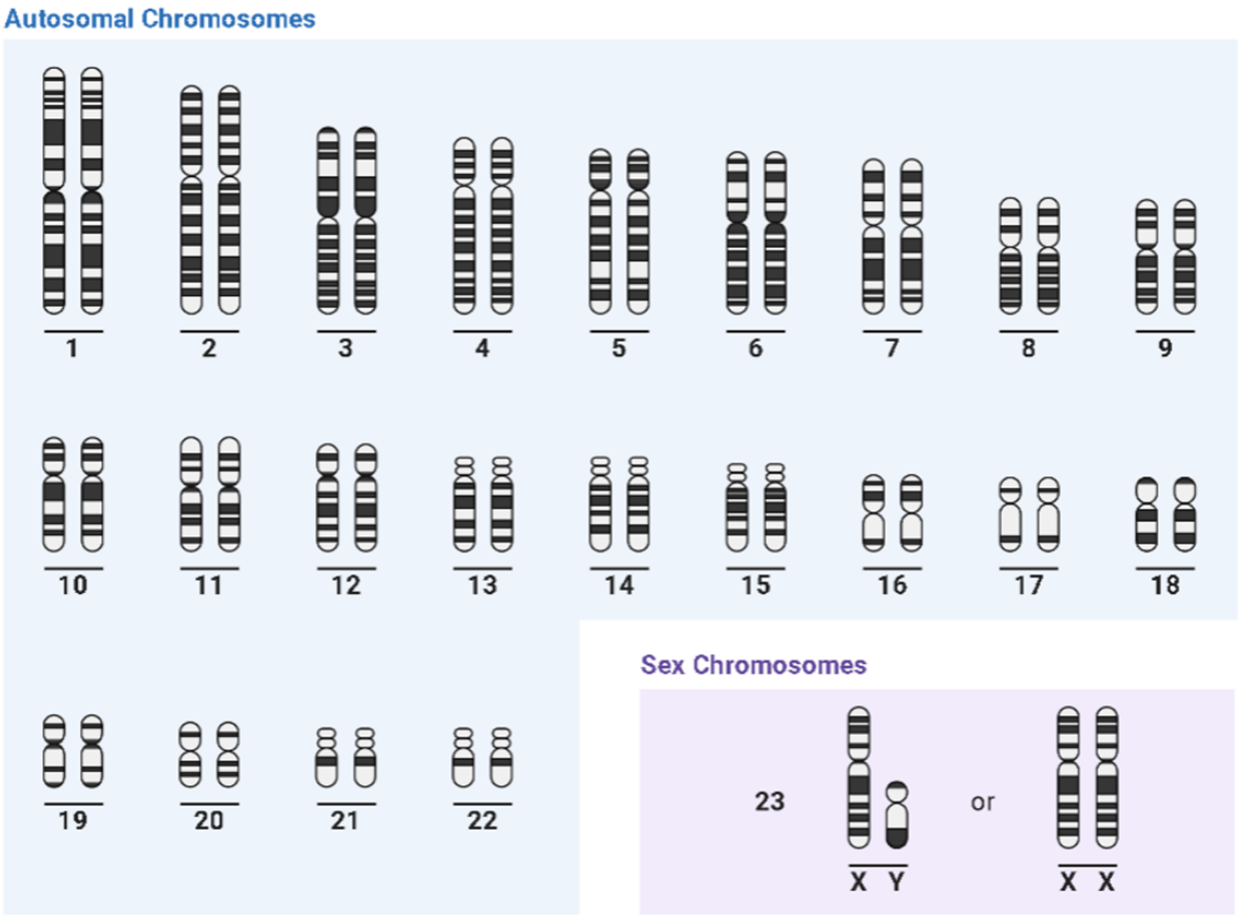

Figure 3: Normal human karyogram

However, DNA is not only located in the cell's nucleus but also mitochondria possess DNA known as mitochondrial DNA(mtDNA). Additionally, Y chromosome DNA (YDNA) is characteristic of the Y chromosome present in men. The difference is that mtDNA is inherited only from maternal lineage and YDNA is inherited only from paternal lineage.

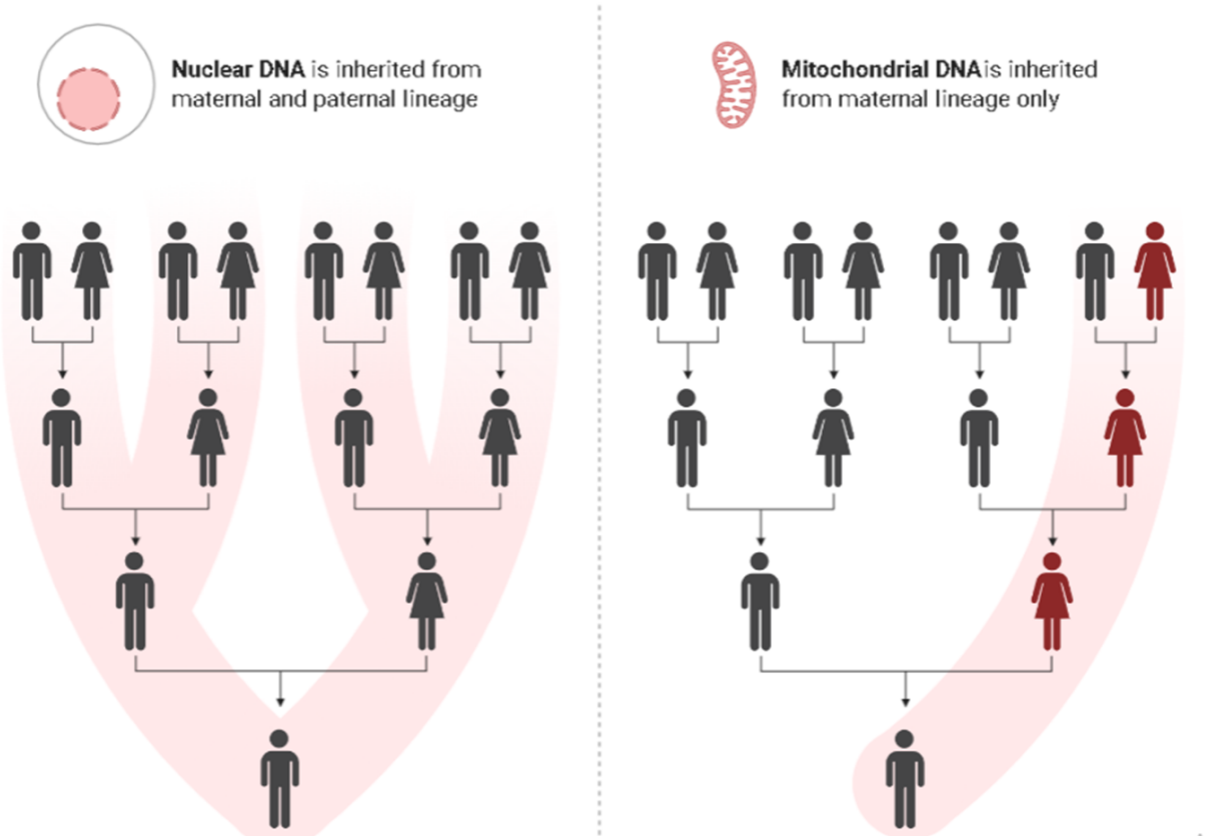

Figure 4: Difference between nuclear and mtDNA

How does information stored in DNA build life?

The primary function of DNA is to provide a cell with the code or instructions on how to make proteins. Proteins will further determine our appearance and traits. Although it may sound simple, this is a very complex process and is a matter of study for molecular biology.

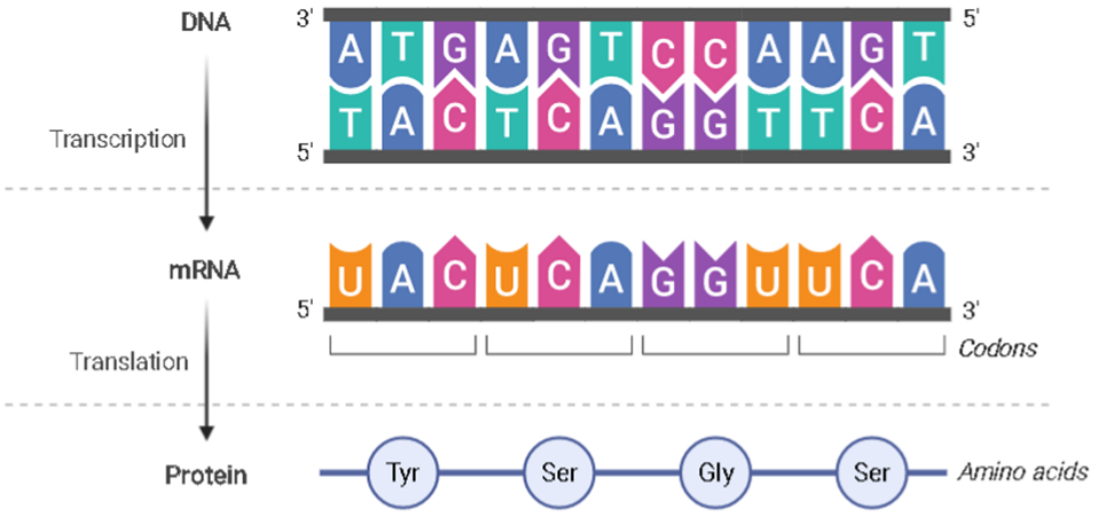

Figure 5: Schematic representation of central dogma

The blueprint of DNA follows the rules of the central dogma. This process explains how to get from DNA to protein via ribonucleic acid (RNA) [6]. As illustrated in Figure 5, it is a process that consists of two stages:

(i) Transcription
(ii) Translation

The synthesis of proteins happens on ribosomes located in the cytoplasm outside the nucleus. But DNA has to go through several steps before synthesis. First, DNA strands have to split apart. Each strand serves as a matrix for synthesizing RNA molecule that has a similar structure to mold DNA molecule but has a single strand. All genetic code contained in DNA is safely transcribed to single-stranded RNA.

RNA carries a message from DNA to cellular machinery in the cytoplasm that reads the message three base pairs at a time and translates them into amino acids that assemble a protein as a final product [6].

Later in this module, we will give a real-life example of the inheritance of traits and how a single mutation can be transmitted from one generation to the next.

What if things in DNA go wrong?

As everywhere in life, errors happen even on a molecular level. These are called mutations and can occur due to errors in DNA replication or be caused by biological or chemical agents. However, it is essential to emphasize that mutations are one of the key evolutionary forces, making us different, despite the huge amount of shared DNA.

The fact is that our DNA is very prone to damage and mistakes. There is an estimate of 175 mutations happening per human genome per generation [7]. Errors occur all the time when our cells undergo replication.

On the other side, our cells possess well-trained and sophisticated proofreading mechanisms that read your DNA and immediately repair any damages. However, sometimes not all damages get repaired. They proceed, causing various diseases sometimes, like cystic fibrosis or sickle cell anemia, caused by a single gene mutation. The next topic will tell more about mutations and genetic variations.

References

1. Mikkelsen, T., Hillier, L., Eichler, E., Zody, M., Jaffe, D., Yang, S. P., ... & Waterston, R. (2005). Initial sequence of the chimpanzee genome and comparison with the human genome. Nature, 437(7055), 69-87.
2. International Chicken Genome Sequencing Consortium. (2004). Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. Nature, 432(7018), 695-716.
3. Kirkness, E. F., Bafna, V., Halpern, A. L., Levy, S., Remington, K., Rusch, D. B., ... & Venter, J. C. (2003). The dog genome: survey sequencing and comparative analysis. Science, 301(5641), 1898-1903.

