GLOSSARY OF GENOMICS AND BIOINFORMATICS TERMS*

A

Accession number (GenBank)

The accession number is the unique identifier assigned to the entire sequence record when the record is submitted to GenBank. The GenBank accession number is a combination of letters and numbers that are usually in the format of one letter followed by five digits (e.g., M12345) or two letters followed by six digits (e.g., AC123456). The accession number for a particular record will not change even if the author submits a request to change some of the information in the record. Take note that an accession number is a unique identifier for a complete sequence record, while a Sequence Identifier, such as a Version, GI, or ProteinID, is an identification number assigned just to the sequence data. The NCBI Entrez System is searchable by accession number using the Accession [ACCN] search field.

Accession number (RefSeq)

This accession number is the unique identification number for a complete RefSeq sequence record. RefSeq accession numbers are written in the following format: two letters followed by an underscore and six digits (e.g., NT_123456). The first two letters of the RefSeq accession number indicate the type of sequence included in the record as described below:

* NT_123456 constructed genomic contigs
* NM_123456 mRNAs (actually the cDNA sequences constructed from mRNA)
* NP_123456 proteins
* NC_123456 chromosomes

Adenine (A)

A nitrogenous base, one member of the base pair AT (adenine-thymine).

See also: base pair, nucleotide

*This glossary was partially compiled based on publicly available glossary of genomics terms at the Human Genome website: http://www.ornl.gov/sci/techresources/Human_Genome

Allele

Alternative form of a genetic locus; a single allele for each locus is inherited from each parent (e.g., at a locus for eye color the allele might result in blue or brown eyes).

Alternative splicing

Different ways of combining a gene's exons to make variants of the complete protein.

Amino acid

Any of a class of 20 molecules that are combined to form proteins in living things. The sequence of amino acids in a protein and hence protein function are determined by the genetic code.

Amplification

An increase in the number of copies of a specific DNA fragment; can be in vivo or in vitro.

See also: cloning, polymerase chain reaction

Annotation

Adding pertinent information such as gene coded for, amino acid sequence, or other commentary to the database entry of raw sequence of DNA bases.

Antisense

Nucleic acid that has a sequence exactly opposite to an mRNA molecule; binds to the mRNA molecule to prevent a protein from being made.

Apoptosis

Programmed cell death, the body's normal method of disposing of damaged, unwanted, or unneeded cells.

Arrayed library

Individual primary recombinant clones (hosted in phage, cosmid, YAC, or other vector) that are placed in two-dimensional arrays in microtiter dishes. Each primary clone can be identified by the identity of the plate and the clone location (row and column) on that plate. Arrayed libraries of clones can be used for many applications, including screening for a specific gene or genomic region of interest.

Assembly
Putting sequenced fragments of DNA into their correct chromosomal positions.

Autosome
A chromosome not involved in sex determination. The diploid human genome consists of a total of 46 chromosomes: 22 pairs of autosomes, and 1 pair of sex chromosomes (the X and Y chromosomes).

B

Bacterial artificial chromosome (BAC)
A vector used to clone DNA fragments (100- to 300-kb insert size; average, 150 kb) in Escherichia coli cells. Based on naturally occurring F-factor plasmid found in the bacterium E. coli.

Base
One of the molecules that form DNA and RNA molecules.

Base pair (bp)
Two nitrogenous bases (adenine and thymine or guanine and cytosine) held together by weak bonds. Two strands of DNA are held together in the shape of a double helix by the bonds between base pairs.

Base sequence
The order of nucleotide bases in a DNA molecule; determines structure of proteins encoded by that DNA.

Base sequence analysis
A method, sometimes automated, for determining the base sequence.

Bioinformatics
The science of managing and analyzing biological data using advanced computing techniques. Especially important in analyzing genomic and transciptomics research data.

BLAST
A computer program that identifies homologous (similar) genes in different organisms, such as human, fruit fly, or nematode.

C

Cancer
Diseases in which abnormal cells divide and grow unchecked. Cancer can spread from its original site to other parts of the body and can be fatal.

Candidate gene
A gene located in a chromosome region suspected of being involved in a disease.

Capillary array
Gel-filled silica capillaries used to separate fragments for DNA sequencing. The small diameter of the capillaries permit the application of higher electric fields, providing high speed, high throughput separations that are significantly faster than traditional slab gels.

Carcinogen
Something which causes cancer to occur by causing changes in a cell's DNA.

cDNA or complementary DNA
DNA that is synthesized in the laboratory from a messenger RNA template

cDNA library
A collection of DNA sequences that code for genes. The sequences are generated in the laboratory from mRNA sequences.

CDS
The coding sequence or the portion of a nucleotide sequence that makes up the triplet codons that actually code for amino acids.

Cell
The basic unit of any living organism that carries on the biochemical processes of life.

Chromosomal deletion
The loss of part of a chromosome's DNA.

Chromosomal inversion

Chromosome segments that have been turned 180 degrees. The gene sequence for the segment is reversed with respect to the rest of the chromosome.

Chromosome

The self-replicating genetic structure of cells containing the cellular DNA that bears in its nucleotide sequence the linear array of genes. In prokaryotes, chromosomal DNA is circular, and the entire genome is carried on one chromosome. Eukaryotic genomes consist of a number of chromosomes whose DNA is associated with different kinds of proteins.

