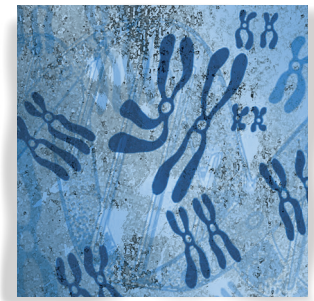
A glossary of relevant genetic terms

Deborah J. Morris-Rosendahl, PhD



Allele
One of two or more alternate forms of a gene or marker at a particular locus on a chromosome.

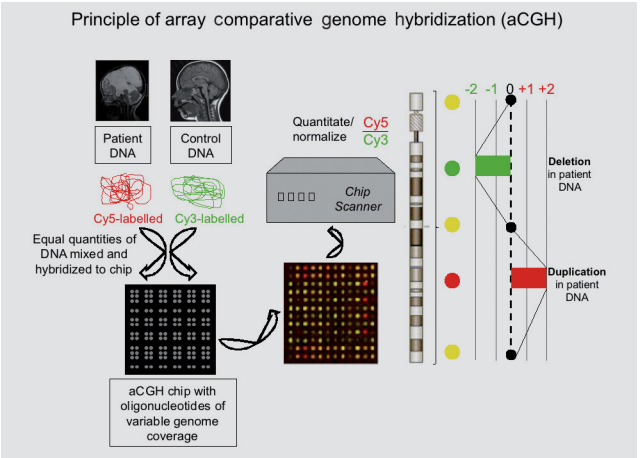
Anticipation (genetic)
Apparent earlier age of onset and increased severity of a disease in successive generations, eg, Huntington's disease.

cDNA
Single-stranded complementary DNA, ie, a DNA molecule synthesized from a RNA template by reverse transcription of RNA.

Common disorder common variant (CDCV) hypothesis
A theory that many common diseases are caused by common alleles that individually have little effect, but in concert confer a high risk.

Complex disease
A disorder in which the cause is considered to be a combination of genetic effects and environmental influences.

Comparative genome hybridization (CGH)
CGH is a molecular-cytogenetic method for the analysis of copy number changes (gains or losses) in the DNA content of a given individual's DNA.



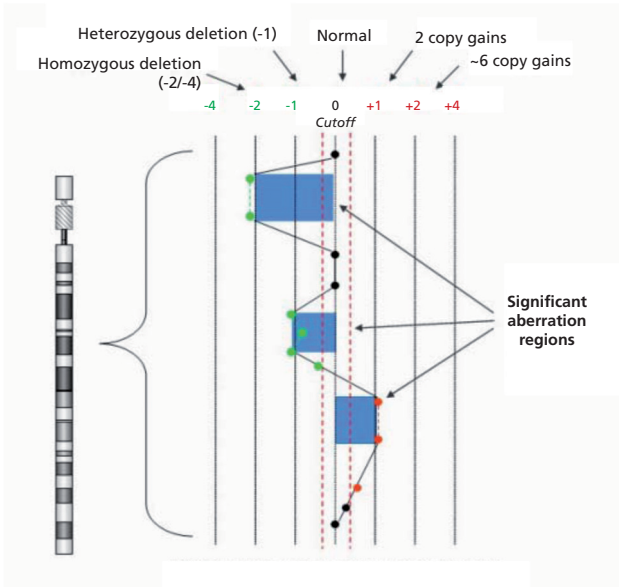
Compound heterozygosity
Heterozygosity for two different mutant alleles of a gene, often the case for autosomal recessive disorders.

Copy number variation (CNV)
A segment of DNA in which copy number differences have been found by comparison of two or more genomes. The segment may range from one kilobase to several megabases in size. The variation is usually due to deletion or duplication.