4. Al Aboud, N. M., Basit, H., & Al-Jindan, F. A. (2019). Genetics, DNA Damage, and Repair.
5. Ghannam, J. Y., Wang, J., & Jan, A. (2020). Biochemistry, DNA Structure. StatPearls [Internet].
6. Cobb, M. (2017). 60 years ago, Francis Crick changed the logic of biology. PLoS biology, 15(9), e2003243.
7. Nachman, M. W., & Crowell, S. L. (2000). Estimate the mutation rate per nucleotide in humans. Genetics, 156(1), 297-304

Figures 3, 4 and 5 generated in BioRender

Topic 2.2: Genetic Variation

Every person on the Earth has a unique genetic makeup, except for identical twins. Although identical twins have the same genetic makeup only at birth, as it is affected and changed by environmental factors and lifestyle, that's a different story [1].

Humans are a relatively young species, so we did not have too much time to accumulate a large amount of genetic variation. Each human being possesses 3 billion base pairs within the genome. Out of that, any two individuals have on average 3 million base pairs that are different. It may seem like a huge number, but the fact is that it comprises only 0.1% of the entire human genome [2].

What is genetic variation?

The 3 billion base pairs of our DNA are tightly packed in chromosomes and constitute approximately 30,000 genes along with intergenic regions. Genes contain information on how our body will produce proteins. Triplets of the above-mentioned base pairs correspond to specific amino acids that produce proteins required for your body to develop and function. Now, let's get back to genetic variations. Genetic variations are slight changes in the genetic sequences of two given individuals that make them different. These changes occur in the translation of DNA base pairs into proteins either at random or due to environmental factors. These changes may affect

your health and well-being, but it does not have to be the case—more on that in the following sections [3].

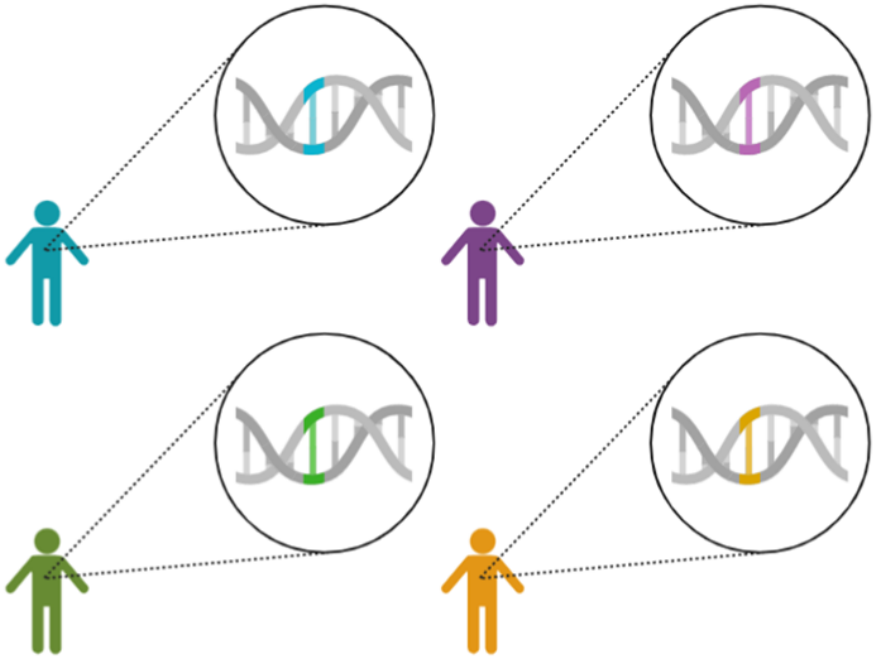

Figure 6: Genetic variation amongst people.

We have discussed the central dogma in detail, the process of producing proteins. Now we will discuss how a mutation in the input of the central dogma (the DNA) can alter its output (the protein)

Major types of genetic variations

A common genetic variation is a change at the DNA level that affects a single nucleotide or more nucleotides. It is called a mutation and may cause an individual to be more prone to certain conditions like sickle cell anemia, hemophilia, etc. However, not all mutations are dangerous, as we have three types of gene mutations based on the effect they make (Figure 7)[4]:

- Silent mutations – do not change the overall protein sequence and therefore do not impact health.

- Nonsense mutations – cause immediate termination of protein sequence, which results in a shorter and truncated protein that can be entirely dysfunctional.
- Missense mutations – cause a change in the amino acid sequence of a protein. In some cases, the new amino acid is very similar to the original amino acid and therefore causes a slight change in the protein function. In other cases, the new amino acid is very different to the original amino acid and this causes major changes in the protein function.

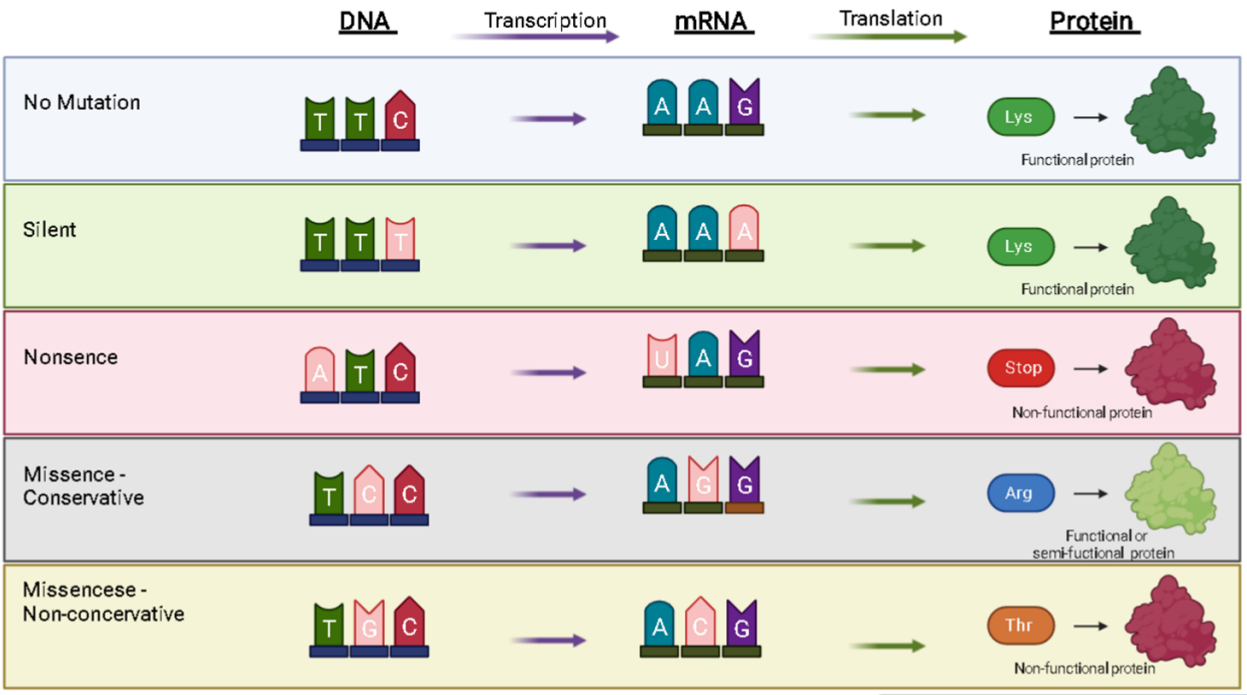

Figure 7: Point mutation and their effects on proteins.

The above-mentioned variants fall under the classification of substitution point mutation as only a nucleotide is changed at a time. However, there are other types of genetic variations :

- Insertions (a portion of DNA inserted elsewhere)
- Deletions (a portion of DNA deleted)
- Duplication (a portion of DNA duplicated)
- Transversions and Transitions (interchanges of DNA segments)
- Chromosomal aberrations and rearrangements (rearrangements of large chunks of a chromosome or chromosomes)
- Single nucleotide polymorphisms (SNPs) (single base-pair is substituted by another one)

- Short tandem repeats (STRs) (short DNA sequence tandemly repeated)
- Copy number variations (the number of copies of a particular gene varies between individuals).

All these types of genetic variations have important implications in health and evolutionary genetics. Moreover, they have various applications in diagnostics, medicine, forensics, and pharmacogenomics [3].

Positive and negative mutation

Genetic variations can have positive, negative, or no effects. In other words, if a change or mutation happens, it does not have to be always negative, causing a disease. It can also be positive, therefore driving the adaptation or survival of the organism and whole species in a given environment.

An example of an advantageous mutation is one in the *CCR5* gene, which may protect against an infection from human immunodeficiency virus (HIV) as this mutation makes it more difficult for the HIV virus to invade the cells and thus prevents autoimmune deficiency syndrome (AIDS) [5].

A negative variation can be seen in sickle cell anemia. This is a hereditary condition characterized by sickle/crescent-shaped red blood cells that get stuck in small blood vessels and slow or block blood flow and oxygen supply. Symptoms include pain, swelling, and tiredness [6]. However, there is a benefit to this mutation as individuals with this mutation are more resistant to malaria, which is caused by a parasite Plasmodium and spread by mosquitoes, all because of the shape of their blood cells affected by the mutation [7].

Either positive or negative, the fact is that genetic variations arise from mutations, and mutations are a key evolutionary force and the ultimate source of what makes us different from each other. Other evolutionary forces that drive changes in populations are:

· gene flow (movement of genes between different groups of organisms),

· genetic drift (random changes of frequency of genetic alleles),

· natural selection (ability to survive better than others in the particular environment due to genetic makeup) [8].

Single Nucleotide Polymorphisms (SNPs)

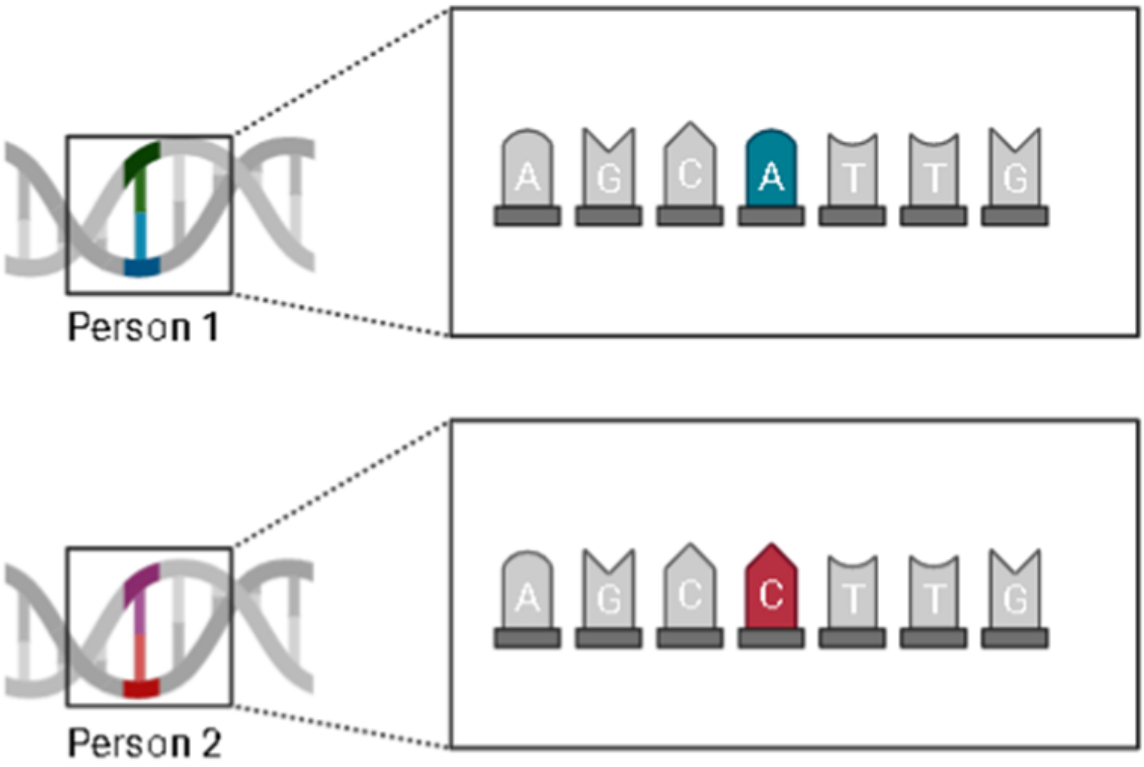

Figure 8: Single-nucleotide polymorphisms.