Chromosome region p

A designation for the short arm of a chromosome.

Chromosome region q

A designation for the long arm of a chromosome.

Clone

An exact copy made of biological material such as a DNA segment (e.g., a gene or other region), a whole cell, or a complete organism.

Clone bank

See: genomic library

Cloning

Using specialized DNA technology to produce multiple, exact copies of a single gene or other segment of DNA to obtain enough material for further study. This process, used by researchers in the Human Genome Project, is referred to as cloning DNA. The resulting cloned (copied) collections of DNA molecules are called clone libraries. A second type of cloning exploits the natural process of cell division to make many copies of an entire cell. The genetic makeup of these cloned cells, called a cell line, is identical to the original cell. A third type of cloning produces complete, genetically identical animals such as the famous Scottish sheep, Dolly.

Cloning vector

DNA molecule originating from a virus, a plasmid, or the cell of a higher organism into which another DNA fragment of appropriate size can be integrated without loss of the vector's capacity for self-replication; vectors introduce foreign

DNA into host cells, where the DNA can be reproduced in large quantities. Examples are plasmids, cosmids, and yeast artificial chromosomes; vectors are often recombinant molecules containing DNA sequences from several sources.

Codon
See: genetic code

Comparative genomics
The study of human genetics by comparisons with model organisms such as mice, the fruit fly, and the bacterium E. coli.

Complementary DNA (cDNA)
DNA that is synthesized in the laboratory from a messenger RNA template.

Complementary sequence
Nucleic acid base sequence that can form a double-stranded structure with another DNA fragment by following base-pairing rules (A pairs with T and C with G). The complementary sequence to GTAC for example, is CATG.

Computational biology
See: bioinformatics

Conserved sequence
A base sequence in a DNA molecule (or an amino acid sequence in a protein) that has remained essentially unchanged throughout evolution.

Constitutive ablation
Gene expression that results in cell death.

Contig
Group of cloned (copied) pieces of DNA representing overlapping regions of a particular chromosome.

Contig map
A map depicting the relative order of a linked library of overlapping clones representing a complete chromosomal segment.

Cosmid

Artificially constructed cloning vector containing the cos gene of phage lambda. Cosmids can be packaged in lambda phage particles for infection into E. coli; this permits cloning of larger DNA fragments (up to 45kb) than can be introduced into bacterial hosts in plasmid vectors.

Crossing over

The breaking during meiosis of one maternal and one paternal chromosome, the exchange of corresponding sections of DNA, and the rejoining of the chromosomes. This process can result in an exchange of alleles between chromosomes.

Cytogenetics

The study of the physical appearance of chromosomes.

Cytological band

An area of the chromosome that stains differently from areas around it.

Cytological map

A type of chromosome map whereby genes are located on the basis of cytological findings obtained with the aid of chromosome mutations.

Cytosine (C)

A nitrogenous base, one member of the base pair GC (guanine and cytosine) in DNA.

D

Deletion

A loss of part of the DNA from a chromosome; can lead to a disease or abnormality.

Deletion map

A description of a specific chromosome that uses defined mutations --specific deleted areas in the genome-- as 'biochemical signposts,' or markers for specific areas.

Deoxyribonucleotide

See: nucleotide

Deoxyribose
A type of sugar that is one component of DNA (deoxyribonucleic acid).

Diploid
A full set of genetic material consisting of paired chromosomes, one from each parental set. Most animal cells except the gametes have a diploid set of chromosomes. The diploid human genome has 46 chromosomes.

Directed mutagenesis
Alteration of DNA at a specific site and its reinsertion into an organism to study any effects of the change.

Directed sequencing
Successively sequencing DNA from adjacent stretches of chromosome.

Disease-associated genes
Alleles carrying particular DNA sequences associated with the presence of disease.

DNA (deoxyribonucleic acid)
The molecule that encodes genetic information. DNA is a double-stranded molecule held together by weak bonds between base pairs of nucleotides. The four nucleotides in DNA contain the bases adenine (A), guanine (G), cytosine (C), and thymine (T). In nature, base pairs form only between A and T and between G and C; thus the base sequence of each single strand can be deduced from that of its partner.

DNA probe
See: probe

DNA repair genes
Genes encoding proteins that correct errors in DNA sequencing.

DNA replication
The use of existing DNA as a template for the synthesis of new DNA strands. In humans and other eukaryotes, replication occurs in the cell nucleus.

DNA sequence

The relative order of base pairs, whether in a DNA fragment, gene, chromosome, or an entire genome.

See also: base sequence analysis

Domain

A discrete portion of a protein with its own function. The combination of domains in a single protein determines its overall function.

Double helix

The twisted-ladder shape that two linear strands of DNA assume when complementary nucleotides on opposing strands bond together.

E

Electrophoresis

A method of separating large molecules (such as DNA fragments or proteins) from a mixture of similar molecules. An electric current is passed through a medium containing the mixture, and each kind of molecule travels through the medium at a different rate, depending on its electrical charge and size. Agarose and acrylamide gels are the media commonly used for electrophoresis of proteins and nucleic acids.

Electroporation

A process using high-voltage current to make cell membranes permeable to allow the introduction of new DNA; commonly used in recombinant DNA technology.