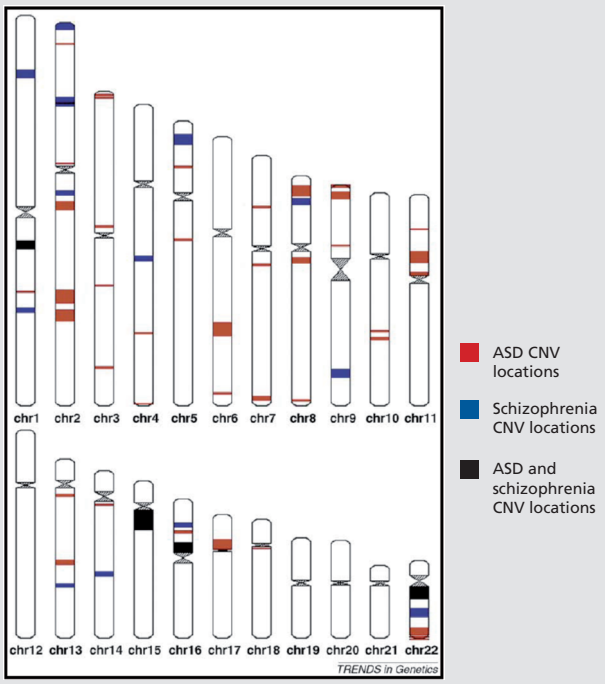
Author affiliations: Institute for Human Genetics, University Clinic Freiburg, Germany

Address for correspondence: Priv.-Doz. Dr. Deborah J. Morris-Rosendahl, Universitaetsklinikum Freiburg, Institut fuer Humangenetik Breisacherstr. 33, D-79106 Freiburg, Germany
(e-mail: deborah.morris-rosendahl@uniklinik-freiburg.de)

Copy number variation (CNV)



CNV association findings in schizophrenia and autism spectrum disorder (ASD)



Adapted from: Merikangas AK, Corvin AP, Gallagher L. Copy number variants in neurodevelopmental disorders: promises and challenges. *Trends Genet.* 2009;25:536-544.

Deep resequencing

A technique for sequencing a gene in several thousand subjects, typically using one of the new high-throughput sequencing technologies.

Epigenetics

Heritable changes to DNA structure that do not alter the underlying DNA sequence, eg, DNA methylation.

Epigenomics

The application of epigenetics to the whole genome.

Exome

The approximately 1% of the human genome that comprises all exons and therefore the entire protein-coding region of the genome.

Genetic association

The nonrandom occurrence of a genetic marker (usually a particular allele of a polymorphism) with a trait, which suggests an association between the genetic marker (or marker close to it) and disease pathogenesis.

Genome

In eukaryotes, the basic (monoploid) chromosome set, consisting of a species-specific number of linkage groups and the genes contained therein. For example, in humans, the genome consists of the 24 different chromosomes (22 autosomes, X and Y chromosomes). The mitochondrial DNA is usually considered to be a separate "mitochondrial" genome.

Genome-wide association study (GWAS)

A test for the association between genetic polymorphisms spread evenly over the entire genome, and a disease. Usually at least 300 000 markers are required to adequately cover the genome.

Genotype

The genetic constitution with respect to the alleles at one or more pairs of genetic loci under observation. The genotype of an individual is the sum total of the genetic information contained on the chromosomes, as distinguished from the individual's phenotype (idiotype).

Brief report

Haploid
A single genome or set of chromosomes (eg, in human gametes, n=23), compared to the normal diploid (double) set of chromosomes (n=46).

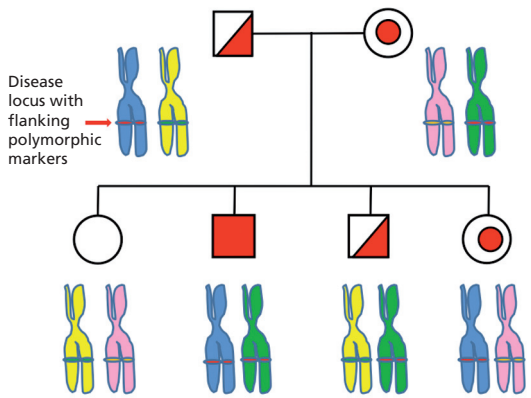
Haplotype
A combination of alleles at closely linked gene loci that are inherited together.

Hemizygous
When one or more genes is present in only one, instead of two copies, eg, men are hemizygous for most genes on the X and Y chromosomes.

