Human Gene Mutation: Mechanisms and Consequences

10

Stylianos E. Antonarakis and David N. Cooper

Abstract A wide variety of different types of pathogenic mutation occur in the human genome, with many diverse mechanisms responsible for their generation. These types of mutation include single base-pair substitutions in coding, regulatory and splicing-relevant regions of human genes, and also micro-deletions, micro-insertions, duplications, repeat expansions, combined micro-insertions/deletions ("indels"), inversions, gross deletions and insertions, and complex rearrangements. A major goal of molecular genetic medicine is to be able to predict the nature of the clinical phenotype through ascertainment of the genotype. However, the extent to which this is feasible in medical genetics is very much disease, gene, and mutation dependent. The study of mutations in human genes is nevertheless of paramount importance for our understanding of the pathophysiology of inherited disorders and for optimizing diagnostic testing, as well as in guiding the design of new therapeutic approaches.

Contents

S.E. Antonarakis
Department of Genetic Medicine and Development
University of Geneva Medical School
and University Hospitals of Geneva
1 rue Michel-Servet, 1211 Geneva, Switzerland
e-mail: Stylianos.Antonarakis@unige.ch

D.N. Cooper (✉)
Institute of Medical Genetics, Cardiff University,
Heath Park, Cardiff CF14 4XN, UK
e-mail: cooperdn@cardiff.ac.uk

10.1 Introduction

A wide variety of different types of pathogenic mutation occur in the human genome, with many diverse mechanisms being responsible for their generation. These types of mutation include single base-pair substitutions in coding, regulatory and splicing-relevant regions of human genes, and also micro-deletions, micro-insertions, duplications, repeat expansions, combined micro-insertions/deletions ("indels"), inversions, gross deletions and insertions, and complex rearrangements. A major goal of molecular genetic medicine is to be able to predict the nature of the clinical phenotype through ascertainment of the genotype. However, the extent to which this is feasible in medical genetics is very much disease, gene, and mutation dependent. The study of mutations in human genes is nevertheless of paramount importance for our understanding of the pathophysiology of inherited disorders and for optimizing diagnostic testing, as well as in guiding the design of new therapeutic approaches.

The first description of the exact molecular defect in a human disease (sickle cell mutation, a substitution

M.R. Speicher et al. (eds.), *Vogel and Motulsky's Human Genetics: Problems and Approaches*,
DOI 10.1007/978-3-540-37654-5_10, © Springer-Verlag Berlin Heidelberg 2010

from Glu to Val at the 6th codon of the β-globin gene) was identified by Ingram in 1956, who found that the difference between hemoglobin A and hemoglobin S lies in a single tryptic peptide (158). His analysis was made possible by methods developed by Sanger for determining the structure of insulin and Edman's stepwise degradation of peptides. Since then, continuous advances have potentiated the identification of numerous disease-related genes and the discovery of thousands of underlying pathologic lesions. Single base-pair substitutions (68%) and micro-deletions (16.4%) are the most frequently encountered mutations in the human genome, the remainder comprising an assortment of micro-insertions (6.6%), indels (1.5%), gross deletions (5.6%), gross insertions and duplications (1.0%), inversions, repeat expansions (0.23%), and complex rearrangements (0.8%). Characterized mutations occur not only in coding sequences, but also in promoter regions, splice junctions, introns and untranslated regions, and any other functional region of the genome. Mutations can interfere with any stage in the pathway of expression from gene activation to synthesis and secretion of the mature protein product. This chapter attempts to provide an overview of the nature of mutations causing human genetic disease and then considers their consequences for the clinical phenotype. The interested reader is also referred to the third edition of this work for an in-depth discussion of mutation rates and factors influencing the generation of mutations. Two online databases contain information on disease-related (pathogenic) mutations: *Mendelian Inheritance in Man* (http://www.ncbi.nlm.nih.gov/Omim/) and the *Human Gene Mutation Database* (http://www.hgmd.org).

10.2 Neutral Variation/DNA Polymorphisms

The term *polymorphism* has been defined (336) as a "Mendelian trait that exists in the population in at least two phenotypes, neither of which occurs at a frequency of less than 1%." Polymorphisms are not rare. Indeed, there is enormous variation in the DNA sequences of any two randomly chosen human haploid genomes. Clearly, not all variations within a gene result in the abnormal expression of protein products. Indeed, single nucleotide substitutions/polymorphisms (SNP) occur in 1/~600–1,200 nucleotides in intervening sequences and flanking DNA (13, 75, 103, 244, 285, 342). These substitutions represent the most common forms of DNA polymorphism that can be used as markers for specific regions of the human genome. Similarly, some single nucleotide substitutions in the coding regions of genes may also be normal (nonpathogenic) polymorphic variants even if they result in nonsynonymous substitutions of the polypeptide product (247). For example, there are three common forms of the β-globin (*HBB*) gene on chromosome 11p. These forms differ at five nucleotides, one of which lies within the first exon of the gene and results in a synonymous codon. The average human gene contains >120 biallelic polymorphisms, 46 of which occur with a frequency >5% and 5 within the coding region (78).

Some polymorphisms entail the alteration of an encoded amino acid, e.g., the Lewis *Le* alleles of the *FUT3* gene (245), whereas others may introduce a stop codon that serves to inactivate the gene in question, e.g., the secretor *se* allele of the *FUT2* gene present in 20% of the population (176). However, not all polymorphisms are SNPs. Examples of other types of gene-associated polymorphism in the human genome include triplet repeat copy number (e.g., in the *FMR1* gene; see Sect. 9.2.1.3), gross gene deletion (e.g., *GSTM1* and *GSTT1*) (273), gene duplication (e.g., *HBG2*) (318) intragenic duplication (e.g., *IVL*) (132), micro-insertion/deletion (e.g., *PAI1*) (84), indel (e.g., *APOE*) (120), gross insertion (e.g., the inserted *Alu* sequence in intron 16 of the *ACE* gene) (277), inversion (e.g., the 48 kb Xq28 inversion involving the *EMD* and *FLN1* genes) (307)) and gene fusion (e.g., between the *RCP* and *GCP* visual pigment genes) (241)). It can be seen that the mutational spectrum of polymorphisms in the human genome is qualitatively different to that underlying human disease; they may vary in terms of location and frequency but otherwise they display remarkable similarities indicative of the same underlying mutational mechanisms.

It is likely, however, that some SNPs, both frequent and rare, alter the risk for common complex human phenotypes. A public SNP database now contains more than 10 million entries (dbSNP; http://www.ncbi.nlm.nih.gov/SNP/index.html). An international project has recently been completed, termed the "HapMap project" (8, 68, 144), the goal of which was to define the

patterns of common SNP genetic variation in a sample of 270 DNAs from individuals of European, African, Chinese, and Japanese origin (http://www.hapmap.org/). The data obtained from this project constitute approximately 2.8 million SNPs and are publicly available. The results of this project are likely to contribute significantly to our understanding of both common and rare human genetic disorders and traits.

Another form of polymorphic variation in our genome is the presence of variable numbers of tandem repeats (VNTRs). The repeat unit can be 10–60 nucleotides in length and many different alleles may exist at a given locus (164, 358). The combination of a VNTR and single nucleotide substitutions within the repeat unit results in an extremely high level of polymorphic variability that can be used as a unique bar-code to distinguish different individuals (165). The introduction of the polymerase chain reaction (PCR) (286) permitted the rapid detection and analysis of variation in short sequence repeats (SSR), e.g., $(GT)_n$ repeats (212, 347) . These are common polymorphisms that occur on average once every 50 kilobases (kb) of genomic DNA. The SSRs also display many alleles and the repeat unit can be two, three, four, five or more nucleotides. Poly(A) tracts may also be polymorphic, exhibiting variation in the number of A residues (101); many of these polymorphisms are localized at the ends of *Alu* repetitive elements. Another kind of polymorphism in the human genome involves the presence or absence of retrotransposons (i.e., *Alu* or LINE repetitive elements or pseudogenes) at specific locations (10, 71) . Furthermore, duplicational polymorphisms have also been reported for some human genes, e.g., *HBA1*, *PRB1-4*, *HBZ*, *CYP21/C4A/C4B* (39, 71).

The use of comparative genomic hybridization against BAC or oligonucleotide arrays has revealed extensive copy number polymorphisms/variation of sizable genomic regions (CNP or CNV) (156, 294, 300). Details of more than 1,400 such genomic variants may be found in the following databases: *Human Structural Variation Database*, http://paralogy.gs.washington.edu/structuralvariation; *Database of Genomic Variants*, http://projects.tcag.ca/variation. A CNV map of the human genome of the 270 "HapMap" individuals has revealed a total of 1,440 CNV such regions which cover some 360 megabases (12% of the genome) (274). The functional significance, if any, of the majority of these polymorphic variants is however unknown. Copy number variants may predispose to phenotypic variability. For example, it has recently been observed that copy number variation of the orthologous rat and human *Fcgr3/FCGR3B* genes is a determinant of susceptibility to immunologically mediated glomerulonephritis (5). Copy number variants in the *CCL3L1* and *DEFB4* genes have also been found to be associated with increased susceptibility to HIV infection and Crohn's disease, respectively (111, 126).

Deletional polymorphisms are also remarkably frequent in the human genome: a typical individual has been estimated to be hemizygous for some 30–50 deletions >5 kb, spanning >550 kb in total and encompassing >250 known or predicted genes (67, 222). Since such deletions appear to be in linkage disequilibrium with neighboring SNPs, we may surmise that they share a common evolutionary history (145).

Human DNA polymorphisms have proven extremely useful in developing linkage maps, for mapping monogenic and polygenic complex disorders, for determining the origin of aneuploidies and chromosomal abnormalities, for distinguishing normal from mutant chromosomes in genetic diagnoses, for performing forensic, paternity, and transplantation studies, for studying the evolution of the genome, the loss of heterozygosity in certain malignancies, the detection of uniparental disomy, the instability of the genome in certain tumors, recombination at the level of the genome, the study of allelic expression imbalance, and the development of haplotype maps of the genome. However, in studying the role of a candidate gene in a given disorder, it is imperative to distinguish between pathogenic mutations that cause a clinical phenotype and the polymorphic variability of the normal genome.

10.3 Disease-Causing Mutations

10.3.1 The Nature of Mutation

Figure 10.1a depicts the frequencies of the various mutation types responsible for molecularly characterized human genetic disorders, as recorded in the *Human Gene Mutation Database* (HGMD) (http://www.hgmd.org) and elsewhere (196, 312)). HGMD records each mutation *once* regardless of the number of independent occurrences of that lesion. Figure 10.1b shows the frequency of the first mutation per disease recorded in *Mendelian Inheritance in Man* (MIM)

10

(http://www.ncbi.nlm.nih.gov/Omim) and by Antonarakis and McKusick (11). As of March 31, 2009, HGMD contained some 88,317 different mutations in 3,337 human genes, whereas MIM contained examples of allelic variants in 2,514 human genes.

10.3.1.1 Nucleotide Substitutions

Single nucleotide substitutions are the most frequent pathologic mutations in the human genome (Fig. 10.1). Most of these alterations occur during DNA replication, which is an accurate, yet error-prone, multistep process. The accuracy of DNA replication depends on the fidelity of the replicative step and the efficiency of the subsequent error correction mechanisms (214). Analysis of more than 7,000 missense and nonsense mutations associated with human disease has indicated that the most common nucleotide substitution for T (thymine) is to C (cytosine), for C it is to T, for A (adenine) it is to G (guanine) and for G it is to A (195). Transitions are therefore much more common than transversions. Some 61% of the missense/nonsense mutations currently logged in HGMD are transitions (T to C, C to T, A to G, G to A), whilst 39% are transversions (T to A or G, A to T or C, G to C or T, C to G or A).

Among single nucleotide substitutions there is one that clearly predominates, and it represents the most common type of mutational lesion: CpG dinucleotides mutate to TpG at a frequency that is about 5 times that of mutations in all other dinucleotides (15, 195, 361, 363). This substitution, which generates TG when it occurs on one DNA strand and CA ("CG to TG or CA rule") when it is on the other, is a major cause of human genetic disease. This phenomenon was first observed in the factor VIII (*F8*) gene in cases of hemophilia A (361), but it was soon noted in studies of many other genes (74). In hemophilia A, CG to TG or CA mutations account for 46% of point mutations in unrelated patients (14). In the HGMD (312) (http://www.hgmd.org), such mutations currently account for ~20% of the total number of missense/nonsense mutations. Among CpG dinucleotide mutations, transitions to TG or CA account for ~90% of substitutions. The mechanism of this common type of mutation appears to be methylation-mediated deamination of 5-methylcytosine (5mC). In eukaryotic genomes, 5mC occurs predominantly in CpG dinucleotides, most of which appear to be methylated (see (70) for review). 5mC then undergoes spontaneous nonenzymatic deamination to form thymine (Fig. 10.2). There is a bias in terms of the origin of CpG to TpG mutations: most occur in male germ cells

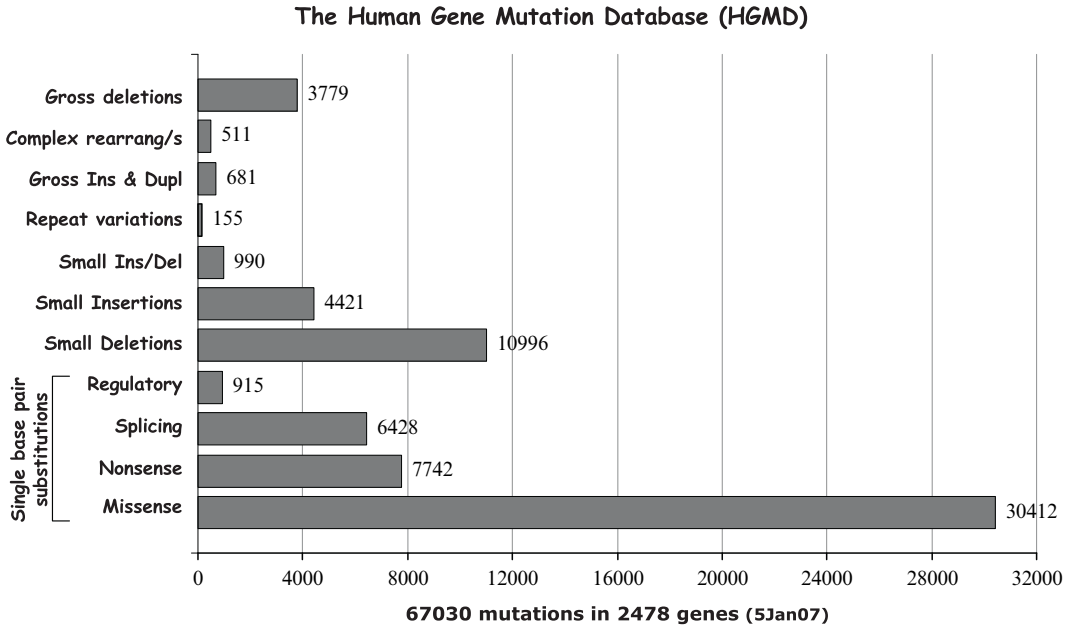

Fig. 10.1 Spectrum of different types of human gene mutations logged in the Human Gene Mutation Database as of January 2007

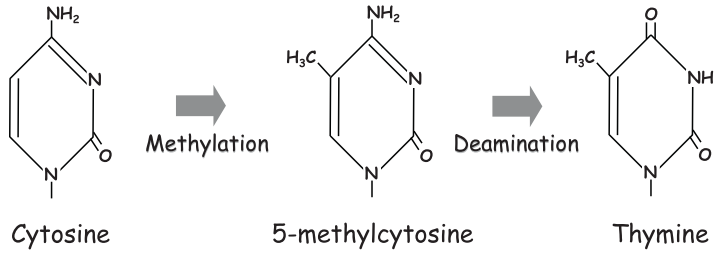

Fig. 10.2 Schematic representation of cytosine, 5'-methylocystosine, and thymine, and the chemical events involved in the mutational transformation of cytosine to thymine

(the male/female ratio is 7 to 1). One reason for this may be that sperm DNA is heavily methylated, whereas oocyte DNA is comparatively undermethylated (99). Another reason may be the considerably higher number of germline cell divisions in males than in females (154).

In a recent analysis, the average direct estimate of the combined rate of all mutations was 1.8×10^{-8} per nucleotide per generation. Single nucleotide substitutions were found to be approximately 25 times more common than all other mutations, whilst deletions were approximately three times as common as insertions; complex mutations were very rare, and the CpG context was found to increase substitution rates by an order of magnitude (185). Rates of different kinds of mutations were also found to be strongly correlated across different loci (185).

10.3.1.2 Micro-Deletions and Micro-Insertions

Deletions or insertions of a few nucleotides are also fairly common as a cause of human inherited disease. Most of these are less than 20 bp in length. Indeed, the majority of micro-deletions involve <5 nucleotides. In HGMD, the deletion of 1 bp accounts for 48% of small deletions, whilst an additional 30% involve 2 or 3 nucleotides. The majority of micro-deletions recorded (78%) result in an alteration of the reading frame. Most micro-deletions occur in regions that contain direct repeats of 2 bp or more. The most common length of direct repeat is 3 bp (48% of direct repeats associated with short deletions (15)). The most plausible mechanism for small deletions mediated by the presence of direct repeats is the slipped mispairing model (104) (Fig. 10.3). In addition, deletions of one or a few nucleotides frequently occur in runs of the same nucleotide, e.g., a poly(T) region (198). Finally, inverted repeats

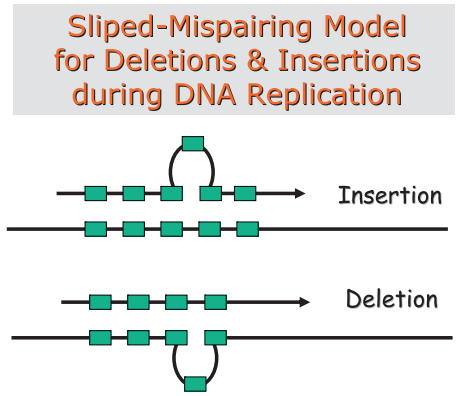

Fig. 10.3 Schematic representation of the slipped mispairing model for deletions and insertions during DNA replication

and "symmetric elements" are also frequently found in the immediate vicinity of micro-deletions (73, 289). Krawczak and Cooper (193) identified a consensus sequence – TG(A/G)(A/G)(G/T)(A/C) – which they claimed represented a deletion hotspot.

Micro-insertions (again up to 20 nucleotides) are rarer than micro-deletions; thus, in HGMD there are three times as many micro-deletions as micro-insertions (Fig. 10.1a). Nearly half of these involve the insertion of only 1 nucleotide (Fig. 10.4). As is the case with micro-deletions, most micro-insertions lead to alterations of the reading frame and are located in regions containing direct or inverted repeats or runs of the same nucleotide. Details of possible mechanisms of generation during replication can be found in elsewhere (72). There are as yet insufficient data available to estimate the frequency ratio of micro-insertions or micro-deletions in male or female germ cells. In the case of such lesions in factor VIII (*F8*) gene, 56% of micro-deletions/-insertions have been reported to occur in DNA regions harboring direct repeats or runs of the same nucleotide (14).

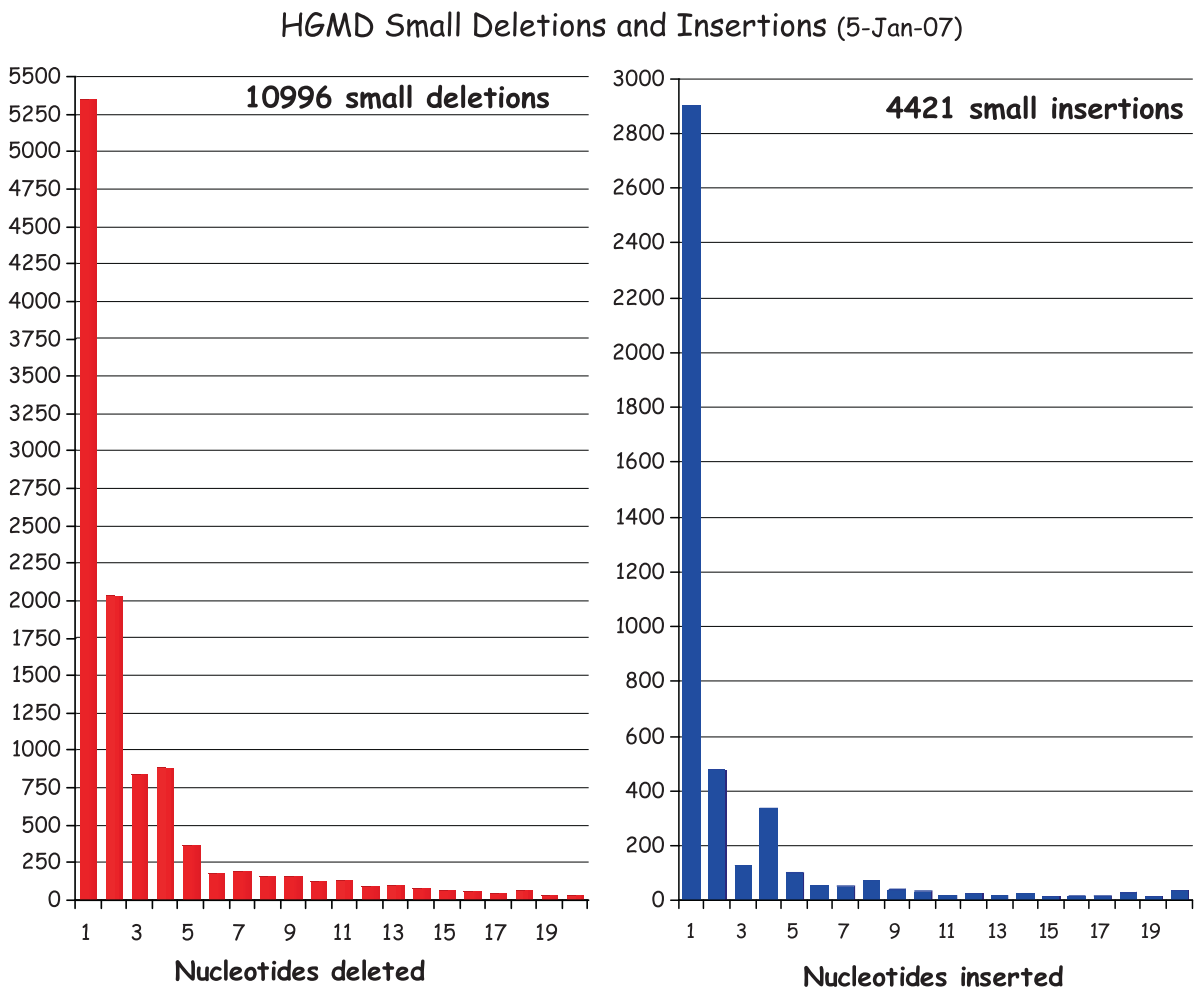

Fig. 10.4 Size distribution of short (<20 bp pathogenic human gene deletions and insertions (HGMD; http://www.hgmd.org, 5 January 2007)

HGMD data (3,767 micro-deletions and 1,960 micro-insertions) were used to perform a meta-analysis of micro-deletions and micro-insertions causing inherited disease, both defined as involving ≤20 bp DNA (23). A positive correlation was noted between the micro-deletion and micro-insertion frequencies for 564 genes for which both micro-deletions and micro-insertions have been reported. This is consistent with the view that the propensity of a given gene/sequence to undergo micro-deletion is related to its propensity to undergo micro-insertion. While micro-deletions and micro-insertions of 1 bp constitute, respectively, 48 and 66% of the corresponding totals, the relative frequency of the remaining lesions correlates negatively with the length of the DNA sequence deleted or inserted. Many micro-deletions and micro-insertions of >1 bp can potentially be explained in terms of slippage mutagenesis, involving the addition or removal of one copy of a mono-, di-, or trinucleotide tandem repeat. The frequency of in-frame 3 and 6 bp micro-insertions and micro-deletions was, however, found to be significantly lower than that of mutations of other lengths, suggesting that some of these in-frame lesions may not have come to clinical attention. Various sequence motifs were found to be overrepresented in the vicinity of both micro-insertions and micro-deletions, including the heptanucleotide CCCCCTG that shares homology with the complement of the 8-bp human minisatellite conserved sequence/

chi-like element (GCWGGWGG). The "indel hotspot" GTAAGT (and its complement ACTTAC) were also found to be overrepresented in the vicinity of both micro-insertions and micro-deletions, thereby providing a first example of a mutational hotspot that is common to different types of gene lesion. Other motifs overrepresented in the vicinity of micro-deletions and micro-insertions included DNA polymerase pause sites and topoisomerase cleavage sites. Several novel micro-deletion/micro-insertion hotspots were noted, and some of these exhibited sufficient similarity to one another to justify terming them "super-hotspot" motifs. Analysis of DNA sequence complexity also demonstrated that a combination of slipped mispairing mediated by direct repeats, and secondary structure formation promoted by symmetric elements, can account for the majority of micro-deletions and micro-insertions. Thus, micro-insertions and micro-deletions exhibit strong similarities in terms of the characteristics of their flanking DNA sequences, implying that they are generated by very similar underlying mechanisms.

