Talking Glossary of Genetics and Genomics

Glossary featuring terms from genetics and genomics explained in an easy-to-understand way by leading scientists and professionals at the National Genome Research Institute.

-A-

Adenine
Adenine (A) is one of the four nucleotide bases in DNA, with the other three being cytosine (C), guanine (G) and thymine (T). Within a double-stranded DNA molecule, adenine bases on one strand pair with thymine bases on the opposite strand. The sequence of the four nucleotide bases encodes DNA's information.

Allele
An allele is one of two or more versions of DNA sequence (a single base or a segment of bases) at a given genomic location. An individual inherits two alleles, one from each parent, for any given genomic location where such variation exists. If the two alleles are the same, the individual is homozygous for that allele. If the alleles are different, the individual is heterozygous.

Alternative Splicing
Alternative splicing is a cellular process in which exons from the same gene are joined in different combinations, leading to different, but related, mRNA transcripts. These mRNAs can be translated to produce different proteins with distinct structures and functions — all from a single gene.

Amino Acid
An amino acid is the fundamental molecule that serves as the building block for proteins. There are 20 different amino acids. A protein consists of one or more chains of amino acids (called polypeptides) whose sequence is encoded in a gene. Some amino acids can be synthesized in the body, but others (essential amino acids) cannot and must be obtained from a person's diet.

Aneuploidy
Aneuploidy is an abnormality in the number of chromosomes in a cell due to loss or duplication. In humans, aneuploidy would be any number of chromosomes other than the usual 46.

Animal Model
An animal model is a non-human species used in biomedical research because it can mimic aspects of a biological process or disease found in humans. Animal models (e.g., mice, rats, zebrafish and others) are sufficiently like humans in their anatomy, physiology or response to a pathogen that researchers can extrapolate the results of animal model studies to better understand human physiology and disease. By using animal models, researchers can perform experiments that would be impractical or ethically prohibited with humans.

Anticodon
A codon is a DNA or RNA sequence of three nucleotides (a trinucleotide) that forms a unit of genetic information encoding a particular amino acid. An anticodon is a trinucleotide sequence located at one end of a transfer RNA (tRNA) molecule, which is complementary to a corresponding codon in a messenger RNA (mRNA) sequence. Each time an amino acid is added to a growing polypeptide during protein synthesis, a tRNA anticodon pairs with its complementary codon on the mRNA molecule, ensuring that the appropriate amino acid is inserted into the polypeptide.

Antisense
Antisense is the non-coding DNA strand of a gene. In a cell, antisense DNA serves as the template for producing messenger RNA (mRNA), which directs the synthesis of a protein.

Autism
Autism is a condition related to brain development that can cause significant social, communication and behavioral challenges. Symptoms usually appear before the age of three. The exact cause of autism is not entirely known, although genetics clearly plays an important role. Autism is one of a group of related developmental conditions sometimes called the autism spectrum that affect people differently and to varying degrees.

Autosomal Dominant Disorder
Autosomal dominant is a pattern of inheritance characteristic of some genetic disorders. "Autosomal" means that the gene in question is located on one of the numbered, or non-sex, chromosomes. "Dominant" means that a single copy of the mutated gene (from one parent) is enough to cause the disorder. A child of a person affected by an autosomal dominant condition has a 50% chance of being affected by that condition via inheritance of a dominant allele. By contrast, an autosomal recessive disorder requires two copies of the mutated gene (one from each parent) to cause the disorder. Huntington's disease is an example of an autosomal dominant genetic disorder.

Autosomal Recessive Disorder
Autosomal recessive is a pattern of inheritance characteristic of some genetic disorders. "Autosomal" means that the gene in question is located on one of the numbered, or non-sex, chromosomes. "Recessive" means that two copies of the mutated gene (one from each parent) are required to cause the disorder. In a family where both parents are carriers and do not have the disease, roughly a quarter of their children will inherit two disease-causing alleles and have the disease. By contrast, an autosomal dominant disorder requires only a single copy of the mutated gene from one parent to cause the disorder. Sickle cell anemia is an example of an autosomal recessive genetic disorder.

Autosome
An autosome is one of the numbered chromosomes, as opposed to the sex chromosomes. Humans have 22 pairs of autosomes and one pair of sex chromosomes (XX or XY). Autosomes are numbered roughly in relation to their sizes. The largest autosome — chromosome 1 — has approximately 2,800 genes; the smallest autosome — chromosome 22 — has approximately 750 genes.

-B-

Base Pair
A base pair consists of two complementary DNA nucleotide bases that pair together to form a "rung of the DNA ladder." DNA is made of two linked strands that wind around each other to resemble a twisted ladder — a shape known as a double helix. Each strand has a backbone made of alternating sugar (deoxyribose) and phosphate groups. Attached to each sugar is one of four bases: adenine (A), cytosine (C), guanine (G) or thymine (T). The two strands are held together by hydrogen bonds between pairs of bases: adenine pairs with thymine, and cytosine pairs with guanine.

Bioinformatics
Bioinformatics, as related to genetics and genomics, is a scientific subdiscipline that involves using computer technology to collect, store, analyze and disseminate biological data and information, such as DNA and amino acid sequences or annotations about those sequences. Scientists and clinicians use databases that organize and index such biological information to increase our understanding of health and disease and, in certain cases, as part of medical care.

Birth Defect
A birth defect (also called a congenital defect) is a physical or physiological abnormality present in a baby at birth. Birth defects can be caused by genetic factors, prenatal events during pregnancy or a combination of both. Some birth defects are easy to see (such as an extra or missing finger), while others (such as an enzyme deficiency) are identified through special tests.

BRCA1/BRCA2

BRCA1 and *BRCA2* are the first two genes found to be associated with inherited forms of breast cancer and ovarian cancer. People with mutations in either *BRCA1* or *BRCA2* have a much higher risk for developing breast, ovarian or other types of cancer than those without mutations in the genes. Both BRCA1 and BRCA2 normally act as tumor suppressors, meaning they help to regulate cell division. Most people have two active copies of these genes. When one of the two copies becomes inactive due to an inherited mutation, a person's cells are left with only one copy. If this remaining copy also becomes inactivated, then uncontrolled cell growth results, which leads to breast, ovarian or other types of cancer.

-C-

Cancer

Cancer is a disease in which some of the body's cells grow uncontrollably. There are many different types of cancer, and each begins when a single cell acquires a genomic change (or mutation) that allows the cell to divide and multiply unchecked. Additional mutations can cause the cancer to spread to other sites. Such mutations can be caused by errors during DNA replication or result from DNA damage due to environmental exposures (such as tobacco smoke or the sun's ultraviolet rays). In certain cases, mutations in cancer genes are inherited, which increases a person's risk of developing cancer.

Cancer-Susceptibility Gene

A cancer-susceptibility gene is a gene that, when changed (or mutated), gives an individual an increased risk for developing cancer. Individuals who have inherited mutations in certain cancer-susceptibility genes have a lifetime risk of cancer that is significantly higher than the general population (e.g., *BRCA1/BRCA2*). Individuals in the high-risk category may benefit from more frequent cancer screens. There are also many gene variants associated with a small increase in risk. In some cases, environmental factors may also play a role.

Candidate Gene

The term candidate gene refers to a gene that is believed to be related to a particular trait, such as a disease or a physical attribute. Because of its genomic location or its known function, the gene is suspected to play a role in that trait, thus making it a candidate for additional study.

Carcinogen

A carcinogen is a substance, organism or agent capable of causing cancer. Carcinogens may occur naturally in the environment (such as ultraviolet rays in sunlight and certain viruses) or may be generated by humans (such as automobile exhaust fumes and cigarette smoke). Most carcinogens work by interacting with a cell's DNA to produce mutations.

Carrier

A carrier, as related to genetics, is an individual who "carries" and can pass on to its offspring a genomic variant (allele) associated with a disease (or trait) that is inherited in an autosomal recessive or sex-linked manner, and who does not show symptoms of that disease (or features of that trait). The carrier has inherited the variant allele from one parent and a normal allele from the other parent. Any offspring of carriers is at risk of inheriting a variant allele from their parents, which would result in that child having the disease (or trait).

Carrier Screening

Carrier screening involves testing to see if a person "carries" a genetic variation (allele) associated with a specific disease or trait. A carrier has inherited a normal and a variant allele for a disease- or trait-associated gene, one from each parent. Most typically, carrier screening is performed to look for recessively inherited diseases when the suspected carrier has no symptoms of the disease, but that person's offspring could have the disease if the other parent is a carrier of a harmful variant in the same gene.

Cell-Free DNA Testing

Cell-free DNA testing is a laboratory method that involves analyzing free (i.e., non-cellular) DNA contained within a biological sample, most often to look for genomic variants associated with a hereditary or genetic disorder. For example, prenatal cell-free DNA testing is a non-invasive method used during pregnancy that examines the fetal DNA that is naturally present in the maternal bloodstream. Cell-free DNA testing is also used for the detection and characterization of some cancers and to monitor cancer therapy.