Heterozygous
Having different alleles for one or more genes in homologous chromosome segments, as opposed to being homozygous with identical alleles at these loci.

Linkage
Genetic linkage refers to the observation that two or more genes located on the same chromosome are inherited together. The ratio of being transmitted together versus being separated allows an estimate of their distance from each other (recombination fraction).

Linkage analysis

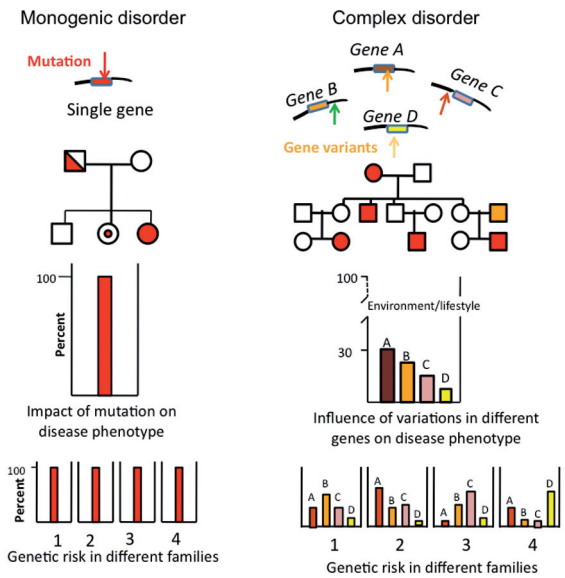


Linkage disequilibrium (LD)
Alleles at different loci that are inherited together more frequently or less frequently than expected by their individual frequencies are said to show linkage disequilibrium.

Methylation (of DNA)
The attachment of a methyl group to DNA. In vertebrates, this typically occurs at CpG sites (cytosine-phosphate-guanine sites) in the DNA sequence, resulting in the conversion of cytosine to 5-methylcytosine.

Monogenic disorder
Disorder caused by one or more mutations in a single gene, eg, cystic fibrosis (mutations in the *CFTR* gene). Such disorders are also sometimes referred to Mendelian diseases.

Monogenic vs complex disease



Penetrance
The frequency (in percent) with which a dominant or homozygous recessive gene or gene combination manifests itself in the phenotype of the carriers.

Pharmacogenetics
A branch of genetics which deals with the genetic variability in individual responses to drugs and drug metabolism.

Phenocopy
A nonhereditary, phenotypic modification (caused by special environmental conditions) that mimics a similar phenotype caused by a gene mutation.

Phenotype
The observable properties (structural and functional) of an organism, produced by the interaction between the

organism's genotype and the environment in which it finds itself.

Pleiotropy
Genes or mutations that result in the production of multiple, apparently unrelated, effects at the phenotypic level. For example, patients with phenylketonuria, caused by mutations in the *PAH* (phenylalanine hydroxylase) gene, have reduced hair and skin pigmentation in addition to mental retardation, resulting from toxic levels of phenylalanine.

Polymorphism (genetic)
A chromosome or DNA variant that is observed in natural populations. A gene locus is defined as polymorphic if a rare allele has a frequency of 0.01 (1%) or more.

Positional cloning
Finding disease genes based on knowledge of their chromosomal location (usually found via linkage analysis in families with the disease) as opposed to knowledge of the function of the gene or protein encoded by the gene.

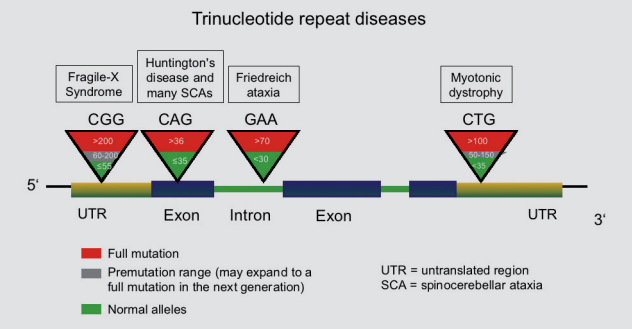
Second- or next-generation sequencing (also referred to as high-throughput sequencing)
New techniques that have increased the speed and decreased the cost of DNA sequencing by two orders of magnitude, enabling the sequencing of the entire genomes of many individuals.