A similar analysis of micro-deletions and micro-insertions in 19 human genes yielded evidence for an elevated micro-deletion rate at YYYTG and an elevated micro-insertion rate at TACCRC and ATMMGCC (186). Kondrashov and Rogozin (186) also found that ~45% of micro-deletions led to the removal of a repeated sequence, an event they termed "deduplication" in order to highlight the identity of the deleted sequence and the sequence abutting the site of deletion.

10.3.1.3 Expansion/Copy Number Variation of Trinucleotide (and Other) Repeat Sequences

Another mechanism of human gene mutation causing hereditary disease is the instability of certain trinucleotide repeats and their expansion in affected genes (44, 218, 282). A growing number of disorders (in excess of 150 are now recorded in HGMD), the majority of which involve neuromuscular tissues, have been found to be due to, or associated with, the expansion of repeat sequences; of these, 23 are expansions of triplet repeats. The first such disease was fragile X, a common cause of male mental retardation, which mapped to chromosome Xq27.3. Table 10.1 lists some examples of these disorders, which include Huntington disease, myotonic dystrophy, spinobulbar muscular atrophy, spinocerebellar ataxia 1, spinocerebellar ataxia 3 or Machado-Joseph disease, the fragile E site, and dentatorubral pallidoluysian atrophy. Genetic "anticipation" (the earlier onset and increasingly severe phenotype in successive generations) is a common phenomenon in these disorders (141). The trinucleotide involved is usually either CAG or CGG, but occasionally CTG, GCG, or GAA. It can be located in the 5' untranslated region (UTR), as in the case of the *FMR1* gene underlying fragile X, within the coding region (as in Huntington disease, SCA1, SCA3, and Kennedy disease) encoding poly(Gln), in an intron, as in Friedreich ataxia (*FXN*) and myotonic dystrophy type 2 (*ZNF9*), or in the 3' UTR, as in myotonic dystrophy type 1 (*DMPK*; Table 10.1, Fig. 10.5). The expansion of the triplet repeat either prevents its expression (329), results in a dominant gain–of–function mutation mediated by the longer poly(Gln) peptide (150), or alters the RNA processing of other genes (211, 288).

The trinucleotide repeats are usually polymorphic in human populations. Rarely, however, the number of trinucleotide repeats lies within a high-risk category that is termed "premutation." In such a case, the premutation exhibits a high probability of further expansion (instability) to yield disease-related alleles ("full mutation"). In fragile X, for example, the normal polymorphic alleles of the CGG repeat contain between 10 and 50 triplets, the premutation between 50 and 200, and the full mutation more than 200 triplets (119). Expansion of premutations to full mutations only occurs during female meiotic transmission. The probability of repeat expansion correlates with repeat copy number in the premutated allele. Since the premutation must precede the appearance of a full mutation, all mothers of affected children carry either a full mutation or a premutation (119). Premutation alleles may also be associated with late-onset movement disorders and premature ovarian failure (69, 163)

The precise mechanism of repeat expansion is unclear, although it is known that DNA polymerase progression is blocked by CTG and CGG repeats and the resultant idling of the polymerase could serve to catalyze slippage, leading to repeat expansion (171). In the case of spinocerebellar ataxia 1 (SCA1), interruption of the CAG repeat with a CAT unit is associated with more stable trinucleotide repeat (56). More details about these "dynamic mutations" can be found in the appropriate sections covering individual disorders, and have also been treated by Wells (348). Short

Table 10.1 Various examples of disorders of trinucleotide and other repeat expansions

	Disorder	Inheritance	Gene	Chr	OMIM#	Repeat	Normal	Mutant	Repeat location	**Mutation** type	**Parental** gender bias
1	Fragile X syndrome	XLD	*FMR1*	Xq27.3	309,550	CGG	6–52	60–200 premutation 230-1000 full mut	5'UTR	LOF, FraX	Maternal
2	Fragile E mental retardation	XLD	*FMR2*	Xq28	309,548	GCC	7–35	130–150 premutation 230–750 full mut	5'UTR	LOF, FraX	ND
3	Myotonic dystrophy	AD	*DMPK*	19q13	160,900	CTG	5–37	50–3,000	3'UTR	?Dom negative	Maternal
4	Spinobulbar muscular atrophy	XLR	*AR*	Xq13-21	313,700	CAG	11–33	38–66	Coding	GOF, LOF	ND
5	Huntington disease	AD	*HD*	4p16.3	143,100	CAG	6–39	36–121	Coding	GOF	Paternal
6	Dentatorubro-pallidoluysian atrophy	AD	*DRPLA*	12p13.31	125,370	CAG	6–35	51–88	Coding	GOF	Paternal
7	Spinocerebellar ataxia 1	AD	SCA1/*ATX1*	6p23	601,556	CAG	6–39	41–81	Coding	GOF	Paternal
8	Spinocerebellar ataxia 2	AD	SCA2/*ATX2*	12q24.1	601,517	CAG	14–31	35–64	Coding	GOF	Paternal
9	Spinocerebellar ataxia 3	AD	SCA3/*MJD1*	14q32.1	109,150	CAG	12–41	40–84	Coding	GOF	Paternal
10	Spinocerebellar ataxia 6/ Episodic ataxia type 2	AD	*CACNA1A*	19p13	601,011	CAG	7–18	20–23 EA2 21–27 SCA6	Coding	ND	ND
11	Spinocerebellar ataxia 7	AD	*SCA7*	3p12-13	164,500	CAG	7–17	38–130	Coding	GOF	Paternal
12	Friedreich ataxia	AR	*FRDA1*	9q13-21.1	229,300	GAA	6–34	80 premutation 112-1700 full mut	Intron 1	LOF, FraX	Maternal
13	Progressive myoclonus epilepsy 1	AR	*CSTB*	21q22.3	601,145	CCCCG-CCCCGCG	2–3	35–80	5' flanking	LOF	Paternal
14	Synpolydactyly	AD	*HOXD13*	2q31-q32	142,989	(GCG)n(GCT)n(GCA)n	15	22–29	Coding	ND	??
15	Oculopharyngeal muscular dystrophy	AD	*PABP2*	14q11.2-q13	602,279	GCG	6	7–13	Coding	ND	??

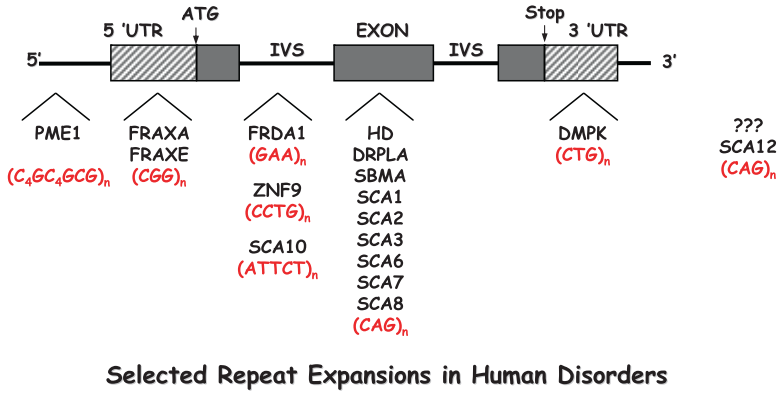

Fig. 10.5 Location of the repeat expansion in selected human disorders

expansions of GCG trinucleotide codons encoding Ala have been observed in the *HOXD13* gene causing dominant polydactyly, and in the *PABP2* gene causing oculopharyngeal muscular dystrophy (37, 237). These mutations may be due to unequal crossing-over rather than polymerase slippage. Generally speaking, it is likely that repeat instability is a consequence of the resolution of unusual secondary structure intermediates during DNA replication, repair and recombination (264).

A repeat expansion of 12 nucleotides (CCCCGCCCCGCG) in the 5' flanking region of the *CSTB* gene causes one form of the recessive progressive myoclonus epilepsy (EPM1) (201). This indicates that repeat sequences other than trinucleotides can become expanded and cause human disorders. This particular expansion silences the *CSTB* gene, probably because it alters the spacing of transcription factor binding sites from each other and/or the transcriptional initiation site (202).

A tetranucleotide repeat expansion $(CCTG)_n$ in intron 1 of the *ZNF9* gene causes myotonic dystrophy type 2 (211). This expansion can be between 75 and 11,000 repeats in length. The expansion of the pentanucleotide repeat $(ATTCT)_n$ is responsible for the phenotype of spinocerebellar ataxia 10 (SCA10). The expansion occurs in intron 9 of the *SCA10* gene and can be up to 22.5 kb in length (221). Expansions of even longer repeats have been reported. In Usher syndrome type 1C, for example, there is an expansion of a 45-bp VNTR in intron 5 of the *USH1C* gene (9 tandem repeats instead of the usual less than 6 such repeats); this expansion has been predicted to inhibit transcription of the gene (332). There are also cases in which a large repeat expansion is not associated with a particular phenotype, e.g., the expansion of an AT-rich 33-mer repeat in the dictamycin-sensitive fragile site 16B (364).

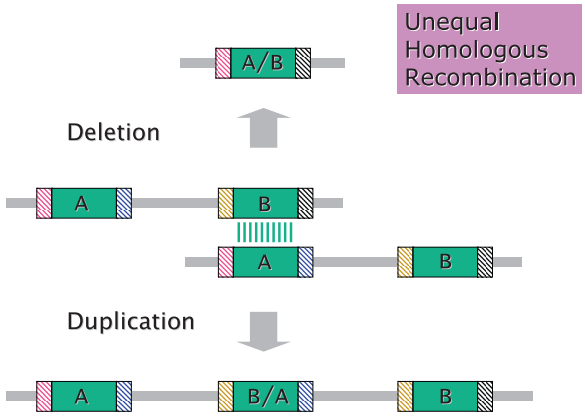

Fig. 10.6 Homologous unequal combination between similar regions of sequences A and B. The recombination events cause either deletions or duplications. In the case of a deletion, a hybrid sequence is generated, with the first part from sequence A and the second, from sequence B. The middle sequence in the duplication product is also a hybrid sequence, with the first part from sequence B and the second, from sequence A

10.3.1.4 Gross Deletions

Gross deletions are common causes of certain disorders and rare in others. In most of the X–linked disorders, for example, large deletions account for about 5% of molecular defects. In other disorders, however, such as steroid sulfatase deficiency, large deletions of the *STS* gene account for 84% of patients (24). The same is true for disorders such as Duchenne muscular dystrophy, growth hormone deficiency and α-thalassemia (92, 242, 335).

A considerable number of large deletions are probably generated by mispairing of homologous sequences and unequal recombination (Fig. 10.6). One of the best examples of homologous unequal recombination is the case of α-globin genes on chromosome 16p. As a result

10

of a recent evolutionary duplication of the α-globin genes, extensive regions of sequence homology exist between the two closely linked α-genes. Unequal crossover results in either deletion of one α-gene or the creation of a fusion hybrid gene (106). The reciprocal product chromosomes carry three α-genes and are not associated with a clinical phenotype (127). Another example of a fusion gene resulting from an unequal crossover is the case of hemoglobin Lepore characterized by a hybrid gene between the δ- and β-globin genes on chromosome 11p (21). In the case of steroid sulfatase deficiency, the deletion can be as large as one megabase (Mb) (299). In Kallmann syndrome, translocation can occur as a result of unequal mispairing of X- and Y-homologous sequences (138).

A number of common genetic disorders are due to large deletions (or duplications) caused by unequal crossing-over of homologous sequences. Figure 10.7 depicts various examples, which include a 1.5-Mb deletion of 17p12 in hereditary neuropathy with liability to pressure palsies (HNPP) (276), deletion of 1.5 Mb of 17q11.2 in neurofibromatosis type 1 (98), deletion of 1.6 Mb of 7q11.23 in Williams syndrome (115), deletion of 5 Mb of 17p11.2 in Smith-Magenis syndrome (169), deletion of either 3 Mb or, more rarely, 1.5 Mb of 22q11 in DiGeorge and velo-cardio-facial syndrome (102, 298), and 4-Mb deletions of 15q in Prader-Willi and Angelman syndromes (54). A recurrent deletion of ~0.5 Mb of 17q21.3, which may be mediated by a common inversion polymorphism, has also been described (188, 301, 305, 311). For a review of chromosomal "duplicons," the low copy repeats that mediate deletions and duplications, see (168). It has been estimated that approximately 5% of the human genome is duplicated either intra- or inter-chromosomally (22). The large deletions or duplications (see below) due to duplicon crossover are also termed "genomic disorders." A recent review of such genomic disorders may be found in (303).

In many cases of large deletion, homologous unequal crossover occurs between repetitive elements such as *Alu* sequences. The *Alu* repeat is the most abundant repetitive element, with about 1.5×10^6 copies in the human genome (86, 203). The element is about 300 bp in length and consists of two similar regions separated by a short A-rich region. Unequal crossover can occur between *Alu* sequences oriented either in opposite directions or in the same direction.

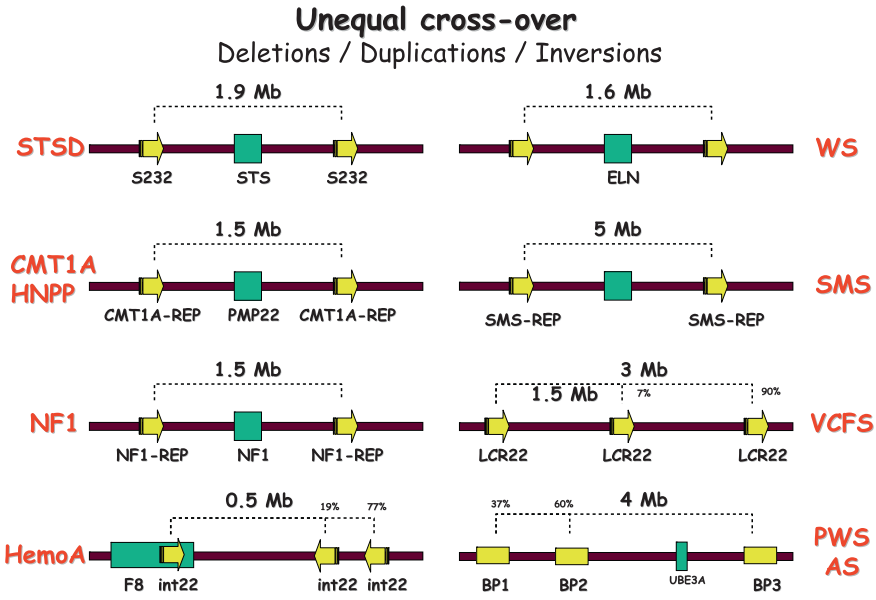

Fig. 10.7 Genes, duplicons, and diseases. Unequal crossover between homologous sequences (duplicons) produce either deletions or duplications of the DNA between the duplicons. The duplicons are shown by *arrows* or *clear boxes*. Genes included in the duplications/deletions are shown as *dark boxes* (*AS* Angelman syndrome *CMTA1* Charcot-Marie-Tooth type 1, *HemoA* hemophilia A, *HNPP* hereditary neuropathy with liability to pressure palsies, *NF1* neurofibromatosis 1, *PWS* Prader-Willi syndrome, *SMS* Smith-Magenis syndrome, *VCFS* velo-cardio-facial syndrome, *STSD* steroid sulfatase deficiency, *WS* Williams syndrome)

In addition, unequal crossings over have been noted between *Alu* elements and nonrepetitive DNA sequences without homology to *Alu*s. The best examples of *Alu-Alu* recombination occur in the genes encoding the low-density lipoprotein receptor (*LDLR*), which underlies familial hypercholesterolemia, and complement component 1 inhibitor (*C1I*) (205, 313)). All but one of the breakpoints associated with *LDLR* gene deletions occur within *Alu* repeats. By contrast, deletions in other *Alu*-rich genes (e.g., *GLA1*) do not necessarily involve *Alu* repetitive elements (189).

Nonhomologous (illegitimate) recombination occurs between two DNA sites that share minimal sequence homology of a few basepairs. This type of recombination during meiosis or alternatively, slipped mispairing during DNA replication mediated by short (2–8) nucleotide direct repeats flanking the deletions is a common finding in many instances of large gene deletions (281). Such deletions have been studied, for example, in hemophilia A; a compilation of 46 junctions from large deletions revealed that about 50% shared 2- to 6-bp homology at the breakpoint junction, as compared with only 17% in which the deletion was due to *Alu-Alu* recombination (356). Similar results have been reported from the intron 7 deletion hotspot in the Duchenne muscular dystrophy (*DMD*) gene; 8/9 deletion breakpoints examined were found to be flanked by DNA sequences with minimal homology (223).

It has also been proposed that alternative DNA conformations may trigger genomic rearrangements through recombination-repair activities. Distance measurements have indicated the significant proximity of alternating purine-pyrimidine and oligo(purine.pyrimidine) tracts to breakpoint junctions in 222 gross deletions and translocations, respectively, involved in human diseases. In 11 deletions analyzed, breakpoints could be explained by non-B DNA structure formation (20).

The Gross Rearrangement Breakpoint Database (GRaBD; http://www.uwcm.ac.uk/uwcm/mg/grabd/). This database was established primarily for the analysis of the sequence context of translocation and deletion breakpoints in a search for characteristics that might have rendered these sequences prone to rearrangement (3). GRaBD, which contains 397 germline and somatic DNA breakpoint junction sequences derived from 219 different rearrangements underlying human inherited disease and cancer, represents a large but not comprehensive collection of sequenced gross gene rearrangement breakpoint junctions. Analysis of these breakpoints has extended our understanding of illegitimate recombination by highlighting the importance of secondary structure formation between single-stranded DNA ends at breakpoint junctions. For example, potential secondary structure was noted between the 5' flanking sequence of the first breakpoint and the 3' flanking sequence of the second breakpoint in 49% of rearrangements, and between the 5' flanking sequence of the second breakpoint and the 3' flanking sequence of the first breakpoint in 36% of rearrangements (58). In addition, deletion breakpoints were found to be AT rich, whereas translocation breakpoints were GC rich. Alternating purine-pyrimidine sequences were found to be significantly overrepresented in the vicinity of deletion breakpoints, while polypyrimidine tracts were over-represented at translocation breakpoints (2).

10.3.1.5 Large Insertions (Via Retrotransposition)

A less common, but nevertheless still fascinating, mechanism of human gene mutation is the de novo insertion of repetitive elements via retrotransposition. The phenomenon was first observed in humans in the factor VIII (*F8*) gene in two unrelated de novo cases of severe hemophilia A (175). Truncated LINE (long interspersed) repetitive elements were introduced into exon 14 of the factor VIII (*F8*) gene, where they caused disruption of the reading frame. The inserted elements contained a poly(A) tract and caused a target site duplication of more than 12 nucleotides. Further analysis of these insertions revealed that, in one case, the inserted element was an exact but truncated copy of a full-length LINE element, with open reading frames found at chromosome 22q11 (97). The master source gene produces an mRNA that is probably reverse transcribed (possibly via a reverse transcriptase encoded by itself) and the double stranded nucleic acid is then reinserted into an A-rich region of the genome (Fig. 10.8). LINEs probably integrate into genomic DNA by a process called target-primed reverse transcription (251). The proposed mechanism of LINE retrotransposition is as follows: an active LINE is transcribed in the nucleus and is subsequently transported to, and translated in, the cytoplasm. The two LINE proteins, ORF1 and ORF2, complex with their encoding LINE transcript in ribonucleoprotein particles. The complex is then transported to recipient DNA sequences where target-primed

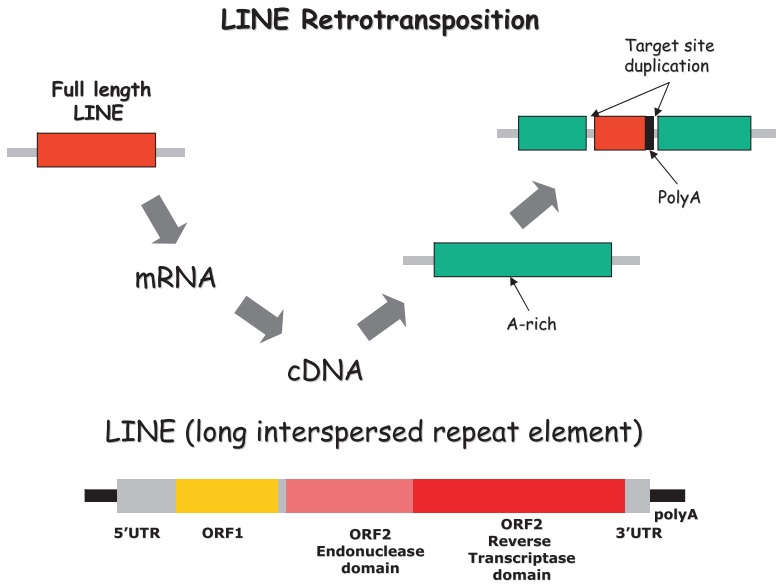

Fig. 10.8 Schematic representation of LINE retrotransposition. A master retrotransposon (full-length LINE from one chromosomal location is transcribed to double-stranded DNA and inserted into an adenine-rich region of another chromosomal location. The transposon has a poly(A) tail and produces a target site duplication

reverse transcription occurs. The new, integrated LINE copy is usually truncated at its 5' end. Over evolutionary time, L1s have shaped mammalian genomes through a number of different mechanisms. First, they have greatly expanded the genome both by their own retrotransposition and by providing the machinery necessary for the retrotransposition of other mobile elements, such as *Alu* sequences or SVA elements (49). Secondly, they have shuffled non-L1 sequence throughout the genome by a process termed transduction. Accidents of retrotransposition can cause disease and a number of such insertions have been reported to date (174, 251). It is noteworthy that insertions of these elements within introns of genes or flanking regions are probably not associated with disease, but instead represent rare, private polymorphisms (355).