Centimorgan (cM)

A centimorgan (abbreviated cM) is a unit of measure for the frequency of genetic recombination. One centimorgan is equal to a 1% chance that two markers on a chromosome will become separated from one another due to a recombination event during meiosis (which occurs during the formation of egg and sperm cells). On average, one centimorgan corresponds to roughly 1 million base pairs in the human genome.

Central Dogma

The central dogma of molecular biology is a theory stating that genetic information flows only in one direction, from DNA to RNA to protein (or RNA to protein).

Centromere

The centromere appears as a constricted region of a chromosome and plays a key role in helping the cell divide up its DNA during division (mitosis and meiosis). Specifically, it is the region where the cell's spindle fibers attach. Following attachment of the spindle fibers to the centromere, the two identical sister chromatids that make up the replicated chromosome are pulled to opposite sides of the dividing cell, such that the two resulting daughter cells end up with identical DNA.

Chromatid

A chromatid is one of the two identical halves of a chromosome that has been replicated in preparation for cell division. The two "sister" chromatids are joined at a constricted region of the chromosome called the centromere. During cell division, spindle fibers attach to the centromere and pull each of the sister chromatids to opposite sides of the cell. Soon after, the cell divides in two, resulting in daughter cells with identical DNA.

Chromatin

Chromatin refers to a mixture of DNA and proteins that form the chromosomes found in the cells of humans and other higher organisms. Many of the proteins — namely, histones — package the massive amount of DNA in a genome into a highly compact form that can fit in the cell nucleus.

Chromosome

Chromosomes are threadlike structures made of protein and a single molecule of DNA that serve to carry the genomic information from cell to cell. In plants and animals (including humans), chromosomes reside in the nucleus of cells. Humans have 22 pairs of numbered chromosomes (autosomes) and one pair of sex chromosomes (XX or XY), for a total of 46. Each pair contains two chromosomes, one coming from each parent, which means that children inherit half of their chromosomes from their mother and half from their father. Chromosomes can be seen through a microscope when the nucleus dissolves during cell division.

Cloning

Cloning, as it relates to genetics and genomics, involves using scientific methods to make identical, or virtually identical, copies of an organism, cell or DNA sequence. The phrase "molecular cloning" typically refers to isolating and copying a particular DNA segment of interest for further study.

Codominance

Codominance, as it relates to genetics, refers to a type of inheritance in which two versions (alleles) of the same gene are expressed separately to yield different traits in an individual. That is, instead of one trait being dominant over the other, both traits appear, such as in a plant or animal that has more than one pigment color.

Codon

A codon is a DNA or RNA sequence of three nucleotides (a trinucleotide) that forms a unit of genomic information encoding a particular amino acid or signaling the termination of protein synthesis (stop signals). There are 64 different codons: 61 specify amino acids and 3 are used as stop signals.

Complex Disease

A complex disease (or condition), when discussed in the context of genetics, reflects a disorder that results from the contributions of multiple genomic variants and genes in conjunction with significant influences of the physical and social environment. For this reason, complex diseases are also called multifactorial diseases. This stands in contrast to a "simple" genetic disease that is more directly caused by mutations in a single gene. Common examples of complex genetic diseases include heart disease, diabetes and cancer.

Congenital

Congenital refers to a condition or trait that exists at birth. Congenital conditions or traits may be hereditary or result from an action or exposure occurring during pregnancy or at birth, or they may be due to a combination of these factors.

Contig

A contig (as related to genomic studies; derived from the word "contiguous") is a set of DNA segments or sequences that overlap in a way that provides a contiguous representation of a genomic region. For example, a clone contig provides a physical map of a set of cloned segments of DNA across a genomic region, while a sequence contig provides the actual DNA sequence of a genomic region.

Copy DNA (cDNA)

cDNA (short for copy DNA; also called complementary DNA) is synthetic DNA that has been transcribed from a specific mRNA through a reaction using the enzyme reverse transcriptase. While DNA is composed of both coding and non-coding sequences, cDNA contains only coding sequences. Scientists often synthesize and use cDNA as a tool in gene cloning and other research experiments.

Copy Number Variation (CNV)

Copy number variation (abbreviated CNV) refers to a circumstance in which the number of copies of a specific segment of DNA varies among different individuals' genomes. The individual variants may be short or include thousands of bases. These structural differences may have come about through duplications, deletions or other changes and can affect long stretches of DNA. Such regions may or may not contain a gene(s).

CRISPR

CRISPR (short for "clustered regularly interspaced short palindromic repeats") is a technology that research scientists use to selectively modify the DNA of living organisms. CRISPR was adapted for use in the laboratory from naturally occurring genome editing systems found in bacteria.

Crossing Over

Crossing over, as related to genetics and genomics, refers to the exchange of DNA between paired homologous chromosomes (one from each parent) that occurs during the development of egg and sperm cells (meiosis). This process results in new combinations of alleles in the gametes (egg or sperm) formed, which ensures genomic variation in any offspring produced.

Cystic Fibrosis (CF)

Cystic fibrosis (abbreviated CF) is a genetic disorder that causes mucus to build up in certain organs of the body, particularly the lungs and pancreas, resulting in breathing problems, respiratory infections and faulty digestion. Caused by a mutation in a single gene (called *CFTR*), the disorder is inherited as an autosomal recessive trait, meaning that an affected individual inherits two mutated copies of the gene, one from each parent. In the past, CF was almost always fatal in childhood. Today, however, with improvements in screening and treatments, individuals with CF may live into their 30s or 40s, or even longer.

Cytogenetics

Cytogenetics is a branch of biology focused on the study of chromosomes and their inheritance, especially as applied to medical genetics. Chromosomes are microscopic structures containing DNA that reside within the nucleus of a cell. During cell division, these structures become condensed and are visible with a microscope. Special staining techniques can be used to assess the number and structure of a person's chromosomes as part of diagnostic testing. The number and/or structure of chromosomes is known to be altered in certain genetic diseases.

Cytosine

Cytosine (C) is one of the four nucleotide bases in DNA, with the other three being adenine (A), guanine (G) and thymine (T). Within a double-stranded DNA molecule, cytosine bases on one strand pair with guanine bases on the opposite strand. The sequence of the four nucleotide bases encodes DNA's information.

-D-

Data Science
Data science involves the study of large, complex data sets that arise from various types of research projects. With respect to genomic studies, such work requires expertise in quantitative scientific disciplines such as bioinformatics, computational biology and biostatistics.

Deletion
A deletion, as related to genomics, is a type of mutation that involves the loss of one or more nucleotides from a segment of DNA. A deletion can involve the loss of any number of nucleotides, from a single nucleotide to an entire piece of a chromosome.

Deoxyribonucleic Acid (DNA)
Deoxyribonucleic acid (abbreviated DNA) is the molecule that carries genetic information for the development and functioning of an organism. DNA is made of two linked strands that wind around each other to resemble a twisted ladder — a shape known as a double helix. Each strand has a backbone made of alternating sugar (deoxyribose) and phosphate groups. Attached to each sugar is one of four bases: adenine (A), cytosine (C), guanine (G) or thymine (T). The two strands are connected by chemical bonds between the bases: adenine bonds with thymine, and cytosine bonds with guanine. The sequence of the bases along DNA's backbone encodes biological information, such as the instructions for making a protein or RNA molecule.

Diploid
Diploid is a term that refers to the presence of two complete sets of chromosomes in an organism's cells, with each parent contributing a chromosome to each pair. Humans are diploid, and most of the body's cells contain 23 chromosomes pairs. Human gametes (egg and sperm cells), however, contain a single set of chromosomes and are said to be haploid.

DNA Fingerprinting
DNA fingerprinting is a laboratory technique used to determine the probable identity of a person based on the nucleotide sequences of certain regions of human DNA that are unique to individuals. DNA fingerprinting is used in a variety of situations, such as criminal investigations, other forensic purposes and paternity testing. In these situations, one aims to "match" two DNA fingerprints with one another, such as a DNA sample from a known person and one from an unknown person.

DNA Replication
DNA replication is the process by which the genome's DNA is copied in cells. Before a cell divides, it must first copy (or replicate) its entire genome so that each resulting daughter cell ends up with its own complete genome.

DNA Sequencing
DNA sequencing refers to the general laboratory technique for determining the exact sequence of nucleotides, or bases, in a DNA molecule. The sequence of the bases (often referred to by the first letters of their chemical names: A, T, C, and G) encodes the biological information that cells use to develop and operate. Establishing the sequence of DNA is key to understanding the function of genes and other parts of the genome. There are now several different methods available for DNA sequencing, each with its own characteristics, and the development of additional methods represents an active area of genomics research.

Dominant Traits and Alleles
Dominant, as related to genetics, refers to the relationship between an observed trait and the two inherited versions of a gene related to that trait. Individuals inherit two versions of each gene, known as alleles, from each parent. In the case of a dominant trait, only one copy of the dominant allele is required to express the trait. The effect of the other allele (the recessive allele) is masked by the dominant allele. Typically, an individual who carries two copies of a dominant allele exhibits the same trait as those who carry only one copy. This contrasts to a recessive trait, which requires that both alleles be present to express the trait.