Embryonic stem (ES) cells

An embryonic cell that can replicate indefinitely, transform into other types of cells, and serve as a continuous source of new cells.

Endonuclease

See: restriction enzyme

Enzyme

A protein that acts as a catalyst, speeding the rate at which a biochemical reaction proceeds but not altering the direction or nature of the reaction.

Epistasis

One gene interfers with or prevents the expression of another gene located at a different locus.

Escherichia coli

Common bacterium that has been studied intensively by geneticists because of its small genome size, normal lack of pathogenicity, and ease of growth in the laboratory.

Eukaryote

Cell or organism with membrane-bound, structurally discrete nucleus and other well-developed subcellular compartments. Eukaryotes include all organisms except viruses, bacteria, and bluegreen algae.

Evolutionarily conserved sequence

A base sequence in a DNA molecule (or an amino acid sequence in a protein) that has remained essentially unchanged throughout evolution

Exogenous DNA

DNA originating outside an organism that has been introducted into the organism.

Exon

The protein-coding DNA sequence of a gene.

See also: intron

Exonuclease

An enzyme that cleaves nucleotides sequentially from free ends of a linear nucleic acid substrate.

Expressed gene

See: gene expression

Expressed sequence tag (EST)

A short strand of DNA that is a part of a cDNA molecule and can act as identifier of a gene. Used in locating and mapping genes.

F

Fluorescence in situ hybridization (FISH)
A physical mapping approach that uses fluorescein tags to detect hybridization of probes with metaphase chromosomes and with the less-condensed somatic interphase chromatin.

Full gene sequence
The complete order of bases in a gene. This order determines which protein a gene will produce.

Functional genomics
The study of genes, their resulting proteins, and the role played by the proteins in the body's biochemical processes.

G

Gamete
Mature male or female reproductive cell (sperm or ovum) with a haploid set of chromosomes (23 for humans).

GC-rich area
Many DNA sequences carry long stretches of repeated G and C which often indicate a gene-rich region.

Gel electrophoresis
See: electrophoresis

Gene
The fundamental physical and functional unit of heredity. A gene is an ordered sequence of nucleotides located in a particular position on a particular chromosome that encodes a specific functional product (i.e., a protein or RNA molecule).
See also: gene expression

Gene amplification
Repeated copying of a piece of DNA; a characteristic of tumor cells.

Gene chip technology

Development of cDNA microarrays from a large number of genes. Used to monitor and measure changes in gene expression for each gene represented on the chip.

Gene expression

The process by which a gene's coded information is converted into the structures present and operating in the cell. Expressed genes include those that are transcribed into mRNA and then translated into protein and those that are transcribed into RNA but not translated into protein (e.g., transfer and ribosomal RNAs).

Gene family

Group of closely related genes that make similar products.

Gene library

See: genomic library

Gene locus (pl. loci)

Gene's position on a chromosome or other chromosome marker; also, the DNA at that position. The use of locus is sometimes restricted to mean expressed DNA regions

Gene mapping

Determination of the relative positions of genes on a DNA molecule (chromosome or plasmid) and of the distance, in linkage units or physical units, between them.

Gene name

Official name assigned to a gene. According to the Guidelines for Human Gene Nomenclature developed by the HUGO Gene Nomenclature Committee, it should be brief and describe the function of the gene.

Gene ontology

A controlled vocabulary of terms relating to molecular function, biological process, or cellular components developed by the Gene Ontology Consortium. A controlled vocabulary allows scientists to use consistent terminology when describing the roles of genes and proteins in cells.

Gene pool
All the variations of genes in a species.

Gene prediction
Predictions of possible genes made by a computer program based on how well a stretch of DNA sequence matches known gene sequences

Gene product
The biochemical material, either RNA or protein, resulting from expression of a gene. The amount of gene product is used to measure how active a gene is; abnormal amounts can be correlated with disease-causing alleles.

Gene testing
See: genetic testing, genetic screening

Gene therapy
An experimental procedure aimed at replacing, manipulating, or supplementing nonfunctional or misfunctioning genes with healthy genes.

Gene transfer
Incorporation of new DNA into and organism's cells, usually by a vector such as a modified virus. Used in gene therapy.

Genetic code
The sequence of nucleotides, coded in triplets (codons) along the mRNA, that determines the sequence of amino acids in protein synthesis. A gene's DNA sequence can be used to predict the mRNA sequence, and the genetic code can in turn be used to predict the amino acid sequence.

Genetic engineering
Altering the genetic material of cells or organisms to enable them to make new substances or perform new functions.

Genetic engineering technology
See: recombinant DNA technology

Genetic illness
Sickness, physical disability, or other disorder resulting from the inheritance of one or more deleterious alleles.

Genetic map
See: linkage map

Genetic marker
A gene or other identifiable portion of DNA whose inheritance can be followed.

Genetic material
See: genome

Genetic polymorphism
Difference in DNA sequence among individuals, groups, or populations (e.g., genes for blue eyes versus brown eyes).

Genetic predisposition
Susceptibility to a genetic disease. May or may not result in actual development of the disease.

Genetic screening
Testing a group of people to identify individuals at high risk of having or passing on a specific genetic disorder.

