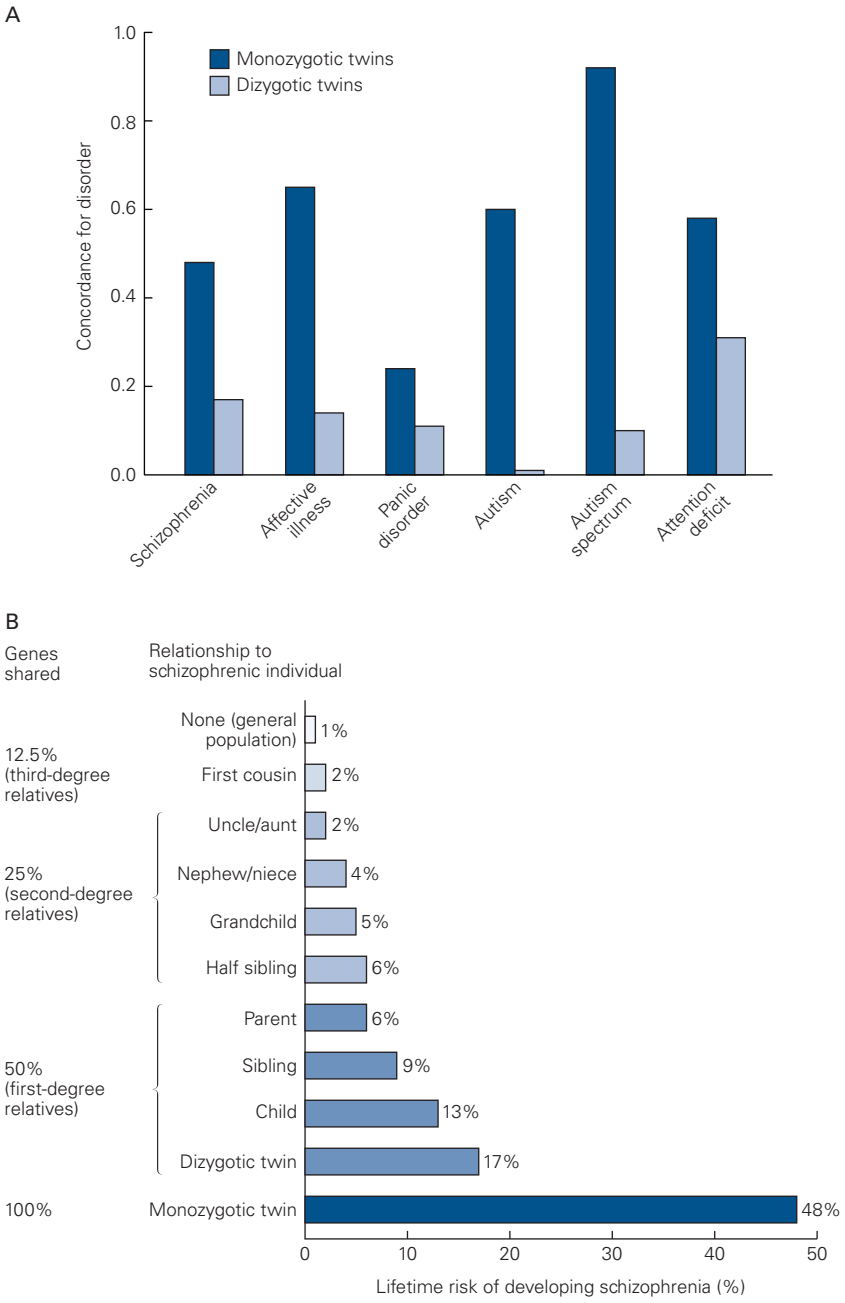

Figure 2–1 Familial risk of psychiatric disorders provides evidence of heritability.

A. Correlations between monozygotic twins for psychiatric disorders are considerably greater than those between dizygotic twins. Monozygotic twins share nearly all genes and have a high (but not 100%) risk of sharing the disease state. Dizygotic twins share 50% of their genetic material. A score of zero represents no correlation (the average result for two random people), whereas a score of 1.0 represents a perfect correlation. (Adapted from McGue and Bouchard 1998.)

B. The risk of developing schizophrenia is greater in close relatives of a schizophrenic patient. Like dizygotic twins, parents and children, as well as brothers and sisters, share 50% of their genetic material. If only a single gene accounted for schizophrenia, the risk should be the same for parents, siblings, children, and dizygotic twins of patients. The variation between family members shows that more complex genetic and environmental factors are in play. (Adapted, with permission, from Gottesman II. 1991.)