Double Helix

Double helix, as related to genomics, is a term used to describe the physical structure of DNA. A DNA molecule is made up of two linked strands that wind around each other to resemble a twisted ladder in a helix-like shape. Each strand has a backbone made of alternating sugar (deoxyribose) and phosphate groups. Attached to each sugar is one of four bases: adenine (A), cytosine (C), guanine (G) or thymine (T). The two strands are connected by chemical bonds between the bases: adenine bonds with thymine, and cytosine bonds with guanine.

Down Syndrome (Trisomy 21)

Down syndrome (also called Trisomy 21) is a genetic condition caused by an error in the process that replicates and then divides up the pairs of chromosomes during cell division, resulting in the inheritance of an extra full or partial copy of chromosome 21 from a parent. This extra chromosomal DNA causes the intellectual disabilities and physical features characteristic of Down syndrome, which vary among individuals.

Duplication

Duplication, as related to genomics, refers to a type of mutation in which one or more copies of a DNA segment (which can be as small as a few bases or as large as a major chromosomal region) is produced. Duplications occur in all organisms. For example, they are especially prominent in plants, although they can also cause genetic diseases in humans. Duplications have been an important mechanism in the evolution of the genomes of humans and other organisms.

-E-

Electrophoresis

Electrophoresis is a laboratory technique used to separate DNA, RNA or protein molecules based on their size and electrical charge. An electric current is used to move the molecules through a gel or other matrix. Pores in the gel or matrix work like a sieve, allowing smaller molecules to move faster than larger molecules. To determine the size of the molecules in a sample, standards of known sizes are separated on the same gel and then compared to the sample.

Environmental Factors

Environmental factors, as related to genetics, refers to exposures to substances (such as pesticides or industrial waste) where we live or work, behaviors (such as smoking or poor diet) or stressful situations (such as racism) that can increase an individual's risk of disease. Genetic studies often take environmental factors into consideration, as these exposures can increase an individual's risk of genetic damage or disease.

Epigenetics

Epigenetics (also sometimes called epigenomics) is a field of study focused on changes in DNA that do not involve alterations to the underlying sequence. The DNA letters and the proteins that interact with DNA can have chemical modifications that change the degrees to which genes are turned on and off. Certain epigenetic modifications may be passed on from parent cell to daughter cell during cell division or from one generation to the next. The collection of all epigenetic changes in a genome is called an epigenome.

Epistasis

Epistasis is a circumstance where the expression of one gene is modified (e.g., masked, inhibited or suppressed) by the expression of one or more other genes.

Eugenics

Eugenics is a discredited belief that selective breeding for certain inherited human traits can improve the "fitness" of future generations. For eugenicists, "fitness" corresponded to a narrow view of humanity and society that developed directly from the ideologies and practices of scientific racism, colonialism, ableism and imperialism.

Evolution

Evolution, as related to genomics, refers to the process by which living organisms change over time through changes in the genome. Such evolutionary changes result from mutations that produce genomic variation, giving rise to individuals whose biological functions or physical traits are altered. Those individuals who are best at adapting to their surroundings leave behind more offspring than less well-adapted individuals. Thus, over successive generations (in some cases spanning millions of years), one species may evolve to take on divergent functions or physical characteristics or may even evolve into a different species.

Exome

An exome is the sequence of all the exons in a genome, reflecting the protein-coding portion of a genome. In humans, the exome is about 1.5% of the genome.

Exon

An exon is a region of the genome that ends up within an mRNA molecule. Some exons are coding, in that they contain information for making a protein, whereas others are non-coding. Genes in the genome consist of exons and introns.

-F-

Family History

A family history, as related to medicine, is a record of the diseases and health conditions of an individual and that person's biological family members, both living and deceased. A family history can help determine whether someone has an increased genetic risk of having or developing certain diseases, disorders or conditions. It is often recorded by drawing a pedigree (a family tree) that illustrates the relationships among individuals.

Fibroblast

A fibroblast is a type of cell that contributes to the formation of connective tissue, a fibrous cellular material that supports and connects other tissues or organs in the body. Fibroblasts secrete collagen proteins that help maintain the structural framework of tissues. They also play an important role in healing wounds. Obtained from a person through a simple skin biopsy, fibroblasts can be grown in the laboratory for use in genetic and other scientific studies of that individual.

First-Degree Relative

A first-degree relative is a family member who shares about half of their genetic information with specific other individuals in their family. First-degree relatives include an individual's parents, siblings and offspring.

Fluorescence In Situ Hybridization (FISH)

Fluorescence in situ hybridization (abbreviated FISH) is a laboratory technique used to detect and locate a specific DNA sequence on a chromosome. In this technique, the full set of chromosomes from an individual is affixed to a glass slide and then exposed to a "probe"—a small piece of purified DNA tagged with a fluorescent dye. The fluorescently labeled probe finds and then binds to its matching sequence within the set of chromosomes. With the use of a special microscope, the chromosome and sub-chromosomal location where the fluorescent probe bound can be seen.

Founder Effect

A founder effect, as related to genetics, refers to the reduction in genomic variability that occurs when a small group of individuals becomes separated from a larger population. Over time, the resulting new subpopulation will have genotypes and physical traits resembling the initial small, separated group, and these may be very different from the original larger population. A founder effect can also explain why certain inherited diseases are found more frequently in some limited population groups. In some cases, a founder effect can play a role in the emergence of new species.

Fragile X Syndrome

Fragile X syndrome is a genetic condition that affects a person's development, in particular their ability to learn and their social behavior. The syndrome results from mutations in a gene on the X chromosome. Because males have only one copy of the X chromosome, they are more likely to show severe symptoms if they inherit the mutated gene compared to females (who have two copies of the X chromosome).

Frameshift Mutation

A frameshift mutation in a gene refers to the insertion or deletion of nucleotide bases in numbers that are not multiples of three. This is important because a cell reads a gene's code in groups of three bases when making a protein. Each of these "triplet codons" corresponds to one of 20 different amino acids used to build a protein. If a mutation disrupts this normal reading frame, then the entire gene sequence following the mutation will be incorrectly read. This can result in the addition of the wrong amino acids to the protein and/or the creation of a codon that stops the protein from growing longer.

Fraternal Twins

Fraternal twins (also called dizygotic twins) result from the fertilization of two separate eggs with two different sperm during the same pregnancy. Fraternal twins may not have the same sex or appearance. They share half their genomes, just like any other siblings. In contrast, identical twins (or monozygotic twins) result from the fertilization of a single egg by a single sperm, with the fertilized egg then splitting into two. As a result, identical twins share the same genomes and are always the same sex.

-G-

Gamete

A gamete is a reproductive cell of an animal or plant. In animals, female gametes are called ova or egg cells, and male gametes are called sperm. Ova and sperm are haploid cells, with each cell carrying only one copy of each chromosome. During fertilization, a sperm and ovum unite to form a new diploid organism.

Gender

Gender is an evolving term, and there are different definitions. In a general sense, gender involves how a person identifies along a broad, fluid spectrum, rather than strictly by the sex assignment they received at birth. Sex refers to the physical differences (e.g., chromosome composition, anatomy, and physiology) among people who are born male, female, or intersex. The term gender may also be used to refer to the social or cultural constructs of "roles" or "norms" typically associated with being masculine or feminine. Gender does not always directly relate to sex assigned at birth.

Gene

The gene is considered the basic unit of inheritance. Genes are passed from parents to offspring and contain the information needed to specify physical and biological traits. Most genes code for specific proteins, or segments of proteins, which have differing functions within the body. Humans have approximately 20,000 protein-coding genes.

Gene Amplification

Gene amplification refers to an increase in the number of copies of a gene in a genome. Cancer cells, for example, sometimes produce multiple copies of a gene(s) in response to signals from other cells or the environment.

Gene–Environment Interaction

Gene–environment interaction refers to the interplay of genes (and, more broadly, genome function) and the physical and social environment. These interactions influence the expression of phenotypes. For example, most human traits and diseases are influenced by how one or more genes interact in complex ways with environmental factors, such as chemicals in the air or water, nutrition, ultraviolet radiation from the sun, and social context.

Gene Expression

Gene expression is the process by which the information encoded in a gene is used to either make RNA molecules that code for proteins or to make non-coding RNA molecules that serve other functions. Gene expression acts as an "on/off switch" to control when and where RNA molecules and proteins are made and as a "volume control" to determine how much of those products are made. The process of gene expression is carefully regulated, changing substantially under different conditions. The RNA and protein products of many genes serve to regulate the expression of other genes.

Gene Mapping

Gene mapping refers to the process of determining the location of genes on chromosomes. Today, the most efficient approach for gene mapping involves sequencing a genome and then using computer programs to analyze the sequence to identify the location of genes.

Gene Pool

A gene pool refers to the combination of all the genes (including alleles) present in a reproducing population or species. A large gene pool has extensive genomic diversity and is better able to withstand environmental challenges. Inbreeding contributes to a smaller gene pool, making populations or species less able to adapt and survive when faced with environmental challenges.