Gene symbol
Symbols for human genes are usually designated by scientists who discover the genes. The symbols are created using the Guidelines for Human Gene Nomenclature developed by the HUGO Gene Nomenclature Committee. Gene symbols usually consist of no more than six upper case letters or combination of uppercase letters and Arabic numbers. Gene symbols should start with the first letters of the gene name. For example, the gene symbol for insulin is "INS." A gene symbol must be submitted to HUGO for approval before it can be considered an official gene symbol.

Genetic testing
Analyzing an individual's genetic material to determine predisposition to a particular health condition or to confirm a diagnosis of genetic disease.

Genetics
The study of inheritance patterns of specific traits.

Genome

All the genetic material in the chromosomes of a particular organism; its size is generally given as its total number of base pairs.

Genomic library

A collection of clones made from a set of randomly generated overlapping DNA fragments that represent the entire genome of an organism.

See also: library, arrayed library

Genomic sequence

See: DNA

Genomics

The study of genes and their function.

Genotype

The genetic constitution of an organism, as distinguished from its physical appearance (its phenotype).

Germ cell

Sperm and egg cells and their precursors. Germ cells are haploid and have only one set of chromosomes (23 in all), while all other cells have two copies (46 in all).

Germ line

The continuation of a set of genetic information from one generation to the next.

Germ line gene therapy

An experimental process of inserting genes into germ cells or fertilized eggs to cause a genetic change that can be passed on to offspring. May be used to alleviate effects associated with a genetic disease.

Germ line genetic mutation

See: mutation

Guanine (G)

A nitrogenous base, one member of the base pair GC (guanine and cytosine) in DNA.

See also: base pair, nucleotide

H

Haploid
A single set of chromosomes (half the full set of genetic material) present in the egg and sperm cells of animals and in the egg and pollen cells of plants. Human beings have 23 chromosomes in their reproductive cells.
See also: diploid

Haplotype
A way of denoting the collective genotype of a number of closely linked loci on a chromosome.

Hemizygous
Having only one copy of a particular gene. For example, in humans, males are hemizygous for genes found on the Y chromosome.

Hereditary cancer
Cancer that occurs due to the inheritance of an altered gene within a family.
See also: sporadic cancer

Heterozygosity
The presence of different alleles at one or more loci on homologous chromosomes.

Heterozygote
See: heterozygosity

Highly conserved sequence
DNA sequence that is very similar across several different types of organisms.

High-throughput sequencing
A fast method of determining the order of bases in DNA.

Homeobox
A short stretch of nucleotides whose base sequence is virtually identical in all the genes that contain it. Homeoboxes have been found in many organisms from fruit flies to human beings. In the fruit fly, a homeobox appears to determine when particular groups of genes are expressed during development.

Homolog
A member of a chromosome pair in diploid organisms or a gene that has the same origin and functions in two or more species.

Homologous chromosome
Chromosome containing the same linear gene sequences as another, each derived from one parent.

Homologous recombination
Swapping of DNA fragments between paired chromosomes.

Homology
Similarity in DNA or protein sequences between individuals of the same species or among different species.

Homozygote
An organism that has two identical alleles of a gene.
See also: heterozygote

Homozygous
See: homozygote

Human artificial chromosome (HAC)
A vector used to hold large DNA fragments.
See also: chromosome, DNA

Human gene therapy
See: gene therapy

Hybridization
The process of joining two complementary strands of DNA or one each of DNA and RNA to form a double-stranded molecule.

I

Immunotherapy
Using the immune system to treat disease, for example, in the development of vaccines. May also refer to the therapy of diseases caused by the immune system.

Imprinting

A phenomenon in which the disease phenotype depends on which parent passed on the disease gene. For instance, both Prader-Willi and Angelman syndromes are inherited when the same part of chromosome 15 is missing. When the father's complement of 15 is missing, the child has Prader-Willi, but when the mother's complement of 15 is missing, the child has Angelman syndrome.

In situ hybridization

Use of a DNA or RNA probe to detect the presence of the complementary DNA sequence in cloned bacterial or cultured eukaryotic cells.

In vitro

Studies performed outside a living organism such as in a laboratory.

In vivo

Studies carried out in living organisms.

Inherit

In genetics, to receive genetic material from parents through biological processes.

Insertion

A chromosome abnormality in which a piece of DNA is incorporated into a gene and thereby disrupts the gene's normal function.

See also: chromosome, DNA, gene, mutation

Insertional mutation

See: insertion

Interphase

The period in the cell cycle when DNA is replicated in the nucleus; followed by mitosis.

Intron

DNA sequence that interrupts the protein-coding sequence of a gene; an intron is transcribed into RNA but is cut out of the message before it is translated into protein. Many introns are shown to contain small RNA-coding sequences

See also: exon

Isoenzyme

An enzyme performing the same function as another enzyme but having a different set of amino acids. The two enzymes may function at different speeds.

K

Karyotype

A photomicrograph of an individual's chromosomes arranged in a standard format showing the number, size, and shape of each chromosome type; used in low-resolution physical mapping to correlate gross chromosomal abnormalities with the characteristics of specific diseases.

Kilobase (kb)

Unit of length for DNA fragments equal to 1000 nucleotides.