Single-nucleotide polymorphisms (SNPs) are the most common type of genetic variation, occurring once at about every 1,000 bases. SNP is the DNA sequence variation when a single

nucleotide in the individual's genome differs from the same nucleotide in the genome of another individual. For example, let's take a look into the following two nuclear DNA sequences:

Sequence 1: AAGCCTA

Sequence 2: AAGCTTA

This single nucleotide shift of cytosine from sequence 1 to thymine in sequence 2 represents single nucleotide polymorphism [3]. SNPs are the most frequent type of polymorphisms, and they have their advantages and disadvantages. There are many advantages of analyzing SNPs:

- They are very abundant, which makes them highly available for analysis and, therefore, very useful.
- Genotyping can be automated.
- Their detection will work with highly degraded DNA samples
- Available for simple interpretation because there are no artifacts
- Have a broad spectrum of applications :
    - Diagnostics/risk profiling
    - Ethnic profiling/ancestry analysis
    - Drug response prediction
    - Gene discovery, mapping, and function identification
    - Comparing differences between individuals (forensics applications) [9].

However, there are also some disadvantages to be considered:

- They may provide low information content compared to other polymorphisms because they are mainly bi-allelic, meaning they include only either of two variants. However, there are also multiallelic SNPs (having multiple alleles).
- It may represent a challenge in determining the exact genotype (variants structure) when interpreting mixture in forensics applications for a kinship investigation.

- SNPs have lower mutation rates than STRs, which may present limitations in their interpretation.

Generally, most genetic variations are “harmless” or often called “neutral,” meaning that they do not affect our ability to survive or adapt. Still, some of them are very important and play crucial roles affecting our health and well-being [10].

At BioCertica, we analyze your genetic makeup and look for specific SNPs and their correlation with particular traits and conditions. Examples of SNPs we analyze are SNPs of the *MC4R* gene. This gene comes in different allele combinations.

The A allele is associated with obesity and increased waist circumference. People with an AA variant of the *MC4R* gene have a 1.8 cm larger waist than average; adults with an AG variant have a 0.9 cm larger waist, while those with GG variants have a regular waist. This gene does not confer them a risk of obesity. Allele C is associated with increased body mass index (BMI), especially in children. CC variant carriers tend to have f.44 higher BMI units; CT variants have 0.22 higher units, while TT variants do not contribute to the overall BMI value [11] [12]. These SNPs are related to the ability to manage obesity, which is one of the features we offer in our Weight management DNA kit.

How are SNPs related to disease prevention?

Some traits and health conditions are products of genes and environmental interactions. Some human characteristics can be changed by human intervention, and others cannot. Here we may introduce the most beneficial application of SNPs – for disease risk identification and disease prevention. Let us explain this through a truly tragic story that happened in the past.

You may have heard of Ekaterina Gordeeva and Sergei Grinkov, former young married Russian figure skaters, who had won two Olympic medals in the pairs competition. They expected to write many success stories together. But the happy story ended in November 1995, when 28-year-old Sergei suddenly collapsed and died during a practice session. He was physically fit,

a non-smoker, and was not an alcohol consumer, and there had been no warning signs. The question is: What happened to him?

Analyses showed that he was born with a mutation in a single gene, which affects the formation of blood clots. This mutation, called *platelet antigen-2 (PLA-2)*, can cause clots to form in the wrong place at the wrong time. Due to this mutation, Grinkov had an increased relative risk for a sudden heart attack [13]. It appears that *PLA-2* mutations interact negatively with cholesterol in the blood. Thus, someone with this mutation may reduce their risk of a heart attack by maintaining a low cholesterol diet and exercising regularly.

While genetic mutations are sporadic, we continue to learn more about how SNPs affect disease risk.

References

[1] A. Chatterjee and I. M. Morison, "Monozygotic twins: genes are not the destiny?," *Bioinformation*, vol. 7, no. 7, p. 369, Dec. 2011.

[2] J. F. Crow, "Unequal by nature: a geneticist's perspective on human differences | American Academy of Arts and Sciences." [Online]. Available: https://www.amacad.org/publication/unequal-nature-geneticists-perspective-human-differences. [Accessed: 03-Mar-2022].

[3] T. Strachan and A. Read, *Human Molecular Genetics*, Fifth. CRC Press Taylor & Francis Group, 2019.

[4] N. Chander and S. Viselli, *Lippincott's Illustrated Reviews: Cell and Molecular Biology*, First. 2010.

[5] L. Lopalco, "CCR5: From Natural Resistance to a New Anti-HIV Strategy," *Viruses*, vol. 2, no. 2, p. 574, Feb. 2010.

[6] M. H. Steinberg, "Overview of Sickle Cell Anemia Pathophysiology," *Sick. Cell Anemia From Basic Sci. to Clin. Pract.*, pp. 49–73, Jan. 2016.

[7] H. Cao and M. A. Vickers, “Oxidative stress, malaria, sickle cell disease, and innate immunity,” *Trends Immunol.*, vol. 42, no. 10, pp. 849–851, Oct. 2021.

[8] S. Sunyaev, F. A. Kondrashov, P. Bork, and V. Ramensky, “Impact of selection, mutation rate and genetic drift on human genetic variation,” *Hum. Mol. Genet.*, vol. 12, no. 24, pp. 3325–3330, Dec. 2003.

[9] N. J. Schork, D. Fallin, and J. S. Lanchbury, “Single nucleotide polymorphisms and the future of genetic epidemiology,” *Clin. Genet.*, vol. 58, no. 4, pp. 250–264, 2000.

[10] A. Amorim and L. Pereira, “Pros and cons in the use of SNPs in forensic kinship investigation: a comparative analysis with STRs,” *Forensic Sci. Int.*, vol. 150, no. 1, pp. 17–21, May 2005.

[11] R. J. F. Loos *et al.*, “Common variants near MC4R are associated with fat mass, weight and risk of obesity,” *Nat. Genet.*, vol. 40, no. 6, pp. 768–775, Jun. 2008.

[12] J. C. Chambers *et al.*, “Common genetic variation near MC4R is associated with waist circumference and insulin resistance,” *Nat. Genet.*, vol. 40, no. 6, pp. 716–718, Jun. 2008.

[13] T. Bickel, P. Gunasekaran, G. Murtaza, R. Gopinathannair, S. Gunda, and D. Lakkireddy, “Sudden Cardiac Death in Famous Athletes, Lessons Learned, Heterogeneity in Expert Recommendations and Pitfalls of Contemporary Screening Strategies,” *J. Atr. Fibrillation*, vol. 12, no. 4, Dec. 2019.

Figures generated using BioRender

Topic 2.3: Inheritance and Environmental Influences

Inheritance

When we talk about biological inheritance, we refer to transmitting genes from one generation to the next. Have you ever wondered how this process occurs? Well, it all begins with gametes that are formed through meiosis, which is the cell division by which reproductive cells divide. Gametes are the reproductive cells with a single copy of chromosomes (haploid), and in humans, these cells are the sperm and egg cells.

According to variants they inherit from their parents, individuals can be homozygous (inherit the same variants) or heterozygous (inherit two different variants). In the case of homozygous individuals, the gamete that they produce can only contain that particular gene variant. As seen in Figure 8, the individual on the left has the genotype AA, and therefore all their gametes will contain the A gene variant. The same is true for the individual on the right with the B gene variant. Should these two individuals have a child, the only possible genotype is AB, a combination of their parents [1].

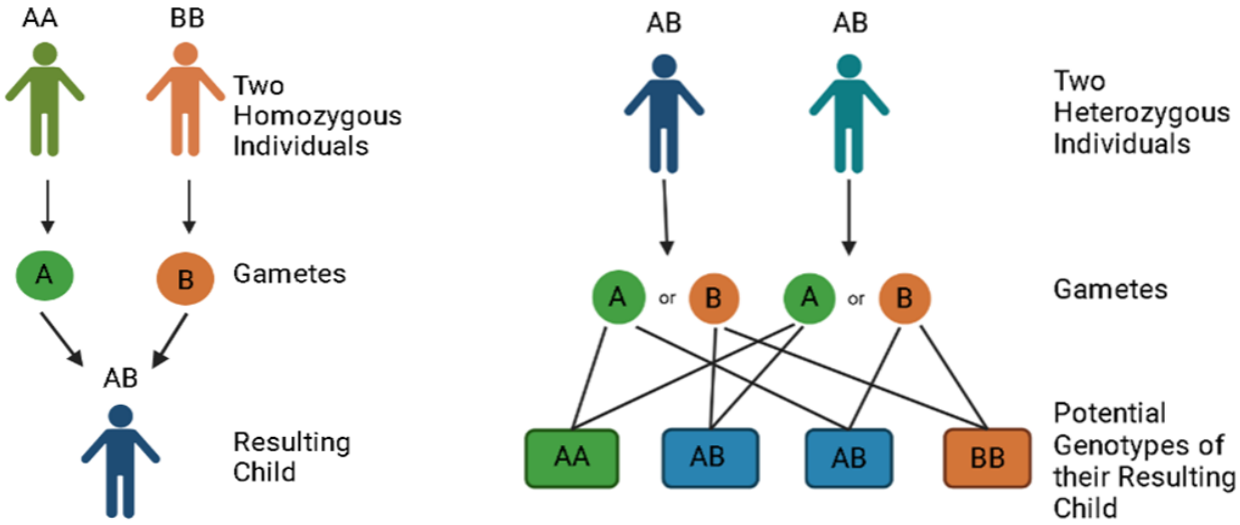

Figure 8: Example of homozygous and heterozygous inheritance

As seen in Figure 8 above, we see the situation in the case of heterozygous individuals on the right. They produce gametes with either one variant (A) or the other (B). Statistically, each variant has a 50% chance of being inherited. Suppose two heterozygous individuals were to have a child, based on the potential combinations of the resulting gametes. In that case, there is

a 50% chance that the child will be heterozygous like their parents. The other possible genotypes are AA or BB, each with a 25% chance of occurring [2].

Mendelian Genetics – Dominant and Recessive traits

We cannot fully explain the basics of genetics without the mention of Gregor Mendel. Gregor Mendel was an Augustinian monk whose work published in 1866 formed the foundational understanding of genetics. Mendel conducted a series of hybridization experiments with *Pisum sativum*, the garden pea plant.

In these experiments, he crossbred a plant with specific contrasting traits with a plant with a different contrasting trait and then observed the characterizations in the subsequent generation of plants. In addition, he also observed the ratios of the traits in the second generation, which led him to further conclusions [1].

Although Mendel looked at many traits, we will use his experiment with the colours of the plant flowers as an example (Figure 9). Mendel started this experiment by crossing a plant with violet flowers and a plant with white flowers. The offspring of this cross were all plants that produced violet flowers. Mendel then allowed these plants to self-pollinate and observed the next generation. About 75% of the plant had violet flowers, but the remaining 25% had white flowers. He concluded that there must be an element in the violet flowers that overwrites the white. Based on this, we now know the concept of dominant and recessive genes.

If an allele is dominant (represented by the capital letter when showing genotypes), then the phenotype is present in the homozygous dominant (RR) and the heterozygous (Rr) form. On the other hand, a recessive allele is only present in its homozygous recessive form (rr). This research led Mendel to the understanding of dominant and recessive traits [3].