Gene Regulation

Gene regulation is the process used to control the timing, location and amount in which genes are expressed. The process can be complicated and is carried out by a variety of mechanisms, including through regulatory proteins and chemical modification of DNA. Gene regulation is key to the ability of an organism to respond to environmental changes.

Gene Therapy

Gene therapy is a technique that uses a gene(s) to treat, prevent or cure a disease or medical disorder. Often, gene therapy works by adding new copies of a gene that is broken, or by replacing a defective or missing gene in a patient's cells with a healthy version of that gene. Both inherited genetic diseases (e.g., hemophilia and sickle cell disease) and acquired disorders (e.g., leukemia) have been treated with gene therapy.

Genetic Ancestry

Genetic ancestry refers to information about the people that an individual is biologically descended from, including their genetic relationships. Genetic information can be combined with historical information to infer where an individual's distant ancestors lived.

Genetic Code

Genetic code refers to the instructions contained in a gene that tell a cell how to make a specific protein. Each gene's code uses the four nucleotide bases of DNA — adenine (A), cytosine (C), guanine (G) and thymine (T) — in various ways to spell out three-letter "codons" that specify which amino acid is needed at each position within a protein.

Genetic Counseling

Genetic counseling refers to guidance relating to genetic disorders that a specialized healthcare professional (genetic counselor) provides to an individual or family. A genetic counselor might provide information about how a genetic condition could affect an individual or family and/or interpret genetic tests designed to help estimate the risk of a disease. The genetic counselor conveys information to address the concerns of the individual or family, helps them make an informed decision about their medical situation and provides psychological counseling to help them adapt to their condition or risk.

Genetic Discrimination

Genetic discrimination refers to the unequal treatment of individuals based on an aspect of their genetic code (or genome), such as the risk for a genetic disorder. Genetic discrimination can involve such genomic information being used against individuals in a variety of circumstances, such as employment, health or disability insurance, or education. In a healthcare context, it can refer to people being treated in a discriminatory way based on their genetic status rather than by more relevant criteria.

Genetic Drift

Genetic drift is a mechanism of evolution characterized by random fluctuations in the frequency of a particular version of a gene (allele) in a population. Though it primarily affects small, isolated populations, the effects of genetic drift can be strong, sometimes causing traits to become overwhelmingly frequent or to disappear from a population.

Genetic Engineering

Genetic engineering (also called genetic modification) is a process that uses laboratory-based technologies to alter the DNA makeup of an organism. This may involve changing a single base pair (A-T or C-G), deleting a region of DNA or adding a new segment of DNA. For example, genetic engineering may involve adding a gene from one species to an organism from a different species to produce a desired trait. Used in research and industry, genetic engineering has been applied to the production of cancer therapies, brewing yeasts, genetically modified plants and livestock, and more.

Genetic Epidemiology

Genetic epidemiology is a field of science focused on the study of how genetic factors influence human traits, such as human health and disease. In many cases, the interaction of genes with the environment is also measured. Genetic epidemiologists seek to understand the causes, distribution and control of inherited disease in groups and the multifactorial causes of genetic diseases in populations.

Genetic Imprinting

Genomic imprinting is the process by which only one copy of a gene in an individual (either from their mother or their father) is expressed, while the other copy is suppressed. Unlike genomic mutations that can affect the ability of inherited genes to be expressed, genomic imprinting does not affect the DNA sequence itself. Instead, gene expression is silenced by the epigenetic addition of chemical tags to the DNA during egg or sperm formation. Epigenetic tags on imprinted genes usually stay in place for the life of the individual.

Genetic Information Nondiscrimination Act (GINA)

The Genetic Information Nondiscrimination Act (abbreviated GINA) is federal legislation in the United States that protects individuals against discrimination based on their personal genetic information, as it applies to health insurance and employment. These protections are intended to encourage Americans to take advantage of genetic testing as part of their medical care. GINA was signed into law on May 22, 2008.

Genetic Map

A genetic map (also called a linkage map) shows the relative location of genetic markers (reflecting sites of genomic variants) on a chromosome. A genetic map is based on the concept of genetic linkage: the closer two markers are to each other on a chromosome, the greater the probability that they will be inherited together. By studying inheritance patterns, the relative order and location of genetic markers along a chromosome can be established.

Genetic Testing

Genetic testing is the use of a laboratory test to examine an individual's DNA for variations, typically performed in the context of medical care, ancestry studies or forensics. In a medical setting, the results of a genetic test can be used to confirm or rule out a suspected genetic disease. Results may also be used to determine the likelihood of parents passing on a genetic mutation to their offspring. Genetic testing may be performed prenatally or after birth. Genetic testing is also used to study the genomes of tumors in cancer cases.

Genetics

Genetics is the branch of biology concerned with the study of inheritance, including the interplay of genes, DNA variation and their interactions with environmental factors.

Genome

The genome is the entire set of DNA instructions found in a cell. In humans, the genome consists of 23 pairs of chromosomes located in the cell's nucleus, as well as a small chromosome in the cell's mitochondria. A genome contains all the information needed for an individual to develop and function.

Genome-Wide Association Study (GWAS)

A genome-wide association study (abbreviated GWAS) is a research approach used to identify genomic variants that are statistically associated with a risk for a disease or a particular trait. The method involves surveying the genomes of many people, looking for genomic variants that occur more frequently in those with a specific disease or trait compared to those without the disease or trait. Once such genomic variants are identified, they are typically used to search for nearby variants that contribute directly to the disease or trait.

Genomic Medicine

Genomic medicine is a medical discipline that involves using a person's genomic information as part of their clinical care. Other similar terms include individualized medicine, personalized medicine and precision medicine. For some conditions, genomic information can be used to help diagnose disease, predict outcomes and guide treatment.

Genomic Variation

Genomic variation refers to DNA sequence differences among individuals or populations. Some variants influence biological function (such as a mutation that causes a genetic disease), while others have no biological effects.

Genomics

Genomics is a field of biology focused on studying all the DNA of an organism — that is, its genome. Such work includes identifying and characterizing all the genes and functional elements in an organism's genome as well as how they interact.

Genotype

A genotype is a scoring of the type of variant present at a given location (i.e., a locus) in the genome. It can be represented by symbols. For example, BB, Bb, bb could be used to represent a given variant in a gene. Genotypes can also be represented by the actual DNA sequence at a specific location, such as CC, CT, TT. DNA sequencing and other methods can be used to determine the genotypes at millions of locations in a genome in a single experiment. Some genotypes contribute to an individual's observable traits, called the phenotype.

Germ Line
Germ line refers to the sex cells (eggs and sperm) that sexually reproducing organisms use to pass on their genomes from one generation to the next (parents to offspring). Egg and sperm cells are called germ cells, in contrast to the other cells of the body, which are called somatic cells.

Gigabase (Gb)
A gigabase (abbreviated Gb) is a unit of measurement used to help designate the length of DNA. One gigabase is equal to 1 billion bases.

GMO (Genetically Modified Organism)
GMO (short for "genetically modified organism") is a plant, animal or microbe in which one or more changes have been made to the genome, typically using high-tech genetic engineering, in an attempt to alter the characteristics of an organism. Genes can be introduced, enhanced or deleted within a species, across species or even across kingdoms. GMOs may be used for a variety of purposes, such as making human insulin, producing fermented beverages and developing pesticide resistance in crop plants.

Guanine
Guanine (G) is one of the four nucleotide bases in DNA, with the other three being adenine (A), cytosine (C) and thymine (T). Within a double-stranded DNA molecule, guanine bases on one strand pair with cytosine bases on the opposite strand. The sequence of the four nucleotide bases encodes DNA's information.

-H-

Haploid
Haploid refers to the presence of a single set of chromosomes in an organism's cells. Sexually reproducing organisms are diploid (having two sets of chromosomes, one from each parent). In humans, only the egg and sperm cells are haploid.

Haplotype
A haplotype is a physical grouping of genomic variants (or polymorphisms) that tend to be inherited together. A specific haplotype typically reflects a unique combination of variants that reside near each other on a chromosome.

Heterozygous
Heterozygous, as related to genetics, refers to having inherited different versions (alleles) of a genomic marker from each biological parent. Thus, an individual who is heterozygous for a genomic marker has two different versions of that marker. By contrast, an individual who is homozygous for a marker has identical versions of that marker.

Histone
A histone is a protein that provides structural support for a chromosome. Each chromosome contains a long molecule of DNA, which must fit into the cell nucleus. To do that, the DNA wraps around complexes of histone proteins, giving the chromosome a more compact shape. Histones also play a role in the regulation of gene expression.

Homologous Recombination
Homologous recombination is a type of genetic recombination in which nucleotide sequences are exchanged between two similar or identical molecules of DNA. During the formation of egg and sperm cells (meiosis), paired chromosomes from the male and female parents align so that similar DNA sequences can cross over, or be exchanged, from one chromosome to the other. This exchanging of DNA is an important source of the genomic variation seen among offspring.

Homozygous
Homozygous, as related to genetics, refers to having inherited the same versions (alleles) of a genomic marker from each biological parent. Thus, an individual who is homozygous for a genomic marker has two identical versions of that marker. By contrast, an individual who is heterozygous for a marker has two different versions of that marker.