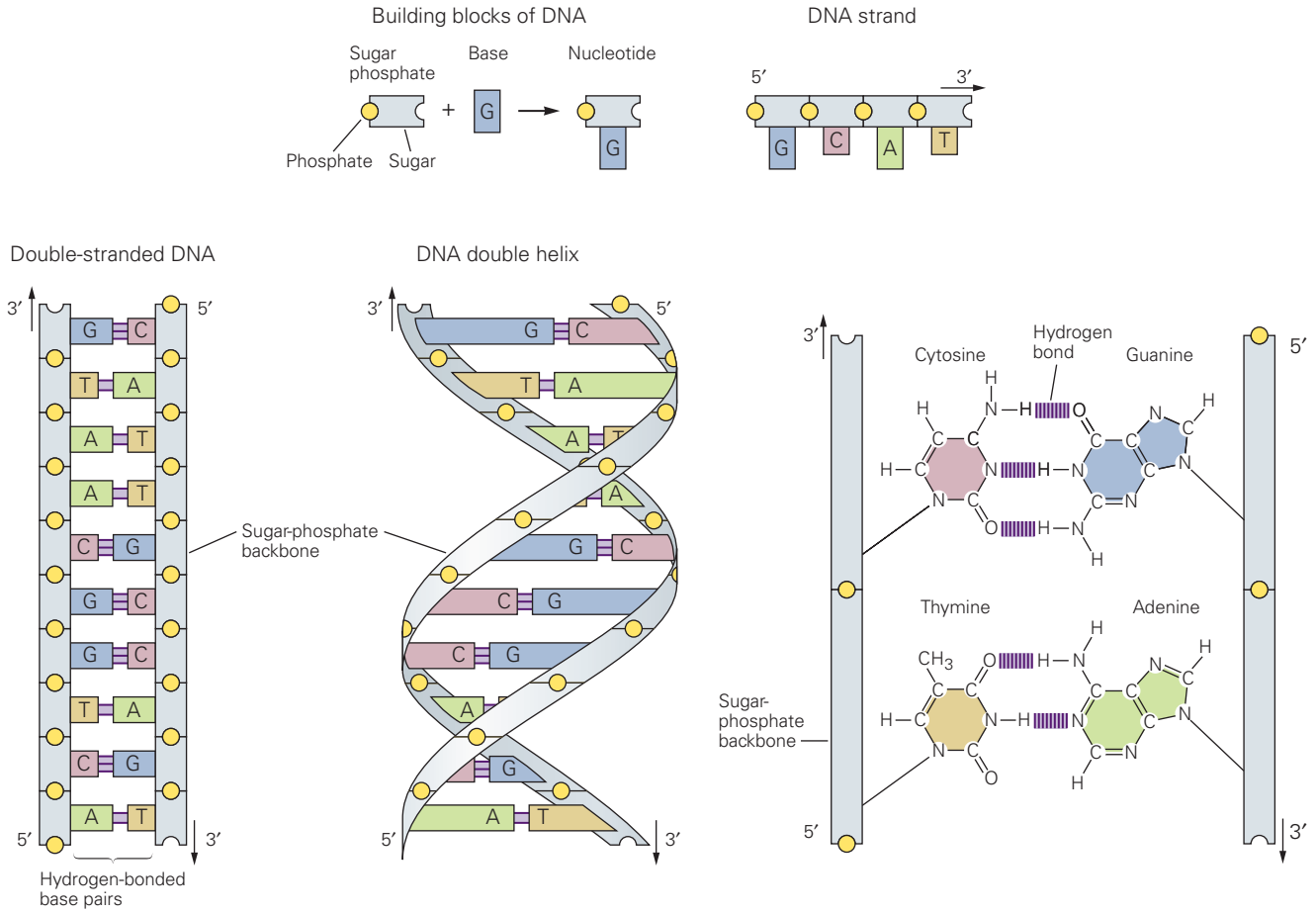

Figure 2–2 Structure of DNA. Four different nucleotide bases, adenine (A), thymine (T), cytosine (C), and guanine (G), are assembled on a sugar phosphate backbone in the double-stranded DNA helix. (Adapted from Alberts et al. 2002.)

Many functional RNA transcripts do not encode proteins. In fact, in the human genome, over 40,000 noncoding transcripts have been characterized as compared with approximately 20,000 protein-coding genes. Such genes include ribosomal RNAs (rRNAs) and transfer RNAs (tRNAs), essential components of the machinery for mRNA translation. Additional noncoding RNAs (ncRNA) include *long noncoding RNAs (lncRNAs)*, arbitrarily defined as longer than 200 bp in length, that do not encode proteins but can have roles in gene regulation; *small noncoding RNAs* of several types, including small nuclear RNAs (snRNAs), that guide mRNA splicing; and microRNAs (miRNAs) that pair with complementary sequences in specific mRNAs to inhibit their translation.

Each cell in the body contains the DNA for every gene but only expresses a specific subset of the genes as RNAs. The part of the gene that is transcribed into RNA is flanked by noncoding DNA regions that may be bound by other proteins, including *transcription factors*, to regulate gene expression. These sequence motifs include *promoters*, *enhancers*, *silencers*, and *insulators*, which together allow accurate expression of the RNA in the right cells at the right time. Promoters are typically found close to the beginning of the region to be transcribed; enhancers, silencers, and insulators may reside at a distance from the gene being regulated. Each type of cell has a unique complement of DNA-binding proteins that interact with promoters and other regulatory sequences to regulate gene expression and the resulting cellular properties.