Knockdown

Deactivation of specific genes via RNA interference mechanism; used in laboratory organisms to study gene function

Knockout

Deactivation of specific genes via specific DNA binding mechanisms; used in laboratory organisms to study gene function.

L

Library

An unordered collection of clones (i.e., cloned DNA from a particular organism) whose relationship to each other can be established by physical mapping.

Localize

Determination of the original position (locus) of a gene or other marker on a chromosome.

Locus (pl. loci)

The position on a chromosome of a gene or other chromosome marker; also, the DNA at that position. The use of locus is sometimes restricted to mean expressed DNA regions.

Long-range restriction mapping

Restriction enzymes are proteins that cut DNA at precise locations. Restriction maps depict the chromosomal positions of restriction-enzyme cutting sites. These are used as biochemical "signposts," or markers of specific areas along the chromosomes. The map will detail the positions where the DNA molecule is cut by particular restriction enzymes.

M

Macrorestriction map

Map depicting the order of and distance between sites at which restriction enzymes cleave chromosomes.

Mapping

See: gene mapping, linkage map, physical map

Marker

See: genetic marker

Mass spectrometry

An instrument used to identify chemicals in a substance by their mass and charge.

Megabase (Mb)

Unit of length for DNA fragments equal to 1 million nucleotides and roughly equal to 1 cM (centimorgan).

Meiosis

The process of two consecutive cell divisions in the diploid progenitors of sex cells. Meiosis results in four rather than two daughter cells, each with a haploid set of chromosomes.

See also: mitosis

Messenger RNA (mRNA)

RNA that serves as a template for protein synthesis.

Metaphase

A stage in mitosis or meiosis during which the chromosomes are aligned along the equatorial plane of the cell.

Microarray

Sets of miniaturized chemical reaction areas that may also be used to test DNA or RNA fragments, antibodies, or proteins.

Micronuclei

Chromosome fragments that are not incorporated into the nucleus at cell division.

microRNAs (miRNA)

single-stranded RNA molecules of 19-24 nucleotides in length, which regulate gene expression. miRNAs are encoded by genes from whose DNA they are transcribed but miRNAs are not translated into protein (non-coding RNA); instead each primary transcript (a pri-miRNA) is processed into a short stem-loop structure called a pre-miRNA and finally into a functional miRNA. Mature miRNA molecules are partially complementary to one or more messenger RNA (mRNA) molecules, and their main function is to down-regulate gene expression via mRNA degradation and/or translation inhibition. They were first described in 1993 by Lee and colleagues in the Victor Ambros lab, yet the term microRNA was only introduced in 2001 in a set of three articles in *Science*.

Mitochondrial DNA

The genetic material found in mitochondria, the organelles that generate energy for the cell. Not inherited in the same fashion as nucleic DNA.

Mitosis

The process of nuclear division in cells that produces daughter cells that are genetically identical to each other and to the parent cell.

See also: meiosis

Model organisms

A laboratory animal or other organism useful for research.

Modeling

The use of statistical analysis, mathematical equations, computer simulation.

Monogenic disorder

A disorder caused by mutation of a single gene.

Monogenic inheritance
See: monogenic disorder

Monosomy
Possessing only one copy of a particular chromosome instead of the normal two copies.

Mouse model
See: model organisms

Multiplexing
A laboratory approach that performs multiple sets of reactions in parallel (simultaneously); greatly increasing speed and throughput.

Murine
Organism in the genus Mus. A rat or mouse.

Mutagen
An agent that causes a permanent genetic change in a cell. Does not include changes occurring during normal genetic recombination.

Mutagenicity
The capacity of a chemical or physical agent to cause permanent genetic alterations.
See also: somatic cell genetic mutation

Mutation
Any heritable change in DNA sequence.
See also: polymorphism

N

Nitrogenous base
A nitrogen-containing molecule having the chemical properties of a base. DNA contains the nitrogenous bases adenine (A), guanine (G), cytosine (C), and thymine (T).

Non-coding RNA (ncRNA)

ncRNA is transcribed from DNA, but not translated into protein. It often plays a regulatory role. Some probably most ancient ncRNAs encompass catalytic functions and are called ribozymes

Northern blot

A gel-based laboratory procedure that locates mRNA sequences on a gel that are complementary to a piece of DNA used as a probe.

Nuclear transfer

A laboratory procedure in which a cell's nucleus is removed and placed into an oocyte with its own nucleus removed so the genetic information from the donor nucleus controls the resulting cell. Such cells can be induced to form embryos. This process was used to create the cloned sheep "Dolly".

Nucleic acid

A large molecule composed of nucleotide subunits.

See also: DNA

Nucleolar organizing region

A part of the chromosome containing rRNA genes.

Nucleotide

A subunit of DNA or RNA consisting of a nitrogenous base (adenine, guanine, thymine, or cytosine in DNA; adenine, guanine, uracil, or cytosine in RNA), a phosphate molecule, and a sugar molecule (deoxyribose in DNA and ribose in RNA). Thousands of nucleotides are linked to form a DNA or RNA molecule.

Nucleus

The cellular organelle in eukaryotes that contains most of the genetic material.

O

Oligo

See: oligonucleotide

Oligonucleotide

A molecule usually composed of 25 or fewer nucleotides; used as a DNA synthesis primer.