Human Genome Project

The Human Genome Project was a large international, collaborative effort that mapped and sequenced the human genome for the first time. Conducted from 1990 to 2003, the project was historic in its scope and scale as well as its groundbreaking approach for the free release of genomic data well ahead of publication, leading to a new ethos for data sharing in biomedical research.

Human Genome Reference Sequence

A human genome reference sequence is an accepted representation of the human genome sequence that is used by researchers as a standard for comparison to DNA sequences generated in their studies. The scientists responsible for assembling and updating such reference sequences aim to provide the highest-quality, best possible consensus representations of the sequence and structural diversity found in the human genome among populations. The genome reference sequence provides a general framework and is not the DNA sequence of a single person.

Huntington's Disease

Huntington's disease is a rare inherited disorder associated with the progressive loss of brain and muscle function. Symptoms usually develop during middle age and may include uncontrolled movements, loss of intellectual abilities and various emotional and psychiatric symptoms. Huntington's disease is inherited as an autosomal dominant trait, meaning that a single mutated copy of the responsible gene (called *HTT*) is sufficient to cause the disease.

Hybridization

Hybridization, as related to genomics, is the process in which two complementary single-stranded DNA and/or RNA molecules bond together to form a double-stranded molecule. The bonding is dependent on the appropriate base-pairing across the two single-stranded molecules. Hybridization is an important process in various research and clinical laboratory techniques.

-I-

Identical Twins

Identical twins (also called monozygotic twins) result from the fertilization of a single egg by a single sperm, with the fertilized egg then splitting into two. Identical twins share the same genomes and are always of the same sex. In contrast, fraternal (dizygotic) twins result from the fertilization of two separate eggs with two different sperm during the same pregnancy. They share half of their genomes, just like any other siblings. Fraternal twins may not be of the same sex or have similar appearances.

In Situ Hybridization

In situ hybridization is a laboratory technique used to localize a sequence of DNA or RNA in a biological sample. In this technique, a biological sample consisting of tissue sections, cells or chromosomes from an individual is affixed to a glass slide and then exposed to a "probe"—a small piece of single-stranded DNA tagged with a chemical or fluorescent dye. The labeled probe finds and then binds to its matching sequence within the biological sample. The location of the bound probe can then be seen with the use of a microscope.

Inherited

Inherited, as related to genetics, refers to a trait or variants encoded in DNA and passed from parent to offspring during reproduction. Inheritance is determined by the rules of Mendelian genetics.

Insertion

An insertion, as related to genomics, is a type of mutation that involves the addition of one or more nucleotides into a segment of DNA. An insertion can involve the addition of any number of nucleotides, from a single nucleotide to an entire piece of a chromosome.

Intergenic Regions

Intergenic regions are the stretches of DNA located between genes. In humans, intergenic regions are non-protein-coding and comprise a large majority of the genome. Some intergenic DNA is known to regulate the expression of nearby genes.

Intron

An intron is a region that resides within a gene but does not remain in the final mature mRNA molecule following transcription of that gene and does not code for amino acids that make up the protein encoded by that gene. Most protein-coding genes in the human genome consist of exons and introns.

Inversion

An inversion in a chromosome occurs when a segment breaks off and reattaches within the same chromosome, but in reverse orientation. DNA may or may not be lost in the process.

-K-

Karyotype

A karyotype is an individual's complete set of chromosomes. The term also refers to a laboratory-produced image of a person's chromosomes isolated from an individual cell and arranged in numerical order. A karyotype may be used to look for abnormalities in chromosome number or structure.

Kilobase (kb)

A kilobase (abbreviated kb) is a unit of measurement used to help designate the length of DNA or RNA. One kilobase is equal to 1,000 bases.

Knockout

A knockout, as related to genomics, refers to the use of genetic engineering to inactivate or remove one or more specific genes from an organism. Scientists create knockout organisms to study the impact of removing a gene from an organism, which often allows them to then learn something about that gene's function.

-L-

Linkage

Linkage, as related to genetics and genomics, refers to the closeness of genes or other DNA sequences to one another on the same chromosome. The closer two genes or sequences are to each other on a chromosome, the greater the probability that they will be inherited together.

Locus

A locus, as related to genomics, is a physical site or location within a genome (such as a gene or another DNA segment of interest), somewhat like a street address. The plural of locus is loci.

LOD Score

An LOD (short for "logarithm of the odds") score is a statistical estimate of the relative probability that two loci (e.g., a disease-associated gene and another sequence of interest, such as a variant or another gene) are located near each other on a chromosome and are therefore likely to be inherited together.

LOH (Loss of Heterozygosity)

LOH (short for "loss of heterozygosity") refers to a type of mutation that results in the loss of one copy of a segment of DNA (typically containing a gene or group of genes). For most parts of the genome, human cells have two copies of any genomic segment — one from each parent — so in the case of LOH, only one copy would still be present.

Lyonization

Lyonization (also called X-inactivation) refers to the normal phenomenon in which one of the two X chromosomes in every cell of a female individual is inactivated during embryonic development. This inactivation prevents females from having twice as many X chromosome gene products as males, who possess only a single copy of the X chromosome. Lyonization is named after Mary F. Lyon, the British geneticist who discovered the phenomenon.

-M-

Mapping
Mapping refers to the process of determining the relative locations of landmarks or markers (such as genes, variants and other DNA sequences of interest) within a chromosome or genome. Historically, there have been two approaches for mapping: physical mapping, which established maps based on physical distances between landmarks, and genetic mapping, which established maps based on the frequency with which two landmarks are inherited together. Today, the most efficient approach for mapping involves sequencing a genome and then using computer programs to analyze the sequence to identify the locations of landmarks.

Marker
A marker (largely synonymous with the word "landmark" and often referred to as a genomic marker or a genetic marker) is a DNA sequence, typically with a known location in a genome. Markers can reflect random sequences, genomic variants or genes. Markers are used as signposts (or landmarks) in the construction of DNA and genome maps. Markers can also be used to track inheritance of traits or disease risk in families.

Megabase (Mb)
A megabase (abbreviated Mb) is a unit of measurement used to help designate the length of DNA. One megabase is equal to 1 million bases.

Meiosis
Meiosis is a type of cell division in sexually reproducing organisms that reduces the number of chromosomes in gametes (the sex cells, or egg and sperm). In humans, body (or somatic) cells are diploid, containing two sets of chromosomes (one from each parent). To maintain this state, the egg and sperm that unite during fertilization must be haploid, with a single set of chromosomes. During meiosis, each diploid cell undergoes two rounds of division to yield four haploid daughter cells — the gametes.

Mendel, Johann (Gregor)
Gregor Mendel was an Austrian monk in the 19th century who worked out the basic laws of inheritance through experiments with pea plants. In his monastery garden, Mendel performed thousands of crosses with pea plants, discovering how characteristics are passed down from one generation to the next — namely, dominant and recessive traits. Mendel's early experiments provided the basis of modern genetics.

Mendelian Inheritance
Mendelian inheritance refers to certain patterns of how traits are passed from parents to offspring. These general patterns were established by the Austrian monk Gregor Mendel, who performed thousands of experiments with pea plants in the 19th century. Mendel's discoveries of how traits (such as color and shape) are passed down from one generation to the next introduced the concept of dominant and recessive modes of inheritance.

Messenger RNA (mRNA)
Messenger RNA (abbreviated mRNA) is a type of single-stranded RNA involved in protein synthesis. mRNA is made from a DNA template during the process of transcription. The role of mRNA is to carry protein information from the DNA in a cell's nucleus to the cell's cytoplasm (watery interior), where the protein-making machinery reads the mRNA sequence and translates each three-base codon into its corresponding amino acid in a growing protein chain.

Metagenomics
Metagenomics is the study of the structure and function of entire nucleotide sequences isolated and analyzed from all the organisms (typically microbes) in a bulk sample. Metagenomics is often used to study a specific community of microorganisms, such as those residing on human skin, in the soil or in a water sample.

Metaphase
Metaphase is a stage during the process of cell division (mitosis or meiosis). Normally, individual chromosomes are spread out in the cell nucleus. During metaphase, the nucleus dissolves and the cell's chromosomes condense and move together, aligning in the center of the dividing cell. At this stage, the chromosomes are distinguishable when viewed through a microscope. Metaphase chromosomes are used in karyotyping, a laboratory technique for identifying chromosomal abnormalities.

Methylation

Methylation is a chemical modification of DNA and other molecules that may be retained as cells divide to make more cells. When found in DNA, methylation can alter gene expression. In this process, chemical tags called methyl groups attach to a particular location within DNA where they turn a gene on or off, thereby regulating the production of proteins that the gene encodes.

Microarray Technology

Microarray technology is a general laboratory approach that involves binding an array of thousands to millions of known nucleic acid fragments to a solid surface, referred to as a "chip." The chip is then bathed with DNA or RNA isolated from a study sample (such as cells or tissue). Complementary base pairing between the sample and the chip-immobilized fragments produces light through fluorescence that can be detected using a specialized machine. Microarray technology can be used for a variety of purposes in research and clinical studies, such as measuring gene expression and detecting specific DNA sequences (e.g., single nucleotide polymorphisms, or SNPs).