The brain expresses a greater number of genes than any other organ in the body, and within the brain, diverse populations of neurons express different groups of genes. The selective gene expression controlled by promoters, other regulatory sequences, and the DNA-binding proteins that interact with them

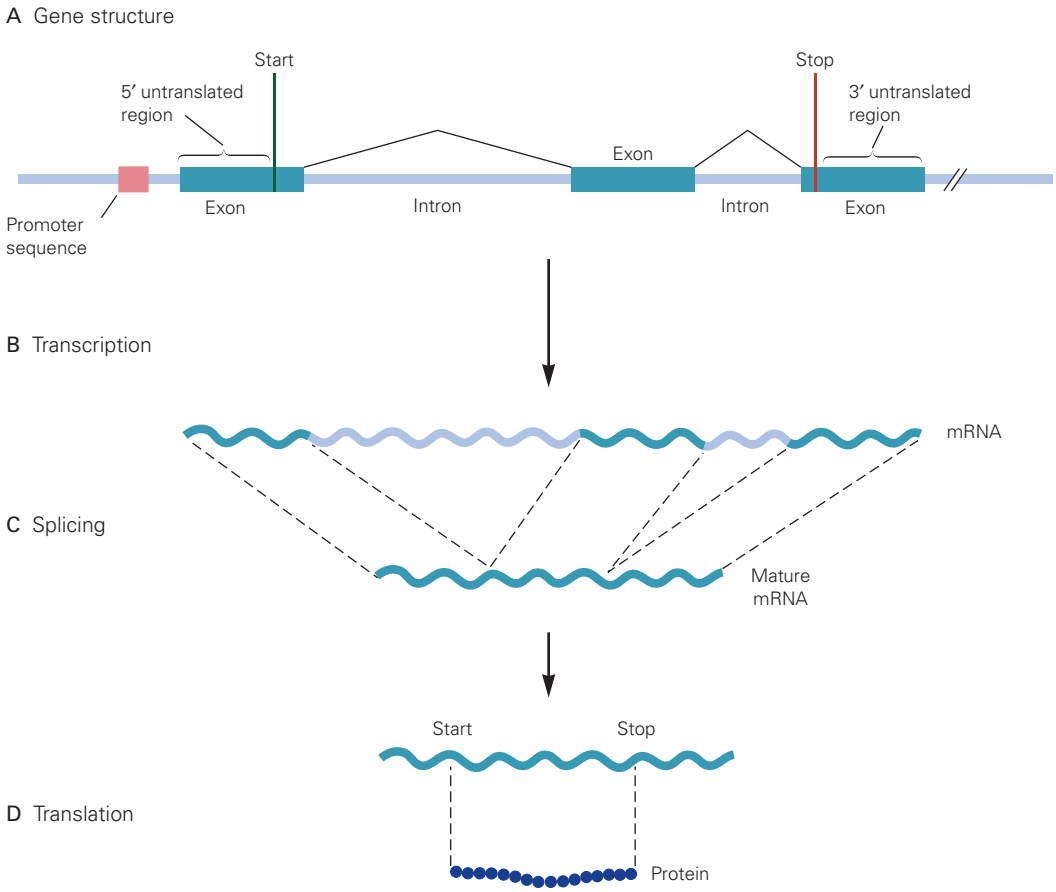

Figure 2–3 Gene structure and expression.

A. A gene consists of coding regions (exons) separated by noncoding regions (introns). Its transcription is regulated by noncoding regions such as promoters and enhancers that are frequently found near the beginning of the gene.

B. Transcription leads to production of a primary single-stranded RNA transcript that includes both exons and introns.

C. Splicing removes introns from the immature transcript and ligates the exons into a mature messenger RNA (**mRNA**), which is exported from the nucleus of the cell.

D. Translation of the mature mRNA produces a protein product.

permits a fixed number of genes to generate a vastly larger number of neuronal cell types and connections in the brain.

Although genes specify the initial development and properties of the nervous system, the experience of an individual and the resulting activity in specific neural circuits can itself alter the expression of genes. In this way, environmental influences are incorporated into the structure and function of neural circuits. Some of the principal goals of genetic studies are to unravel the ways that individual genes affect biological processes, the ways that networks of genes influence each other's activity, and the ways that genes interact with the environment.

Genes Are Arranged on Chromosomes

The genes in a cell are arranged in an orderly fashion on long, linear stretches of DNA called *chromosomes*. Each gene in the human genome is reproducibly located at a characteristic position (locus) on a specific chromosome, and this genetic "address" can be used to associate biological traits with a gene's effects. Most multicellular animals (including worms, fruit flies, and mice, as well as humans) are *diploid*; every somatic cell carries two complete sets of chromosomes, one from the mother and the other from the father.

Humans have about 20,000 genes but only 46 chromosomes: 22 pairs of autosomes (chromosomes