Single nucleotide polymorphism (SNP)
Heritable polymorphism resulting from a single base pair change. SNPs generally have only two alleles.

Structural variant
Structural genomic variation includes any genetic variant that alters chromosomal structure, including inversions, translocations, duplications and deletions. Duplications and deletions, collectively known as CNVs (see *copy number variation*) are the most common form of structural variation in the human genome.

Synonymous nucleotide change/non-synonymous nucleotide change
A change in the DNA sequence which does not result in the change in the amino acid sequence, eg, GTT>GTC both code for Valine (Val or V). A nonsynonymous change results in the coding of a different amino acid (eg, GTT>GAT results in Val>Asp).

Trinucleotide repeat expansion
An increased number of contiguous trinucleotide repeats (eg, CAG, CGG) in the DNA sequence from one generation to the next. When the expansion extends into the pathological range, this type of mutation causes diseases such as Huntington's disease, fragile X syndrome, myotonic dystrophy, and many forms of spinocerebellar ataxia.



X-inactivation
The random, early embryological, inactivation of one of the X chromosomes in females, so that the expression of X-chromosomal genes is the same as that in males.

Useful genetic databases

National Center for Biotechnology information (NCBI)
http://www.ncbi.nlm.nih.gov/
Provides links to many other databases, including many of the databases below.

Online Mendelian Inheritance in Man (OMIM)
http://www.ncbi.nlm.nih.gov/sites/entrez?db=omim
A comprehensive, authoritative, and timely compendium of human genes and genetic phenotypes.

Genome Database (GDB)
http://gdbwww.gdb.org/
Gene and protein sequence database.

UCSC Genome Browser/Bioinformatics site
http://genome.ucsc.edu/index.html
Provides the reference sequence and working draft assemblies for a large number of genomes. The browser

has many useful tools, eg, for searching for sequences within a genome, and comparing sequences within and between genomes.

ENSEMBL Database
http://www.ensembl.org
A genome database for vertebrates and other species, providing gene sequence data as well as chromosomal localization overviews and some information regarding transcripts and proteins.

db-SNP Polymorphism Repository
http://www.ncbi.nlm.nih.gov/SNP/
Database for single nucleotide polymorphisms and other classes of minor genetic variation.

The SNP Consortium Ltd. (TSC)
http://snp.cshl.org/
A non-profit foundation organized to develop up to 300 000 single nucleotide polymorphisms (SNPs) distributed evenly throughout the human genome and to make the information related to these SNPs available to the public without intellectual property restrictions.

European Bioinformatics Institute (EBI)
http://www2.ebi.ac.uk/
A centre for research and services in bioinformatics, which is part of the European Molecular Biology Laboratory (EMBL).

Gene Cards Database
http://bioinformatics.weizmann.ac.il/cards/
Summarizes most available information on a particular gene, with links to many other databases, eg, protein databases

International HapMap Project
http://hapmap.ncbi.nlm.nih.gov/
A partnership of scientists and funding agencies from Canada, China, Japan, Nigeria, the United Kingdom, and the USA to develop a public resource that will help researchers find genes associated with human disease and response to pharmaceuticals.

The Human Genome Variation Database
http://hgvbase.cgb.ki.se/
A reference for the nomenclature of genetic variation; also provides links to various mutation databases

The Human Gene Mutation Database
http://archive.uwcm.ac.uk/uwcm/mg/hgmd0.html
Database of published mutations for different disease genes.

The Pharmacogenomics and Pharmacogenetics Knowledgebase
http://pharmgkb.org/do/serve?id=home.welcome

The Human Variome Project
http://www.humanvariomeproject.org/
Web site of the global initiative to collect and curate all human genetic variation affecting human health.

Useful glossary references

Rieger R, Michaelis A, Green MM. *Glossary of Genetics, Classical and Molecular*. Berlin, Germany: Springer-Verlag; 1991.

Passarge E. *Colour Atlas of Genetics*. Stuttgart, Germany: Thieme Verlag; 2007.