Similar retrotranspositions that involve members of the *Alu* sequence family have also been reported in several genes (examples include *Alu* insertions into the *NF1* gene causing type 1 neurofibromatosis, into the factor IX (*F9*) gene causing hemophilia B, and into the cholinesterase (*BCHE*) gene in a case of acholinesterasemia) (238, 333, 341). It is likely that LINEs provide the molecular machinery necessary for the retrotransposition of *Alu*s. One study using mutation analysis of the *F9* gene has estimated the frequency of retrotransposition to be such that it occurs somewhere in the genome of about 1 in every 17 children born (208).

In an analysis of 199 unrelated families with proven mutations in *BTK* X-linked agammaglobulinemia, two families with retrotransposon insertions at exactly the same nucleotide within the coding region of the *BTK* gene have been identified. These insertions, of an SVA element and an *Alu*Y sequence, respectively, occurred 12 bp before the end of exon 9. Both had the typical hallmarks of a retrotransposon insertion, including target site duplication and a long poly A tail. The occurrence of two retrotransposon sequences at precisely the same site suggests that this site may be especially vulnerable to insertional mutagenesis (65).

Some 17% of a collection of gross insertions, all ≥276 bp in length, were due to LINE-1 (L1) retrotransposition involving different types of elements (L1 trans-driven *Alu*, L1 direct, and L1 trans-driven SVA) (49). A meta-analysis of 48 recent L1-mediated retrotranspositional events known to have caused human genetic disease revealed that 26 were L1 *trans*-driven *Alu* insertions, 15 were direct L1 insertions, four were L1 *trans*-driven SVA insertions, and three were associated with simple poly(A) insertions (52). The systematic study of these lesions, when combined with previous in vitro and genome-wide analyzes, allowed several conclusions regarding L1-mediated retrotransposition to be drawn: (a) ~25% of L1 insertions are associated with the 3' transduction of adjacent genomic sequences, (b) ~25% of the new L1 inserts are full length, (c) poly(A) tail length correlates inversely with the age of the element, and (d) the

length of target site duplication in vivo is rarely longer than 20 bp. This analysis also suggested that some 10% of L1-mediated retrotranspositional events are associated with significant genomic deletions in humans.

Interestingly, Audrezet et al. (19) reported an indel in the *CFTR* gene that involved the insertion of a short 41-bp sequence with partial homology to a retrotranspositionally-competent LINE-1 element. These authors dubbed such insertions of ultra-short LINE-1 elements "hyphen elements."

10.3.1.6 Large Insertion of Repetitive and Other Elements

The insertion of non-retrotransposons, namely beta-satellite repeats, has been observed in the human genome. The insertion of 18 copies of the 68-bp monomer of the beta satellite repeat in exon 11 of the *TMPRSS3* gene on chromosome 21 caused one form of recessive nonsyndromic deafness, DFNB10 (293). This may have been mediated by invasion of the genomic DNA by a small polydispersed circular DNA (spcDNA).

A patient with a sporadic case of Pallister–Hall syndrome has been shown to have experienced a de novo nucleic acid transfer from the mitochondrial to the nuclear genome. This mutation, a 72-bp insertion into exon 14 of the *GLI3* gene, creates a premature stop codon and predicts a truncated protein product. Both the mechanism and the cause of the mitochondrial-nuclear transfer are however unknown (326). A second example of pathologic mitochondrial-nuclear sequence transfer has been subsequently (and retrospectively) identified in the *USH1C* gene, but appears to have arisen via a novel mechanism, "*trans*-replication slippage" (49).

Gross insertions (>20 bp) comprise < 1% of disease-causing mutations. In an attempt to study these insertions in a systematic way, 158 gross insertions ranging in size between 21 bp and approximately 10 kb were identified from the HGMD; their study has revealed extensive diversity in terms of the nature of the inserted DNA sequence and has provided new insights into the underlying mutational mechanisms (49). Some 70% of gross insertions were found to represent sequence duplications of different types (tandem, partial tandem, or complex). In the context of a 26-bp insertion into the *ERCC6* gene, Chen et al. also speculated as to whether they had found evidence for another mechanism of human genetic disease, involving the possible capture of DNA oligonucleotides (49).

10.3.1.7 Inversions

The most common inversion found to date is that associated with the *F8* gene, which occurs via intrachromosomal recombination mediated by a 9.5-kb sequence that is repeated three times in the last megabase of Xqter; once in intron 22 of the *F8* gene and twice about 400 kb telomeric to the first (200, 240) (Fig. 10.9). Most inversions, which are high-frequency independent recurring events, involve the distal sequence. The vast majority of inversions occur in male germ cells (280), perhaps because intrachromosomal recombination is inhibited by the presence of homologous X chromosomes (the male-to-female ratio was estimated

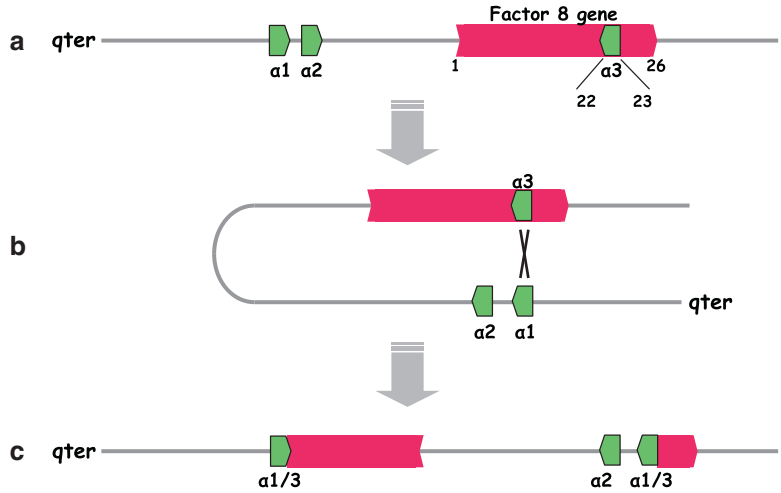

Fig. 10.9 A–C. Common inversion of the factor VIII (*F8*) gene in severe hemophilia A. (**a**) Schematic representation of the most distal 1 megabase of Xq. Regions α_1, α_2, and α_3 are 9.5-kb highly homologous DNA elements. The orientations of these sequences are shown by *arrows*. (**b**) Introchromosomal recombination between elements α_1 and α_3. (**c**) The crossover results in the inversion of exons 1 to 22 of the *F8* gene

to be about 300–1). Almost all mothers of persons with inversion hemophilia A are carriers of the abnormality. DNA diagnosis of the molecular lesion in severe hemophilia A has been greatly facilitated by the frequent occurrence of this common inversion of the *F8* gene (45% of individuals with severe hemophilia A). The frequency of de novo *F8* gene inversion has been estimated at 7.2×10^{-6} per gamete per generation. Another example of inversion has been described in the *IDS* gene (also on Xq) in about 13% of cases of Hunter syndrome (34). Inversions of DNA sequences have also been reported in the β-globin gene cluster on 11p and in the *APOA1-APOC3-APOA4* gene cluster on 11q (167, 172).

A meta-analysis of inversions of ≥5 bp but <1 kb has been performed by Chen et al. (51). Of the 21 mutations studied, 19 were found to be compatible with a model of intrachromosomal serial replication slippage in *trans* (SRStrans) mediated by short inverted repeats. Eighteen (one simple inversion, six inversions involving sequence replacement by upstream or downstream sequence, five inversions involving the partial reinsertion of removed sequence, and six inversions that occurred in a more complicated context) of these were found to be consistent with either two steps of intrachromosomal SRStrans or a combination of replication slippage in *cis* plus intrachromosomal SRStrans. The remaining lesion, a 31-kb segmental duplication associated with a small inversion in the *SLC3A1* gene, was explained in terms of a modified SRS model incorporating the concept of "break-induced replication." This study has therefore lent broad support to the idea that intrachromosomal SRStrans can account for a variety of complex gene rearrangements involving inversions.

10.3.1.8 Duplications

Duplications of whole genes or exons have contributed very significantly to the evolution of the human genome (71). Indeed, most gene clusters (e.g., β-globin, growth hormone, Hox) owe their origin to gene duplications that have occurred during vertebrate evolution. Furthermore, the presence of similar domains in proteins (e.g., immunoglobulin-like domains in many transmembrane proteins) are due to duplications of certain exons.

Occasionally, however, duplications may also be the cause of genetic disorders. The most frequent mechanism of duplication is homologous unequal crossover, as described for large deletions. In fact, most large duplications are generated as the reciprocal product of a deletion resulting from homologous unequal crossover. Duplications are less common, however, than their theoretically reciprocal deletions (see, e.g., (151), for the *DMD* gene). This may be due to the nonpathogenicity of a duplication (e.g., α-globin genes (127)), elimination of duplications as is the case for the *HPRT1* gene, or the fact that not all mechanisms that lead to deletions also produce duplications. A large and common duplication has been identified in cases of Charcot-Marie-Tooth disease type 1A (265). This duplication involves 1.5 Mb of DNA on chromosome 17p containing the peripheral myelin protein 22 (*PMP22*) gene. It results from homologous unequal crossover events between 24-kb repeats that flank the duplicated region. The reciprocal deletion product of this recombination event is responsible for a completely different clinical phenotype: hereditary neuropathy with liability to pressure palsies (Fig. 10.7). Another notable duplication of at least 500 kb that includes the *PLP1* gene is a frequent cause of Pelizaeus-Merzbacher disease (357). The pathogenetic mechanism of these duplications involves unequal crossing-over in meiosis mediated by "duplicons" in the genome (168).

The molecular defect in the majority of cases with ectrodactyly type SHFM3 on chromosome 10q24, is an approximately 0.5-Mb tandem duplication. The exact pathogenetic mechanism of this duplication is unknown (90).

Additional gene duplications causing recognizable syndromes include the *APP* duplication causing early-onset Alzheimer disease (283), the *SNCA* duplication and Parkinson disease (306), and the triplication of an ~605-kb segment containing the *PRSS1* gene in families with hereditary pancreatitis (206).

10.3.1.9 Gene Conversion

Gene conversion is the modification of one of two alleles by the other. It involves the nonreciprocal correction of an "acceptor" gene or DNA sequence by a "donor" sequence, which itself remains physically unchanged. In most known instances of gene conversion as a cause of human genetic disease, the functional gene has been wholly or partially converted to the sequence of a highly homologous and closely linked pseudogene, which therefore acts as the donor sequence. Probable examples include the genes for steroid

21-hydroxylase (*CYP21*) (327), polycystic kidney disease (*PKD1*) (345), neutrophil cytosolic factor p47-*phox* (*NCF1*) (130), immunoglobulin λ-like polypeptide 1 (*IGLL1*) (230), glucocerebrosidase (*GBA*) (108)), von Willebrand factor (*VWF*) (105), and phosphomannomutase (*PMM2*) (291). These gene/pseudogene pairs are all closely linked with the exception of the *VWF* gene (12p13) and its pseudogene (22q11-q13), and the *PMM2* gene (16p13) and its pseudogene (18p). Together, these two exceptions seem to establish a precedent for the occasional occurrence of gene conversion between unlinked loci in the human genome.

10.3.1.10 Insertion-Deletions (Indels)

A relatively rare type of mutation causing human genetic disease is the insertion-deletion, or *indel*, a complex lesion that appears to represent a combination of microdeletion and micro-insertion. One example is provided by the 9 deleted base-pairs encoding codons 39–41 of the α2-globin (*HBA2*) gene that were replaced by eight inserted bases that served to duplicate the adjacent downstream sequence (250). Indels constitute a fairly infrequent type of lesion causing human genetic disease; some 1.5% of lesions in HGMD fall into this category.

Several indel hotspots have been noted in a meta-analysis of HGMD data on 211 different indels underlying genetic disease (57). A GTAAGT motif was found to be significantly overrepresented in the vicinity of the indels studied. The change in complexity consequent to a mutation was also found to be indicative of the type of repeat sequence involved in mediating the event, thereby providing clues as to the underlying mutational mechanism. The majority of indels (>90%) were explicable in terms of a two-step process involving established mutational mechanisms. Indels equivalent to double base-pair substitutions (22% of the total) were found to be mechanistically indistinguishable from the remainder and may therefore be regarded as a special type of indel.

10.3.1.11 Other Complex Defects

Complex mutational events that involve combined gross duplications, deletions, and/or insertions of DNA sequence have been not infrequently observed and together constitute ~1% of entries in HGMD. One example of this type of gene defect is a 10.9-kb deletion coupled with a 95-bp inversion in the factor IX (*F9*) gene causing hemophilia B (178). The molecular characterization of this type of lesion is often extremely complicated and in most cases the underlying mutational mechanisms could not be readily inferred.

Recently, however, a meta-analysis of 21 complex gene rearrangements derived from the HGMD revealed that all but one could be accounted for by a model of serial replication slippage, involving twin or multiple rounds of replication slippage (50). Thus, of the 20 complex gene rearrangements, 19 (seven simple double deletions, one triple deletion, two double mutational events comprising a simple deletion and a simple insertion, six simple indels that may constitute a novel and noncanonical class of gene conversion, and three complex indels) were compatible with the model of serial replication slippage in *cis*; by contrast, the remaining indel in the *MECP2* gene appears to have arisen via interchromosomal replication slippage in *trans*.

10.3.1.12 Molecular Misreading

Long runs of adenines (and perhaps other mononucleotides or dinucleotides) promote a phenomenon termed "molecular misreading," by which DNA replication/RNA transcription and/or translation result in erroneous products with different numbers of (A)s derived from the original DNA sequence. In a family with hypobetalipoproteinemia, a deletion of one C in the A_5CA_3 coding sequence of the *APOB* gene results in a run of $(A)_8$. The patient, however, did not have severe disease, because some ApoB protein was made. This was the result of molecular misreading, in which ~10% of the resulting mRNAs contained $(A)_9$ instead of the expected $(A)_8$; this partially restored the reading frame, thereby templating the synthesis of low amounts of normal ApoB (210). Similarly, a family with mild to moderately severe hemophilia A with a deletion of one T within the coding A_8TA_2 sequence of the *F8* gene has been reported. The partial "correction" of the phenotype was due to restoration of the reading frame because of molecular misreading in which ~5% of the resulting RNAs contained $(A)_{11}$ instead of the expected $(A)_{10}$. In this family, there was also evidence for ribosomal frameshifting during translation of the mutant RNA (360).

Another example of this phenomenon was observed in the *APC* gene. A T-to-A transversion is present in the coding A_3TA_4 sequence of the *APC* gene in 6% of

Ashkenazi Jews, and in about 28% of Ashkenazim with a family history of colorectal cancer. This mutation creates a small hypermutable region, indirectly causing cancer predisposition because there are many somatic cells in which stretches of $(A)_9$ occur instead of the expected $(A)_8$; the $(A)_9$ results in frameshifting and a truncated dysfunctional APC (199). Interestingly, in the neurofibrillary tangles, neuritic plaques, and neuropil threads in the cerebral cortex of Alzheimer disease and Down syndrome, abnormal forms of β-amyloid precursor protein and ubiquitin B have been observed. These aberrant proteins were produced because of +1 frameshifting that resulted from a deletion of AG in a sequence GAGAG that occurred in the coding regions of both genes (*APP* and *UBB*, respectively). This dinucleotide deletion was again the result of molecular misreading during transcription or posttranscriptional editing of RNA (330). This mechanism is likely to yield a considerable quantity of abnormal RNA molecules and protein products in somatic cells (259).

10.3.1.13 Germline Epimutations

Epimutations are modifications of DNA that constitute clonally heritable (yet potentially reversible) alterations in the transcriptional status of a gene that lead to the abnormal silencing of that gene. Epimutations are not mutations in the strictest sense of the word, since they do not alter the gene's nucleotide sequence. However, germline epimutations of the *MLH1* gene have been reported in individuals with multiple cancers (316) and in the *MLH1* and *MSH2* genes in hereditary nonpolyposis colorectal cancer (147). These heritable inactivating epimutations are characterized by mono-allelic hypermethylation of the *MLH1* gene and, to all intents and purposes, are functionally equivalent to conventional mutations.

10.3.1.14 Frequency of Disease-Producing Mutations

Mutation Frequency Within Genes. The frequency of different molecular defects is not the same for every gene and every disorder. It depends very largely on the DNA sequence characteristics of the gene in question (e.g., the presence of repeat units or homologous sequences) and the function of, and evolutionary constraints experienced by, its encoded protein (314). For some genes, deletions predominate; for others, one particular type of lesion such as an inversion may be especially common. Some genes exhibit mainly frameshifts and stop codons associated with a specific disorder, whereas others manifest mainly missense mutations for a given phenotype, or expansions of trinucleotide repeats.

Disease mutations are nonuniformly distributed within genes (229). Such mutations were found to be statistically overrepresented in conserved domains, and underrepresented in variable regions, even after allowing for the amino acid site variability of domains over long-term evolutionary history. This finding suggests that there is a nonadditive influence of amino acid site conservation on the observed intragenic distribution of disease mutations.

Mutation Frequency Within Human Populations. Population genetic considerations are also likely to be very important in determining why some mutations occur frequently, either within a patient cohort or in the population at large (see *Frequency of Inherited Disorders Database*, http://archive.uwcm.ac.uk/uwcm/mg/fidd/; *FINDbase*, http://www.findbase.org/). Selection, migration and genetic drift are all likely to play a part, as well as the mutation rate (114, 320, 365). Thus, the mutational spectrum of the *PAH* gene underlying phenylketonuria appears to result from a range of different factors including founder effect, range expansion and migration, genetic drift and possibly also heterozygote advantage (368). Selection can also serve to maintain deleterious mutations at high frequencies in particular populations by overdominant selection (heterozygote advantage). Good examples of this phenomenon are provided by a reduction in risk of severe malaria associated with female heterozygotes and male hemizygotes for mutations in the X-linked *G6PD* gene (232, 284), for individuals heterozygous for the β-globin (*HBB*) sickle cell mutation, Glu6Val (4), and for individuals heterozygous and homozygous for α^+-thalassemia (351). Intriguingly, however, the protection against malaria afforded by sickle cell disease and α^+-thalassemia when inherited individually is lost when the two conditions are co-inherited (350). Other possible examples of heterozygote advantage include an elevated cortisol response in heterozygous carriers of *CYP21A* mutations (352), higher values for hemoglobin, serum iron and transferrin saturation in women heterozygous for *HFE* gene mutations (82), resistance

to prion infection conferred by a common prion protein (*PRNP*) polymorphism (225), resistance to severe sepsis in heterozygous carriers of the factor V Leiden polymorphism, Arg506Gln (177), and increased keratinocyte cell survival in individuals heterozygous for *GJB2* gene mutations (64). Resistance to cholera toxin (122), protection against bronchial asthma (292), and resistance to *Pseudomonas aeruginosa* infection (269) have all been mooted as possible bases for overdominant selection in heterozygous carriers of *CFTR* gene mutations. However, cystic fibrosis heterozygotes have been shown to secrete chloride at the same rate as individuals lacking *CFTR* gene mutations (149).

A number of genetic diseases are known to be particularly prevalent in Jewish populations (236, 252). The presence of four distinct lysosomal storage diseases at significant frequencies among Ashkenazi Jews has often been thought to provide evidence for a selective advantage accruing to heterozygotes in this population. However, evidence in support of the idea of genetic drift appears to be more compelling (117, 278).

Selection may also act at an extremely early stage to boost the frequency of some mutations that are deleterious at a later stage in development. Gain-of-function missense mutations in the fibroblast growth factor receptor 2 (*FGFR2*) gene responsible for Apert syndrome have been shown to confer a selective advantage on spermatogonial cells by promoting the clonal expansion of mutant cells (128, 129)

10.3.1.15 Chromosomal Distribution of Human Disease Genes

Human disease genes are characterized by the greater lengths of their encoded amino acid sequences, larger numbers of longer introns, broader ranges of tissue expression, and wider phylogenetic distributions (187, 216). Human disease genes are also known to be unevenly distributed between human chromosomes (48, 152). Furthermore, synonymous nucleotide substitutions appear to occur at a higher rate in human disease genes, a finding that may reflect increased mutation rates in the chromosomal regions in which disease genes are found (152). It may be that disease genes are more prevalent in genomic regions that experience elevated rates of mutation (55). Another possible explanation is that the disease gene set may contain a disproportionately lower number of genes expressed in the germline (152). This is because mutations in such genes might be expected to be more effectively repaired by transcription-coupled repair (transcription-coupled repair in the germline appears to account for the strand asymmetry that the human genome exhibits in terms of inherited mutations (133, 217). Strand asymmetries with respect to the mutation rate may, however, also arise through the influence of DNA replication origins (321) and recombination (153, 275).

10.3.1.16 Mutation Nomenclature

Some consistency in the way in which mutations are described is essential for the accurate and unambiguous reporting and curation of mutation data. The most recently published set of guidelines on how to describe mutational changes in human genes is to be found in den Dunnen and Antonarakis' work published in 2001 (91).