Microbiome

The microbiome is the community of microorganisms (such as fungi, bacteria and viruses) that exists in a particular environment. In humans, the term is often used to describe the microorganisms that live in or on a particular part of the body, such as the skin or gastrointestinal tract. These groups of microorganisms are dynamic and change in response to a host of environmental factors, such as exercise, diet, medication and other exposures.

Microsatellite

Microsatellite, as related to genomics, refers to a short segment of DNA, usually one to six or more base pairs in length, that is repeated multiple times in succession at a particular genomic location. These DNA sequences are typically non-coding. The number of repeated segments within a microsatellite sequence often varies among people, which makes them useful as polymorphic markers for studying inheritance patterns in families or for creating a DNA fingerprint from crime scene samples.

Missense Mutation

A missense mutation is a DNA change that results in different amino acids being encoded at a particular position in the resulting protein. Some missense mutations alter the function of the resulting protein.

Mitochondrial DNA

Mitochondrial DNA is the circular chromosome found inside the cellular organelles called mitochondria. Located in the cytoplasm, mitochondria are the site of the cell's energy production and other metabolic functions. Offspring inherit mitochondria — and as a result mitochondrial DNA — from their mother.

Mitosis

Mitosis is the process by which a cell replicates its chromosomes and then segregates them, producing two identical nuclei in preparation for cell division. Mitosis is generally followed by equal division of the cell's content into two daughter cells that have identical genomes.

Monosomy

Monosomy refers to the condition in which only one chromosome from a pair is present in cells rather than the two copies usually found in diploid cells. When cells have one chromosome from a pair plus a portion of the second chromosome, this is referred to as partial monosomy. Monosomy, or partial monosomy, causes certain human diseases such as Turner syndrome and Cri du chat syndrome.

Mosaicism

Mosaicism refers to the presence of cells in a person that have a different genome from the body's other cells. This difference could be due to a specific genomic variant, for example, or the addition or loss of a chromosome. The condition can stem from a genetic error that occurs after fertilization of an egg, during very early embryo development, or it could occur later in development. Mosaicism can affect any type of cell and does not always cause disease.

Mutagen

A mutagen is a chemical or physical agent capable of inducing changes in DNA called mutations. Examples of mutagens include tobacco products, radioactive substances, x-rays, ultraviolet radiation and a wide variety of chemicals. Exposure to a mutagen can produce DNA mutations that cause or contribute to certain diseases.

Mutation

A mutation is a change in the DNA sequence of an organism. Mutations can result from errors in DNA replication during cell division, exposure to mutagens or a viral infection. Germline mutations (that occur in eggs and sperm) can be passed on to offspring, while somatic mutations (that occur in body cells) are not passed on.

-N-

Nanopore DNA Sequencing

Nanopore DNA sequencing is a laboratory technique for determining the exact sequence of nucleotides, or bases, in a DNA molecule. The sequence of the bases (often referred to by the first letters of their chemical names: A, T, C and G) encodes the biological information that cells use to develop and operate. Nanopore DNA sequencing involves reading the code of single DNA strands as they are threaded through extremely tiny pores (nanopores) embedded within a membrane. As the DNA moves through the pore, it creates signals that can be converted to read each base. This approach offers a low-cost, rapid process for studying long stretches of DNA.

Nanotechnology

Nanotechnology (often shortened to nanotech) is the understanding and use of matter on an atomic and molecular scale for industrial purposes. Manipulating matter at nanoscale — between approximately 1 and 100 nanometers — holds potential for novel applications in many fields, including genomics, engineering, computer science and medicine.

Newborn Genetic Screening

Newborn screening is a set of laboratory tests performed on newborn babies to detect a set of known genetic diseases. Typically, this testing is performed on a blood sample obtained from a heel prick when the baby is two or three days old. In the United States, newborn screening is mandatory for a defined set of genetic diseases, although the exact set differs from state to state. Newborn screening tests focus on conditions for which early diagnosis is important to treating or preventing disease.

Next-Generation DNA Sequencing

DNA sequencing establishes the order of the bases that make up DNA. Next-generation DNA sequencing (abbreviated NGS) refers to the use of technologies for sequencing DNA that became available shortly after the completion of the Human Genome Project (which relied on the first-generation method of Sanger sequencing). Faster and cheaper than their predecessors, NGS technologies can sequence an entire human genome in a single day and for less than $1,000.

Non-Coding DNA

Non-coding DNA corresponds to the portions of an organism's genome that do not code for amino acids, the building blocks of proteins. Some non-coding DNA sequences are known to serve functional roles, such as in the regulation of gene expression, while other areas of non-coding DNA have no known function.

Nonsense Mutation

A nonsense mutation occurs in DNA when a sequence change gives rise to a stop codon rather than a codon specifying an amino acid. The presence of the new stop codon results in the production of a shortened protein that is likely non-functional.

Northern Blot

Northern blot is a laboratory analysis method used to study RNA. Specifically, purified RNA fragments from a biological sample (such as blood or tissue) are separated by using an electric current to move them through a sieve-like gel or matrix, which allows smaller fragments to move faster than larger fragments. The RNA fragments are transferred out of the gel or matrix onto a solid membrane, which is then exposed to a DNA probe labeled with a radioactive, fluorescent or chemical tag. The tag allows any RNA fragments containing complementary sequences with the DNA probe sequence to be visualized within the Northern blot.

Nuclear Membrane

The nuclear membrane is a double layer that encloses the cell's nucleus, where the chromosomes reside. The nuclear membrane serves to separate the chromosomes from the cell's cytoplasm and other contents. An array of small holes or pores in the nuclear membrane permits the selective passage of certain materials, such as nucleic acids and proteins, between the nucleus and cytoplasm.

Nucleic Acids

Nucleic acids are large biomolecules that play essential roles in all cells and viruses. A major function of nucleic acids involves the storage and expression of genomic information. Deoxyribonucleic acid, or DNA, encodes the information cells need to make proteins. A related type of nucleic acid, called ribonucleic acid (RNA), comes in different molecular forms that play multiple cellular roles, including protein synthesis.

Nucleolus

The nucleolus is a spherical structure found in the cell's nucleus whose primary function is to produce and assemble the cell's ribosomes. The nucleolus is also where ribosomal RNA genes are transcribed. Once assembled, ribosomes are transported to the cell cytoplasm, where they serve as the sites for protein synthesis.

Nucleopore

A nucleopore is one of a series of openings found in the cell's nuclear membrane. Nucleopores serve as channels for the selective transport of nucleic acids and proteins into and out of the cell nucleus.

Nucleosome

A nucleosome is the basic repeating subunit of chromatin packaged inside the cell's nucleus. In humans, about six feet of DNA must be packaged into a nucleus with a diameter less than a human hair, and nucleosomes play a key role in that process. A single nucleosome consists of about 150 base pairs of DNA sequence wrapped around a core of histone proteins. In forming a chromosome, the nucleosomes repeatedly fold in on themselves to tighten and condense the packaged DNA.

Nucleotide

A nucleotide is the basic building block of nucleic acids (RNA and DNA). A nucleotide consists of a sugar molecule (either ribose in RNA or deoxyribose in DNA) attached to a phosphate group and a nitrogen-containing base. The bases used in DNA are adenine (A), cytosine (C), guanine (G) and thymine (T). In RNA, the base uracil (U) takes the place of thymine. DNA and RNA molecules are polymers made up of long chains of nucleotides.

Nucleus

A nucleus, as related to genomics, is the membrane-enclosed organelle within a cell that contains the chromosomes. An array of holes, or pores, in the nuclear membrane allows for the selective passage of certain molecules (such as proteins and nucleic acids) into and out of the nucleus.

-O-

Oncogene

An oncogene is a mutated gene that has the potential to cause cancer. Before an oncogene becomes mutated, it is called a proto-oncogene, and it plays a role in regulating normal cell division. Cancer can arise when a proto-oncogene is mutated, changing it into an oncogene and causing the cell to divide and multiply uncontrollably. Some oncogenes work like an accelerator pedal in a car, pushing a cell to divide again and again. Others work like a faulty brake in a car parked on a hill, also causing the cell to divide unchecked.

Open Reading Frame

An open reading frame, as related to genomics, is a portion of a DNA sequence that does not include a stop codon (which functions as a stop signal). A codon is a DNA or RNA sequence of three nucleotides (a trinucleotide) that forms a unit of genomic information encoding a particular amino acid or signaling the termination of protein synthesis (stop codon). There are 64 different codons: 61 specify amino acids and 3 are used as stop codons. A long open reading frame is often part of a gene (that is, a sequence directly coding for a protein).

Pedigree

A pedigree, as related to genetics, is a chart that diagrams the inheritance of a trait or health condition through generations of a family. The pedigree particularly shows the relationships among family members and, when the information is available, indicates which individuals have a trait(s) of interest.