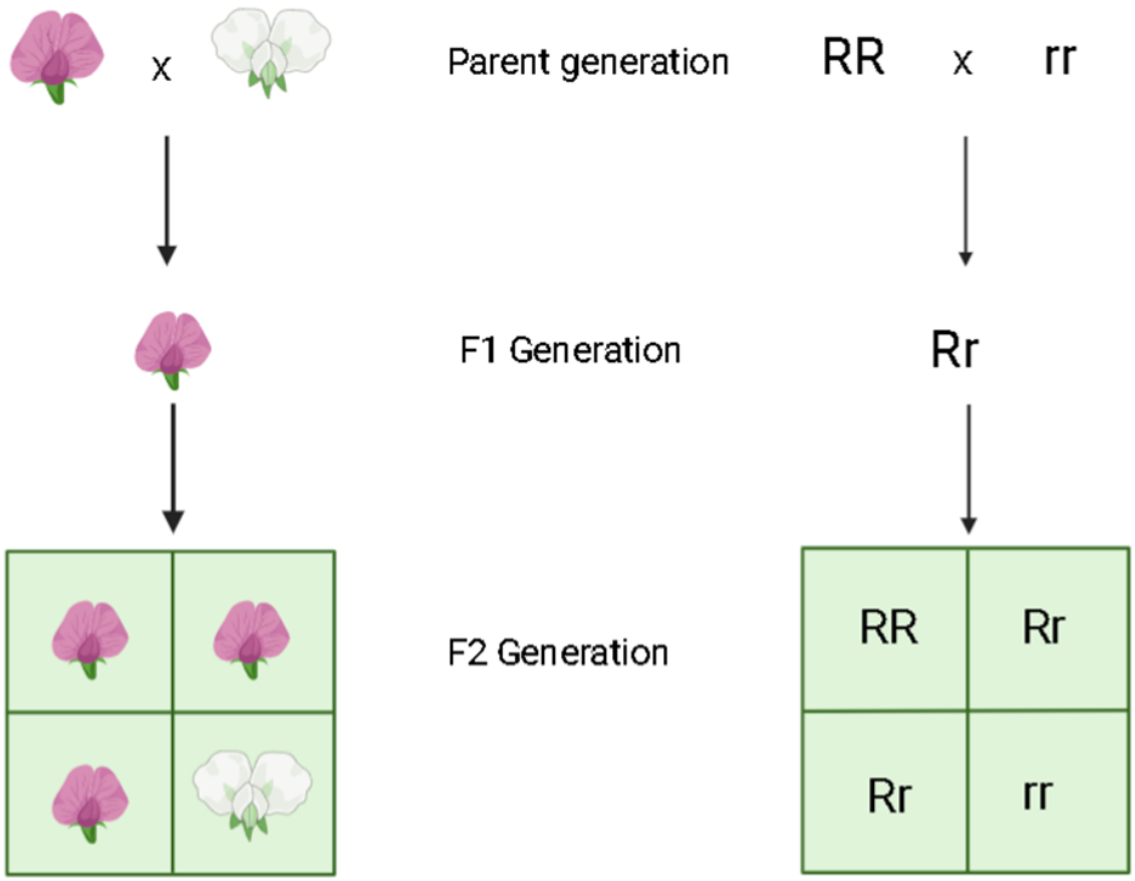

Figure 9: Mendel's cross of violet and white plants

However, the heterozygous genotype may also give rise to an incomplete dominant trait where the resulting phenotype is an intermediate between the dominant phenotype and the recessive phenotype. Clear-cut cases of incomplete dominance are rare; however, snapdragon flowers are a good example of incomplete dominance. In this case, the two homozygous phenotypes are either red or white, and the heterozygous phenotype is pink [2]. Although Mendel's research gave us insight into three main aspects of genetics, his understanding of dominance and recessive traits is what is most important for the purpose of this course. We recommend you watch this video for further information on Mendel's experiments.

Real-life examples of these traits

Mendel's work on peas gave us the foundation but let's face it; we are not pea plants, so let's look at some real-life examples. Figure 10 illustrates dominant and recessive traits that occur in the autosomal chromosomes. An example of an autosomal dominant trait is Huntington's disease. In this neurodegenerative disorder, a multiple trinucleotide repeat of cytosine-adenine-guanine occurs in the *HTT* gene and results in a long string of glutamine residues (amino acid). In the case of Huntington's disease, one mutated version of the *HTT* gene is enough to cause the onset of the disease [4].

On the other hand, sickle cell anemia is an autosomal recessive trait that results from a SNP mutation in the gene for beta-globin, a component of hemoglobin. This mutation causes an amino acid change in the protein, resulting in the hemoglobin sticking to each other and forming a rigid rod within red blood cells. In the case of sickle cell anemia, an individual must bear two mutated gene copies to be affected by the disease. An individual with only a single mutated copy of the gene is known as a carrier as they can pass on the trait but are not affected by it [5].

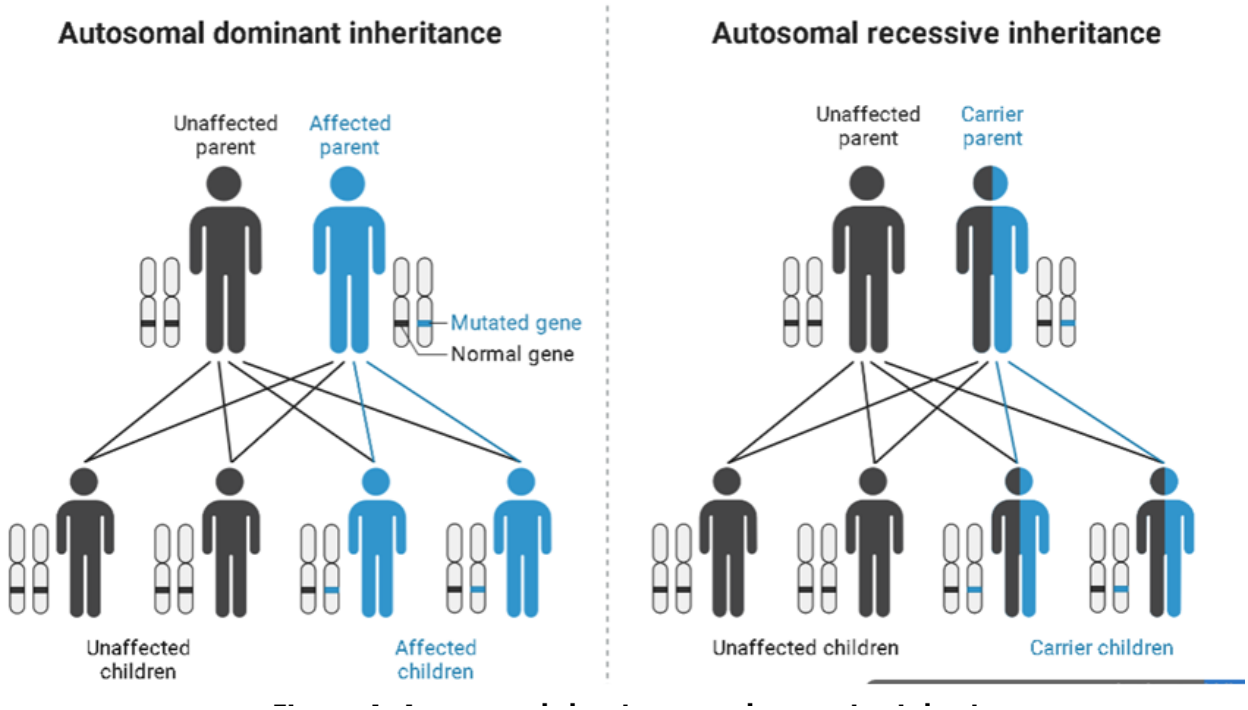

FIgure 10: Autosomal dominant and recessive inheritance

Now after since we have covered autosomal traits, you may ask what about the sex chromosomes? Well, these can get a bit complex. Still, in the case of one of the most famous

examples of disease inheritance documented, we must dive into understanding X-linked recessive inheritance.

As you will remember, females have two X chromosomes, and males have one X chromosome and one Y chromosome. In the case of a mutation on the X chromosome, a male who has that mutation on his X chromosome will be affected by the disease regardless of the trait being dominant or recessive as there is no “backup” copy of the X chromosome. However, in the case of females, X-linked dominant and recessive traits work in the same manner as autosomal dominant and recessive traits. In the case of X-linked recessive inheritance, as seen in Figure 11, the daughters that can result from a carrier female and an unaffected male can either be a carrier or unaffected (50% chance of either occurring). However, in the case of sons, there is a 50% chance that a son will be affected and a 50% chance that he will be unaffected. Furthermore, an affected male will produce carrier daughters and unaffected sons provided that he mates with an unaffected female [1].

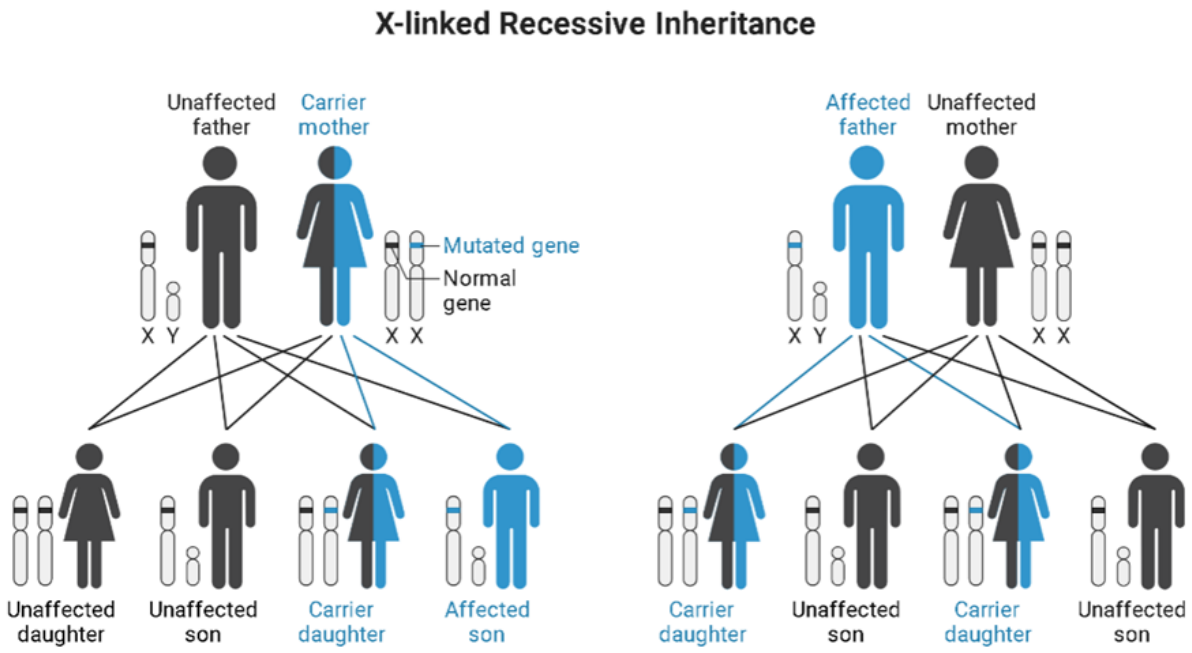

Figure 11: X-linked recessive inheritance

Hemophilia is an X – linked recessive disease characterized by life-threatening bleeding. Hemophilia is known as the “royal disease” because it prominently ran through the descendants of Queen Victoria. Figure 12 is a pedigree showing the inheritance of hemophilia in Queen Victoria’s lineage. A genetic pedigree is an illustrative method to show the inheritance of genetic

traits through generations. As you can see, Queen Victoria was a carrier of the hemophilia gene. Her healthy X chromosome compensated for the mutated one and left her unaffected; however, she passed the mutated X chromosome onto one of her sons, Leopold, and two of her daughters, Alice and Beatrice. As was tradition, her children married into other royal families, and these three took the hemophilia gene with them. This ultimately led to the most famous hemophiliac, Alexei, the son of Tsar Nicholas II and Tsarina Alexandra [6].