10.3.1.17 Mutations in Gene Evolution

Mutations in human gene pathology and evolution represent two sides of the same coin in that those same mutational mechanisms that have frequently been implicated in human pathology have also been involved in potentiating evolutionary change (71). Regardless of whether they are advantageous, disadvantageous, or neutral, these mutational changes and their putative underlying causal mechanisms are very similar. It is now clear that the gene has often been a dynamic entity over evolutionary time, and not a static one. Indeed, during vertebrate evolution, many genes have undergone gross rearrangement as a result of the action of a variety of mutational processes, including insertion, inversion, duplication, repeat expansion, translocation, and deletion. What links pathology and evolution is the underlying genomic architecture with its hitherto largely unexplored vocabulary of structural elements, and different types and patterns of repetitive DNA sequences (303). It can thus be seen that the mutational spectra of germline mutations responsible for inherited disease, somatic mutations underlying tumorigenesis, polymorphisms (either neutral or functionally significant), and differences between orthologous gene sequences exhibit remarkable similarities, implying that they are very likely to have causal mechanisms in common.

10

10.3.2 Consequences of Mutations

10.3.2.1 Mutations Affecting the Amino Acid Sequence of the Predicted Protein, but not Gene Expression

Many missense mutations (i.e., nucleotide substitutions that result in an amino acid substitution) cause hereditary disease in humans. Missense mutations are of importance for understanding the structure or function of a protein, since they usually occur in amino acid residues of structural or functional significance (228). Occasionally, however, not only is the mutated residue not conserved in mouse, but the substituting residue in humans is identical to its wild-type counterpart in the orthologous murine gene (123). It is thought that the most likely explanation for the majority of these cases of fixation of disease mutations in mice is *compensatory mutation*. This hypothesis holds that loss-of-function amino acid substitutions in a protein can be rescued by additional substitutions in the vicinity that compensate structurally for the original change.

It is sometimes difficult to establish a causative link between a missense mutation and a disease phenotype (76). The absence of the mutation in a large sample (usually 200 individuals) from the same ethnic group as the patient serves to exclude the possibility of a common polymorphism. Amino acid substitutions in evolutionarily conserved residues can also be good candidates for true pathogenicity (228). If the function of the protein is known, assessment of the effect of the missense mutation can be performed by in vitro mutagenesis and functional assay. Finally, the introduction of the mutation into an entire organism (e.g., into transgenic mice) and the study of its systemic effects provide one of the best means of assessing its contribution to a particular clinical phenotype. Amino acid substitutions can be shown to reduce or abolish the physiological function of a protein; for example, missense mutations have been identified in factor VIII that abolish thrombin cleavage, which is necessary for its activation (15), interfere with binding to other proteins, such as von Willebrand factor (143), or create or abolish *N*-glycosylation sites (9). In other proteins, mutations have been identified, e.g., in DNA binding domains, catalytic domains, transmembrane domains, ATP-binding regions, receptor-ligand contact sites, and phosphorylation or other chemical modification sites. Missense mutations may also affect protein folding, causing a dramatic change in secondary and tertiary structure such that the protein can no longer fulfill its physiological function.

A classic example of a missense mutation in the active site of an enzyme is provided by α1-antitrypsin Pittsburgh, found in an individual with a fatal bleeding disorder (253). The underlying mutation in the α1-antitrypsin (*SERPINA1*) gene substituted Arg for Met358 within the active site of the molecule. Substitution by Arg served to alter the substrate specificity of α1-antitrypsin by converting its "bait loop" (which is specific for elastase) to one that was specific for thrombin. In effect, the molecule lost its anti-elastase activity and became a serine protease inhibitor capable of inhibiting thrombin and factor Xa.

Mutations involving gains of glycosylation have generally been considered rare, and the pathogenic role of the new carbohydrate chains has never been formally established. Vogt et al. (337), however, identified three children with Mendelian susceptibility to mycobacterial disease who were homozygous with respect to a missense mutation in the *IFNGR2* gene that created a new *N*-glycosylation site in the IFNγR2 chain. The resulting additional carbohydrate moiety was found to be both necessary and sufficient to abolish the cellular response to IFNγ. From 10,047 HGMD mutations in 577 genes encoding proteins trafficked through the secretory pathway, 142 candidate missense mutations (~1.4%) in 77 genes (~13.3%) for potential gain of *N*-glycosylation were identified. Six mutant proteins were shown to bear new *N*-linked carbohydrate moieties. Thus, it may be that an unexpectedly high proportion of mutations causing human genetic disease do so via the creation of new *N*-glycosylation sites. Indeed, the pathogenic effects of these mutations may be a direct consequence of the addition of *N*-linked carbohydrate.

Missense mutations can result in disease by (1) elimination or reduction of the physiological activity/role of the protein; (2) gain of function by which the amino acid substitution creates new functional capabilities of the protein in biochemical and developmental processes in which the protein either does not participate or has a different role; (3) change of the target function of another protein, as in the case of the mutation in the protein C cleavage site at Arg 506 of coagulation factor V, which is associated with thrombophilia (30), or in the case of a mutation in the thrombin cleavage site of factor VIII that eliminates normal

activation of factor VIII (16), or in the case of severe obesity from childhood and R236G in the human pro-opiomelanocortin (*POMC*) gene that disrupts the dibasic cleavage site between beta melanocyte-stimulating hormone (beta-MSH) and beta-endorphin (46); and (4) participation of the mutant polypeptide in protein complexes, which renders the entire complex abnormal or nonfunctional, as in the case of the triple helical structure of certain collagens in which incorporation of one abnormal collagen chain results in "protein suicide" or an abnormal structure that degrades rapidly (41).

Missense mutations have a multitude of different effects on protein structure and function including (a) introduction of larger residues within the hydrophobic protein core leading to adverse interactions between residues, (b) introduction of buried charged residues, (c) disruption of protein-protein interactions, (d) disruption of hydrogen bonding, (e) interference with DNA binding, (f) breakage of disulphide covalent linkages, (g) mutation of catalytic residues, (h) perturbation of metal binding, and (i) disruption of quaternary structure.

Without in-depth analytical studies, missense mutations are often difficult to distinguish from polymorphisms with little or no clinical significance. In the "post-genome era," a substantial amount of human genetic variation will become amenable to high-throughput analysis in the form of single nucleotide polymorphisms (SNPs), and many of these SNPs will directly influence the structure, function, or expression of genes and the RNAs/proteins they encode. Prior knowledge as to which SNPs are most likely to be clinically relevant would greatly enhance the power of studies that aim to identify disease genes through the genotypic screening of patients in both families and populations. Inclusion of structural/functional information could be especially important in the elucidation of multifactorial disease, where genetic heterogeneity and complex interactions between genes and environment have so far limited the success of genetic epidemiological studies (146). Recently, several predictive models have been developed that employ a number of different biophysical parameters to estimate the likely impact of an amino acid substitution on the structure and function of a protein (112, 315, 317, 340, 343). These models have been used to distinguish reasonably successfully between pathologic substitutions, functional polymorphisms, and neutral polymorphisms. Vitkup et al. (334) have concluded that mutations at arginine and glycine residues are together responsible for about 30% of cases of genetic disease, whereas random mutations at tryptophan and cysteine have the highest probability of causing disease.

10.3.2.2 Mutations Affecting Gene Expression

Mutations that do not result in amino acid substitution invariably affect gene expression, i.e., transcription, RNA processing and maturation, translation, or protein stability. Total or partial gene deletions, insertions, inversions, and other gross rearrangements obviously result in the loss of gene expression. These types of mutation are usually less frequent unless the genomic sequence environment of specific genes (e.g., presence of repeats) predisposes to such lesions. Disorders with high frequencies of gross rearrangements include α–thalassemia, Duchenne muscular dystrophy, steroid sulfatase deficiency, and hemophilia A. Some partial gene deletions that eliminate one or a few exons in frame result in milder clinical phenotypes because gene expression is not totally eliminated; the resulting protein may lack an amino acid domain that is not critical for its function (362).

10.3.2.3 Transcription (Promoter) Mutations

Mutations in known promoter motifs usually lead to reduced (or occasionally increased) mRNA levels. Such mutations have been studied in the TATA box of the β-globin (*HBB*) gene (12). Other nucleotide substitutions within DNA motifs that bind transcription factors include those located in the CACCC motif of the β–globin (*HBB*) gene influencing transcription factor EKLF binding (248, 266) , several motifs in the γ-globin (*HBG*) genes (63), the CCAAT motif of the *F9* gene influencing C/EBP binding (80), the SP1 motif of the *LDLR* gene promoter (183), the HNF-1 binding site in the *PROC* gene (29), and the binding site for the transcription factor Oct-1 in the lipoprotein lipase (*LPL*) gene (359). These few examples are only representative of a total of over 370 known promoter mutations listed in HGMD and causing human genetic disease. The importance of these mutants lies in the specific DNA sequences thereby implicated in binding to transcription factors. Although most of the known mutations reduce the levels of mRNA production, some substitutions actually increase it. Examples

include various lesions in the promoters of the Gγ and Aγ globin (*HBG1* and *HBG2*) genes that cause hereditary persistence of fetal hemoglobin due to the inappropriate continuation of γ–globin gene expression into adult life (346). An increase in the distance of promoter elements from the transcriptional start site may also result in gene silencing. Such an example has been found in the promoter elements of the *CSTB* gene in progressive myoclonus epilepsy type 1 (EPM1) (202). Mutations that alter the transcriptional regulation of gene expression have been reviewed elsewhere (295).

The concomitant change in local DNA sequence complexity surrounding a substituted nucleotide is directly related to the likelihood of a regulatory mutation coming to clinical attention (196). This finding is consistent with the view that DNA sequence complexity is a critical determinant of gene regulatory function and may reflect the internal axial symmetry that frequently characterizes transcription factor binding sites.

Polymorphisms in the promoter region that are associated with differential levels of gene expression may predispose to common disorders. For example, a G>A single nucleotide polymorphism (SNP) at nucleotide-6 relative to the transcriptional initiation site of the angiotensin (*AGT*) gene influences the basal level of transcription and may predispose to essential hypertension (159). Listed in HGMD are in excess of 250 disease-associated promoter polymorphisms plus >170 functional promoter polymorphisms that significantly increase or decrease promoter activity but which have not yet been associated with a clinical phenotype.

10.3.2.4 mRNA Splicing Mutants

Single base-pair substitutions in splice junctions constitute at least 10% of all mutations causing human inherited disease. There are, however, a wide variety of mutations within both introns and exons that can affect normal RNA splicing (see (194) for review). The different mechanisms by which disruption of pre-mRNA splicing play a role in human disease were reviewed by Faustino and Cooper in 2003 (110). The most commonly found mutations occur in the invariant dinucleotides GT and AG found at the beginning and end of the donor (5') and acceptor (3') consensus splice sequences (see Fig. 10.10 for the consensus splice elements and Fig. 10.11 for the different kinds of RNA splicing abnormalities). Almost all of these mutations cause either exon skipping or cryptic splice site utilization, resulting in the severe reduction or absence of normally spliced mRNA. In addition, mutations in nucleotides +3, +4, +5, +6, -1 and –2 of the consensus donor splice site have frequently been observed (Fig. 10.12), with variable severity of the RNA splicing defect. Similarly, mutations in positions -3 and the polypyrimidine tract of the consensus acceptor splice site have been noted (Fig. 10.12). In the majority of these cases, some normal splicing occurs and the defect is not severe. Utilization of cryptic splice sites leads to the production of abnormal mature mRNA with premature stop codons or to the inclusion of additional amino acids after translation (see (15) for examples, and references cited therein).

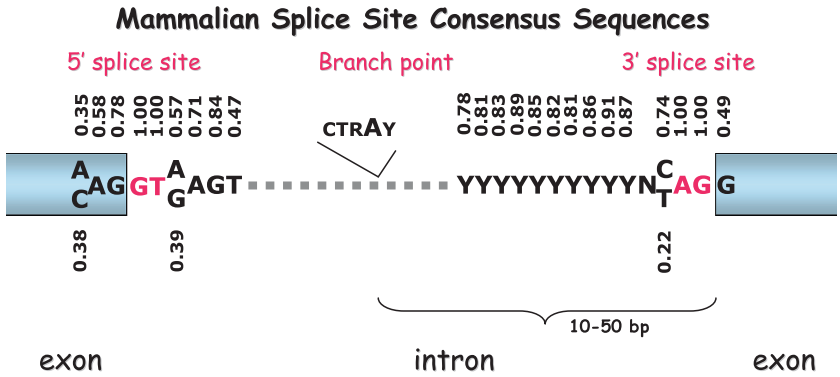

Fig. 10.10 Consensus sequences for the donor (5' splice) and acceptor (3' splice) sites and the branch point. *Numbers* above or below the nucleotides correspond to frequencies of a given nucleotide in a large number of mammalian splice site sequences. Note that the dinucleotides GT and AG (in *red*) at the beginning and end of the intron are invariant

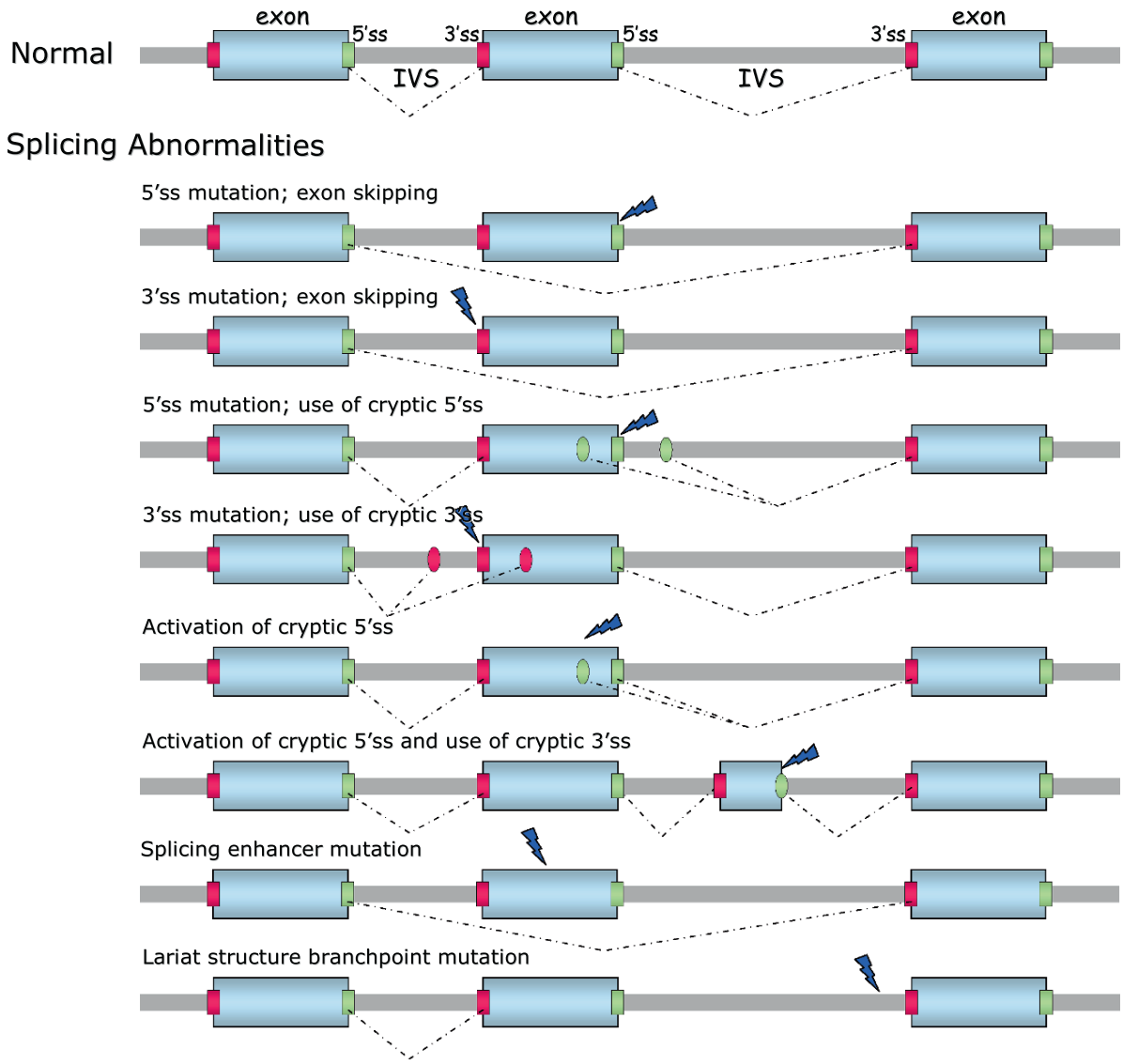

Fig. 10.11 Examples of splicing abnormalities in introns of human genes. Exons are shown as *blue boxes*; introns, as *lines* between exons. *Green squares* denote the normal 5' (donor) splice sites; *red squares* represent the normal 3' (acceptor) splice sites. *Green* and *red circles* denote cryptic 5' and 3' splice sites, respectively. The *broken blue wedge* represents the site of mutation

Employing a neural network for splice site recognition, Krawczak et al. (197) performed a meta-analysis of 478 disease-associated splicing mutations, in 38 different genes, for which detailed laboratory-based mRNA phenotype assessment had been performed. Inspection of the ±50-bp DNA sequence context of the mutations revealed that exon skipping was the preferred phenotype when the immediate vicinity of the affected exon-intron junctions was devoid of alternative splice sites. By contrast, in the presence of at least one such motif, cryptic splice site utilization became more prevalent. This association was, however, confined to donor splice sites. Outside the obligate dinucleotide, the spatial distribution of pathological mutations was found to differ significantly from that of SNPs. Whereas disease-associated lesions clustered at positions -1 and +3 to +6 for donor sites and −3 for acceptor sites, SNPs were found to be almost evenly distributed over all sequence positions considered. When all putative missense mutations in the vicinity of

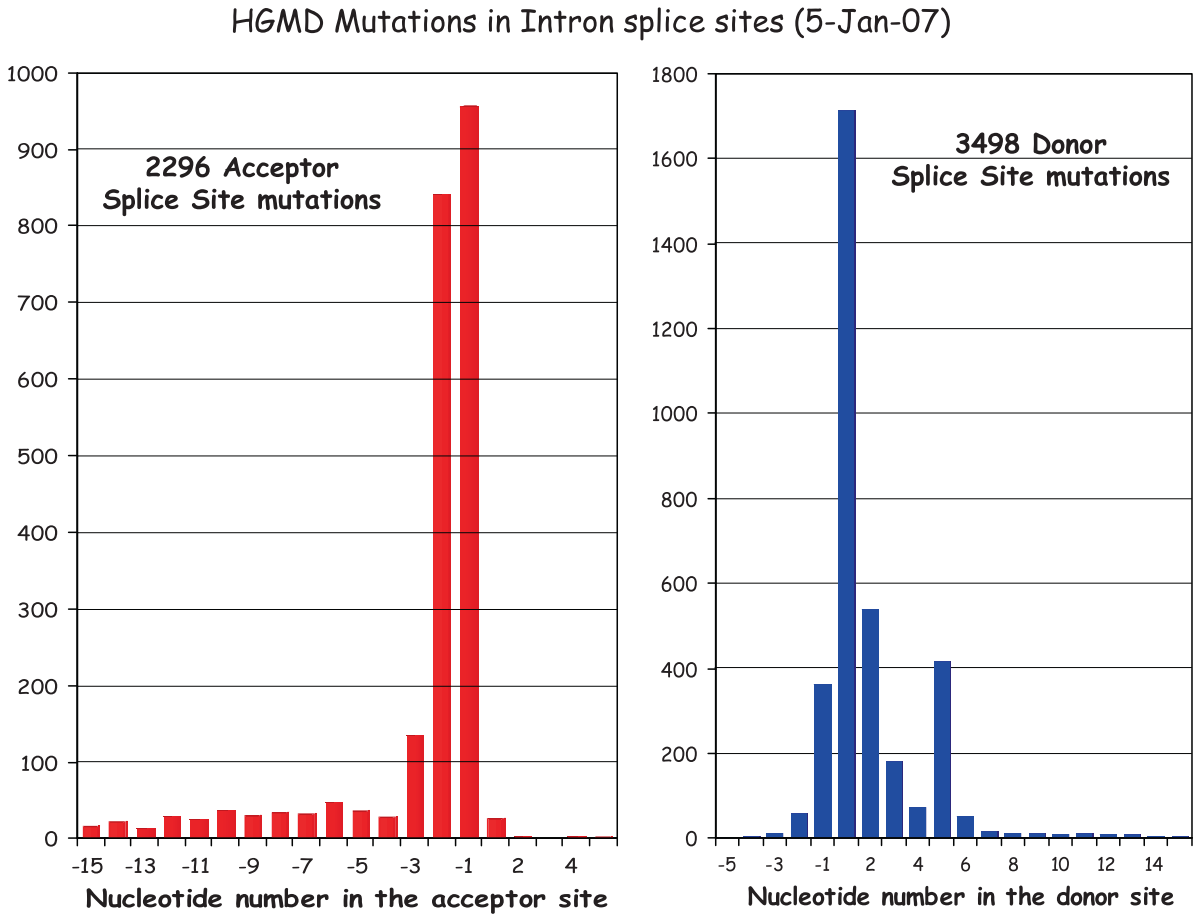

Fig. 10.12 Mutations in the consensus sequences of splice junctions recorded in the HGMD

splice sites were extracted from the HGMD for the 38 studied genes, a significantly higher proportion of changes at donor sites (37/152; 24.3%) than at acceptor splice sites (1/142; 0.7%) was found to reduce the neural network signal emitted by the respective splice site. It is estimated that some 1.6% of disease-causing missense substitutions in human genes are likely to affect the mRNA splicing phenotype.

Other kinds of mutation in introns include those that cause the activation of cryptic splice sites (by altering a sequence so as to make it more similar to an authentic consensus splice site) or by creation of new splice sites (323). In both instances, new intron splice patterns occur with consequent introduction of stop codons or abnormal peptides after translation. These mutations do not completely abolish normal splicing and are therefore not associated with the absence of normal mature mRNA. A mutation in a lariat structure branchpoint (302) has been found in the *L1CAM* gene in a patient with X-linked hydrocephalus (279). By contrast, another mutation in intron 5 of the type 2 neurofibromatosis (*NF2*) gene created a consensus branchpoint sequence and led to the activation of a cryptic exon (87).