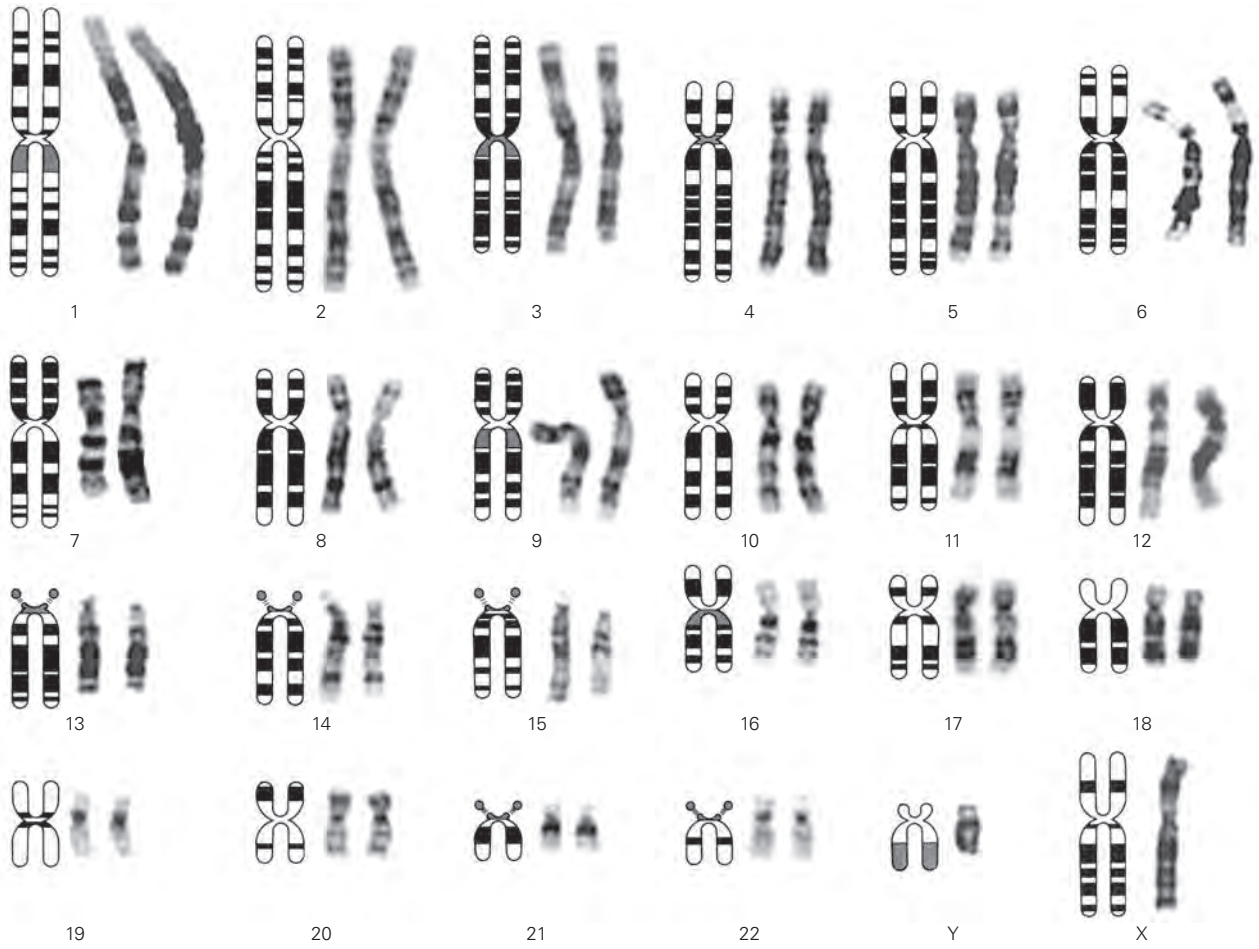

Figure 2–4 A map of normal human chromosomes at metaphase illustrates the distinctive morphology of each chromosome. Characteristic sizes and characteristic light and dark regions allow chromosomes to be distinguished from one another. (Adapted, with permission, from Watson, Tooze, and Kurtz 1983.)

that are present in both males and females) and two sex chromosomes (two X chromosomes in females, one X and one Y chromosome in males) (Figure 2–4). Each parent supplies one copy of each autosome to the diploid offspring. Each parent also supplies one X chromosome to female (XX) offspring, but XY males inherit their single X chromosome from their mothers and their single Y chromosome from their fathers. Sex-linked inheritance was discovered in fruit flies by Thomas Hunt Morgan in 1910. This pattern of sex-linked inheritance associated with the single X chromosome has been highly significant in human genetic studies, where certain X-linked genetic diseases are commonly observed only in males but are genetically transmitted from mothers to their sons.

In addition to the genes on chromosomes, a very small number of an organism's genes are transmitted through *mitochondria*, cytoplasmic organelles that carry out metabolic processes. Mitochondria in all children come from the ovum and therefore are transmitted from the mother to the child. Certain human disorders, including some neuromuscular degenerative diseases, some forms of intellectual disability, and some forms of deafness, are caused by mutations in the mitochondrial DNA.

The Relationship Between Genotype and Phenotype Is Often Complex

The two copies of a particular autosomal gene in an individual are called *alleles*. If the two alleles are identical, the individual is said to be *homozygous* at that locus. If the alleles vary because of mutations, the individual is *heterozygous* at that locus. Males

are *hemizygous* for genes on the X chromosome. A population can have a large number of alleles of a gene; for example, a single gene that affects human eye color, called *OCA2*, can have alleles that encode shades of blue, green, hazel, or brown. Because of this variation, it is important to distinguish the *genotype* of an organism (its genetic makeup) and the *phenotype* (its appearance). In the broad sense, a genotype is the entire set of alleles forming the genome of an individual; in the narrow sense, it is the specific alleles of one gene. By contrast, a phenotype is a description of a whole organism, and is a result of the expression of the organism's genotype in a particular environment.

If a mutant phenotype is expressed only when both alleles of a gene are mutated, the resulting phenotype is called *recessive*. This can occur if individuals are homozygous for the mutant allele or if they are carrying a different damaging allele in a given gene on each of their chromosomes (so-called *compound heterozygote*). Recessive mutations usually result from loss or reduction of a functional protein. Recessive inheritance of mutant traits is commonly observed in humans and experimental animals.