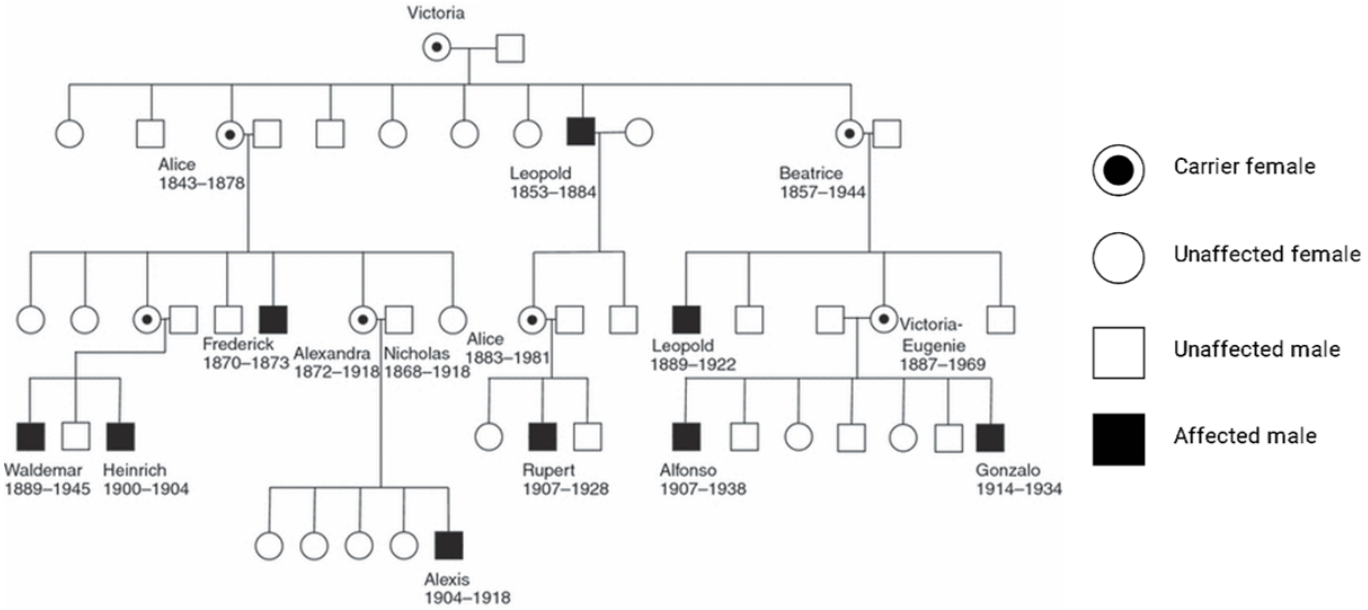

Figure 12: Pedigree showing hemophilia inheritance in Queen Victoria's family

Different types of inheritance

So far, we have mostly been looking at monogenic traits, which arise from a variation in a single gene. One more example of this is cystic fibrosis, which is caused by a mutation in the *cystic fibrosis transmembrane conductance regulator (CFTR)* gene and severely affects the functioning of the lungs [7]. However, this isn't always the case.

Most traits are caused by an accumulation of variation in multiple different genes, and these are called polygenic traits. In Figure 13, we show a few of the genes that have been associated with the development of coronary artery disease [8]. Another type of inheritance is pleiotropy which refers to the traits where a single gene variant causes multiple phenotypes. Here we use the example of phenylketonuria which is a metabolism disorder caused by a mutation of the

phenylalanine hydroxylase (PAH) gene. This disorder has a variety of symptoms from eczema to seizures, and if left untreated, can even cause brain damage [9].

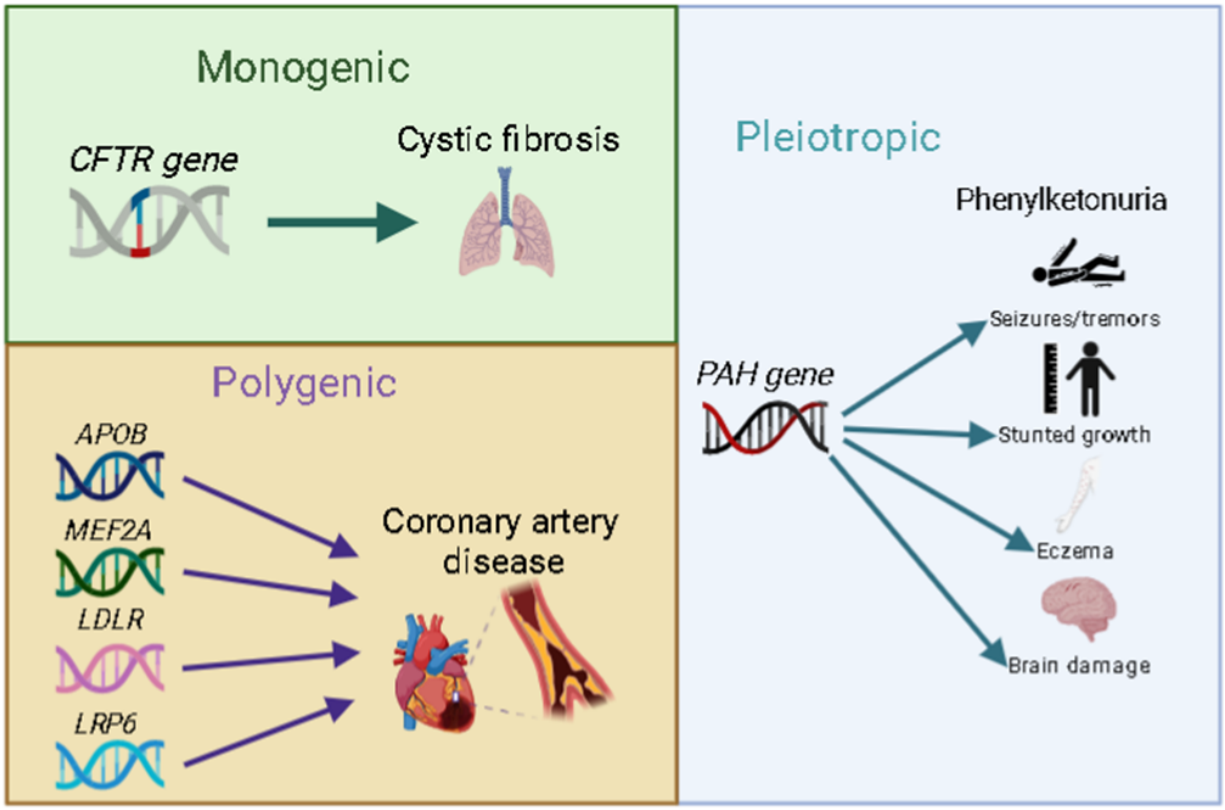

Figure 13: Different types of traits

Environmental influences

Finally, genetics is not always the “all-determining” in some cases. We also have to consider how our lifestyles affect our predisposition to diseases and traits. Some traits are what we call qualitative, the traits with very discrete and easy to measure phenotypes. These traits are mostly determined by genes and have very little influence from their environment. For example, the colours of Mendel’s pea plants are qualitative traits as they are either violet or white.

In contrast, the multifactorial or complex traits are generally polygenic and are determined by their environment. Typically, there are three different classes of multifactorial traits. First is the quantitative traits, which are those that can be measured on a continuum, such as height or weight. Secondly, we have meristic traits, where the phenotype is measured by a limited set of numerical categories, such as the number of eggs a chicken can have in a year or the number of seeds in a pod. Lastly, threshold traits, which seemingly have two phenotypes, with or without; however, the risk of developing the trait is quantitative. For example, type II diabetes is a threshold trait as many genetic and environmental factors influence the risk of developing the disease, and there is a threshold point that defines having diabetes or not [2].

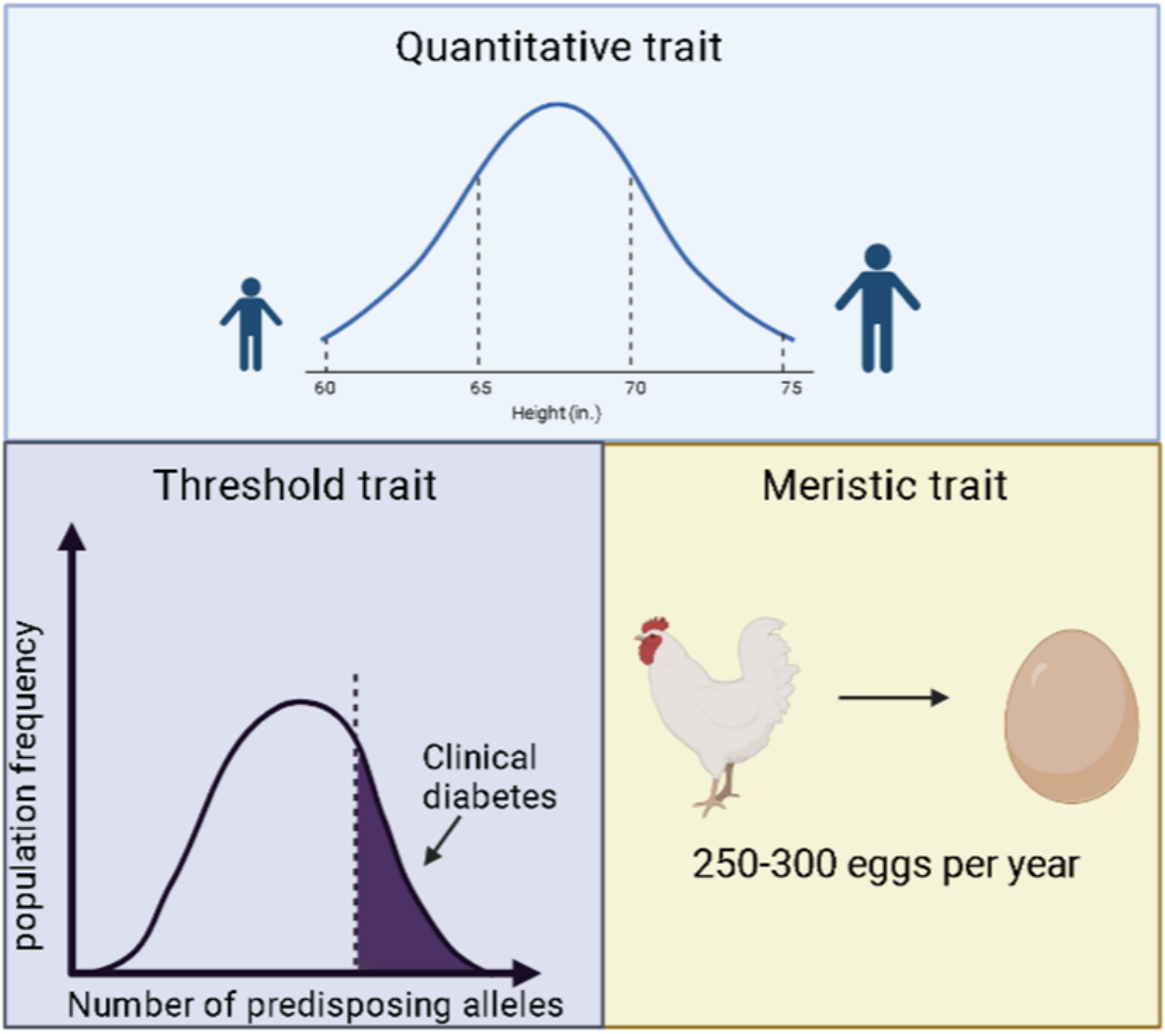

Figure 14: Multifactorial traits

To discuss environmental factors, we should also define the term heritability, a relative measure of the degree to which observed phenotypic differences for a trait are genetic. For a multifactorial trait in a given population, a high heritability estimate indicates that much of the variation can be attributed to genetic factors, with the environment having less impact on the expression of the trait. With a low heritability estimate, environmental factors are likely to have a greater impact on phenotypic variation within the population [2].