Some 98.7% of all splice sites in human genes conform to consensus sequences that include the invariant dinucleotides GT and AG at the 5' and 3' ends of the introns, respectively (40). Noncanonical sequences (e.g., GA-AG, GC-AG, and AT-AC) do, however, occur at human splice junctions, albeit much less frequently (<0.02, 0.69, and 0.05%, respectively. Some of these noncanonical splice sites are nevertheless known to be utilized with high efficiency and may be conserved over quite long stretches of evolutionary time. Such sites have occasionally come to clinical attention when they have harbored mutations causing human inherited

disease (304). Moreover, the utilization of a cryptic noncanonical donor splice site within exon 1 of the *HRPT2* gene in a case of familial isolated primary hyperparathyroidism as a consequence of a causative lesion in intron 1 of the gene has been reported. RNA isolated from EBV-transformed lymphoblastoid cell lines derived from the patients was utilized to demonstrate the consequences at the level of the mRNA phenotype (the loss of 30 bases from the mRNA transcript).

Single base-pair substitutions within "splicing enhancer" sequences may also perturb splicing by promoting exon skipping; examples include a mutation in intron 3 of the growth hormone (*GH1*) gene causing short stature (62) and a mutation in exon 5 of the adenosine deaminase (*ADA*) gene causing ADA deficiency (287). In patients with frontotemporal dementia with parkinsonism, three heterozygous mutations in a cluster of 4 nucleotides +13 to +16 of exon 10 of the microtubule-associated protein tau (*MAPT*) gene destabilized a potential stem-loop structure that is probably involved in regulating the alternative splicing of exon 10. This caused more frequent use of the 5' splice site and an increased proportion of tau transcripts that include exon 10. The increase in exon 10+ mRNA increased the proportion of tau protein containing four microtubule-binding repeats, which is consistent with the neuropathology described in families with this type of frontotemporal dementia (155). One mutation found in the *ATM* gene causing ataxia-telangiectasia, was a deletion of four nucleotides (GTAA) in intron 20 within an intron-splicing processing element (ISPE) that is complementary to U1 snRNA. This element mediates accurate intron processing and interacts specifically with U1 snRNP particles (256). Finally, the intronic prothrombin (*F2*) gene 19911A>G polymorphism influences splicing efficiency by altering a known functional pentamer CAGGG motif (338).

Some nonsense mutations cause skipping of one or more exons, presumably during pre-mRNA splicing in the nucleus; this phenomenon has been termed "nonsense-mediated altered splicing" (NAS) but its underlying mechanism is unclear. The first such mutation was described in the *FBN1* gene in Marfan syndrome (95). It is now recognized that any nucleotide substitution within exons (nonsense, missense or translationally silent synonymous point mutation) that disrupts a splicing enhancer or silencer (ESE enhancer splicing element; CERES composite exonic regulatory element of splicing) or creates an exon splicing silencer (ESS) may affect either the pattern or efficiency of mRNA splicing (32, 43, 47, 213) (Fig. 10.13). In exon 12 of the *CFTR* gene, about one quarter of synonymous variations result in exon skipping, and hence lead to the synthesis of an inactive CFTR protein (257). For a review on the effects of exonic variants in splicing, and additional examples of such pathogenic mutations, see (255). It has been estimated that pathogenic effects of ~20% of mutations in the *MSH2* gene result from missense mutations that disrupt ESE sites and perturb splicing. Similarly, the pathogenic effects of ~16% of missense mutations in the *MLH1* gene are thought to be ESE-related (131).

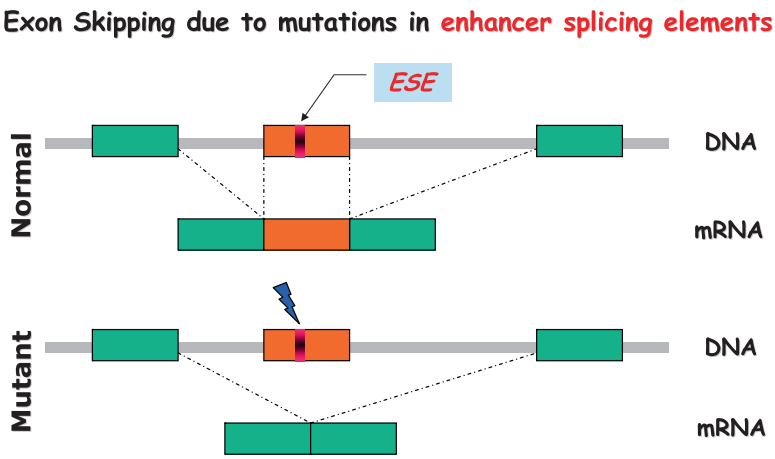

Fig. 10.13 Exon skipping attributable to nonsense, missense, and silent mutations in enhancer splicing elements (*ESE*). This element is shown as a darkened segment of the middle exon

Splice-mediated insertional inactivation involving an *Alu* repeat was first reported by Mitchell et al. in 1991 (231). Analysis of the ornithine δ-aminotransferase (*OAT*) mRNA of a patient with gyrate atrophy revealed a 142 nucleotide insertion at the junction of exons 3 and 4. An *Alu* sequence is normally present in intron 3 of the *OAT* gene, 150 bp downstream of exon 3. The *Alu* sequence found in the cDNA was identical to this one, except that the patient was homozygous for a C→G transversion in the right arm of the *Alu* repeat which served to create a new 5' splice site. This activated an upstream cryptic 3' splice site (the poly(T) complement of the *Alu* poly(A) tail followed by an AG dinucleotide) and a new "exon," containing the majority of the right arm of the *Alu* sequence, was recognized by the splicing apparatus and incorporated into the mRNA. The splice-mediated insertion of an *Alu* sequence in reverse orientation has also been reported in the *COL4A3* gene causing Alport syndrome (182).

A number of "deep intronic" mutations, at some considerable distance from splice sites and known splicing-related sequence elements, have been reported as a cause of human inherited disease (77, 140, 325). Such lesions often create novel splice sites thereby activating cryptic exons ('pseudoexons'). As mutational screening techniques improve, it is anticipated that an increasing number of such lesions will be identified which will turn out to have adverse effects on the mRNA splicing phenotype.

10.3.2.5 RNA Cleavage-Polyadenylation Mutants

A number of examples of RNA cleavage-polyadenylation mutations have now been described (53). Those reported occur in the sequence AAUAAA, which is 10–30 nucleotides upstream of the polyadenylation site and is important for the endonucleolytic cleavage and polyadenylation of the mRNA. Mutation in this sequence of the β-globin (*HBB*) gene results in mild thalassemia (249). In these cases, normal polyadenylation and cleavage occurs at a level about 10% of normal. Alternative AAUAAA sites downstream of the mutated one are used, resulting in larger mRNAs that are highly unstable. Other mutations near the poly(A) cleavage sequence may result in mRNA destabilization; one such mutation has been described 12 bp upstream of the AAUAAA sequence of the *HBB* gene in a patient with β-thalassemia (42).

The G>A mutation at the 3'-terminal nucleotide of the 3' untranslated region (UTR) of the *F2* (prothrombin) gene mRNA gives rise to an elevated prothrombin plasma level and represents a common genetic risk factor for the occurrence of thromboembolic events. This mutation creates an inefficient 3' end cleavage signal and represents a gain-of-function mutation, causing increased cleavage site recognition, increased 3' end processing, and increased mRNA accumulation and protein synthesis (124, 271) .

10.3.2.6 Mutations in miRNA-Binding Sites

Micro-RNAs (miRNAs) post-transcriptionally down-regulate gene expression by binding to complementary sequences on the 3' untranslated regions (UTRs) of their cognate mRNAs, thereby inducing either mRNA degradation or translational repression. Over 400 human miRNAs have so far been identified, but many more probably still remain to be discovered. These miRNAs are each likely to down-regulate a large number of different target mRNAs. Mutations in miRNA-binding sites could in principle cause disease, although in practice only one such lesion has so far been reported: a G A transition in a binding site for miR-189 within the 3' UTR of the *SLITRK1* gene of two apparently unrelated Tourette syndrome patients (1). Experimental confirmation of the functional effect of this mutation came from the demonstration that, in the presence of miRNA-189, in vitro constructs bearing the 3' UTR mutation served to increase repression of a reporter gene by comparison with the wild-type.

An instructive pathogenic mutation was recently found in an miRNA target site. A quantitative trait locus with a major effect on muscle mass of Texel sheep was mapped to a chromosome interval encompassing the myostatin (*GDF8*) gene. The *GDF8* allele of Texel sheep is characterized by a G-to-A transition in the 3' UTR that creates a target site for mir1 and mir206, miRNAs that are highly expressed in skeletal muscle. This causes translational inhibition of the myostatin gene and hence contributes to the muscular hypertrophy of Texel sheep (60). A further example of a functional miRNA target site variation involves an SNP in the 3'UTR of the human *AGTR1* gene; the variant allele is not down-regulated by miR155; remarkably, the variant allele has been associated with hypertension in numerous studies (296).

10.3.2.7 Cap Site Mutations

Transcription of the mRNA is initiated at the so-called cap site, which is protected from exonucleolytic degradation by the addition of α-methylguanine. An A-to-C transversion at the cap site of the β-globin (*HBB*) gene was found in a patient with β-thalassemia (354). It is not, however, clear whether this mutation causes reduced transcription or abnormal initiation of transcription since C is found in 6% of transcriptional initiation sites (190) (the most common nucleotide (76%) at position +1 is A). A functional (C/A) polymorphism of the transcriptional initiation site has been noted in the *APOH* gene; the rarer A allele displayed a carrier frequency of 0.12 and was associated with markedly reduced plasma β2 glycoprotein I (226).

10.3.2.8 Mutations in 5′ Untranslated Regions

Sequence motifs in the 5' UTRs of genes are thought to play a role in controlling the translation of the encoding mRNA. The phenotypic effects of lesions in 5' UTRs and their clinical consequences have been reviewed (45). Mutations in the iron response element (IRE) in the 5' UTR of the ferritin (*FTH1*) gene interfere with the post-transcriptional regulation of ferritin synthesis by decreasing the affinity of IRE for IRE-binding protein (125). By contrast, decreases in the steady state level of β-globin (*HBB*) mRNA have been noted in association with a single base deletion at position +10, a G-to-A substitution at position +22, a C-to-G transversion at position +33, and a 4 bp deletion (AAAC) at position +(40-43) in the *HBB* 5' UTR (18, 148, 297).

10.3.2.9 Mutations in 3′ Regulatory Regions

Sequences in the 3' regulatory regions (3' RRs) of genes are known to be involved in controlling mRNA cleavage/polyadenylation and determining mRNA stability, nuclear export, intracellular localization, and translational efficiency. Although such regions are rich in regulatory elements, relatively few pathologic mutations have been reported (53, 66). Although only ~0.2% of mutations currently logged in HGMD are located within 3' RRs, this is likely to represent a rather conservative estimate of their actual prevalence. A typical example is the G→A transition 69 nucleotides downstream of the polyadenylation site of the δ-globin (*HBD*) gene causing δ-thalassemia (233); the mutation occurs within a GATA motif and serves to increase the binding affinity of the sequence for erythroid-specific DNA binding protein.

In an attempt to study 3' RR mutations systematically, Chen et al. (53) collated 121 3' RR variants in 94 human genes including 17 mutations in the upstream core polyadenylation signal sequence (UCPAS), 79 in the upstream sequence (USS) between the translational termination codon and the UCPAS, 6 in the left arm of the 'spacer' sequence (LAS) between the UCPAS and the pre-mRNA cleavage site (CS), 3 in the right arm of the 'spacer' sequence (RAS) or downstream core polyadenylation signal sequence (DCPAS), and 7 in the downstream sequence (DSS) of the 3'-flanking region. All the UCPAS mutations and the rather unusual cases of *DMPK*, *SCA8*, *FCMD*, and *GLA* mutations were found to exert a significant effect on the mRNA phenotype, and the majority cause monogenic disease. By contrast, most of the remaining variants were polymorphisms, were found to exert a comparatively minor influence on mRNA expression, but may predispose to, protect from, or modify complex clinical phenotypes. The systematic study of these lesions permitted the identification of consistent patterns of secondary structural change that promise to allow the discrimination of nonfunctional USS variants from their functional counterparts.

10.3.2.10 Translational Initiation Mutations

Mutations in the ATG translational initiation codon have been reported in quite a wide variety of disorders (e.g., (270)). Instances of substitutions in all three nucleotides have been observed in β-thalassemia, Norrie disease, albinism, phenylketonuria, McArdle disease, and Albright osteodystrophy, among others. Indeed, a total of 251 mutations within ATG translational initiation codons are recorded in HGMD, representing ~0.6% of all missense and nonsense mutations. Almost invariably, the mutation leads to severe reduction of steady state mRNA levels similar to that associated with nonsense mutations. The mutant mRNA is presumably not translated. The first AUG codon occurs in the context of the so-called Kozak consensus sequence GCCA/GCCAUGG, which is thought to be

recognized by the 40S ribosomal subunit (191). Mutations at the initiator methionine ATG may completely abolish translation; however, there are alternative possibilities, viz. utilization of the mutant ATG with much reduced efficiency or translational initiation at the next available ATG codon. A C/T polymorphism immediately 5' to the ATG codon within the Kozak sequence of the *CD40* gene is thought to influence translation efficiency (162).

Some diseases are caused by mutations that perturb the initiation step of translation by changing the context around the start AUG codon or introducing upstream AUG codons (see (192) for a review). The scanning mechanism provides a framework for understanding the effects of these changes in mRNAs. The scanning mechanism refers to the entry of the small ribosomal subunit at the (usually capped) 5' end of the mRNA and linear migration until an AUG codon is encountered. Mutational mechanisms such as: (a) reinitiation at an internal start codon (e.g., thrombopoietin, *TPO*); and (b) leaky scanning (as in the case of the *Rx/rax* gene underlying the mouse eyeless mutation) probably account for such cases.

Naturally occurring mutations in the GCCA/GCCAUGG motif include (for the numbering of the mutant nucleotide, the A of the AUG codon is +1; see references in (192)): (a) +4 G-to-A in the androgen receptor (*AR*) gene in a family with partial androgen insufficiency; (b) -1 C-to-T transition in the α-tocopherol transfer protein (*TTPA*) gene in a family with vitamin E deficiency; (c) a 2nt deletion causes an A-to-C change at position -3 of the α-globin gene (*HBA*) in a patient with α-thalassemia; (d) -3 A-to-T transversion in the mouse *Pax6* gene causes defects in eye development; (e) -3 G-to-C somatic mutation in the *BRCA1* gene in one case of highly aggressive sporadic breast cancer. It is not surprising that most of the naturally occurring mutations involve positions -3 and +4, the positions wherein experimentally induced mutations have the strongest effect.

10.3.2.11 Termination Codon Mutations

The classic example of a termination codon mutant is the case of the α_2-globin Constant Spring, with a mutation in the normal stop codon; this substitution leads to incorporation of an additional 31 amino acid residues in the α_2-globin polypeptide chain (59). The resulting protein is unstable and does not interact properly with the β–globin chains of hemoglobin. Some 81 mutations within Term codons are recorded in HGMD, representing ~0.2% of all missense/nonsense mutations.

10.3.2.12 Frameshift Mutations

A large number of frameshift mutations have been described in numerous disease-related genes. All lead to altered translational termination with abnormal polypeptide chains after the frameshifts; severe phenotypes are usually seen. Frameshifts occur with microdeletions or micro-insertions and exon skipping. The mechanisms underlying these mutations were discussed earlier in this chapter.

10.3.2.13 Nonsense Mutations

Nonsense mutations obviously cause premature termination of translation and truncated polypeptides. Some 48% of nonsense mutations in HGMD are to codon TGA, with 28% being to TAA and 24%, to TAG. About 55% of the newly created TGA codons are CG-to-TG transitions resulting from the methylation-mediated deamination of 5mC described earlier. Many such mutations have been described in a large number of disease-related genes.

Nonsense mutations are usually associated with a reduction in the steady state level of cytoplasmic mRNA (28). This mechanism of "nonsense-mediated mRNA decay" (NMD) is responsible for the degradation of mRNAs that contain a premature termination codon at a position at least 50 nt upstream of an exon-exon boundary (219), but it is not universal (157). One or more parameters could be affected: the transcription rate, the efficiency of mRNA processing or transport to the cytoplasm, or mRNA stability.

Nonsense mutations account for at least 11% of all described gene lesions causing human inherited disease. In the majority of cases, the resulting disorders are recessive in nature as a consequence of the haploinsufficiency resulting from the NMD-induced absence of the truncated proteins (which ensures that such polypeptides do not interfere with the function of the wild-type protein). Nonsense mutations that do not elicit NMD can, however, give rise to a dominant negative condition (e.g., mutations in the *SOX10* gene causing Waardenburg-Shah syndrome (160)). Since, for NMD to be activated, the nonsense mutation must

reside at least 50-55 nt upstream of an exon-exon boundary, it follows that the precise location of the nonsense mutation could be an important factor in predicting the pathogenicity of that lesion. By way of example, nonsense mutations within the last exon of the human β-globin (*HBB*) gene do not elicit NMD. As a consequence, the truncated β-globin product has near-normal abundance, fails to associate properly with α-globin, and hence gives rise to a dominantly inherited form of α-thalassemia (318). Different nonsense mutations within the same gene may thus be associated with different clinical phenotypes depending upon whether or not NMD is activated. Another example of this is provided by a nonsense mutation (Q37X) in the *DAX1* gene of an adrenal hypoplasia congenita patient; this lesion is associated with a milder clinical phenotype than expected on account of the expression of a partially functional, amino terminal-truncated DAX1 protein synthesized from an alternative in-frame translational start site at Met83 (254).

In practical terms, the observation of greatly reduced or absent cytoplasmic mRNA associated with nonsense mutations has important implications for mutation screening. Thus, attempts to obtain mRNA for RT-PCR and mutation detection may result in amplification of nucleic acid from only the non-nonsense mutation-bearing allele. Nonsense mutations in the factor VIII (*F8*) gene (hemophilia A) and fibrillin (*FBN1*) gene (Marfan syndrome) have been associated with the skipping of exons containing these mutations (95, 240), and this observation has now been extended to other genes; exon skipping is either complete or partial. The mechanism underlying this phenomenon is unknown although a number of intriguing models have been proposed (118).

10.3.2.14 Unstable Protein Mutants

Missense mutations can cause abnormal protein folding and are therefore associated with reduced expression owing to instability of the protein. Reviews of mutations that affect protein stability can be found in (6, 258). For proteins that circulate in body fluids, most mutations are associated with 'CRM-negative' status in which the amount of protein correlates with the amount of activity or "CRM-reduced" status in which the amount of activity is still lower than the amount of protein produced. Many such mutations have been seen in factor VIII causing mild/moderate hemophilia A (14).

The nature of the biophysical properties of amino acid substitutions in p53 that increase their likelihood of coming to clinical attention has been explored (239); these include solvent inaccessibility, the number of adverse steric interactions introduced and a reduction in H-bond number. This study was extended by modeling *in silico* all amino acid replacements that could potentially have arisen from an inherited single base-pair substitution in five human genes encoding arylsulfatase A (*ARSA*), antithrombin III (*SERPINC1*), protein C (*PROC*), phenylalanine hydroxylase (*PAH*), and transthyretin (*TTR*) (317). A total of 9,795 possible mutant structures were modeled and 20 different biophysical parameters assessed. Comparison with the HGMD-derived spectra of 469 clinically detected mutations indicated that several types of mutation-associated change affected protein function, including the energy difference between wild-type and mutant structures, solvent accessibility of the mutated residue, and distance from the binding/active site. These parameters are considered to be important in protein folding, which adds support to the view that many missense mutations come to clinical attention by virtue of their consequences for protein folding and stability (38, 135) .

10.3.2.15 Mutations in Remote Promoter Elements/Locus Control Regions

In the β-globin gene cluster, a regulatory region about 10 kb upstream of the ε-globin (*HBE*) gene has been identified that is capable of directing a high level of position-independent β-globin gene expression (137). This region, termed the locus control region (LCR), is thought to organize the entire 60-kb β-globin gene cluster into an active chromatin domain and to enhance the transcription of individual globin genes (310). A similar LCR is also present in the α-globin gene cluster and other gene clusters (339). Deletions of the LCR in the β-globin gene cluster result in silencing of the β-globin and other genes of the cluster, even though the genes themselves are intact (346). A particular 25-kb deletion, known as Hispanic γδβ-thalassemia, which deletes sequences 9.5–39 kb upstream of the ε-globin gene including the LCR, renders the β-globin gene 60 kb downstream of the deletion nonfunctional (100). This extraordinary effect of the deletion of the LCR is thought to be due to an altered (DNase I-resistant) state of chromatin associated with nonfunctional genes. Several other examples of similar

deletions in the LCR of the α-globin gene cluster have been reported (209).

10.3.2.16 Cellular Consequences of Trinucleotide Repeat Expansions

Trinucleotide repeat expansion has been discussed earlier. In the case of fragile X, the $(CGG)_n$ repeat is located in the 5' UTR of the *FMR1* gene and its expansion to full mutation results in hypermethylation of the promoter region, loss of transcription, and hence silencing of the gene (344). Loss of the encoded protein, FMRP, which is thought to play a role in dendritic mRNA transport and translation, is responsible for the classical fragile X syndrome phenotype. Gene inactivation can also be caused by altering the spacing of promoter elements from the transcriptional start site as in the case of the 12mer repeat expansion in the *CSTB* gene (202).

When the trinucleotide repeat lies within the gene coding region, as in Huntington disease, its expansion results in an abnormal protein with a gain of function owing to the enlargement of the polyglutamine tract. Mutant huntingtin exerts its pathologic effects via abnormal protein aggregation, transcriptional dysregulation, mitochondrial dysfunction, excitotoxicity, and abnormal cellular trafficking, leading to neuronal loss particularly in the dorsal substratum (35).

Another example of a gain-of-function mutation is provided by the expansion of the CTG repeat in the 3' UTR of the *DMPK* gene causing type 1 myotonic dystrophy (DM1). This does not abolish transcription but rather causes nuclear retention of RNA transcripts leading to the transcriptional dysregulation of other genes (83). CTG expansion appears to lead to the sequestration of cellular RNA-binding proteins which in turn gives rise to the abnormal splicing of multiple transcripts. DM1 thus exemplifies a disease whose mechanistic basis lies at the RNA level.

10.3.2.17 Mutations Producing Inappropriate Gene Expression

Hereditary persistence of fetal hemoglobin (HPFH) and hereditary persistence of α-fetoprotein (HPAFP) are two clinical conditions that are prototypes for the inappropriate expression of γ-globin (*HBG1 and HBG2*) and α-fetoprotein (*AFP*) genes, respectively. Normally the levels of fetal hemoglobin (HbF; α2γ2) in adult life are very low, as there is a switch from fetal to adult hemoglobin during the perinatal period. Similarly, AFP is produced at a high level in fetal liver but declines rapidly after birth. In HPFH and HPAFP, however, the levels of HbF and AFP, respectively, are inappropriately high in adult life. This is often due to single nucleotide substitutions in the promoter regions of the *HBG2*, *HBG1*, or *AFP* genes. A considerable number of mutations that occur in the region -114 to -202 of the γ-globin genes have been characterized and presumably cause persistent expression of their corresponding genes (346). A similar situation has been observed with a -119 mutation in the *AFP* gene (224). These mutations occur within DNA-binding motifs for transcriptional regulators.