Peptide

A peptide is a short chain of amino acids (typically 2 to 50) linked by chemical bonds (called peptide bonds). A longer chain of linked amino acids (51 or more) is a polypeptide. The proteins manufactured inside cells are made from one or more polypeptides.

Pharmacogenomics

Pharmacogenomics (also called pharmacogenetics) is a component of genomic medicine that involves using a patient's genomic information to tailor the selection of drugs used in their medical management. In this way, pharmacogenomics aims to provide a more individualized (or precise) approach to the use of available medication in treating patients.

Phenotype

Phenotype refers to an individual's observable traits, such as height, eye color and blood type. A person's phenotype is determined by both their genomic makeup (genotype) and environmental factors.

Physical Map

A physical map, as related to genomics, is a graphical representation of physical locations of landmarks or markers (such as genes, variants and other DNA sequences of interest) within a chromosome or genome. A complete genome sequence is one type of physical map. Physical maps are used to identify genes or other sequences believed to play a role in health conditions or diseases. They are also valuable in providing an organizational framework for generating complete sequences of genomes.

Plasmid

A plasmid is a small circular DNA molecule found in bacteria and some other microscopic organisms. Plasmids are physically separate from chromosomal DNA and replicate independently. They typically have a small number of genes — notably, some associated with antibiotic resistance — and can be passed from one cell to another. Scientists use recombinant DNA methods to splice genes that they want to study into a plasmid. When the plasmid copies itself, it also makes copies of the inserted gene.

Point Mutation

A point mutation occurs in a genome when a single base pair is added, deleted or changed. While most point mutations are benign, they can also have various functional consequences, including changes in gene expression or alterations in encoded proteins.

Polydactyly

Polydactyly is a condition in which a person has more than the normal number of fingers or toes. It can occur in association with other physical anomalies or intellectual impairment, or it may occur as an isolated birth defect. Polydactyly can either be inherited or it can arise sporadically in an individual.

Polygenic Risk Score (PRS)

A polygenic risk score (abbreviated PRS) uses genomic information alone to assess a person's chances of having or developing a particular medical condition. A person's PRS is a statistical calculation based on the presence or absence of multiple genomic variants, without taking environmental or other factors into account.

Polygenic Trait

A polygenic trait is a characteristic, such as height or skin color, that is influenced by two or more genes. Because multiple genes are involved, polygenic traits do not follow the patterns of Mendelian inheritance. Many polygenic traits are also influenced by the environment and are called multifactorial.

Polymerase Chain Reaction (PCR)
Polymerase chain reaction (abbreviated PCR) is a laboratory technique for rapidly producing (amplifying) millions to billions of copies of a specific segment of DNA, which can then be studied in greater detail. PCR involves using short synthetic DNA fragments called primers to select a segment of the genome to be amplified, and then multiple rounds of DNA synthesis to amplify that segment.

Polymorphism
Polymorphism, as related to genomics, refers to the presence of two or more variant forms of a specific DNA sequence that can occur among different individuals or populations. The most common type of polymorphism involves variation at a single nucleotide (also called a single-nucleotide polymorphism, or SNP). Other polymorphisms can be much larger, involving longer stretches of DNA.

Population Genomics
Population genomics is the large-scale application of genomic technologies to study populations of individuals. For example, population genomics research can be used to study human ancestry, migrations and health.

Positional Cloning
Positional cloning is a laboratory approach used to locate the position of a disease-associated gene on a chromosome. Such a strategy can succeed even when nothing is known about the role of the gene's encoded protein in the disease. The technique typically relies on the use of known polymorphic markers whose inheritance can be traced through various members of families affected by the disease.

Precision Medicine
Precision medicine (generally considered analogous to personalized medicine or individualized medicine) is an innovative approach that uses information about an individual's genomic, environmental and lifestyle information to guide decisions related to their medical management. The goal of precision medicine is to provide more a precise approach for the prevention, diagnosis and treatment of disease.

Primer
A primer, as related to genomics, is a short single-stranded DNA fragment used in certain laboratory techniques, such as the polymerase chain reaction (PCR). In the PCR method, a pair of primers hybridizes with the sample DNA and defines the region that will be amplified, resulting in millions and millions of copies in a very short timeframe. Primers are also used in DNA sequencing and other experimental processes.

Proband
A proband is an individual who is affected by a genetic condition or who is concerned they are at risk. Usually, the proband is the first person in a family who brings the concern of a genetic disorder to the attention of healthcare professionals.

Promoter
A promoter, as related to genomics, is a region of DNA upstream of a gene where relevant proteins (such as RNA polymerase and transcription factors) bind to initiate transcription of that gene. The resulting transcription produces an RNA molecule (such as mRNA).

Protein
Proteins are large, complex molecules that play many important roles in the body. They are critical to most of the work done by cells and are required for the structure, function and regulation of the body's tissues and organs. A protein is made up of one or more long, folded chains of amino acids (each called a polypeptide), whose sequences are determined by the DNA sequence of the protein-encoding gene.

Pseudogene
A pseudogene is a segment of DNA that structurally resembles a gene but is not capable of coding for a protein. Pseudogenes are most often derived from genes that have lost their protein-coding ability due to accumulated mutations that have occurred over the course of evolution.

-R-

Race

Race is a social construct used to group people. Race was constructed as a hierarchal human-grouping system, generating racial classifications to identify, distinguish and marginalize some groups across nations, regions and the world. Race divides human populations into groups often based on physical appearance, social factors and cultural backgrounds.

Recessive Traits and Alleles

Recessive, as related to genetics, refers to the relationship between an observed trait and the two inherited versions of a gene related to that trait. Individuals inherit two versions of each gene, known as alleles, from each parent. In the case of a recessive trait, the alleles of the trait-causing gene are the same, and both (recessive) alleles must be present to express the trait. A recessive allele does not produce a trait at all when only one copy is present. This contrasts to a dominant trait, which requires that only one of the two alleles be present to express the trait.

Recombinant DNA Technology

Recombinant DNA technology involves using enzymes and various laboratory techniques to manipulate and isolate DNA segments of interest. This method can be used to combine (or splice) DNA from different species or to create genes with new functions. The resulting copies are often referred to as recombinant DNA. Such work typically involves propagating the recombinant DNA in a bacterial or yeast cell, whose cellular machinery copies the engineered DNA along with its own.

Repressor

A repressor, as related to genomics, is a protein that inhibits the expression of one or more genes. The repressor protein works by binding to the promoter region of the gene(s), which prevents the production of messenger RNA (mRNA). Repressor proteins are essential for the regulation of gene expression in cells.

Restriction Enzyme

A restriction enzyme is a protein isolated from bacteria that cleaves DNA sequences at sequence-specific sites, producing DNA fragments with a known sequence at each end. The use of restriction enzymes is critical to certain laboratory methods, including recombinant DNA technology and genetic engineering.

Restriction Fragment Length Polymorphism (RFLP)

Restriction fragment length polymorphism (abbreviated RFLP) refers to differences (or variations) among people in their DNA sequences at sites recognized by restriction enzymes. Such variation results in different sized (or length) DNA fragments produced by digesting the DNA with a restriction enzyme. RFLPs can be used as genetic markers, which are often used to follow the inheritance of DNA through families.

Retrovirus

A retrovirus is a virus that uses RNA as its genomic material. Upon infection with a retrovirus, a cell converts the retroviral RNA into DNA, which in turn is inserted into the DNA of the host cell. The cell then produces more retroviruses, which infect other cells. Many retroviruses are associated with diseases, including AIDS and some forms of cancer.

Ribonucleic Acid (RNA)

Ribonucleic acid (abbreviated RNA) is a nucleic acid present in all living cells that has structural similarities to DNA. Unlike DNA, however, RNA is most often single-stranded. An RNA molecule has a backbone made of alternating phosphate groups and the sugar ribose, rather than the deoxyribose found in DNA. Attached to each sugar is one of four bases: adenine (A), uracil (U), cytosine (C) or guanine (G). Different types of RNA exist in cells: messenger RNA (mRNA), ribosomal RNA (rRNA) and transfer RNA (tRNA). In addition, some RNAs are involved in regulating gene expression. Certain viruses use RNA as their genomic material.

Ribosome

A ribosome is an intercellular structure made of both RNA and protein, and it is the site of protein synthesis in the cell. The ribosome reads the messenger RNA (mRNA) sequence and translates that genetic code into a specified string of amino acids, which grow into long chains that fold to form proteins.

Risk

Risk, as related to genetics, refers to the probability that an individual will be affected by a particular heritable or genetic disorder. Both a person's genome and environmental exposures can influence risk. An individual's risk may be higher because they inherit a genetic variant (or allele) in one gene or a combination of many variants in different genes that increases susceptibility to or overtly causes a disorder. Other individuals may be at higher risk because they have been exposed to one or more environmental factors that promote the development of a certain disorder.

-S-

Scientific Racism

Scientific racism is a historical pattern of ideologies that generate and perpetuate pseudoscientific racist beliefs that lead to racial bias and discrimination in science and research. Leading scientists across scientific institutions in the 19th and early 20th centuries were proponents of such ideologies. By the mid-20th century, pseudoscientific racist beliefs were widely disproven. However, evidence shows that scientific racism persists in science and research.