If a mutant phenotype results from a combination of one mutant and one wild-type allele, the phenotypic trait and mutant allele are said to be *dominant.* Some mutations are dominant because 50% of the gene product is not enough for a normal phenotype (*haploinsufficiency*). Other dominant mutations lead to the production of an abnormal protein or to the expression of the wild-type gene product at an inappropriate time or place; if this acts antagonistically to the normal protein product, it is called a *dominant negative* mutation.

The difference between genotype and phenotype is evident when considering the consequences of having one normal (wild-type) allele and one mutant allele of the same gene. Recent progress in gene discovery in a range of neurodevelopmental disorders, including autism and epilepsy, has demonstrated that the human genome is more sensitive to haploinsufficiency than previously appreciated. However, while the complete inactivation of both copies of a gene typically has a reliable effect, the severity and manifestation of haploinsufficiency vary to a greater degree from individual to individual, a phenomenon known as *variable*, *partial*, or *incomplete penetrance*.

Genetic variations that disturb development, cell function, or behavior in humans fall on a continuum between common alleles (also referred to as *polymorphisms*), which generally have small individual effects on biology and behavior, and rare variants, which may have larger biological effects (Box 2–1). While these categorizations are useful generalizations, there are nonetheless important cases in which common polymorphisms carry large disease risks; a common variation in the *APOE* gene, present in 16% of the population, results in a four-fold increase in the risk of late-onset Alzheimer disease.

Genes Are Conserved Through Evolution

The nearly complete nucleotide sequence of the human genome was reported in 2001, and the complete nucleotide sequences of many animal genomes have also been decoded. Comparisons between these genomes lead to a surprising conclusion: the unique human species did not result from the invention of unique new human genes.

Humans and chimpanzees are profoundly different in their biology and behavior, yet they share 99% of their protein-coding genes. Moreover, most of the approximately 20,000 genes in humans are also present in other mammals, such as mice, and over half of all human genes are very similar to genes in invertebrates such as worms and flies (Figure 2–5). The conclusion from this surprising discovery is that ancient genes that humans share with other animals are regulated in new ways to produce novel human properties, like the capacity to generate complex thoughts and language.

Because of this conservation of genes throughout evolution, insights from studies of one animal can often be applied to other animals with related genes, an important fact as animal experiments are often possible when experiments on humans are not. For example, a gene from a mouse that encodes an amino acid sequence similar to a human gene usually has a similar function to the *orthologous* human gene.

Approximately one-half of the human genes have functions that have been demonstrated or inferred from orthologous genes in other organisms (Figure 2–6). A set of genes shared by humans, flies, and even unicellular yeasts encodes the proteins for intermediary metabolism; synthesis of DNA, RNA, and protein; cell division; and cytoskeletal structures, protein transport, and secretion.

The evolution from single-cell organisms to multicellular animals was accompanied by an expansion of genes concerned with intercellular signaling and gene regulation. The genomes of multicellular

Box 2–1 Mutation: The Origin of Genetic Diversity

Although DNA replication generally is carried out with high fidelity, spontaneous errors called *mutations* do occur. Mutations can result from damage to the purine and pyrimidine bases in DNA, mistakes during the DNA replication process, and recombinations that occur during meiosis.

Changes in a single DNA base (also called a point mutation) within a coding region fall into five general categories:

1. A *silent mutation* changes a base but does not result in an obvious change in the encoded protein.
2. A *missense mutation* is a point mutation that results in one amino acid in a protein being substituted for another; increasingly these are being categorized using both informatics and empirical evidence into at least two subclasses: mutations that are damaging to protein function and those that may be functionally neutral.
3. A *nonsense mutation*, where a *codon* (a triplet of nucleotides) within the coding region specifying a specific amino acid is replaced by a stop codon, resulting in a shortened (truncated) protein product.
4. A *canonical splice site mutation* changes a nucleotide that specifies the exon/intron boundary.
5. A *frameshift mutation*, in which small insertions or deletions of nucleotides change the reading frame, leading to the production of a truncated or abnormal protein.

In the current literature, mutations falling into the latter four categories (including damaging missense mutations) are often referred to as *likely gene disrupting (LGD)* mutations.

The frequency of mutations greatly increases when the organism is exposed to chemical mutagens or ionizing radiation during experimental genetic studies. Chemical mutagens tend to induce *point mutations* involving changes in a single DNA base pair or the deletion of a few base pairs. Ionizing radiation can induce large insertions, deletions, or translocations.

In humans, point mutations occur at a low spontaneous rate in oocytes and sperm, leading to mutations present in the child but not in either parent, called de novo mutations. Each generation, between 70 and 90 single base changes are introduced into the entire genome (approximately 3 billion base pairs), of which one, on average, will cause a missense or nonsense mutation in a protein-coding gene. The number of de novo point mutations is increased in the children of older fathers, whereas the frequency of larger chromosome abnormalities is increased in the children of older mothers.

With the sequencing of the human genome in 2001 and increasingly high-resolution methods to detect genetic variation, it is also now clear that point mutations are not the only differences in sequence between humans. Certain sequences may be missing or repeated several times on a chromosome, and thus can have different numbers of copies in different individuals. When such variations encompass more than 1000 base pairs, they are called copy number variations (CNVs). Changes in more than a single base and less than 1000 base pairs are referred to as insertion/deletions (indels).

The contribution of any genetic variation to a disease or syndrome may be referred to as *simple* (or *Mendelian*) or *complex*. In general, a simple or Mendelian mutation is one that is sufficient to confer a phenotype without additional genetic risks. This does not imply that everyone with the mutation will show exactly the same phenotype. However, there is typically a highly reliable relationship between a specific disease allele and a phenotype, one that approaches a one-to-one relationship (as seen in sickle cell anemia or Huntington disease).