A very interesting mutational mechanism has been proposed for facioscapulohumeral muscular dystrophy (FSHD), an autosomal dominant myopathy. This disease is characterized by deletions of a tandem 3.3-kb repeat termed D4Z4 on chromosome 4q35. In the general population, the size of the D4Z4 repeat array may vary between 11 and 150 units, whereas FSHD patients carry fewer than 11 repeats (142). Partial deletion of D4Z4 leads to a local change in chromatin structure (267). As a consequence, genes expressed in muscle and located up to 3 Mb upstream of D4Z4 are inappropriately overexpressed. A multiprotein complex binds D4Z4 and appears to mediate the transcriptional repression of neighboring genes. The deletion of an integral number of D4Z4 repeats below a certain threshold reduces the number of bound repression complexes, and consequently decreases transcriptional repression of 4q35 genes including the *ANT1* gene, an excellent candidate for contributing to the pathogenesis of FSHD (121).

10.3.2.18 Position Effect in Human Disorders

In several instances, a DNA alteration is found well outside the putative gene that is primarily involved with a disease. Mutations acting by "positional effect" are those in which the transcription unit and minimal promoter of the gene remain intact but there is a nearby alteration that influences gene expression (180). These positional effect DNA lesions may involve distal promoter regions, enhancer/silencer elements, or changes in the local chromatin environment. The positional effect could be up to several megabases away from the gene of interest. The examples of the LCR in the β-globin gene cluster and the transcriptional repressor

D4Z4 in FSHD are provided elsewhere in this chapter. Most of the position effects are due to chromosomal rearrangements that frequently lead to alteration of the chromatin environment of the gene. Possible mechanisms that may lead to a positional effect include: (a) separation of the transcription unit from distant *cis*-regulatory elements by the rearrangement (enhancer removal results in gene silencing, whereas silencer removal results in inappropriate gene activation); (b) juxtaposition of the gene with an enhancer element from another part of the genome; (c) removal of an insulator or boundary element may also lead to inappropriate gene silencing; (d) enhancer competition of DNA sequences that were juxtaposed to the gene; (e) positional effect variegation in which the chromosomal rearrangement causes the juxtaposition of an euchromatic gene with a region of heterochromatin.

Some examples of positional effect mutations attributable to translocation breakpoints include genes *PAX6* in aniridia (109), *SOX9* in campomelic dysplasia (268, 331), *POU3F4* in X-linked deafness (88), *HOXD* complex in mesomelic dysplasia (308), *FOXL2* in blepharophimosis/ptosis/epicanthus inversus syndrome (BPES) (31, 79), and the *SHH* gene in preaxial polydactyly (207). In these cases, the translocation breakpoints may be in excess of a megabase away from the inappropriately expressed/silenced gene. Indeed, in one example of campomelic dysplasia, the breakpoint maps ~1.3 Mb downstream of the *SOX9* gene, making this the longest range position effect so far found (331). For a recent review of position effect mutations, see (181).

It is likely that, in the majority of cases, the position effect involves a highly conserved *cis*-acting regulatory element. These *conserved noncoding elements* (CNCs; also termed multiple-species conserved sequences (MCS), conserved non-genic sequences (CNGs); the most highly conserved are also called ultraconserved elements (UCEs)) comprise approximately 1–2% of the human genome and represent potential targets for pathogenic mutations (27, 33, 93, 94, 319). An example of such a lesion is provided by the 52-kb deletion of a large noncoding region downstream of the sclerostin (*SOST*) gene in patients with van Buchem disease, leading to altered expression of the *SOST* gene (215). The deletion disrupts a bone-specific enhancer element that drives *SOST* gene expression.

Pathogenic mutation may also occur in nonconserved elements that could become functional after the introduction of the mutant sequence. This pathogenetic mechanism has been described underlying a variant form of α-thalassemia. Affected individuals from Melanesia have a gain-of-function regulatory single-nucleotide polymorphism (rSNP) in a nongenic region between the α-globin genes and their upstream regulatory elements. The rSNP creates a new promoter-like element that interferes with the normal activation of all downstream α-like globin genes (85).

10.3.2.19 Position Effect by an Antisense RNA

An individual with an inherited α-thalassemia has been described who has a deletion that results in a truncated, widely expressed gene (*LUC7L*) becoming juxtaposed to a structurally normal α-globin (*HBA2*) gene. Although it retained all of its local and remote *cis*-regulatory elements, expression of the *HBA2* gene was nevertheless silenced and its CpG island became completely methylated at an early stage during development. The antisense RNA of the *LUC7L* gene appears to have been responsible for the silencing of the *HBA2* gene (324).

10.3.2.20 Abnormal Proteins Due to Fusion of Two Different Genes

The translation of fusion genes results in novel proteins with different or abnormal properties from their parent polypeptides. Fusion genes are either the result of (1) homologous unequal crossing-over, or (2) junction sequences at breakpoints of chromosomal translocations. Hemoglobin Lepore, a fusion of δ- and β-globin genes, is the prime example of the first mechanism. Other examples of abnormal fusion genes caused by unequal crossover include the case of glucocorticoid-suppressible hyperaldosteronism (GSH), an autosomal dominant form of hypertension, caused by oversecretion of aldosterone (262); some GSH patients have hybrid genes between *CYP11B1* and *CYP11B2,* two highly homologous cytochrome P450 genes on 8q22. The hybrid gene contains the regulatory elements of *CYP11B1,* expressed in the adrenal gland, and the 3' coding region of *CYP11B2,* which is essential for aldosterone synthesis. Another example is the case of abnormalities of color vision resulting from fusion of the green and red color pigment (*RCP*, *GCP*) genes (239). Recombination between the Kallmann gene on Xp22.3 (*KALX*) and its homolog (*KALY*) at Yp11.21

results in a fusion gene that is transcriptionally inactive and is associated with Kallmann syndrome secondary to an X;Y translocation.

A growing number of hematologic malignancies are associated with abnormal fusion proteins, the genes of which are found at the breakpoints of chromosomal translocations. One of the first reported examples was the case of fusion of the *BCR* and *ABL* genes in the t(9;22) known as Philadelphia (Ph) chromosome in chronic myelogenous leukemia. The *BCR* gene is on chromosome 22 and the *ABL* gene is on chromosome 9; after the translocation junction, a fusion gene is created with the promoter elements of the *ABL* gene and the 3' half of the *BCR* gene (25). A new abnormal protein is detected in the leukemia cells, the abnormal function of which probably contributes to the malignant phenotype. Another example is the case of Ewing sarcoma (a solid tumor of bone) in which an 11;22 translocation results in a fusion of the *FLI1* gene on 11q24 with the *EWS* gene on 22q12 (89); for a classic review see (272)). Fusion genes can be readily identified by PCR and can serve either as diagnostic indicators for relapse in the disorders concerned or as indicators of the need for an alternative therapeutic regimen.

10.3.2.21 Mutations in Genes Involved in Mismatch Repair Associated with Genomic Instability in the Soma

The study of somatic mutation is extremely important both for the study of cancer (116) and other diseases such as paroxysmal nocturnal hemoglobinuria (107). Mutations that lead to abnormal or abolished function of genes encoding for proteins involved in DNA mismatch repair are of particular importance because they lead to accumulation of mutations throughout the genome. For example, some forms of hereditary nonpolyposis colon cancer (HNPCC), which may account for up to 10% of colon carcinoma, are due to mutations in genes such as *MSH2* or *MLH1* that encode mismatch repair proteins (113, 204, 260) . In families with mutations in these genes, the DNA of tumor tissue shows considerable instability as detected by the generation of new alleles for numerous DNA polymorphic markers (161). One of the genes affected by the genomic instability is that encoding the type II transforming growth factor-β (TGF-β) receptor (*TGFB2R*), which has a run of 10 adenines in its coding region. This run of As is altered, resulting in a frameshift and absence of the receptor, which in turn releases the cell from TGF-β-inhibitory effects and contributes to malignancy (220). The discovery and further study of genes of the mutation repair system will enhance our understanding both of germline and of somatic mutations.

To date, relatively few studies have attempted to compare germline and somatic mutational spectra for the same genes. This notwithstanding, the mutational mechanisms underlying single base-pair substitutions (290, 328), micro-deletions and micro-insertions (134, 166, 328), and even gross gene rearrangements (184, 246) often appear to exhibit similarities between the germline and the soma.

10.3.2.22 Mosaicism

Germline mosaicism is a relatively frequent mechanism of inherited disease and provides an explanation for the inheritance pattern in cases where multiple affected offspring are born to clinically and phenotypically normal parents (367). It arises through the occurrence of a mutation de novo in a germline cell or one of its precursors during the early embryonic development of the parent. Since mitotic divisions predominate in both spermatogenesis and oogenesis, most germline mutations are likely to be mitotic rather than meiotic in origin. *Somatic mosaicism* results from mutations occurring during mitotic cell divisions in the embryo with subsequent clonal expansion of the affected cells (139). The clinical effect of somatic mosaicism depends critically upon the developmental stage at which the mutation occurs. Thus, a mutation that occurs very early on in embryonic development is likely to affect many somatic tissues. By contrast, mutations occurring rather later may give rise to a phenotype that is confined to a single body region or even to a single organ. Somatic mosaicism arising at a very early embryonic stage can involve both somatic cells and germ cells. Such individuals (*gonosomal mosaics*) are at risk of having affected children.

10.3.2.23 Sex Differences in Mutation Rates

Sex differences in mutation rates may have a variety of different underlying causes. For *premeiotic mutations*,

the single most important factors are likely to be the much higher number of cell divisions during spermatogenesis than oogenesis and the fact that the number of male germ cell divisions experienced is age dependent (81). However, the likelihood of a given mutation having originated in a particular parent is often dependent upon the nature of the mutation in question. In general, point mutations tend to display a paternal bias, arising during spermatogenesis, whilst gross deletions tend to occur predominantly in females, having originated during oogenesis (26, 136).

10.3.2.24 Concepts of Dominance and Recessiveness in Relation to the Underlying Mutations

A genetic character is held to be *dominant* if it is manifest in the heterozygous state and *recessive* if it is not. Thus, for a truly dominant condition, homozygotes should be clinically and phenotypically indistinguishable from heterozygotes (349). If this is not so, and the homozygote is more seriously affected, then the respective alleles may be regarded as *semidominant* (366).

In general, most recessive alleles are loss-of-function alleles and include gross gene deletions and rearrangements, frameshift mutations, nonsense mutations, etc. By contrast, dominant alleles are often associated with gain of function, resulting either from dominant negative mutations (which interfere with and hence abrogate the function of the wild-type allele) or from dominant positive mutations (which confer increased, constitutive, novel or toxic activity upon the mutant protein). Examples of dominant negative mutations are to be found in the *GH1* (61) and *KIT* (309) genes, whilst dominant positive mutations have been reported in the *PMP22* (263), *GNAS1* (7), *DMPK* (234), and *SERPINA1* (253) genes. It should be noted that loss-of-function mutations (e.g., *TERT* (17) and *RUNX2* (179)) can also be associated with dominantly inherited conditions in cases where a 50% reduction in the level of the protein product is sufficient to impede function.

For X-linked diseases it is probably inappropriate to use the terms dominant and recessive, since males are hemizygous and females often display variable expressivity of their heterozygous mutations owing to skewed X-inactivation or clonal expansion (96).

10.4 General Principles of Genotype-Phenotype Correlations

Several general principles have emerged as a result of the intensive study of causative mutations in genetic disorders. The following discussion highlights some of these principles. The reader is encouraged to use the Online Mendelian Inheritance in Man (OMIM) database at http://www3.ncbi.nlm.nih.gov/Omim for further information or for specific genes and clinical phenotypes. Wolf's review (353) provides an excellent guide to the complex issues inherent in the study of the relationship between mutant genotype and clinical phenotype.

Mutations in the Same Gene may be Responsible for More than One Disorder. There are many examples to illustrate the principle that mutations in a single gene can cause different and distinct clinical phenotypes ("allelic heterogeneity"). Historically, the first example is that of the β-globin (*HBB*) gene on 11pter. Mutations of this gene cause β-thalassemia, sickle cell disease, and methemoglobinemia. The *L1CAM* gene on Xq28 has been shown to be mutated in hydrocephalus and stenosis of aqueduct of Sylvius, MASA syndrome (*m*ental retardation, *a*phasia, shuffing gait, *a*dducted thumbs), and spastic paraplegia 1. The *COL1A2* gene on 7q21-q22 is involved in four different clinical forms of osteogenesis imperfecta (types II, III, IV, and atypical) and in Ehlers-Danlos syndrome type VII B. The fibroblast growth factor receptor 2 (*FGFR2*) gene is mutated in three different craniosynostosis syndromes, namely Pfeiffer, Crouzon, and Jackson-Weiss. The *COL2A1* gene is implicated in Stickler syndrome type 1, SED congenita, Kneist dysplasia, achondrogenesis-hypochondrogenesis type 2, precocious osteoarthritis, Wagner syndrome type 2, and SMED Strudwick type. In a survey of 1014 genes causing disorders in OMIM, 165 genes were associated with two disorders, 52 genes with three disorders, 24 genes with four disorders, and 19 genes with five or more disorders (11).

One Disorder May Be Caused by Mutations in More than One Gene. There are a plethora of similar clinical phenotypes caused by mutations in different genes. This observation, also known as "nonallelic" or "locus" heterogeneity, is well understood, thanks to linkage analyzes for genetic disorders and the search for mutations in different genes. Thus, tuberous sclerosis, a relatively common autosomal dominant disorder, is

caused by lesions in at least two different loci: *TSC1* on 9q34 and *TSC2* on 16p13.3. Approximately 60% of TSC families show linkage to the *TSC2* locus and 40% to the *TSC1* locus. Hereditary nonpolyposis colon cancer has been associated with mutations in five different genes. *MLH1* on 3p, *MSH2* on 2p16, *PMS1* on 2q31-q33, *PMS2* on 7p22 and *MSH6* on 2p16. Retinitis pigmentosa has so far been associated with a total of 23 different genes, and the list is still growing. We expect that disorders of complex or polygenic phenotypes, such as hypertension, atherosclerosis, diabetes, schizophrenia, and manic-depressive illness, will be associated with a considerable number of genes scattered throughout the genome.

One and the Same Mutation May Give Rise to Different Clinical Phenotypes ("Polypheny"). The clinical phenotype does not only depend on the one mutation in the responsible gene; it can be modified by the action of any of the other ~25,000–30,000 genes in the genome (353). The environment can also have an important role in the full development of the clinical phenotype. The classic sickle cell disease mutation in the β-globin (*HBB*) gene (Glu6Val) may be associated with severe or mild sickle cell disease. The amelioration of the severe clinical phenotype in this case can be attributed to the increased expression of γ-globin genes and the presence of high levels of HbF. The genomic environment of the β-globin gene cluster may therefore modify the severity of sickle cell disease, as may genetic variation originating from other loci, e.g., the α-globin genes (73). Another example of this phenomenon has recently been provided by studies of certain craniosynostoses. Both Pfeiffer and Crouzon syndromes can be associated with the same C342Y or C342R mutations in the *FGFR2* gene.

The clinical phenotype associated with the D178N missense mutation in the prion protein (*PRNP*) gene is critically dependent upon the presence of the Met or Val 129 polymorphic allele to which it is coupled. When D178N lies in *cis* to the Met129 allele, fatal familial insomnia (FFI) results, whereas D178N coupled to the Val129 allele is associated with Creutzfeldt-Jakob disease (261). The Met/Val 129 polymorphism also exerts an effect in *trans* through the normal allele, since FFI is more severe and of longer duration in patients homozygous for either the Met or the Val allele.

One of the best examples of the contribution of the environment to the clinical phenotype of single gene disorders is that of phenylketonuria resulting from phenylalanine hydroxylase (PAH) deficiency. Individuals homozygous or compound heterozygous for mutations in the *PAH* gene develop severe mental handicap if fed a normal diet. However, the cognitive status remains normal if these individuals are fed with a special, "phenylalanine-free" diet.

Mutations in More than One Gene May Be Required to Express a Given Clinical Phenotype (Digenic Inheritance; Triallelic Inheritance). Digenic inheritance refers to clinical phenotypes caused by the co-inheritance of mutations in two unlinked genes. Thus one form of retinitis pigmentosa is due to the co-inheritance of mutations in the *RDS* gene on 6p *and* the *ROM* gene on 11q (170). Individuals with either one or the other mutation,do not suffer from the disease. In similar vein, digenic inheritance of mutations in the *MITF* and *TYR* genes has been reported as a cause of Waardenburg syndrome type 2 in conjunction with ocular albinism (235). This phenomenon may be common in polygenic disorders and in disorders with "low penetrance."

Triallelic inheritance refers to clinical phenotypes with apparent recessive mode of inheritance caused by the co-inheritance of three mutant alleles, two in one gene and one in another gene. An example of triallelic inheritance is provided by the Bardet–Biedl syndrome. There are pedigrees in which affected individuals have two mutant alleles in the *BBS6* gene and one mutant allele in the *BBS2* gene. Other pedigrees have two mutant alleles in the *BBS2* gene and one mutant allele in *BBS6* (173). This type of inheritance indicates that some forms of BBS have a complex pattern of inheritance. As above, this phenomenon may be relevant in polygenic disorders and in disorders with "low penetrance."

Different Mutations in the Same Gene May Give Rise to Distinct Dominant and Recessive Forms of the Same Disease. von Willebrand factor (vWF) deficiency is a relatively common monogenic disease of blood coagulation. Many mutations have been studied in the *VWF* gene on chromosome 12p. A proportion of mutations (usually deletions, nonsense codons, or frameshift mutations) cause vWF deficiency with a recessive mode of inheritance; other mutations (mostly missense substitutions), however, are associated with a dominant mode of inheritance of the vWF deficiency (243).

Whereas the majority of hitherto characterized growth hormone (*GH1*) gene lesions (including gross deletions and missense/nonsense mutations) that underlie familial short stature are inherited in autosomal recessive fashion, there is a group of intron 3 splicing mutations

that are characterized by a dominant mode of inheritance (62). These lesions result in the in-frame skipping of exon 3 encoding 40 amino acids, including a Cys residue. The dominant negative nature of this mutation is thought to be explicable in terms of the participation of the resulting free unpaired cysteine residue in an illegitimate intermolecular disulfide linkage, leading to dimerization of the mutant molecule with a normal GH molecule and inhibition of GH secretion.

10.5 Why Study Mutation?

The sequencing of the human genome is now essentially complete and its annotation well under way. Full exploitation of the emerging data, specifically in relation to understanding the etiology of inherited disease and disease predisposition, is likely to be hampered by our ignorance of the basic processes underlying inter-individual, inter-population, and inter-species genetic diversity, however. At the population level, such an understanding is seen as essential for any meaningful interpretation of the prevalence/incidence patterns observed for diseases with a genetic basis. Within families, it is a prerequisite for being able to explain how inter-individual variation arises and how variable phenotypic expression can be associated with identical gene lesions. Thus, for human genome sequence data to be useful in the context of molecular medicine, they must eventually be related to the genetic variation underlying human inherited disease. To this end, the meta-analysis of pathological germline mutations in human genes should facilitate:

1. The assessment of the spectrum of known genetic variation underlying human inherited disease.
2. The identification of factors determining the propensity of DNA sequences to undergo germline mutation.
3. The optimization of mutational screening strategies.
4. Improvements in our ability to predict the clinical phenotype from knowledge of the mutant genotype.
5. The identification of disease states that exhibit incomplete mutational spectra, prompting the search for, and detection of, novel gene lesions associated with different clinical phenotypes (227).
6. Extrapolation toward the genetic basis of other, more complex traits and diseases (36).
7. Improvements in our understanding of the function of a given protein.
8. Meaningful comparison between the mechanisms of mutagenesis underlying both inherited and somatic disease.
9. Studies of human genetic diseases in their evolutionary context (303).

Acknowledgments The authors wish to thank Professor Michael Krawczak (Institut für Medizinische Informatik und Statistik, Christian-Albrechts-Universität Kiel, Germany) for his contributions to earlier published versions of this chapter and Peter Stenson for provision of HGMD data.