Nature vs. Nurture

One of the oldest and the most famous debates in psychology is the one known as “nature vs. nurture,” questioning whether our genes (i.e., nature) or environmental factors (i.e. nurture) have a more significant impact on our appearance, behaviour, and personality [10]. Just to underline, the concept of nature vs. nurture applies to a myriad of human traits.

Scientists nowadays agree that both genetics and environmental factors affect a wide variety of our traits and behaviours. Indeed, nowadays, the actual question is not *if* but *how* these two factors interplay and affect our lives. There is a particular trade-off between these two for any trait, as some are more genetically determined while others are more prone to environmental effects. The conclusion is that we should consider both sides important and contributing [11], [12].

For example, height is a trait hugely influenced by genetics. A child from “a tall family” may have inherited genes that would predispose him to be tall in the future. However, if a healthy and balanced diet does not follow his growth and development, he may never reach his maximum height determined by his genes. Since genetics is not enough in such cases for final appearance, we have to consider height as a product of nature and nurture interaction.

Furthermore, we must also mention epigenetics, the study of the effects of reversible chemical modifications to DNA and histones on the pattern of gene expression. Epigenetic modifications do not alter the nucleotide sequence of DNA. This field of study is becoming a largely important field of research, and we refer you to our article on epigenetics for a more in-depth dive into this topic.

References

[1] T. Strachan and A. Read, *Human Molecular Genetics*, Fifth. CRC Press Taylor & Francis Group, 2019.

[2] K. • Cummings, S. • Palladino, W. S. Klug, M. R. Cummings, C. A. Spencer, and M. A. Palladino, *GLOBAL EDITION Concepts of Genetics ELEVENTH EDITION*. 2016.

[3] "Gregor Mendel and the Principles of Inheritance | Learn Science at Scitable." [Online]. Available: https://www.nature.com/scitable/topicpage/gregor-mendel-and-the-principles-of-inheritance-593/. [Accessed: 26-Feb-2022].

[4] F. O. Walker, "Huntington's disease," *Lancet*, vol. 369, no. 9557, pp. 218–228, Jan. 2007.

[5] S. Roy, "Sickle Cell Anemia, an Autosomal Recessive Disorder: A Review," 2014.

[6] P. Green, "The 'Royal Disease,'" *J. Thromb. Haemost.*, vol. 8, no. 10, pp. 2214–2215, Oct. 2010.

[7] S. Gallati, "Disease-modifying genes and monogenic disorders: experience in cystic fibrosis," *Appl. Clin. Genet.*, vol. 7, p. 133, Jul. 2014.

[8] A. V. Khera and S. Kathiresan, "Genetics of coronary artery disease: discovery, biology and clinical translation," *Nat. Rev. Genet.*, vol. 18, no. 6, p. 331, Jun. 2017.

[9] N. Blau, F. J. Van Spronsen, and H. L. Levy, "Phenylketonuria," *Lancet*, vol. 376, no. 9750, pp. 1417–1427, Oct. 2010.

[10] M. Levitt, "Perceptions of nature, nurture and behaviour," *Life Sci. Soc. Policy*, vol. 9, no. 1, Dec. 2013.

[11] "The Tangled Tale of Genes and Environment: Moore's The Dependent Gene: The Fallacy of 'nature VS. Nurture.'" [Online]. Available: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2223161/. [Accessed: 26-Feb-2022].

[12] M. E. Hayiou-Thomas, "Going beyond nature versus nurture," *Trends Cogn. Sci.*, vol. 10, no. 9, pp. 399–400, Sep. 2006.

Figures generated using BioRender.

Topic 2.4: DNA testing - a powerful tool for discovering your nature

Now when you understand the nature vs. nurture concept, you know that you can influence certain environmental factors shaping you as an individual, as long as you may manage your lifestyle, habits, exercise routines, social life, etc.

However, you may ask yourself whether it is possible to get insight into your nature and influence it? And the good news is that it is possible to screen our genome through genetic testing and gather the information that will eventually help us estimate the effect of genetics on our life.

What is genetic / DNA testing?

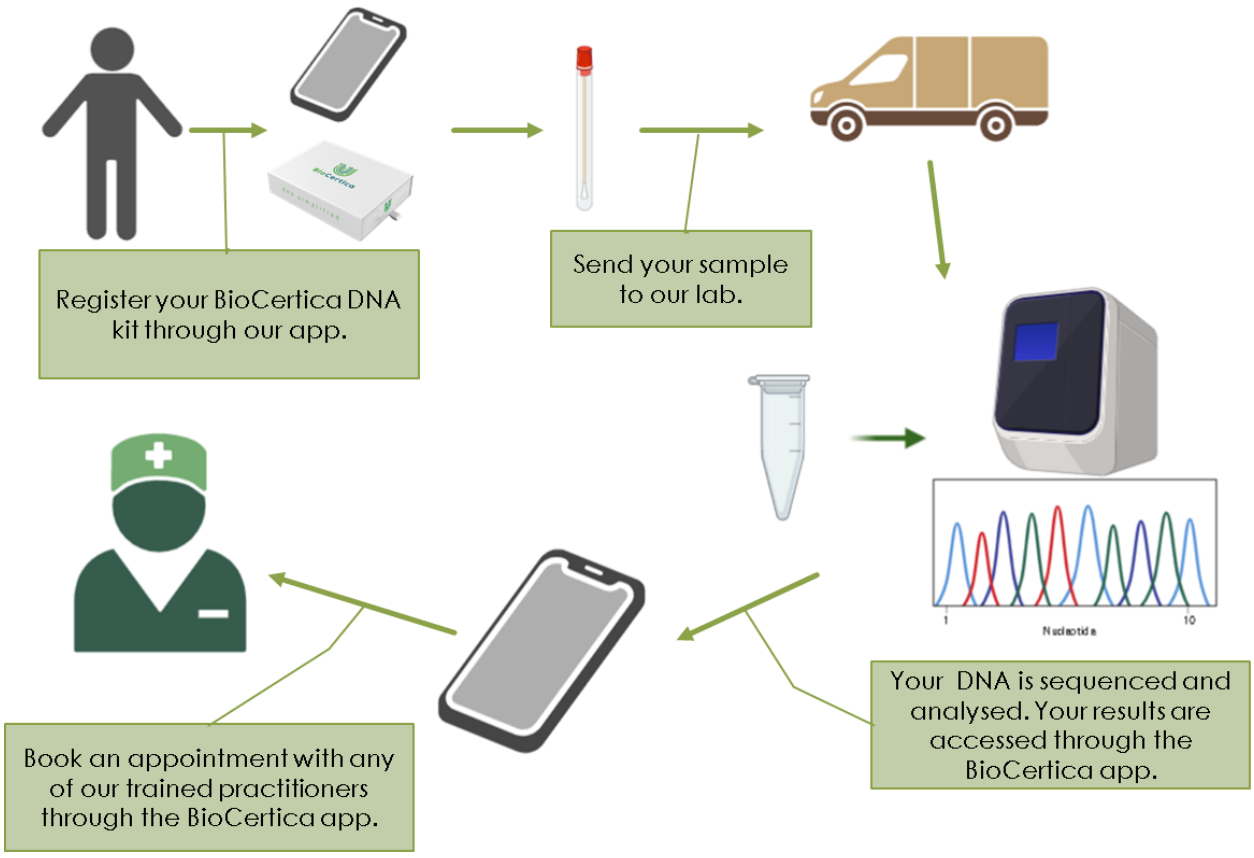

Figure 15: Genetic testing procedure at BioCertica

DNA testing, also known as genetic testing or screening, identifies changes in an individual's DNA sequence [1]. In that way, people can learn more about their DNA and what makes them different from other people.

Genetic variants you inherit from your parents tell you the "nature" part of the nature vs. nurture concept. Furthermore, they will tell you if you have an increased or decreased risk for certain health conditions [2, 3].

Genetic tests give you more information regarding the genetic contribution to a given health condition or a trait. Imagine a person who is in doubt whether to take specific health treatment or not, or an average person that cares about his health and well-being, or imagine a just married couple who are planning to have a child. These are examples where genetic tests can help improve health and well-being. However, whether positive or negative, the results of DNA tests are important and beneficial..

Benefits of genetic tests include the following:

- They help people better understand themselves and their health, as with understanding their genetic makeup, they know their risks for genetic diseases.
- They help people to learn about their origin and ancestors
- They help people better take care of themselves and improve their lifestyle and well-being. For example:
    - They help people to create their personalized fitness and wellness plans.
    - They may help people create better nutrition and exercising plan that will suit body needs
    - They give people insight into weight loss tendency or likelihood for injury
    - They can provide insight into mental health, cardiovascular health, sleeping habits, and more [4].

What is being tested?

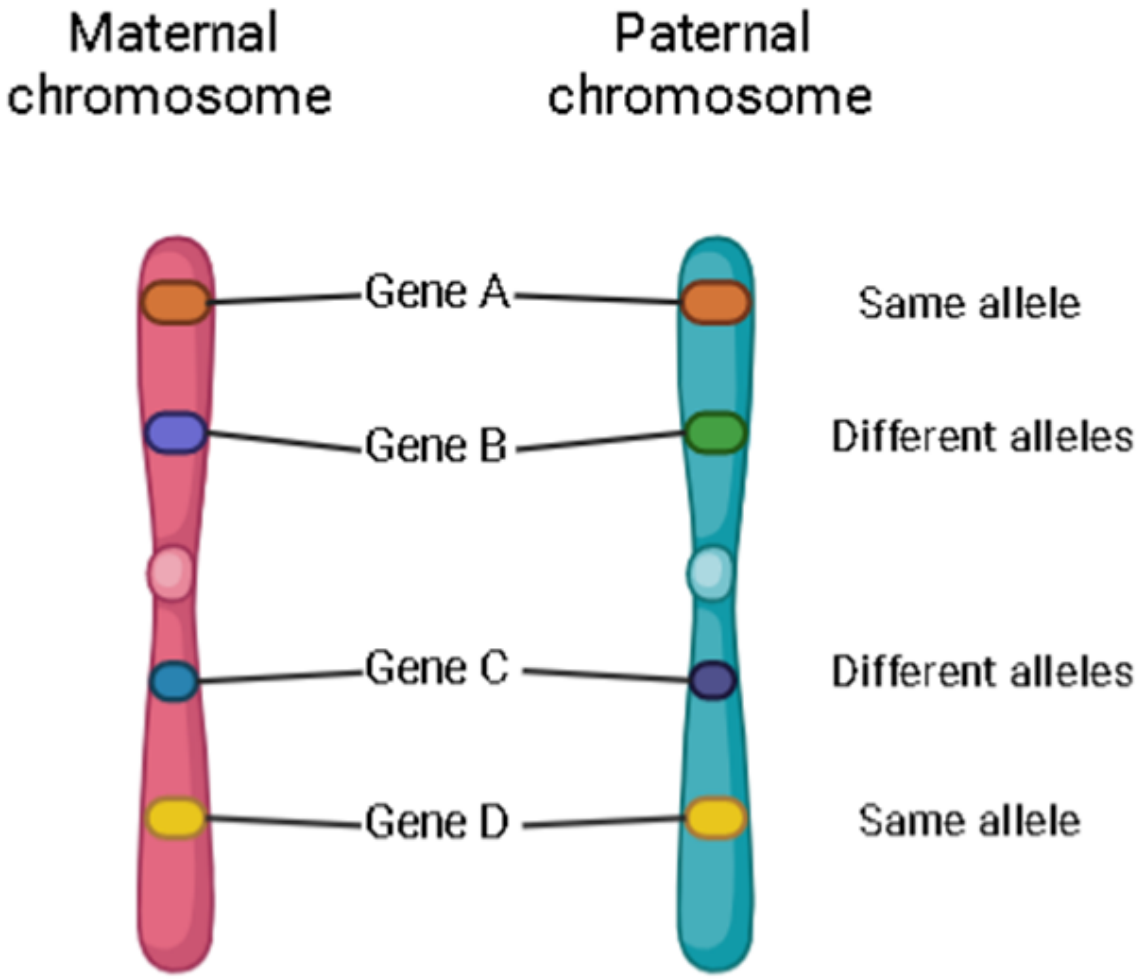

Figure 16: Example of SNPs and traits

Identifying specific alleles forms the basis of genetic testing. Common variations exist in DNA sequences, and you learned about these in Topic 2. These common variations are termed polymorphisms. The most common variations, those that we at BioCertica currently test for, are single nucleotide polymorphisms (or SNPs, for short) where one specific nucleotide substitutes another, potentially giving rise to differences in the proteins coded for. We test which alleles individuals have for which specific SNPs.