In contrast, a complex genetic disorder is one in which genetic risk factors change the probability of a disease but are not fully causal. This genetic contribution may involve rare mutations, common polymorphisms, or both, and is typically quite heterogeneous, with multiple different genes and alleles having the capacity to increase risk or play a protective role. Most complex disorders also involve a contribution from the environment.

animals, such as worms, flies, mice, and humans, typically encode thousands of transmembrane receptors, many more than are present in unicellular organisms. These transmembrane receptors are used in cell-to-cell communication in development, in signaling between neurons, and as sensors of environmental stimuli. The genome of a multicellular animal also encodes 1,000 or more different DNA-binding proteins that regulate the expression of other genes.

Many of the transmembrane receptors and DNA-binding proteins in humans are related to specific orthologous genes in other vertebrates and invertebrates. By enumerating the shared genetic heritage of the animals, we can infer that the basic molecular

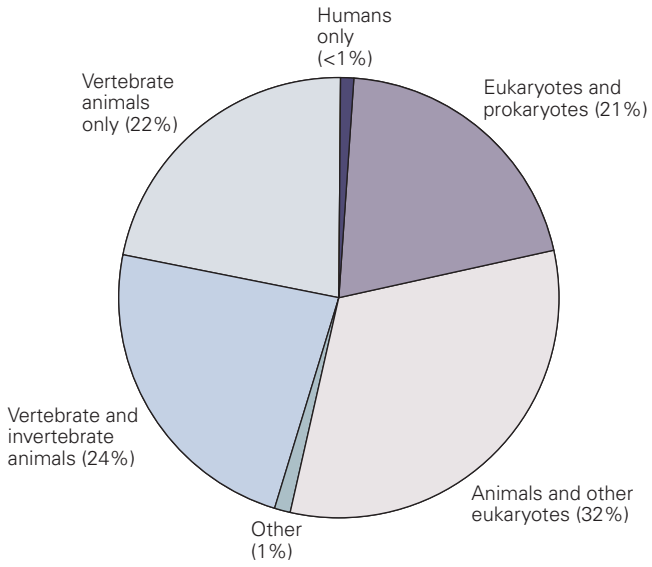

Figure 2–5 **Most human genes are related to genes in other species.** Less than 1% of human genes are specific to humans; other genes may be shared by all living things, by all eukaryotes, by animals only, or by vertebrate animals only. (Adapted, with permission, from Lander et al. 2001. Copyright © 2001 Springer Nature.)

pathways for neuronal development, neurotransmission, electrical excitability, and gene expression were present in the common ancestor of worms, flies, mice, and humans. Moreover, studies of animal and human genes have demonstrated that the most important genes in the human brain are those that are most conserved throughout animal phylogeny. Differences between mammalian genes and their invertebrate counterparts most often result from gene duplication in mammals or subtle changes in gene expression and function, rather than the creation of entirely new genes.

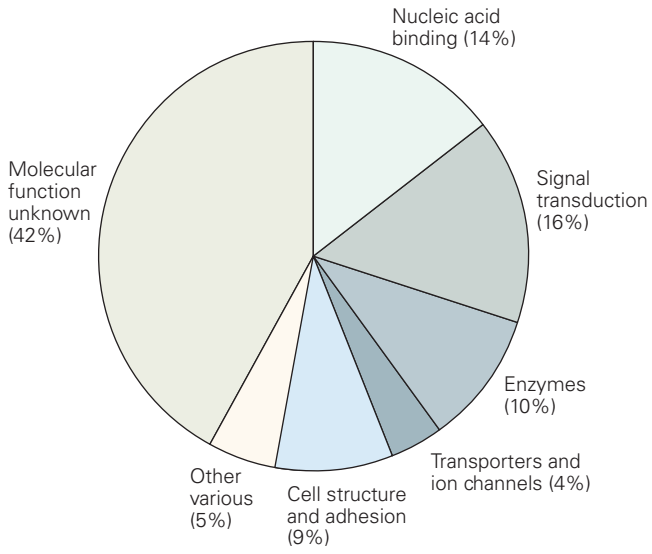

Figure 2–6 **The predicted molecular functions of 26,383 human genes.** (Adapted, with permission, from Venter et al. 2001.)

Genetic Regulation of Behavior Can Be Studied in Animal Models

Because of the evolutionary conservation between human and animal genes, studies in animal models of the relationships between the genes, proteins, and neural circuits that underlie behavior are likely to yield insight into these relationships in humans. Two important strategies have been applied with great success in the study of gene function.

In *classical genetic analysis*, organisms are first subjected to mutagenesis with a chemical or irradiation that induces random mutations and then screened for heritable changes that affect the behavior of interest, say, sleep. This approach does not impose a bias as to the kind of gene involved; it is a random search of all possible mutations that cause detectable changes. Genetic tracing of heritable changes allows the identification of the individual genes that are altered in the mutant organism. Thus the pathway of discovery in classical genetics moves from phenotype to genotype, from organism to gene. In *reverse genetics*, a specific gene of interest is targeted for alteration, a genetically modified animal is produced, and the animals with these altered genes are studied. This strategy is both focused and biased—one begins with a specific gene—and the pathway of discovery moves from genotype to phenotype, from gene to organism.