References

1. Abelson JF, Kwan KY, O'Roak BJ, Baek DY, Stillman AA, Morgan TM, Mathews CA, Pauls DL, Rasin MR, Gunel M, Davis NR, Ercan-Sencicek AG, Guez DH, Spertus JA, Leckman JF, LSt D, Kurlan R, Singer HS, Gilbert DL, Farhi A, Louvi A, Lifton RP, Sestan N, State MW (2005) Sequence variants in SLITRK1 are associated with Tourette's syndrome. Science 310:317–320
2. Abeysinghe SS, Chuzhanova N, Krawczak M, Ball EV, Cooper DN (2003) Translocation and gross deletion breakpoints in human inherited disease and cancer. I nucleotide composition and recombination-associated motifs. Hum Mutat 22:229–244
3. Abeysinghe SS, Stenson PD, Krawczak M, Cooper DN (2004) Gross rearrangement breakpoint database (GRaBD). Hum Mutat 23:219–221
4. Aidoo M, Terlouw DJ, Kolczak MS, McElroy PD, ter Kuile FO, Kariuki S, Nahlen BL, Lal AA, Udhayakumar V (2002) Protective effects of the sickle cell gene against malaria morbidity and mortality. Lancet 359:1311–1312
5. Aitman TJ, Dong R, Vyse TJ, Norsworthy PJ, Johnson MD, Smith J, Mangion J, Roberton-Lowe C, Marshall AJ, Petretto E, Hodges MD, Bhangal G, Patel SG, Sheehan-Rooney K, Duda M, Cook PR, Evans DJ, Domin J, Flint J, Boyle JJ, Pusey CD, Cook HT (2006) Copy number polymorphism in Fcgr3 predisposes to glomerulonephritis in rats and humans. Nature 439:851–855
6. Alber T (1989) Mutational effects on protein stability. Annu Rev Biochem 58:765–798
7. Aldred MA, Trembath RC (2000) Activating and inactivating mutations in the human GNAS1 gene. Hum Mutat 16: 183–189
8. Altshuler D, Brooks LD, Chakravarti A, Collins FS, Daly MJ, Donnelly P (2005) A haplotype map of the human genome. Nature 437:1299–1320
9. Aly AM, Higuchi M, Kasper CK, Kazazian HH Jr, Antonarakis SE, Hoyer LW (1992) Hemophilia A due to mutations that create new N-glycosylation sites. Proc Natl Acad Sci USA 89:4933–4937
10. Anagnou NP, O'Brien SJ, Shimada T, Nash WG, Chen MJ, Nienhuis AW (1984) Chromosomal organization of the

human dihydrofolate reductase genes: dispersion, selective amplification, and a novel form of polymorphism. Proc Natl Acad Sci USA 81:5170–5174
11. Antonarakis SE, McKusick VA (2000) OMIM passes the 1, 000-disease-gene mark. Nat Genet 25:11
12. Antonarakis SE, Irkin SH, Cheng TC, Scott AF, Sexton JP, Trusko SP, Charache S, Kazazian HH Jr (1984) beta-Thalassemia in American Blacks: novel mutations in the "TATA" box and an acceptor splice site. Proc Natl Acad Sci USA 81:1154–1158
13. Antonarakis SE, Kazazian HH Jr, Orkin SH (1985) DNA polymorphism and molecular pathology of the human globin gene clusters. Hum Genet 69:1–14
14. Antonarakis SE, Kazazian HH, Tuddenham EG (1995) Molecular etiology of factor VIII deficiency in hemophilia A. Hum Mutat 5:1–22
15. Antonarakis SE, Krawczak M, Cooper DN (2001) The nature and mechanisms of human gene mutation. In: Scriver CR, Beaudet AL, Valle D et al (eds) The metabolic and molecular bases of inherited disease. McGraw-Hill, New York, pp 343–377
16. Arai M, Inaba H, Higuchi M, Antonarakis SE, Kazazian HH Jr, Fujimaki M, Hoyer LW (1989) Direct characterization of factor VIII in plasma: detection of a mutation altering a thrombin cleavage site (arginine-372-histidine). Proc Natl Acad Sci USA 86:4277–4281
17. Armanios M, Chen JL, Chang YP, Brodsky RA, Hawkins A, Griffin CA, Eshleman JR, Cohen AR, Chakravarti A, Hamosh A, Greider CW (2005) Haploinsufficiency of telomerase reverse transcriptase leads to anticipation in autosomal dominant dyskeratosis congenita. Proc Natl Acad Sci USA 102:15960–15964
18. Athanassiadou A, Papachatzopoulou A, Zoumbos N, Maniatis GM, Gibbs R (1994) A novel beta-thalassaemia mutation in the 5' untranslated region of the beta-globin gene. Br J Haematol 88:307–310
19. Audrezet MP, Chen JM, Raguenes O, Chuzhanova N, Giteau K, Le Marechal C, Quere I, Cooper DN, Ferec C (2004) Genomic rearrangements in the CFTR gene: extensive allelic heterogeneity and diverse mutational mechanisms. Hum Mutat 23:343–357
20. Bacolla A, Jaworski A, Larson JE, Jakupciak JP, Chuzhanova N, Abeysinghe SS, O'Connell CD, Cooper DN, Wells RD (2004) Breakpoints of gross deletions coincide with non-B DNA conformations. Proc Natl Acad Sci USA 101:14162–14167
21. Baglioni C (1962) The fusion of two peptide chains in hemoglobin Lepore and its interpretation as a genetic deletion. Proc Natl Acad Sci USA 48:1880–1886
22. Bailey JA, Gu Z, Clark RA, Reinert K, Samonte RV, Schwartz S, Adams MD, Myers EW, Li PW, Eichler EE (2002) Recent segmental duplications in the human genome. Science 297:1003–1007
23. Ball EV, Stenson PD, Abeysinghe SS, Krawczak M, Cooper DN, Chuzhanova NA (2005) Microdeletions and microinsertions causing human genetic disease: common mechanisms of mutagenesis and the role of local DNA sequence complexity. Hum Mutat 26:205–213
24. Ballabio A, Carrozzo R, Parenti G, Gil A, Zollo M, Persico MG, Gillard E, Affara N, Yates J, Ferguson-Smith MA et al (1989) Molecular heterogeneity of steroid sulfatase deficiency: a multicenter study on 57 unrelated patients, at DNA and protein levels. Genomics 4:36–40
25. Bartram CR, de Klein A, Hagemeijer A, van Agthoven T, Geurts van Kessel A, Bootsma D, Grosveld G, Ferguson-Smith MA, Davies T, Stone M et al (1983) Translocation of c-ab1 oncogene correlates with the presence of a Philadelphia chromosome in chronic myelocytic leukaemia. Nature 306:277–280
26. Becker J, Schwaab R, Moller-Taube A, Schwaab U, Schmidt W, Brackmann HH, Grimm T, Olek K, Oldenburg J (1996) Characterization of the factor VIII defect in 147 patients with sporadic hemophilia A: family studies indicate a mutation type-dependent sex ratio of mutation frequencies. Am J Hum Genet 58:657–670
27. Bejerano G, Pheasant M, Makunin I, Stephen S, Kent WJ, Mattick JS, Haussler D (2004) Ultraconserved elements in the human genome. Science 304:1321–1325
28. Benz EJ, Forget BG, Hillman DG, Cohen-Solal M, Pritchard J, Cavallesco C, Prensky W, Housman D (1978) Variability in the amount of beta-globin mRNA in beta0 thalassemia. Cell 14:299–312
29. Berg LP, Scopes DA, Alhaq A, Kakkar VV, Cooper DN (1994) Disruption of a binding site for hepatocyte nuclear factor 1 in the protein C gene promoter is associated with hereditary thrombophilia. Hum Mol Genet 3:2147–2152
30. Bertina RM, Koeleman BP, Koster T, Rosendaal FR, Dirven RJ, de Ronde H, van der Velden PA, Reitsma PH (1994) Mutation in blood coagulation factor V associated with resistance to activated protein C. Nature 369:64–67
31. Beysen D, Raes J, Leroy BP, Lucassen A, Yates JR, Clayton-Smith J, Ilyina H, Brooks SS, Christin-Maitre S, Fellous M, Fryns JP, Kim JR, Lapunzina P, Lemyre E, Meire F, Messiaen LM, Oley C, Splitt M, Thomson J, Peer YV, Veitia RA, De Paepe A, De Baere E (2005) Deletions involving long-range conserved nongenic sequences upstream and downstream of FOXL2 as a novel disease-causing mechanism in blepharophimosis syndrome. Am J Hum Genet 77:205–218
32. Blencowe BJ (2000) Exonic splicing enhancers: mechanism of action, diversity and role in human genetic diseases. Trends Biochem Sci 25:106–110
33. Boffelli D, Nobrega MA, Rubin EM (2004) Comparative genomics at the vertebrate extremes. Nat Rev Genet 5:456–465
34. Bondeson ML, Dahl N, Malmgren H, Kleijer WJ, Tonnesen T, Carlberg BM, Pettersson U (1995) Inversion of the IDS gene resulting from recombination with IDS-related sequences is a common cause of the Hunter syndrome. Hum Mol Genet 4:615–621
35. Borrell-Pages M, Zala D, Humbert S, Saudou F (2006) Huntington's disease: from huntingtin function and dysfunction to therapeutic strategies. Cell Mol Life Sci 63: 2642–2660
36. Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disase, future approaches for complex disease. Nat Genet 33 (Suppl):228–237
37. Brais B, Bouchard JP, Xie YG, Rochefort DL, Chretien N, Tome FM, Lafreniere RG, Rommens JM, Uyama E, Nohira O, Blumen S, Korczyn AD, Heutink P, Mathieu J, Duranceau A, Codere F, Fardeau M, Rouleau GA (1998) Short GCG

expansions in the PABP2 gene cause oculopharyngeal muscular dystrophy. Nat Genet 18:164–167
38. Bross P, Corydon TJ, Andresen BS, Jorgensen MM, Bolund L, Gregersen N (1999) Protein misfolding and degradation in genetic diseases. Hum Mutat 14:186–198
39. Buckland PR (2003) Polymorphically duplicated genes: their relevance to phenotypic variation in humans. Ann Med 35:308–315
40. Burset M, Seledtsov IA, Solovyev VV (2000) Analysis of canonical and non-canonical splice sites in mammalian genomes. Nucleic Acids Res 28:4364–4375
41. Byers P (2001) Disorders of collagen biosynthesis and structure. In: Scriver CR, Beaudet AL, Valle D et al (eds) The metabolic and molecular bases of inherited disease. McGraw-Hill, New York, pp 5241–5286
42. Cai SP, Eng B, Francombe WH, Olivieri NF, Kendall AG, Waye JS, Chui DH (1992) Two novel beta-thalassemia mutations in the 5' and 3' noncoding regions of the beta-globin gene. Blood 79:1342–1346
43. Cartegni L, Chew SL, Krainer AR (2002) Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285–298
44. Caskey CT, Pizzuti A, Fu YH, Fenwick RG Jr, Nelson DL (1992) Triplet repeat mutations in human disease. Science 256:784–789
45. Cazzola M, Skoda RC (2000) Translational pathophysiology: a novel molecular mechanism of human disease. Blood 95:3280–3288
46. Challis BG, Pritchard LE, Creemers JW, Delplanque J, Keogh JM, Luan J, Wareham NJ, Yeo GS, Bhattacharyya S, Froguel P, White A, Farooqi IS, O'Rahilly S (2002) A missense mutation disrupting a dibasic prohormone processing site in pro-opiomelanocortin (POMC) increases susceptibility to early-onset obesity through a novel molecular mechanism. Hum Mol Genet 11:1997–2004
47. Chao HK, Hsiao KJ, Su TS (2001) A silent mutation induces exon skipping in the phenylalanine hydroxylase gene in phenylketonuria. Hum Genet 108:14–19
48. Chelala C, Auffray C (2005) Sex-linked recombination variation and distribution of disease-related genes. Gene 346:29–39
49. Chen JM, Chuzhanova N, Stenson PD, Ferec C, Cooper DN (2005) Meta-analysis of gross insertions causing human genetic disease: novel mutational mechanisms and the role of replication slippage. Hum Mutat 25:207–221
50. Chen JM, Chuzhanova N, Stenson PD, Ferec C, Cooper DN (2005) Complex gene rearrangements caused by serial replication slippage. Hum Mutat 26:125–134
51. Chen JM, Chuzhanova N, Stenson PD, Ferec C, Cooper DN (2005) Intrachromosomal serial replication slippage in trans gives rise to diverse genomic rearrangements involving inversions. Hum Mutat 26:362–373
52. Chen JM, Stenson PD, Cooper DN, Ferec C (2005) A systematic analysis of LINE-1 endonuclease-dependent retrotranspositional events causing human genetic disease. Hum Genet 117:411–427
53. Chen JM, Ferec C, Cooper DN (2006) A systematic analysis of disease-associated variants in the 3' regulatory regions of human protein-coding genes II: the importance of mRNA secondary structure in assessing the functionality of 3' UTR variants. Hum Genet 120:301–333
54. Christian SL, Fantes JA, Mewborn SK, Huang B, Ledbetter DH (1999) Large genomic duplicons map to sites of instability in the Prader-Willi/Angelman syndrome chromosome region (15q11–q13). Hum Mol Genet 8:1025–1037
55. Chuang JH, Li H (2004) Functional bias and spatial organization of genes in mutational hot and cold regions in the human genome. PLoS Biol 2:E29
56. Chung MY, Ranum LP, Duvick LA, Servadio A, Zoghbi HY, Orr HT (1993) Evidence for a mechanism predisposing to intergenerational CAG repeat instability in spinocerebellar ataxia type I. Nat Genet 5:254–258
57. Chuzhanova NA, Anassis EJ, Ball EV, Krawczak M, Cooper DN (2003) Meta-analysis of indels causing human genetic disease: mechanisms of mutagenesis and the role of local DNA sequence complexity. Hum Mutat 21:28–44
58. Chuzhanova N, Abeysinghe SS, Krawczak M, Cooper DN (2003) Translocation and gross deletion breakpoints in human inherited disease and cancer II: potential involvement of repetitive sequence elements in secondary structure formation between DNA ends. Hum Mutat 22:245–251
59. Clegg JB, Weatherall DJ, Milner PF (1971) Haemoglobin Constant Spring–a chain termination mutant? Nature 234:337–340
60. Clop A, Marcq F, Takeda H, Pirottin D, Tordoir X, Bibe B, Bouix J, Caiment F, Elsen JM, Eychenne F, Larzul C, Laville E, Meish F, Milenkovic D, Tobin J, Charlier C, Georges M (2006) A mutation creating a potential illegitimate microRNA target site in the myostatin gene affects muscularity in sheep. Nat Genet 38:813–818
61. Cogan JD, Phillips JA 3rd, Schenkman SS, Milner RD, Sakati N (1994) Familial growth hormone deficiency: a model of dominant and recessive mutations affecting a monomeric protein. J Clin Endocrinol Metab 79:1261–1265
62. Cogan JD, Prince MA, Lekhakula S, Bundey S, Futrakul A, McCarthy EM, Phillips JA 3rd (1997) A novel mechanism of aberrant pre-mRNA splicing in humans. Hum Mol Genet 6:909–912
63. Collins FS, Stoeckert CJ Jr, Serjeant GR, Forget BG, Weissman SM (1984) G gamma beta+ hereditary persistence of fetal hemoglobin: cosmid cloning and identification of a specific mutation 5' to the G gamma gene. Proc Natl Acad Sci USA 81:4894–4898
64. Common JE, Di WL, Davies D, Kelsell DP (2004) Further evidence for heterozygote advantage of GJB2 deafness mutations: a link with cell survival. J Med Genet 41:573–575
65. Conley ME, Partain JD, Norland SM, Shurtleff SA, Kazazian HH Jr (2005) Two independent retrotransposon insertions at the same site within the coding region of BTK. Hum Mutat 25:324–325
66. Conne B, Stutz A, Vassalli JD (2000) The 3' untranslated region of messenger RNA: a molecular 'hotspot' for pathology? Nat Med 6:637–641
67. Conrad DF, Andrews TD, Carter NP, Hurles ME, Pritchard JK (2006) A high-resolution survey of deletion polymorphism in the human genome. Nat Genet 38:75–81
68. Consortium. TIH (2003) The International hapmap project. Nature 426:789–796
69. Conway GS, Hettiarachchi S, Murray A, Jacobs PA (1995) Fragile X premutations in familial premature ovarian failure. Lancet 346:309–310

70. Cooper DN (1983) Eukaryotic DNA methylation. Hum Genet 64:315–333
71. Cooper DN (1999) Human gene evolution. Bios Scientific, Oxford
72. Cooper DN, Krawczak M (1991) Mechanisms of insertional mutagenesis in human genes causing genetic disease. Hum Genet 87:409–415
73. Cooper DN, Krawczak M (1993) Human gene mutation. Bios Scientific, Oxford
74. Cooper DN, Youssoufian H (1988) The CpG dinucleotide and human genetic disease. Hum Genet 78:151–155
75. Cooper DN, Smith BA, Cooke HJ, Niemann S, Schmidtke J (1985) An estimate of unique DNA sequence heterozygosity in the human genome. Hum Genet 69:201–205
76. Cotton RG, Scriver CR (1998) Proof of "disease causing" mutation. Hum Mutat 12:1–3
77. Coutinho G, Xie J, Du L, Brusco A, Krainer AR, Gatti RA (2005) Functional significance of a deep intronic mutation in the ATM gene and evidence for an alternative exon 28a. Hum Mutat 25:118–124
78. Crawford DC, Akey DT, Nickerson DA (2005) The patterns of natural variation in human genes. Annu Rev Genomics Hum Genet 6:287–312
79. Crisponi L, Deiana M, Loi A, Chiappe F, Uda M, Amati P, Bisceglia L, Zelante L, Nagaraja R, Porcu S, Ristaldi MS, Marzella R, Rocchi M, Nicolino M, Lienhardt-Roussie A, Nivelon A, Verloes A, Schlessinger D, Gasparini P, Bonneau D, Cao A, Pilia G (2001) The putative forkhead transcription factor FOXL2 is mutated in blepharophimosis/ptosis/epicanthus inversus syndrome. Nat Genet 27:159–166
80. Crossley M, Brownlee GG (1990) Disruption of a C/EBP binding site in the factor IX promoter is associated with haemophilia B. Nature 345:444–446
81. Crow JF (2000) The origins, patterns and implications of human spontaneous mutation. Nat Rev Genet 1:40–47
82. Datz C, Haas T, Rinner H, Sandhofer F, Patsch W, Paulweber B (1998) Heterozygosity for the C282Y mutation in the hemochromatosis gene is associated with increased serum iron, transferrin saturation, and hemoglobin in young women: a protective role against iron deficiency? Clin Chem 44:2429–2432
83. Davis BM, McCurrach ME, Taneja KL, Singer RH, Housman DE (1997) Expansion of a CUG trinucleotide repeat in the 3' untranslated region of myotonic dystrophy protein kinase transcripts results in nuclear retention of transcripts. Proc Natl Acad Sci USA 94:7388–7393
84. Dawson SJ, Wiman B, Hamsten A, Green F, Humphries S, Henney AM (1993) The two allele sequences of a common polymorphism in the promoter of the plasminogen activator inhibitor-1 (PAI-1) gene respond differently to interleukin-1 in HepG2 cells. J Biol Chem 268:10739–10745
85. De Gobbi M, Viprakasit V, Hughes JR, Fisher C, Buckle VJ, Ayyub H, Gibbons RJ, Vernimmen D, Yoshinaga Y, de Jong P, Cheng JF, Rubin EM, Wood WG, Bowden D, Higgs DR (2006) A regulatory SNP causes a human genetic disease by creating a new transcriptional promoter. Science 312:1215–1217
86. Deininger PL, Batzer MA (1999) Alu repeats and human disease. Mol Genet Metab 67:183–193
87. De Klein A, Riegman PH, Bijlsma EK, Heldoorn A, Muijtjens M, den Bakker MA, Avezaat CJ, Zwarthoff EC (1998) A G–>A transition creates a branch point sequence and activation of a cryptic exon, resulting in the hereditary disorder neurofibromatosis 2. Hum Mol Genet 7:393–398
88. de Kok YJ, Vossenaar ER, Cremers CW, Dahl N, Laporte J, Hu LJ, Lacombe D, Fischel-Ghodsian N, Friedman RA, Parnes LS, Thorpe P, Bitner-Glindzicz M, Pander HJ, Heilbronner H, Graveline J, den Dunnen JT, Brunner HG, Ropers HH, Cremers FP (1996) Identification of a hot spot for microdeletions in patients with X-linked deafness type 3 (DFN3) 900 kb proximal to the DFN3 gene POU3F4. Hum Mol Genet 5:1229–1235
89. Delattre O, Zucman J, Plougastel B, Desmaze C, Melot T, Peter M, Kovar H, Joubert I, de Jong P, Rouleau G et al (1992) Gene fusion with an ETS DNA-binding domain caused by chromosome translocation in human tumours. Nature 359:162–165
90. de Mollerat XJ, Gurrieri F, Morgan CT, Sangiorgi E, Everman DB, Gaspari P, Amiel J, Bamshad MJ, Lyle R, Blouin JL, Allanson JE, Le Marec B, Wilson M, Braverman NE, Radhakrishna U, Delozier-Blanchet C, Abbott A, Elghouzzi V, Antonarakis S, Stevenson RE, Munnich A, Neri G, Schwartz CE (2003) A genomic rearrangement resulting in a tandem duplication is associated with split hand-split foot malformation 3 (SHFM3) at 10q24. Hum Mol Genet 12:1959–1971
91. den Dunnen JT, Antonarakis SE (2001) Nomenclature for the description of human sequence variations. Hum Genet 109:121–124
92. den Dunnen JT, Bakker E, Breteler EG, Pearson PL, van Ommen GJ (1987) Direct detection of more than 50% of the Duchenne muscular dystrophy mutations by field inversion gels. Nature 329:640–642
93. Dermitzakis ET, Reymond A, Lyle R, Scamuffa N, Ucla C, Deutsch S, Stevenson BJ, Flegel V, Bucher P, Jongeneel CV, Antonarakis SE (2002) Numerous potentially functional but non-genic conserved sequences on human chromosome 21. Nature 420:578–582
94. Dermitzakis ET, Reymond A, Antonarakis SE (2005) Conserved non-genic sequences - an unexpected feature of mammalian genomes. Nat Rev Genet 6:151–157
95. Dietz HC, Valle D, Francomano CA, Kendzior RJ Jr, Pyeritz RE, Cutting GR (1993) The skipping of constitutive exons in vivo induced by nonsense mutations. Science 259:680–683
96. Dobyns WB, Filauro A, Tomson BN, Chan AS, Ho AW, Ting NT, Oosterwijk JC, Ober C (2004) Inheritance of most X-linked traits is not dominant or recessive, just X-linked. Am J Med Genet 129:136–143
97. Dombroski B, Mathias S, Nanthakumar E, Scott A, Kazazian H Jr (1991) Isolation of an active human transposable element. Science 254:1805–1808
98. Dorschner MO, Sybert VP, Weaver M, Pletcher BA, Stephens K (2000) NF1 microdeletion breakpoints are clustered at flanking repetitive sequences. Hum Mol Genet 9:35–46
99. Driscoll DJ, Migeon BR (1990) Sex difference in methylation of single-copy genes in human meiotic germ cells: implications for X chromosome inactivation, parental imprinting, and origin of CpG mutations. Somat Cell Mol Genet 16:267–282