Genetic testing applications

Novel advancements in genomics technologies have made DNA tests extraordinarily cheap compared to the past and widely available nowadays and used for various purposes. In the following few paragraphs, we discuss various aspects and applications of genetic tests.

Ancestry

One of the most interesting things for all people may be to discover their ancestors and origin. People may be curious to see if they have distant relatives from Africa, Asia, or Europe. This is exactly what a DNA test can reveal. In a matter of only a few weeks, you can go thousands of years back into the past and discover your lineage [5, 6].

Genetic risk calculation for common disorders

Probably the most interesting and valuable application of DNA tests is for estimating genetic risk for certain common conditions. Assessing genetic risk makes it easier to employ preventive measures for a given condition, even before its onset, thus avoiding the beginning of the disease altogether or improving chances for successful treatment.

The whole thing is done by estimating polygenic risk scores (PRS). PRS is also known under the names genomic scores (GS) and polygenic scores (PGS) or genomic risk scores (GRS). All these refer to the same term. You will have opportunity to learn more about PRS in the next module. For now, it will be enough to say that PRS is a statistical method employed as a part of genetic testing that gives an estimated relative risk score for the developing a certain disease or trait. In other words, a PRS test can tell you how likely is someone to develop particular disease based on their genetics.

Pharmacogenomics

Pharmacogenomics is a relatively new and young scientific branch combining genomics and pharmacology. It explains how people respond differently to the same medication or treatment based on their genetic makeup [7]. Therefore, genetic tests may be helpful here to determine the proper medication or therapy or a dosage that will be helpful to a person affected by a specific condition, knowing their genetics.

Forensics

The most used type of DNA test in forensics is the DNA fingerprint test. This laboratory technique strives to link a DNA sample found on the crime scene and a suspect in a criminal investigation. This is done by comparing their DNAs. In that way, scientists can help determine whether the accused person committed a crime [8].

Paternity testing

Also, a popular type of DNA test is a paternity or maternity test, available to compare a child's DNA molecules to the DNA of a potential parent to establish parenthood [9]. Besides being beneficial to solve parenthood uncertainties, it is helpful to provide support to the child.

Prenatal diagnostics

Prenatal DNA diagnostics tests are becoming more popular nowadays. Women decide to take care of their pregnancy and go for either a traditional approach via analysis of blood biomarkers or by choosing invasive tests like amniocentesis [10, 11].

Also popular is a non-invasive prenatal test (NIFTY) that uses a sequencing approach to isolate and examine a baby's DNA from the mother's blood. In these ways, it is possible to discover whether a future mother is bearing a baby affected by a wide variety of genetic disorders such as Down syndrome, Edwards syndrome, Patau syndrome, Klinefelter syndrome, and many others.

What genetic testing can't tell you?

Besides many benefits, genetic tests also have their risks and limitations and limited amount of information they can provide. Actually, physical risks are very small, particularly for these tests

requiring blood samples or buccal smear. Probably the greatest risk is in prenatal diagnostic testing that carry real risk of losing the pregnancy, because they require a sample of amniotic fluid or tissue from around the fetus. Other risks are associated with the emotional, social, or financial consequences of the test results.

Moreover, genetic testing can provide only limited amounts of information on inherited condition, and cannot often determine if a person will develop symptoms of a disorder, how severe the symptoms will be, or whether the disorder will progress over the time.

Meet Biocertica DNA test

Figure 17: BioCertica DNA kit

There is no doubt that DNA is a vital tool nowadays, as its discovery has revolutionized our lives in many ways, and it will continue to do so. Not only is it a blueprint of life, but it also provides a broad spectrum of answers essential to improve our lives.

At BioCertica, we analyze your whole genome. What distinguishes us from others is the wide variety of DNA kits available and the fact that your identity and data are safeguarded and secured, being always available to you. Current DNA kits available include:

- Bio Ancestry Test
- Bio Nutrition and Well-being
- Bio Weight Management
- Bio Fitness
- Bio Cardiovascular Health
- Bio Mental Health
- Bio Skin Care
- Bio Traits

As depicted in Figure 15 above, the first step after sample collection is to activate the test kit online at BioCertica mobile app by scanning the QR code on a camera phone. Users will immediately get instructions on how to register their kit.

Upon registration, they will need to schedule a courier to pick up their samples and deliver them to the lab as soon as possible. Finally, they will get results within 4 to 6 weeks after sample collection. Of course, we will notify them once your results are ready. Genetic testing is a complex process consisting of many steps like DNA sample collection, isolation, quality controls, genotyping etc.

References

1. Burke, W. (2002). Genetic testing. *New England Journal of Medicine*, *347*(23), 1867-1875.
2. Eichler, E. E. (2019). Genetic variation, comparative genomics, and the diagnosis of disease. *New England Journal of Medicine*, *381*(1), 64-74.
3. Orr, N., & Chanock, S. (2008). Common genetic variation and human disease. *Advances in genetics*, *62*, 1-32.

4. Genetic Testing FAQ. (n.d.). Retrieved August 26, 2021, from Genetic Testing FAQ
5. Royal, C. D., Novembre, J., Fullerton, S. M., Goldstein, D. B., Long, J. C., Bamshad, M. J., & Clark, A. G. (2010). Inferring genetic ancestry: opportunities, challenges, and implications. *The American Journal of Human Genetics*, *86*(5), 661-673.
6. Elliott, C., & Brodwin, P. (2002). Identity and genetic ancestry tracing. *BMJ*, *325*(7378), 1469-1471.
7. Spear, B. B., Heath-Chiozzi, M., & Huff, J. (2001). Clinical application of pharmacogenetics. *Trends in molecular medicine*, *7*(5), 201-204.
8. Shrivastava, P., Trivedi, V. B., Singh, A. K., & Mishra, N. (2012). Application of DNA fingerprinting technology in forensic investigation. *Int J Sci Res*, *2*, 1-4.
9. Pena, S. D., & Chakraborty, R. (1994). Paternity testing in the DNA era. *Trends in Genetics*, *10*(6), 204-209.
10. Levy, B., & Stosic, M. (2019). Traditional prenatal diagnosis: past to present. *Prenatal Diagnosis*, 3-22.
11. Alfirevic, Z., Navaratnam, K., & Mujezinovic, F. (2017). Amniocentesis and chorionic villus sampling for prenatal diagnosis. *Cochrane Database of Systematic Reviews*, (9).

Figure 1 generated with BioRender

Topic 2.5: From DNA sample to results

In the last topic, we explained that genetic testing is a complex and step-wise process, from sampling over DNA extraction and various quality checks to finally obtaining and analyzing results. We divide the whole procedure into the following steps:

- Collecting the sample
- DNA extraction

- DNA quality control, storage, and management
- Genotyping
- Obtaining and analyzing genotyping results
- Generating report

Sample collection

Figure 18: BioCertica Saliva Sample Collection Kit

By definition, DNA sampling represents the method of obtaining biological material to undergo further DNA procedures and mainstream application. It is often the first and most crucial step in conducting research or performing any analysis in molecular biology. The sample collection technique mainly depends on a sampling DNA kit and typically involves samples from buccal swabs, saliva, or blood [1, 2].

There are several ways of collecting DNA samples, most of them being non-invasive and painless. DNA can be obtained from literally all living material: tissue, nail, blood, saliva, semen, or hair. At BioCertica, we rely on saliva sample collection.

Saliva is an extracellular fluid in the mouth produced by salivary glands. Human saliva consists of 98% of water but also contains electrolytes, mucus, white blood cells, epithelial cells., enzymes, antimicrobial agents, and lysozymes [3]. The easiest way to collect a saliva sample is to simply drool into the collection container you get within the ordered DNA test kit [4].

In order to ensure the DNA test integrity and quality, all stages of sample collection have to be conducted under strict protocols. After collection, samples are usually sent via couriers to the laboratory appointed for DNA testing. If you are curious to read more about tips one should pay attention to when collecting DNA samples, and to learn more about other sampling methods, feel free to refer to Additional Reading 2: DNA sample collection.

DNA extraction

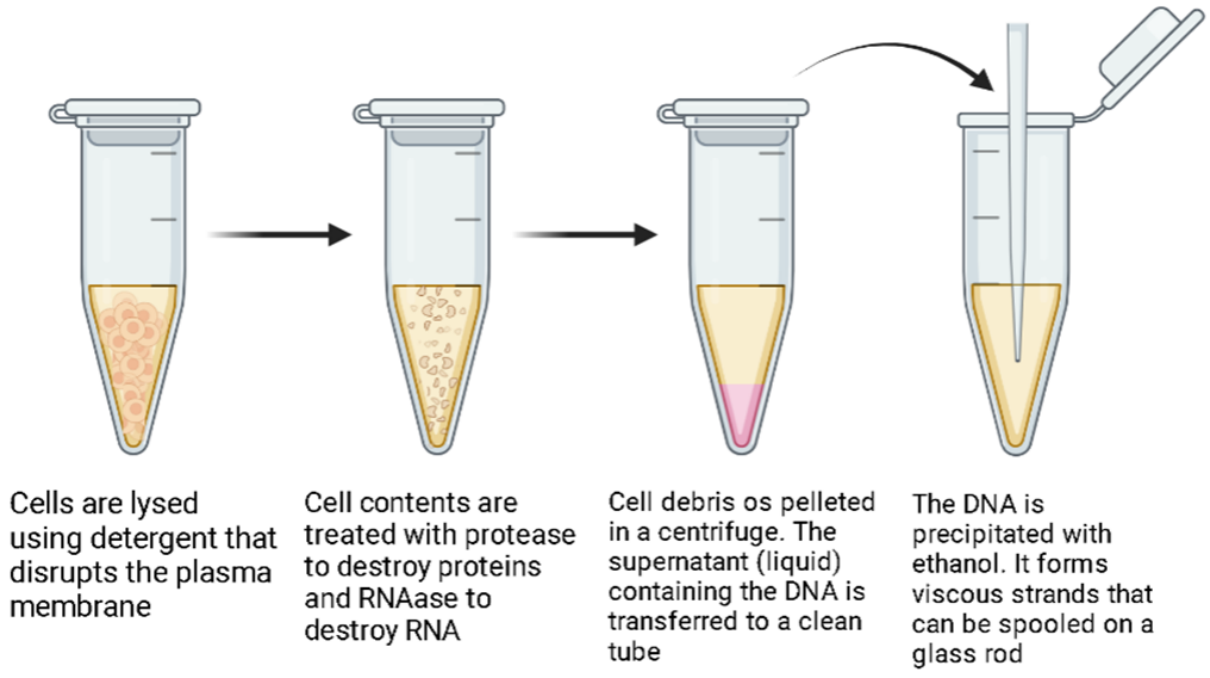

Figure19: DNA extraction steps

Since scientist Friedrich Miescher performed the first DNA isolation or extraction experiment in early 1869, DNA extraction has become the essential tool of molecular biology and genetics and applied fields like forensics, medical diagnostics, and agriculture [5].