These two experimental strategies and their more subtle variations form the basis of experimental genetics. Gene manipulation by classical and reverse genetics is conducted in experimental animals, not in humans.

A Transcriptional Oscillator Regulates Circadian Rhythm in Flies, Mice, and Humans

The first large-scale studies of the influence of genes on behavior were initiated by Seymour Benzer and his colleagues around 1970. They used random mutagenesis and classical genetic analysis to identify mutations that affected learned and innate behaviors in the fruit fly *Drosophila melanogaster*: circadian (daily) rhythms, courtship behavior, movement, visual perception, and memory (Box 2–2 and Box 2–3). These induced mutations have had an immense influence on our understanding of the role of genes in behavior.

Box 2–2 Generating Mutations in Experimental Animals

Random Mutagenesis in Flies

Genetic analysis of behavior in the fruit fly (*Drosophila*) is carried out on flies in which individual genes have been mutated. Mutations can be made by chemical mutagenesis or by insertional mutagenesis, strategies that can affect any gene in the genome. Similar random mutagenesis strategies are used to create mutations in the nematode worm *Caenorhabditis elegans*, zebra fish, and mice.

Chemical mutagenesis, for example with ethyl methanesulfonate (EMS), typically creates random point mutations in genes. Insertional mutagenesis occurs when mobile DNA sequences called *transposable elements* randomly insert themselves into other genes.

The most widely used transposable elements in *Drosophila* are the P elements. P elements may be modified to carry genetic markers for eye color, which makes them easy to track in genetic crosses, and they may also be modified to alter expression of the gene into which they are inserted.

To cause P element transposition, *Drosophila* strains that carry P elements are crossed to those that do not. This genetic cross leads to destabilization and transposition of the P elements in the resulting offspring. The mobilization of the P element causes its transposition into a new location in a random gene.

Targeted Mutagenesis in Mice

Advances in molecular manipulation of mammalian genes have permitted precise replacement of a known normal gene with a mutant version. The process of generating a strain of mutant mice involves two separate manipulations. A gene on a chromosome is replaced by homologous recombination in a special cell line known as embryonic stem cells, and the modified cell line is incorporated into the germ cell population of the embryo (Figure 2–7).

The gene of interest must first be cloned. The gene is mutated, and a selectable marker, usually a drug-resistance gene, is then introduced into the mutated fragment. The altered gene is then introduced into embryonic stem cells, and clones of cells that incorporate the altered gene are isolated. DNA samples of each clone are tested to identify a clone in which the mutated gene has been integrated into the homologous (normal) site, rather than some other random site.

When a suitable clone has been identified, the cells are injected into a mouse embryo at the blastocyst stage (3 to 4 days after fertilization), when the embryo consists of approximately 100 cells. These embryos are then reintroduced into a female that has been hormonally prepared for implantation and allowed to come to term. The resulting embryos are chimeric mixtures between the stem cell line and the host embryo.

Embryonic stem cells in the mouse have the capability of participating in all aspects of development, including the germline. The injected cells can become germ cells and pass on the altered gene to future generations of mice. This technique has been used to generate mutations in various genes crucial to development or function in the nervous system.

Restricting Gene Knockout and Regulating Transgenic Expression

To improve the utility of gene knockout technology, methods have been developed that restrict deletions to cells in a specific tissue or at specific points in an animal's development. One method of regional restriction exploits the Cre/loxP system. The Cre/loxP system is a site-specific recombination system, derived from the P1 phage, in which the phage enzyme Cre recombinase catalyzes recombination between 34 bp loxP recognition sequences, which are normally not present in animal genomes.

The loxP sequences can be inserted into the genome of embryonic stem cells by homologous recombination such that they flank one or more exons of a gene of interest (called a floxed gene). When the stem cells are injected into an embryo, one can eventually breed a mouse in which the gene of interest is floxed and still functional in all cells of the animal.

A second line of transgenic mice can then be generated that expresses Cre recombinase under the control of a neural promoter sequence that is normally expressed in a restricted brain region. By crossing the Cre transgenic line of mice with the line of mice with the floxed gene of interest, the gene will only be deleted in those cells that express the Cre transgene.

In the example shown in Figure 2–8A, the gene encoding the NR1 (or GluN1) subunit of the *N*-methyl-D-aspartate (NMDA) glutamate receptor has been flanked with loxP elements and then crossed with a mouse line expressing Cre recombinase under control of the CaMKII promoter, which normally is expressed in forebrain neurons. In this particular line, expression was fortuitously limited to the CA1 region of the hippocampus, resulting in selective deletion of the NR1 subunit in this brain region (Figure 2–8B). Because the CaMKII promoter only activates gene transcription postnatally, early developmental changes are minimized by this strategy.

In addition to regional restriction of gene expression, control over the timing of gene expression gives the investigator an additional degree of flexibility and can exclude the possibility that any abnormality observed in the phenotype of the mature animal is the result of a developmental defect produced by the transgene. This can be done in mice by constructing a gene that can be turned on or off with a drug.