See also: nucleotide

Oncogene

A gene, one or more forms of which is associated with cancer. Many oncogenes are involved, directly or indirectly, in controlling the rate of cell growth.

Open reading frame (ORF)

The sequence of DNA or RNA located between the start-code sequence (initiation codon) and the stop-code sequence (termination codon).

Operon

A set of genes transcribed under the control of an operator gene.

Overlapping clones

See: genomic library

P

P1-derived artificial chromosome (PAC)

One type of vector used to clone DNA fragments (100- to 300-kb insert size; average, 150 kb) in Escherichia coli cells. Based on bacteriophage (a virus) P1 genome.

See also: cloning vector

Peptide

Two or more amino acids joined by a bond called a “peptide bond”.

See also: polypeptide

Phage

A virus for which the natural host is a bacterial cell.

Pharmacogenomics

The study of the interaction of an individual's genetic makeup and response to a drug.

Phenotype

The physical characteristics of an organism or the presence of a disease that may or may not be genetic.

See also: genotype

Physical map

A map of the locations of identifiable landmarks on DNA (e.g., restriction-enzyme cutting sites, genes), regardless of inheritance. Distance is measured in base pairs. For the human genome, the lowest-resolution physical map is the banding patterns on the 24 different chromosomes; the highest-resolution map is the complete nucleotide sequence of the chromosomes.

Plasmid

Autonomously replicating extra-chromosomal circular DNA molecules, distinct from the normal bacterial genome and nonessential for cell survival under nonselective conditions. Some plasmids are capable of integrating into the host genome. A number of artificially constructed plasmids are used as cloning vectors.

Pleiotropy

One gene that causes many different physical traits such as multiple disease symptoms.

Pluripotency

The potential of a cell to develop into more than one type of mature cell, depending on environment.

Polymerase chain reaction (PCR)

A method for amplifying a DNA base sequence using a heat-stable polymerase and two 20-base primers, one complementary to the (+) strand at one end of the sequence to be amplified and one complementary to the (-) strand at the other end. Because the newly synthesized DNA strands can subsequently serve as additional templates for the same primer sequences, successive rounds of primer annealing, strand elongation, and dissociation produce rapid and highly specific amplification of the desired sequence. PCR also can be used to detect the existence of the defined sequence in a DNA sample.

Polymerase, DNA or RNA

Enzyme that catalyzes the synthesis of nucleic acids on preexisting nucleic acid templates, assembling RNA from ribonucleotides or DNA from deoxyribonucleotides.

Polymorphism

Difference in DNA sequence among individuals that may underlie differences in health. Genetic variations occurring in more than 1% of a population would be considered useful polymorphisms for genetic linkage analysis.

See also: mutation

Polypeptide

A protein or part of a protein made of a chain of amino acids joined by a peptide bond.

Population genetics

The study of variation in genes among a group of individuals.

Positional cloning

A technique used to identify genes, usually those that are associated with diseases, based on their location on a chromosome.

Premature chromosome condensation (PCC)

A method of studying chromosomes in the interphase stage of the cell cycle.

Primer

Short preexisting polynucleotide chain to which new deoxyribonucleotides can be added by DNA polymerase.

Probe

Single-stranded DNA or RNA molecules of specific base sequence, labeled either radioactively or immunologically, that are used to detect the complementary base sequence by hybridization.

Prokaryote

Cell or organism lacking a membrane-bound, structurally discrete nucleus and other subcellular compartments. Bacteria are examples of prokaryotes.

See also: chromosome, eukaryote

Promoter

A DNA site to which RNA polymerase will bind and initiate transcription.

Pronucleus

The nucleus of a sperm or egg prior to fertilization.

See also: nucleus, transgenic

Protein

A large molecule composed of one or more chains of amino acids in a specific order; the order is determined by the base sequence of nucleotides in the gene that codes for the protein. Proteins are required for the structure, function, and regulation of the body's cells, tissues, and organs; and each protein has unique functions. Examples are hormones, enzymes, and antibodies.

Proteome

Proteins expressed by a cell or organ at a particular time and under specific conditions.

Proteomics

The study of the full set of proteins encoded by a genome.

Pseudogene

A sequence of DNA similar to a gene but nonfunctional; probably the remnant of a once-functional gene that accumulated mutations.

Purine

A nitrogen-containing, double-ring, basic compound that occurs in nucleic acids. The purines in DNA and RNA are adenine and guanine.

See also: base pair

Pyrimidine

A nitrogen-containing, single-ring, basic compound that occurs in nucleic acids. The pyrimidines in DNA are cytosine and thymine; in RNA, cytosine and uracil.

See also: base pair

R

Radiation hybrid

A hybrid cell containing small fragments of irradiated human chromosomes. Maps of irradiation sites on chromosomes for the human, rat, mouse, and other genomes provide important markers, allowing the construction of very precise STS maps indispensable to studying multifactorial diseases.

Recessive gene

A gene which will be expressed only if there are 2 identical copies or, for a male, if one copy is present on the X chromosome.

Reciprocal translocation

When a pair of chromosomes exchange exactly the same length and area of DNA. Results in a shuffling of genes.

Recombinant clone

Clone containing recombinant DNA molecules.