DNA extraction is a routine method used to isolate DNA from the cell's nucleus or mitochondria [6]. DNA is located in the nucleus that is surrounded by its nuclear membrane. To isolate DNA, we have to break down both cell and nuclear membranes. Then we have to remove cell debris. The final step will be to precipitate and purify DNA. DNA extraction consists of three stages [6]:

1. Breakdown (or so-called lysis) of cell and nuclear membrane that is achieved by adding detergents.
2. Removal of cell debris and proteins is done by adding proteases. Proteases are enzymes that digest proteins.
3. DNA purification and precipitation are done by adding ice-cold alcohol ethanol or isopropanol.

There are many different methods for DNA extraction. At BioCertica, we rely on the silica-column extraction method. If you are interested in learning more about this technique and other DNA extraction procedures, then refer to Additional Reading 3: DNA extraction.

Quality evaluation

Usually, after extraction, we should check the quality of extracted DNA, which is done using spectrophotometry or agarose gel electrophoresis. In that sense, we can check the quality of DNA by calculating the concentration, yield, and purity of extracted DNA.

Spectrophotometry is a method that measures how much a chemical substance can absorb light [7]. Every substance absorbs or transmits light over a specific range of wavelengths. Based on that information, we can determine the composition of the substance. The concentration of the DNA is estimated by measuring the absorbance at 260 nm. The next step is the evaluation of other parameters like the dilution factor. It helps us to determine whether the obtained sample is pure DNA or contains contaminants [8].

Besides DNA spectrophotometry, **electrophoresis** is another popular and effective method used widely to examine or analyze extracted samples. We use spectrophotometry to reveal DNA samples' absorbance, concentration, and purity. On the other hand, we use electrophoresis to separate fragments of nucleic acids or proteins, applying an electric current and electromagnetic field. It helps reveal the size, charge, or binding affinity of molecules of interest [9, 10].

Storage and Management

After our DNA successfully passes all necessary quality checks, we may use it immediately for genotyping or store it for later use. To achieve the best results in the mainstream application, extracted DNA samples must be fully preserved and stored under conditions that limit degradation.

Samples are usually stored in buffers with a pH above 8.5. Buffers are solutions used to maintain the pH of the solution relatively stable. They can resist the pH change due to the addition of acidic or basic components. Finally, samples are stored at -20℃, and repeated freezing and thawing of DNA should be avoided [11].

We should keep the sample safe and away from contamination with deoxyribonuclease (DNase) and heavy metals. They will both boost degradation. Human skin is the most frequent source of contamination as it contains DNase, a DNA-degrading enzyme. Even very low concentrations of DNAse at short exposure may result in substantial sample degradation. Therefore, it is recommended to avoid any contact between samples and fingers by either using sterilized solutions and tubes or wearing gloves [12].

Genotyping

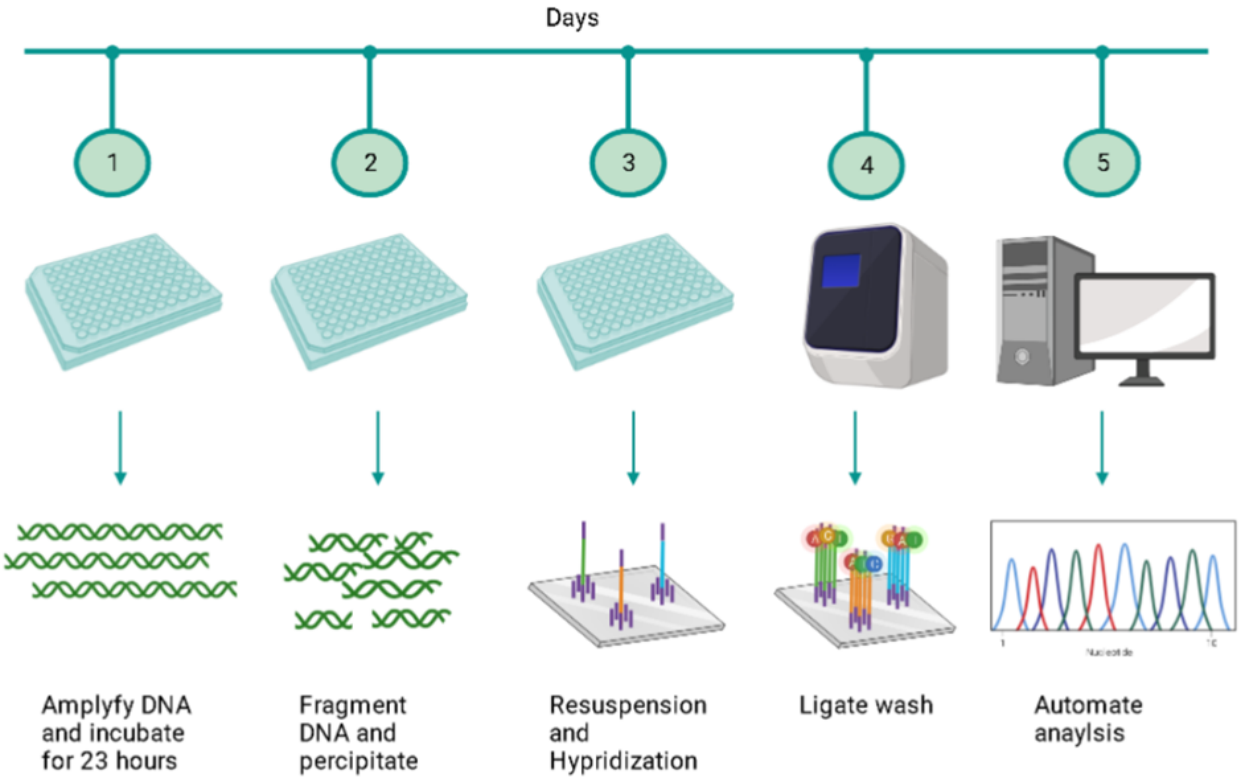

Figure 20: Stepwise illustration of the genotyping process (Based on Thermofisher Scientific-Axiom Genotyping Solution)

Samples are stored until the application in genotyping, which compares the genetic makeup of the sample of interest to the reference database and finds differences. Here we will briefly cover the method of genotyping we use. However, you can read more about genotyping, its applications, and the main genotyping technologies in Additional Reading 4: Genotyping: What is it and why is it important? and Additional Reading 5: Main genotyping technologies.

At Biocertica, we use Affymetrix Axiom genotyping technology and PMDA Plus Array on the GeneTitan machine. This technology is cost-effective and enables complete automation of sample preparation and other steps. Also, this assay contains 96 individual arrays and allows genotyping of up to 800 000 SNPs, with a throughput of ~200 samples per month [13].

The whole process consists of the following steps (depicted in Figure 20 above):

- DNA amplification: Once DNA is extracted and passes quality control of DNA, we can proceed with DNA amplification. The amplification step aims to enrich the DNA for the successful completion of the next steps. A total of 100 ng of genomic DNA is amplified.
- Fragmentation and precipitation: After DNA amplification, the DNA is fragmented and precipitated. During these steps, the aim is to break DNA into shorter fragments of 25-125 base pairs. This will enable the fragments to attach to hybridization probes on the Affymetrix PMDA (Precision Medicine Diversity Array). Additionally, quality control checks of obtained fragments are performed.
- Resuspension and hybridization: After the DNA samples are fragmented and precipitated, it is resuspended by mixing and prepared for the hybridization step. The DNA samples are hybridized onto the Affymetrix PMDA array plate for 23 hours.
- Staining: Following hybridization, the bound target is washed under stringent conditions to remove the non-specific background caused by random ligation events. Each polymorphic nucleotide is queried via a multi-color ligation event carried out on the array surface. After ligation and staining, the plate is imaged on GeneTitan MC Instrument to obtain intensity data from which genotype information is received.

Afterward, the results are ready to analyze and generate the report.

Obtaining genotyping data

After genotyping on the GeneTitan machine, we obtain raw signal intensity values that have to be converted to meaningful genotype data. These intensity value are stored as a set of 5 different files, including the CEL file - an Affymetrix Probe GeneChip results file, that contains information on the probe set's intensity values, where a probe represents genes [14].

In other words, it contains information about single nucleotide polymorphisms (SNPs). Information about probes gets extracted from image DNA microarray data by an image analysis software called Affymetrix and a process known as a genotype calling.

Genotype calling is a process where intensity CEL files are converted to VCF files containing the user's genotype data. It takes intensity data from CEL files, performs classification of every SNP, assigns genotype, and obtains genotype data files used for all other processes.

Variant Call Format (VCF) files contain meta-information about an SNP like a chromosomal position in the genome, SNP identifier such as rs number, the reference base non-reference bases, quality score, etc. [15].

Extracted data may contain thousands of data points, making it large in size. After obtaining our SNPs data, we are ready to process them further, including SNPs selection and report generation.

SNPs selection and report generation

If you look at any of our DNA tests, you will see several traits we test for (vitamin and mineral levels, blood pressure, etc.). You may ask now, how do we choose and select traits to test for in our DNA tests? Moreover, how do we select SNPs for our reports?

Well, we look at available genetic variants and their confirmed associations to diseases or traits available at public databases like Polygenic Score (PGS) Catalog and GWAS Catalog as well as other relevant genetic databases on NCBI.

All the traits we include in our reports adhere to the strict quality protocol they have to pass, ensuring that the information we provide you is accurate and up to date according to the latest scientific research.
In order to generate the final genetic report, we use polygenic risk score (PRS) - a comprehensive statistical method used to estimate the relative genetic risks of being affected by a certain condition or inheriting a certain trait. In the next module, you will have the opportunity to learn everything about PRS.

References

1. Mullegama, S. V., Alberti, M. O., Au, C., Li, Y., Toy, T., Tomasian, V., & Xian, R. R. (2019). Nucleic acid extraction from human biological samples. *Biobanking*, 359-383.
2. National Research Council. (1998). Evaluating human genetic diversity.
3. Young, J. A., & Schneyer, C. A. (1981). Composition of saliva in Mammalia. *Australian Journal of Experimental Biology and Medical Science*, *59*(1), 1-53.
4. Granger, D. A., & Taylor, M. K. (Eds.). (2020). *Salivary bioscience: foundations of interdisciplinary saliva research and applications*. Springer Nature.
5. Tan, S. C., & Yiap, B. C. (2009). DNA, RNA, and protein extraction: the past and the present. Journal of Biomedicine and Biotechnology, 2009.
6. Alonso, A. (2013). DNA Extraction and Quantification. Encyclopedia of Forensic Sciences (Second Edition). 214-218
7. Morris, R. (2015). Spectrophotometry. Current Protocols Essential Laboratory Techniques, 11(1), 2-1.
8. Vo, K. (2015, July 22). Spectrophotometry. Retrieved from 2.1.5: Spectrophotometry - Chemistry LibreTexts
9. Metzker, M. L. (2005). Emerging technologies in DNA sequencing. Genome Research, 15(12), 1767-1776.
10. Johansson, B. G. (1972). Agarose gel electrophoresis. Scandinavian Journal of Clinical and Laboratory Investigation, 29(sup124), 7-19.
11. DNA Extraction from Buccal Swabs. Retrieved from DNA Extraction from Buccal Swabs | Thermo Fisher Scientific - BA
12. Surzycki, S. (2012). Basic techniques in molecular biology. Springer Science & Business Media.
13. Shapero, M., Purdy, M., Kurapati, R., Law, J., Lee, H., … & Bellon, L. SNP genotyping using Affymetrix Axiom® Genotyping Solution, Application Brief, Thermofisher Scientific.
14. Affymetrix® CEL Data File Format. (n.d.). Retrieved from Affymetrix CEL data file format
15. IGSR: The International Genome Sample Resource. (n.d.). Retrieved from VCF (Variant Call Format) version 4.0 | 1000 Genomes (internationalgenome.org)