One starts by creating two lines of mice. Line 1 carries a particular transgene that is under control of the promoter tetO, which is ordinarily found only in bacteria. This promoter cannot by itself turn on the gene; it needs to be activated by a specific transcriptional regulator. Thus the second line of mice expresses a second transgene that encodes a hybrid transcription factor,

(*continued*)

Box 2–2 Generating Mutations in Experimental Animals (continued)

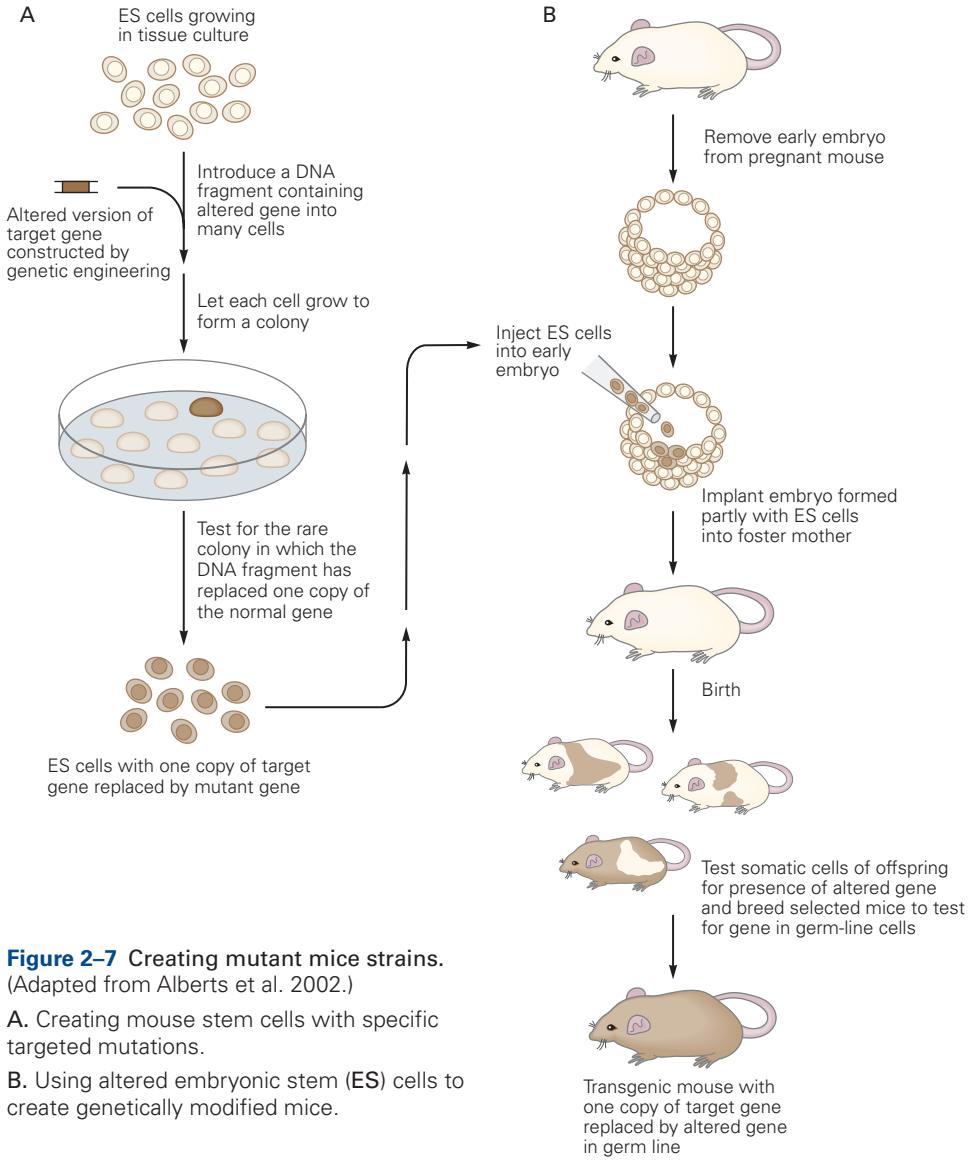

Figure 2–7 Creating mutant mice strains. (Adapted from Alberts et al. 2002.)

A. Creating mouse stem cells with specific targeted mutations.

B. Using altered embryonic stem (ES) cells to create genetically modified mice.

the tetracycline transactivator (tTA), which recognizes and binds to the tetO promoter. Expression of tTA can be placed under the control of a promoter in the mouse genome that normally drives gene transcription in only specific classes of neurons or specific brain regions.

When the two lines of mice are mated, some of the offspring will carry both transgenes. In these mice, the tTA binds to the tetO promoter and activates the downstream transgene. What makes the tTA transcription factor particularly useful is that it becomes inactivated when it binds certain antibiotics, such as tetracycline, allowing transgene expression to be regulated by administering antibiotics to mice. One can also generate mice that express a mutant form of tTA called reverse tTA (rtTA). This transactivator will not bind to tetO unless the animal is fed doxycycline. In this case, the transgene is always turned off unless the drug is given (Figure 2–9).

Altering Gene Function by RNA Interference and CRISPR

Finally, genes can be inactivated by targeting them with modern molecular tools. One such method is RNA interference, which takes advantage of the fact that most double-stranded RNAs in eukaryotic cells are routinely destroyed; the whole RNA is destroyed even if only part of it is double-stranded. By introducing a short RNA sequence that artificially causes a selected mRNA to become double-stranded, researchers can activate this process to reduce the mRNA levels for specific genes.

Another experimental tool is CRISPR, a method in which components of a bacterial immune system are deployed in nonbacterial cells to directly attack a selected DNA sequence. To target a gene with CRISPR,