Secondary Genomic Finding

A secondary genomic finding refers to a genomic variant, found through the analysis of a person's genome, that is of potential medical value yet is unrelated to the initial reason for examining the person's genome. In certain cases, a secondary genomic finding might offer clinicians the chance to identify a previously unrecognized risk for disease that could change the medical management of that patient and potentially prevent or more effectively treat the disease.

Sex Chromosome

A sex chromosome is a type of chromosome involved in sex determination. Humans and most other mammals have two sex chromosomes, X and Y, that in combination determine the sex of an individual. Females have two X chromosomes in their cells, while males have one X and one Y.

Sex-Linked

Sex-linked, as related to genetics, refers to characteristics (or traits) that are influenced by genes carried on the sex chromosomes. In humans, the term often refers to traits or disorders influenced by genes on the X chromosome, as it contains many more genes than the smaller Y chromosome. Males, who have only a single copy of the X chromosome, are more likely to be affected by a sex-linked disorder than females, who have two copies. In females, the presence of a second, non-mutated copy may cause different, milder, or no symptoms of a sex-linked disorder.

Shotgun Sequencing

Shotgun sequencing is a laboratory technique for determining the DNA sequence of an organism's genome. The method involves randomly breaking up the genome into small DNA fragments that are sequenced individually. A computer program looks for overlaps in the DNA sequences, using them to reassemble the fragments in their correct order to reconstitute the genome.

Single Nucleotide Polymorphism (SNP)

A single nucleotide polymorphism (abbreviated SNP, pronounced snip) is a genomic variant at a single base position in the DNA. Scientists study if and how SNPs in a genome influence health, disease, drug response and other traits.

Somatic Cells

Somatic cells are the cells in the body other than sperm and egg cells (which are called germ cells). In humans, somatic cells are diploid, meaning they contain two sets of chromosomes, one inherited from each parent. DNA mutations in somatic cells can affect an individual, but they cannot be passed on to their offspring.

Southern Blot

Southern blot analysis is a laboratory method used to study DNA. Specifically, purified DNA from a biological sample (such as blood or tissue) is digested with a restriction enzyme(s), and the resulting DNA fragments are separated by using an electric current to move them through a sieve-like gel or matrix, which allows smaller fragments move faster than larger fragments. The DNA fragments are transferred out of the gel or matrix onto a solid membrane, which is then exposed to a DNA probe labeled with a radioactive, fluorescent or chemical tag. The tag allows any DNA fragments containing complementary sequences with the DNA probe sequence to be visualized within the Southern blot. The method is named for its creator, British molecular biologist Edwin Southern.

Stop Codon

A stop codon is a sequence of three nucleotides (a trinucleotide) in DNA or messenger RNA (mRNA) that signals a halt to protein synthesis in the cell. There are 64 different trinucleotide codons: 61 specify amino acids and 3 are stop codons (i.e., UAA, UAG and UGA).

Substitution

Substitution, as related to genomics, is a type of mutation in which one nucleotide is replaced by a different nucleotide. The term can also refer to the replacement of one amino acid in a protein with a different amino acid.

Susceptibility

Susceptibility, as related to genetics, refers to the state of being predisposed to, or sensitive to, developing a certain disease. An individual's disease susceptibility is influenced by a combination of genetic and environmental factors.

Syndrome

A syndrome, as related to genetics, is a group of traits or conditions that tend to occur together and characterize a recognizable disease. Some syndromes have a genetic cause.

-T-

Tandem Repeat

A tandem repeat is a sequence of two or more DNA bases that is repeated numerous times in a head-to-tail manner on a chromosome. Tandem repeats are generally present in non-coding DNA. In some cases, tandem repeats can serve as genetic markers to track inheritance in families. They can also be useful for DNA fingerprinting in forensic studies.

Telomere

A telomere is a region of repetitive DNA sequences at the end of a chromosome. Telomeres protect the ends of chromosomes from becoming frayed or tangled. Each time a cell divides, the telomeres become slightly shorter. Eventually, they become so short that the cell can no longer divide successfully, and the cell dies.

Thymine

Thymine (T) is one of the four nucleotide bases in DNA, with the other three being adenine (A), cytosine (C) and guanine (G). Within a double-stranded DNA molecule, thymine bases on one strand pair with adenine bases on the opposite strand. The sequence of the four nucleotide bases encodes DNA's information.

Trait

A trait, as related to genetics, is a specific characteristic of an individual. Traits can be determined by genes, environmental factors or by a combination of both. Traits can be qualitative (such as eye color) or quantitative (such as height or blood pressure). A given trait is part of an individual's overall phenotype.

Transcription

Transcription, as related to genomics, is the process of making an RNA copy of a gene's DNA sequence. This copy, called messenger RNA (mRNA), carries the gene's protein information encoded in DNA. In humans and other complex organisms, mRNA moves from the cell nucleus to the cell cytoplasm (watery interior), where it is used for synthesizing the encoded protein.

Transfer RNA (tRNA)

Transfer RNA (abbreviated tRNA) is a small RNA molecule that plays a key role in protein synthesis. Transfer RNA serves as a link (or adaptor) between the messenger RNA (mRNA) molecule and the growing chain of amino acids that make up a protein. Each time an amino acid is added to the chain, a specific tRNA pairs with its complementary sequence on the mRNA molecule, ensuring that the appropriate amino acid is inserted into the protein being synthesized.

Transgenic

Transgenic refers to an organism or cell whose genome has been altered by the introduction of one or more foreign DNA sequences from another species by artificial means. Transgenic organisms are generated in the laboratory for research purposes.

Translation

Translation, as related to genomics, is the process through which information encoded in messenger RNA (mRNA) directs the addition of amino acids during protein synthesis. Translation takes place on ribosomes in the cell cytoplasm, where mRNA is read and translated into the string of amino acid chains that make up the synthesized protein.

Translocation

A translocation, as related to genetics, occurs when a chromosome breaks and the (typically two) fragmented pieces re-attach to different chromosomes. The detection of chromosomal translocations can be important for the diagnosis of certain genetic diseases and disorders.

Tumor Suppressor Gene

A tumor suppressor gene encodes a protein that acts to regulate cell division, keeping it in check. When a tumor suppressor gene is inactivated by a mutation, the protein it encodes is not produced or does not function properly, and as a result, uncontrolled cell division may occur. Such mutations may contribute to the development of a cancer.

-U-

Uracil

Uracil (U) is one of the four nucleotide bases in RNA, with the other three being adenine (A), cytosine (C) and guanine (G). In RNA, uracil pairs with adenine. In a DNA molecule, the nucleotide thymine (T) is used in place of uracil.

-V-

Variant of Uncertain Significance (VUS)

When analysis of a patient's genome identifies a variant, but it is unclear whether that variant is actually connected to a health condition, the finding is called a variant of uncertain significance (abbreviated VUS). In many cases, these variants are so rare in the population that little information is available about them. Typically, more information is required to determine if the variant is disease related. Such information may include more extensive population data, functional studies, and tracing the variant in other family members who have or do not have the same health condition.

Vector

A vector, as related to molecular biology, is a DNA molecule (often plasmid or virus) that is used as a vehicle to carry a particular DNA segment into a host cell as part of a cloning or recombinant DNA technique. The vector typically assists in replicating and/or expressing the inserted DNA sequence inside the host cell.

Virus

A virus is an infectious microbe consisting of a segment of nucleic acid (either DNA or RNA) surrounded by a protein coat. A virus cannot replicate alone; instead, it must infect cells and use components of the host cell to make copies of itself. Often, a virus ends up killing the host cell in the process, causing damage to the host organism. Well-known examples of viruses causing human disease include AIDS, COVID-19, measles and smallpox.

-X-

X Chromosome

The X chromosome is one of the two sex chromosomes that are involved in sex determination. Humans and most other mammals have two sex chromosomes (X and Y) that in combination determine the sex of an individual. Females have two X chromosomes in their cells, while males have one X and one Y.

X-Linked

X-linked, as related to genetics, refers to characteristics or traits that are influenced by genes on the X chromosome. Humans and most other mammals have two sex chromosomes, X and Y. Females have two X chromosomes in their cells, while males have one X and one Y. In the case of an X-linked disease, it is usually males that are affected because they have a single copy of the X chromosome that carries the disease-causing mutation. In females, the presence of a second, non-mutated copy may cause different, milder, or no symptoms of a sex-linked disorder.

-Y-

Y Chromosome

The Y chromosome is one of the two sex chromosomes that are involved in sex determination. Humans and most other mammals have two sex chromosomes (X and Y) that in combination determine the sex of an individual. Females have two X chromosomes in their cells, while males have one X and one Y.

National Human Genome Research Institute
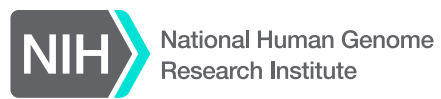
genome.gov

April 2022

The **Forefront**
of **Genomics**