Recombinant DNA molecules

A combination of DNA molecules of different origin that are joined using recombinant DNA technologies.

Recombinant DNA technology

Procedure used to join together DNA segments in a cell-free system (an environment outside a cell or organism). Under appropriate conditions, a recombinant DNA molecule can enter a cell and replicate there, either autonomously or after it has become integrated into a cellular chromosome.

Recombination

The process by which progeny derive a combination of genes different from that of either parent. In higher organisms, this can occur by crossing over.

See also: crossing over, mutation

Regulatory region or sequence

A DNA base sequence that controls gene expression.

Repetitive DNA

Sequences of varying lengths that occur in multiple copies in the genome; it represents much of the human genome.

Reporter gene

See: marker

Resolution

Degree of molecular detail on a physical map of DNA, ranging from low to high.

Restriction enzyme, endonuclease

A protein that recognizes specific, short nucleotide sequences and cuts DNA at those sites. Bacteria contain over 400 such enzymes that recognize and cut more than 100 different DNA sequences.

Restriction fragment length polymorphism (RFLP)

Variation between individuals in DNA fragment sizes cut by specific restriction enzymes; polymorphic sequences that result in RFLPs are used as markers on both physical maps and genetic linkage maps. RFLPs usually are caused by mutation at a cutting site.

Restriction-enzyme cutting site

A specific nucleotide sequence of DNA at which a particular restriction enzyme cuts the DNA. Some sites occur frequently in DNA (e.g., every several hundred base pairs); others much less frequently (rare-cutter; e.g., every 10,000 base pairs).

Retroviral infection

The presence of retroviral vectors, such as some viruses, which use their recombinant DNA to insert their genetic material into the chromosomes of the host's cells. The virus is then propogated by the host cell.

Reverse transcriptase

An enzyme used by retroviruses to form a complementary DNA sequence (cDNA) from their RNA. The resulting DNA is then inserted into the chromosome of the host cell.

Ribonucleotide

See: nucleotide

Ribose

The five-carbon sugar that serves as a component of RNA.

See also: ribonucleic acid, deoxyribose

Ribosomal RNA (rRNA)

A class of RNA found in the ribosomes of cells.

Ribosomes

Small cellular components composed of specialized ribosomal RNA and protein; site of protein synthesis.

See also: RNA

RNA (Ribonucleic acid)

A chemical found in the nucleus and cytoplasm of cells; it plays an important role in protein synthesis and other chemical activities of the cell. The structure of RNA is similar to that of DNA. There are several classes of RNA molecules, including messenger RNA, transfer RNA, ribosomal RNA, and other small RNAs, each serving a different purpose.

RNA interference (RNAi)

A gene-silencing process in which double-stranded RNAs trigger the destruction of specific RNAs.

S

Sanger sequencing

A widely used method of determining the order of bases in DNA.

See also: sequencing, shotgun sequencing

Satellite

A chromosomal segment that branches off from the rest of the chromosome but is still connected by a thin filament or stalk.

Scaffold

In genomic mapping, a series of contigs that are in the right order but not necessarily connected in one continuous stretch of sequence.

Segregation

The normal biological process whereby the two pieces of a chromosome pair are separated during meiosis and randomly distributed to the germ cells.

Sequence

See: base sequence

Sequence assembly

A process whereby the order of multiple sequenced DNA fragments is determined.

Sequence tagged site (STS)

Short (200 to 500 base pairs) DNA sequence that has a single occurrence in the human genome and whose location and base sequence are known. Detectable by polymerase chain reaction, STSs are useful for localizing and orienting the mapping and sequence data reported from many different laboratories and serve as landmarks on the developing physical map of the human genome. Expressed sequence tags (ESTs) are STSs derived from cDNAs.

Sequencing

Determination of the order of nucleotides (base sequences) in a DNA or RNA molecule or the order of amino acids in a protein.

Sequencing technology

The instrumentation and procedures used to determine the order of nucleotides in DNA.

Sex chromosome

The X or Y chromosome in human beings that determines the sex of an individual. Females have two X chromosomes in diploid cells; males have an X and a Y chromosome. The sex chromosomes comprise the 23rd chromosome pair in a karyotype.

See also: autosome

Sex-linked

Traits or diseases associated with the X or Y chromosome; generally seen in males.

Shotgun method

Sequencing method that involves randomly sequenced cloned pieces of the genome, with no foreknowledge of where the piece originally came from. This can be contrasted with "directed" strategies, in which pieces of DNA from known chromosomal locations are sequenced. Because there are advantages to both strategies, researchers use both random (or shotgun) and directed strategies in combination to sequence the human genome.

Single nucleotide polymorphism (SNP)

DNA sequence variations that occur when a single nucleotide (A, T, C, or G) in the genome sequence is altered.

Somatic cell

Any cell in the body except gametes and their precursors.

See also: gamete

Somatic cell gene therapy

Incorporating new genetic material into cells for therapeutic purposes. The new genetic material cannot be passed to offspring.

See also: gene therapy

Somatic cell genetic mutation

A change in the genetic structure that is neither inherited nor passed to offspring. Also called acquired mutations.

See also: germ line genetic mutation

Southern blotting

Transfer by absorption of DNA fragments separated in electrophoretic gels to membrane filters for detection of specific base sequences by radio-labeled complementary probes.

Splice site

Location in the DNA sequence where RNA removes the noncoding areas to form a continuous gene transcript for translation into a protein.

Sporadic cancer

Cancer that occurs randomly and is not inherited from parents..

Stem cell

Undifferentiated, primitive cells in the bone marrow that have the ability both to multiply and to differentiate into specific blood cells.

Substitution

In genetics, a type of mutation due to replacement of one nucleotide in a DNA sequence by another nucleotide or replacement of one amino acid in a protein by another amino acid.

See also: mutation

Suppressor gene
A gene that can suppress the action of another gene.

Syndrome
The group or recognizable pattern of symptoms or abnormalities that indicate a particular trait or disease.

Syngeneic
Genetically identical members of the same species.

Synteny
Genes occurring in the same order on chromosomes of different species.
See also: linkage, conserved sequence

Systems biology
A field that seeks to study the relationships and interactions between various parts of a biological system (metabolic pathways, organelles, cells, and organisms) and to integrate this information to understand how biological systems function

T

Tandem repeat sequences
Multiple copies of the same base sequence on a chromosome; used as markers in physical mapping.

Targeted mutagenesis
Deliberate change in the genetic structure directed at a specific site on the chromosome. Used in research to determine the targeted region's function.

Technology transfer
The process of transferring scientific findings from research laboratories to the commercial sector.

Telomerase
The enzyme that directs the replication of telomeres.

Telomere

The end of a chromosome. This specialized structure is involved in the replication and stability of linear DNA molecules.

Teratogenic

Substances such as chemicals or radiation that cause abnormal development of a embryo.

Thymine (T)

A nitrogenous base, one member of the base pair AT (adenine-thymine).

See also: base pair, nucleotide

Toxicogenomics

The study of how genomes respond to environmental stressors or toxicants. Combines genome-wide mRNA expression profiling with protein expression patterns using bioinformatics to understand the role of gene-environment interactions in disease and dysfunction.

Transcription

The synthesis of an RNA copy from a sequence of DNA (a gene); the first step in gene expression.

See also: translation

Transcription factor

A protein that binds to regulatory regions and helps control gene expression.

Transcriptome

The full complement of activated genes, mRNAs, or transcripts in a particular tissue at a particular time

Transfection

The introduction of foreign DNA into a host cell.

See also: cloning vector, gene therapy

Transfer RNA (tRNA)

A class of RNA having structures with triplet nucleotide sequences that are complementary to the triplet nucleotide coding sequences of mRNA. The role of tRNAs in protein synthesis is to bond with amino acids and transfer them to the

ribosomes, where proteins are assembled according to the genetic code carried by mRNA.

Transformation
A process by which the genetic material carried by an individual cell is altered by incorporation of exogenous DNA into its genome.

Transgenic
An experimentally produced organism in which DNA has been artificially introduced and incorporated into the organism's germ line.

Translation
The process in which the genetic code carried by mRNA directs the synthesis of proteins from amino acids.
See also: transcription

Translocation
A mutation in which a large segment of one chromosome breaks off and attaches to another chromosome.
See also: mutation

Transposable element
A class of DNA sequences that can move from one chromosomal site to another.

Trisomy
Possessing three copies of a particular chromosome instead of the normal two copies.

U

Uracil
A nitrogenous base normally found in RNA but not DNA; uracil is capable of forming a base pair with adenine.
See also: base pair, nucleotide

V

Vector
See: cloning vector

Version (GenBank)

Similar to the Protein ID for protein sequences, the version is a nucleotide sequence identification number assigned to each GenBank sequence. The format for this sequence identifier is accession.version (e.g., M12345.1). Whenever the author of a particular sequence record changes the sequence data in any way (even if just a single nucleotide is altered), the version number will be increased by an increment of one, while the accession number base remains constant. For example, M12345.1 would become M12345.2. Each sequence change also results in the assignment of a new GI number [link to GI entry]. Whenever an individual searches an NCBI sequence database, only the most recent version of a record is retrieved. Use NCBI's Sequence Revision History page to view the different GI numbers, version numbers, or update dates associated with a particular GenBank record.

Virus

A noncellular biological entity that can reproduce only within a host cell. Viruses consist of nucleic acid covered by protein; some animal viruses are also surrounded by membrane. Inside the infected cell, the virus uses the synthetic capability of the host to produce progeny virus.

W

Western blot

A technique used to identify and locate proteins based on their ability to bind to specific antibodies.

See also: DNA, Northern blot, protein, RNA, Southern blotting

Wild type

The form of an organism that occurs most frequently in nature.

X

X chromosome

One of the two sex chromosomes, X and Y.

See also: Y chromosome, sex chromosome

Xenograft

Tissue or organs from an individual of one species transplanted into or grafted onto an organism of another species, genus, or family. A common example is the use of pig heart valves in humans.

Y

Y chromosome

One of the two sex chromosomes, X and Y.

See also: X chromosome, sex chromosome

Yeast artificial chromosome (YAC)

Constructed from yeast DNA, it is a vector used to clone large DNA fragments.

See also: cloning vector, cosmid

Z

Zinc-finger protein

A secondary feature of some proteins containing a zinc atom; a DNA-binding protein.

This page intentionally left blank

INDEX