100. Driscoll MC, Dobkin CS, Alter BP (1989) Gamma delta beta-thalassemia due to a de novo mutation deleting the 5' beta-globin gene activation-region hypersensitive sites. Proc Natl Acad Sci USA 86:7470–7474
101. Economou EP, Bergen AW, Warren AC, Antonarakis SE (1990) The polydeoxyadenylate tract of Alu repetitive elements is polymorphic in the human genome. Proc Natl Acad Sci USA 87:2951–2954
102. Edelmann L, Pandita RK, Morrow BE (1999) Low-copy repeats mediate the common 3-Mb deletion in patients with velo-cardio-facial syndrome. Am J Hum Genet 64:1076–1086
103. Editorial (2005) A haplotype map of the human genome. Nature 437:1299–1320
104. Efstratiadis A, Posakony JW, Maniatis T, Lawn RM, O'Connell C, Spritz RA, DeRiel JK, Forget BG, Weissman SM, Slightom JL, Blechl AE, Smithies O, Baralle FE, Shoulders CC, Proudfoot NJ (1980) The structure and evolution of the human beta-globin gene family. Cell 21:653–668
105. Eikenboom JC, Vink T, Briet E, Sixma JJ, Reitsma PH (1994) Multiple substitutions in the von Willebrand factor gene that mimic the pseudogene sequence. Proc Natl Acad Sci USA 91:2221–2224
106. Embury SH, Miller JA, Dozy AM, Kan YW, Chan V, Todd D (1980) Two different molecular organizations account for the single alpha-globin gene of the alpha-thalassemia-2 genotype. J Clin Invest 66:1319–1325
107. Erickson RP (2003) Somatic gene mutation and human disease other than cancer. Mutat Res 543:125–136
108. Eyal N, Wilder S, Horowitz M (1990) Prevalent and rare mutations among Gaucher patients. Gene 96:277–283
109. Fantes J, Redeker B, Breen M, Boyle S, Brown J, Fletcher J, Jones S, Bickmore W, Fukushima Y, Mannens M, Danes S, van Heyningen V, Hanson I (1995) Aniridia-associated cytogenetic rearrangements suggest that a position effect may cause the mutant phenotype. Hum Mol Genet 4:415–422
110. Faustino NA, Cooper TA (2003) Pre-mRNA splicing and human disease. Genes Dev 17:419–437
111. Fellermann K, Stange DE, Schaeffeler E, Schmalzl H, Wehkamp J, Bevins CL, Reinisch W, Teml A, Schwab M, Lichter P, Radlwimmer B, Stange EF (2006) A chromosome 8 gene-cluster polymorphism with low human beta-defensin 2 gene copy number predisposes to Crohn disease of the colon. Am J Hum Genet 79:439–448
112. Ferrer-Costa C, Orozco M, de la Cruz X (2002) Characterization of disease-associated single amino acid polymorphisms in terms of sequence and structure properties. J Mol Biol 315:771–786
113. Fishel R, Lescoe MK, Rao MR, Copeland NG, Jenkins NA, Garber J, Kane M, Kolodner R (1993) The human mutator gene homolog MSH2 and its association with hereditary nonpolyposis colon cancer. Cell 75:1027–1038
114. Flint J, Harding RM, Clegg JB, Boyce AJ (1993) Why are some genetic diseases common? Distinguishing selection from other processes by molecular analysis of globin gene variants. Hum Genet 91:91–117
115. Francke U (1999) Williams-Beuren syndrome: genes and mechanisms. Hum Mol Genet 8:1947–1954
116. Frank SA, Nowak MA (2004) Problems of somatic mutation and cancer. Bioessays 26:291–299
117. Frisch A, Colombo R, Michaelovsky E, Karpati M, Goldman B, Peleg L (2004) Origin and spread of the 1278insTATC mutation causing Tay-Sachs disease in Ashkenazi Jews: genetic drift as a robust and parsimonious hypothesis. Hum Genet 114:366–376
118. Frischmeyer PA, Dietz HC (1999) Nonsense-mediated mRNA decay in health and disease. Hum Mol Genet 8:1893–1900
119. Fu Y-H, Kuhl D, Pizzuti A, Pieretti M, Sutcliffe JS, Richards CS, Verkerk AJMH, Holden J, Fenwick RJ, Warren ST, Oostra BA, Nelson DL, Caskey CT (1991) Variation of the CGG repeat at the Fragile X site results in genetic instability: resolution of the Sherman paradox. Cell 67:1047–1058
120. Fullerton SM, Clark AG, Weiss KM, Nickerson DA, Taylor SL, Stengard JH, Salomaa V, Vartiainen E, Perola M, Boerwinkle E, Sing CF (2000) Apolipoprotein E variation at the sequence haplotype level: implications for the origin and maintenance of a major human polymorphism. Am J Hum Genet 67:881–900
121. Gabellini D, Green MR, Tupler R (2002) Inappropriate Gene Activation in FSHD. A repressor complex binds a chromosomal repeat deleted in dystrophic muscle. Cell 110:339–348
122. Gabriel SE, Brigman KN, Koller BH, Boucher RC, Stutts MJ (1994) Cystic fibrosis heterozygote resistance to cholera toxin in the cystic fibrosis mouse model. Science 266: 107–109
123. Gao L, Zhang J (2003) Why are some human disease-associated mutations fixed in mice? Trends Genet 19:678–681
124. Gehring NH, Frede U, Neu-Yilik G, Hundsdoerfer P, Vetter B, Hentze MW, Kulozik AE (2001) Increased efficiency of mRNA 3' end formation: a new genetic mechanism contributing to hereditary thrombophilia. Nat Genet 28:389–392
125. Girelli D, Corrocher R, Bisceglia L, Olivieri O, De Franceschi L, Zelante L, Gasparini P (1995) Molecular basis for the recently described hereditary hyperferritinemia-cataract syndrome: a mutation in the iron-responsive element of ferritin L-subunit gene (the "Verona mutation"). Blood 86:4050–4053
126. Gonzalez E, Kulkarni H, Bolivar H, Mangano A, Sanchez R, Catano G, Nibbs RJ, Freedman BI, Quinones MP, Bamshad MJ, Murthy KK, Rovin BH, Bradley W, Clark RA, Anderson SA, O'Connell RJ, Agan BK, Ahuja SS, Bologna R, Sen L, Dolan MJ, Ahuja SK (2005) The influence of CCL3L1 gene-containing segmental duplications on HIV-1/AIDS susceptibility. Science 307:1434–1440
127. Goossens M, Dozy AM, Embury SH, Zachariades Z, Hadjiminas MG, Stamatoyannopoulos G, Kan YW (1980) Triplicated alpha-globin loci in humans. Proc Natl Acad Sci USA 77:518–521
128. Goriely A, McVean GA, Rojmyr M, Ingemarsson B, Wilkie AO (2003) Evidence for selective advantage of pathogenic FGFR2 mutations in the male germ line. Science 301:643–646
129. Goriely A, McVean GA, van Pelt AM, O'Rourke AW, Wall SA, de Rooij DG, Wilkie AO (2005) Gain-of-function amino acid substitutions drive positive selection of FGFR2 mutations in human spermatogonia. Proc Natl Acad Sci USA 102:6051–6056

130. Gorlach A, Lee PL, Roesler J, Hopkins PJ, Christensen B, Green ED, Chanock SJ, Curnutte JT (1997) A p47-phox pseudogene carries the most common mutation causing p47-phox- deficient chronic granulomatous disease. J Clin Invest 100:1907–1918
131. Gorlov IP, Gorlova OY, Frazier ML, Amos CI (2003) Missense mutations in hMLH1 and hMSH2 are associated with exonic splicing enhancers. Am J Hum Genet 73:1157–1161
132. Green H, Djian P (1992) Consecutive actions of different gene-altering mechanisms in the evolution of involucrin. Mol Biol Evol 9:977–1017
133. Green P, Ewing B, Miller W, Thomas PJ, Green ED (2003) Transcription-associated mutational asymmetry in mammalian evolution. Nat Genet 33:514–517
134. Greenblatt MS, Grollman AP, Harris CC (1996) Deletions and insertions in the p53 tumor suppressor gene in human cancers: confirmation of the DNA polymerase slippage/misalignment model. Cancer Res 56:2130–2136
135. Gregersen N, Bross P, Jorgensen MM, Corydon TJ, Andresen BS (2000) Defective folding and rapid degradation of mutant proteins is a common disease mechanism in genetic disorders. J Inherit Metab Dis 23:441–447
136. Grimm T, Meng G, Liechti-Gallati S, Bettecken T, Muller CR, Muller B (1994) On the origin of deletions and point mutations in Duchenne muscular dystrophy: most deletions arise in oogenesis and most point mutations result from events in spermatogenesis. J Med Genet 31:183–186
137. Grosveld F, van Assendelft GB, Greaves DR, Kollias G (1987) Position-independent, high-level expression of the human beta-globin gene in transgenic mice. Cell 51:975–985
138. Guioli S, Incerti B, Zanaria E, Bardoni B, Franco B, Taylor K, Ballabio A, Camerino G (1992) Kallmann syndrome due to a translocation resulting in an X/Y fusion gene. Nat Genet 1:337–340
139. Hall JG (1988) Review and hypotheses: somatic mosaicism: observations related to clinical genetics. Am J Hum Genet 43:355–363
140. Harland M, Mistry S, Bishop DT, Bishop JA (2001) A deep intronic mutation in CDKN2A is associated with disease in a subset of melanoma pedigrees. Hum Mol Genet 10:2679–2686
141. Harper PS, Harley HG, Reardon W, Shaw DJ (1992) Anticipation in myotonic dystrophy: new light on an old problem. Am J Hum Genet 51:10–16
142. Hewitt JE, Lyle R, Clark LN, Vallely EM, Wright TJ, Wijmenga C, van Deutekom JCT, Francis F, Sharpe PT, Hofker M, Frants RR, Williamson R (1994) Analysis of the tandem repeat locus D4Z4 associated with facioscapulohumeral muscular dystrophy. Hum Mol Genet 3:1287–1295
143. Higuchi M, Wong C, Kochhan L, Olek K, Aronis S, Kasper CK, Kazazian HH Jr, Antonarakis SE (1990) Characterization of mutations in the factor VIII gene by direct sequencing of amplified genomic DNA. Genomics 6:65–71
144. Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR (2005) Whole-genome patterns of common DNA variation in three human populations. Science 307:1072–1079
145. Hinds DA, Kloek AP, Jen M, Chen X, Frazer KA (2006) Common deletions and SNPs are in linkage disequilibrium in the human genome. Nat Genet 38:82–85
146. Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K (2002) A comprehensive review of genetic association studies. Genet Med 4:45–61
147. Hitchins M, Williams R, Cheong K, Halani N, Lin VA, Packham D, Ku S, Buckle A, Hawkins N, Burn J, Gallinger S, Goldblatt J, Kirk J, Tomlinson I, Scott R, Spigelman A, Suter C, Martin D, Suthers G, Ward R (2005) MLH1 germline epimutations as a factor in hereditary nonpolyposis colorectal cancer. Gastroenterology 129:1392–1399
148. Ho PJ, Rochette J, Fisher CA, Wonke B, Jarvis MK, Yardumian A, Thein SL (1996) Moderate reduction of beta-globin gene transcript by a novel mutation in the 5' untranslated region: a study of its interaction with other genotypes in two families. Blood 87:1170–1178
149. Hogenauer C, Santa Ana CA, Porter JL, Millard M, Gelfand A, Rosenblatt RL, Prestidge CB, Fordtran JS (2000) Active intestinal chloride secretion in human carriers of cystic fibrosis mutations: an evaluation of the hypothesis that heterozygotes have subnormal active intestinal chloride secretion. Am J Hum Genet 67:1422–1427
150. Housman D (1995) Gain of glutamines, gain of function? Nat Genet 10:3–4
151. Hu XY, Ray PN, Murphy EG, Thompson MW, Worton RG (1990) Duplicational mutation at the Duchenne muscular dystrophy locus: its frequency, distribution, origin, and phenotypegenotype correlation. Am J Hum Genet 46:682–695
152. Huang H, Winter EE, Wang H, Weinstock KG, Xing H, Goodstadt L, Stenson PD, Cooper DN, Smith D, Alba MM, Ponting CP, Fechtel K (2004) Evolutionary conservation and selection of human disease gene orthologs in the rat and mouse genomes. Genome Biol 5:R47
153. Hurles M (2005) How homologous recombination generates a mutable genome. Hum Genomics 2:179–186
154. Hurst LD, Ellegren H (1998) Sex biases in the mutation rate. Trends Genet 14:446–452
155. Hutton M, Lendon CL, Rizzu P, Baker M, Froelich S, Houlden H, Pickering-Brown S et al (1998) Association of missense and 5'-splice-site mutations in tau with the inherited dementia FTDP-17. Nature 393:702–705
156. Iafrate AJ, Feuk L, Rivera MN, Listewnik ML, Donahoe PK, Qi Y, Scherer SW, Lee C (2004) Detection of large-scale variation in the human genome. Nat Genet 36: 949–951
157. Inacio A, Silva AL, Pinto J, Ji X, Morgado A, Almeida F, Faustino P, Lavinha J, Liebhaber SA, Romao L (2004) Nonsense mutations in close proximity to the initiation codon fail to trigger full nonsense-mediated mRNA decay. J Biol Chem 279:32170–32180
158. Ingram VM (1956) A specific chemical difference between the globins of normal human and sickle-cell anaemia haemoglobin. Nature 178:792–794
159. Inoue I, Nakajima T, Williams CS, Quackenbush J, Puryear R, Powers M, Cheng T, Ludwig EH, Sharma AM, Hata A, Jeunemaitre X, Lalouel JM (1997) A nucleotide substitution in the promoter of human angiotensinogen is associated with essential hypertension and affects basal transcription in vitro. J Clin Invest 99:1786–1797

160. Inoue K, Khajavi M, Ohyama T, Hirabayashi S, Wilson J, Reggin JD, Mancias P, Butler IJ, Wilkinson MF, Wegner M, Lupski JR (2004) Molecular mechanism for distinct neurological phenotypes conveyed by allelic truncating mutations. Nat Genet 36:361–369
161. Ionov Y, Peinado MA, Malkhosyan S, Shibata D, Perucho M (1993) Ubiquitous somatic mutations in simple repeated sequences reveal a new mechanism for colonic carcinogenesis. Nature 363:558–561
162. Jacobson EM, Concepcion E, Oashi T, Tomer Y (2005) A Graves' disease-associated Kozak sequence single-nucleotide polymorphism enhances the efficiency of CD40 gene translation: a case for translational pathophysiology. Endocrinology 146:2684–2691
163. Jacquemont S, Hagerman RJ, Leehey M, Grigsby J, Zhang L, Brunberg JA, Greco C, Des Portes V, Jardini T, Levine R, Berry-Kravis E, Brown WT, Schaeffer S, Kissel J, Tassone F, Hagerman PJ (2003) Fragile X premutation tremor/ataxia syndrome: molecular, clinical, and neuroimaging correlates. Am J Hum Genet 72:869–878
164. Jeffreys AJ, Wilson V, Thein SL (1985) Hypervariable 'minisatellite' regions in human DNA. Nature 314:67–73
165. Jeffreys AJ, Neumann R, Wilson V (1990) Repeat unit sequence variation in minisatellites: a novel source of DNA polymorphism for studying variation and mutation by single molecule analysis. Cell 60:473–485
166. Jego N, Thomas G, Hamelin R (1993) Short direct repeats flanking deletions, and duplicating insertions in p53 gene in human cancers. Oncogene 8:209–213
167. Jennings MW, Jones RW, Wood WG, Weatherall DJ (1985) Analysis of an inversion within the human beta globin gene cluster. Nucleic Acids Res 13:2897–2907
168. Ji Y, Eichler EE, Schwartz S, Nicholls RD (2000) Structure of chromosomal duplicons and their role in mediating human genomic disorders. Genome Res 10:597–610
169. Juyal RC, Figuera LE, Hauge X, Elsea SH, Lupski JR, Greenberg F, Baldini A, Patel PI (1996) Molecular analyses of 17p11.2 deletions in 62 Smith-Magenis syndrome patients. Am J Hum Genet 58:998–1007
170. Kajiwara K, Berson EL, Dryja TP (1994) Digenic retinitis pigmentosa due to mutations at the unlinked peripherin/RDS and ROM1 loci. Science 264:1604–1608
171. Kang S, Ohshima K, Jaworski A, Wells RD (1996) CTG triplet repeats from the myotonic dystrophy gene are expanded in Escherichia coli distal to the replication origin as a single large event. J Mol Biol 258:543–547
172. Karathanasis SK, Ferris E, Haddad IA (1987) DNA inversion within the apolipoproteins AI/CIII/AIV-encoding gene cluster of certain patients with premature atherosclerosis. Proc Natl Acad Sci USA 84:7198–7202
173. Katsanis N, Ansley SJ, Badano JL, Eichers ER, Lewis RA, Hoskins BE, Scambler PJ, Davidson WS, Beales PL, Lupski JR (2001) Triallelic inheritance in Bardet-Biedl syndrome, a Mendelian recessive disorder. Science 293:2256–2259
174. Kazazian HH Jr (1998) Mobile elements and disease. Curr Opin Genet Dev 8:343–350
175. Kazazian HH Jr, Wong C, Youssoufian H, Scott AF, Phillips DG, Antonarakis SE (1988) Haemophilia A resulting from *de novo* insertion of L1 sequences represents a novel mechanism for mutation in man. Nature 332:164–166
176. Kelly RJ, Rouquier S, Giorgi D, Lennon GG, Lowe JB (1995) Sequence and expression of a candidate for the human Secretor blood group alpha(1, 2)fucosyltransferase gene (FUT2). Homozygosity for an enzyme-inactivating nonsense mutation commonly correlates with the non-secretor phenotype. Biol Chem 270:4640–4649
177. Kerlin BA, Yan SB, Isermann BH, Brandt JT, Sood R, Basson BR, Joyce DE, Weiler H, Dhainaut JF (2003) Survival advantage associated with heterozygous factor V Leiden mutation in patients with severe sepsis and in mouse endotoxemia. Blood 102:3085–3092
178. Ketterling RP, Ricke DO, Wurster MW, Sommer SS (1993) Deletions with inversions: report of a mutation and review of the literature. Hum Mutat 2:53–57
179. Kim HJ, Nam SH, Kim HJ, Park HS, Ryoo HM, Kim SY, Cho TJ, Kim SG, Bae SC, Kim IS, Stein JL, van Wijnen AJ, Stein GS, Lian JB, Choi JY (2006) Four novel RUNX2 mutations including a splice donor site result in the cleidocranial dysplasia phenotype. J Cell Physiol 207:114–122
180. Kleinjan DJ, van Heyningen V (1998) Position effect in human genetic disease. Hum Mol Genet 7:1611–1618
181. Kleinjan DA, van Heyningen V (2005) Long-range control of gene expression: emerging mechanisms and disruption in disease. Am J Hum Genet 76:8–32
182. Knebelmann B, Forestier L, Drouot L, Quinones S, Chuet C, Benessy F, Saus J, Antignac C (1995) Splice-mediated insertion of an Alu sequence in the COL4A3 mRNA causing autosomal recessive Alport syndrome. Hum Mol Genet 4:675–679
183. Koivisto UM, Palvimo JJ, Janne OA, Kontula K (1994) A single-base substitution in the proximal Sp1 site of the human low density lipoprotein receptor promoter as a cause of heterozygous familial hypercholesterolemia. Proc Natl Acad Sci USA 91:10526–10530
184. Kolomietz E, Meyn MS, Pandita A, Squire JA (2002) The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. Genes Chromosomes Cancer 35:97–112
185. Kondrashov AS (2003) Direct estimates of human per nucleotide mutation rates at 20 loci causing Mendelian diseases. Hum Mutat 21:12–27
186. Kondrashov AS, Rogozin IB (2004) Context of deletions and insertions in human coding sequences. Hum Mutat 23:177–185
187. Kondrashov FA, Ogurtsov AY, Kondrashov AS (2004) Bioinformatical assay of human gene morbidity. Nucleic Acids Res 32:1731–1737
188. Koolen DA, Vissers LE, Pfundt R, de Leeuw N, Knight SJ, Regan R, Kooy RF, Reyniers E, Romano C, Fichera M, Schinzel A, Baumer A, Anderlid BM, Schoumans J, Knoers NV, van Kessel AG, Sistermans EA, Veltman JA, Brunner HG, de Vries BB (2006) A new chromosome 17q21.31 microdeletion syndrome associated with a common inversion polymorphism. Nat Genet 38:999–1001
189. Kornreich R, Bishop DF, Desnick RJ (1990) a-galactosidase A gene rearrangements causing Fabry disease. J Biol Chem 265:9319–9326
190. Kozak M (1984) Compilation and analysis of sequences upstream from the translational start site in eukaryotic mRNAs. Nucleic Acids Res 12:857–872

191. Kozak M (1991) Structural features in eukaryotic mRNAs that modulate the initiation of translation. J Biol Chem 266:19867–19870
192. Kozak M (2002) Emerging links between initiation of translation and human diseases. Mamm Genome 13:401–410
193. Krawczak M, Cooper DN (1991) Gene deletions causing human genetic disease: mechanisms of mutagenesis and the role of the local DNA sequence environment. Hum Genet 86:425–441
194. Krawczak M, Reiss J, Cooper DN (1992) The mutational spectrum of single base-pair substitutions in mRNA splice junctions of human genes: causes and consequences. Hum Genet 90:41–54
195. Krawczak M, Ball EV, Cooper DN (1998) Neighboring-nucleotide effects on the rates of germ-line single-base-pair substitution in human genes. Am J Hum Genet 63:474–488
196. Krawczak M, Chuzhanova NA, Stenson PD, Johansen BN, Ball EV, Cooper DN (2000) Changes in primary DNA sequence complexity influence the phenotypic consequences of mutations in human gene regulatory regions. Hum Genet 107:362–365
197. Krawczak M, Thomas NS, Hundrieser B, Mort M, Wittig M, Hampe J, Cooper DN (2006) Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. Hum Mutat 28:150–158
198. Kunkel TA (1985) The mutational specificity of DNA polymerases-alpha and -gamma during in vitro DNA synthesis. J Biol Chem 260:12866–12874
199. Laken SJ, Petersen GM, Gruber SB, Oddoux C, Ostrer H, Giardiello FM, Hamilton SR, Hampel H, Markowitz A, Klimstra D, Jhanwar S, Winawer S, Offit K, Luce MC, Kinzler KW, Vogelstein B (1997) Familial colorectal cancer in Ashkenazim due to a hypermutable tract in APC. Nat Genet 17:79–83
200. Lakich D, Kazazian HH Jr, Antonarakis SE, Gitschier J (1993) Inversions disrupting the factor VIII gene are a common cause of severe haemophilia A. Nat Genet 5:236–241
201. Lalioti MD, Scott HS, Buresi C, Rossier C, Bottani A, Morris MA, Malafosse A, Antonarakis SE (1997) Dodecamer repeat expansion in cystatin B gene in progressive myoclonus epilepsy. Nature 386:847–851
202. Lalioti MD, Scott HS, Antonarakis SE (1999) Altered spacing of promoter elements due to the dodecamer repeat expansion contributes to reduced expression of the cystatin B gene in EPM1. Hum Mol Genet 8:1791–1798
203. Lander ES, Linton LM, Birren B, Nusbaum C, Zody MC, Baldwin J, Devon K et al (2001) Initial sequencing and analysis of the human genome. Nature 409:860–921
204. Leach FS, Nicolaides NC, Papadopoulos N, Liu B, Jen J, Parsons R, Peltomaki P, Sistonen P, Aaltonen LA, Nystrom-Lahti M et al (1993) Mutations of a mutS homolog in hereditary nonpolyposis colorectal cancer. Cell 75: 1215–1225
205. Lehrman MA, Goldstein JL, Russell DW, Brown MS (1987) Duplication of seven exons in LDL receptor gene caused by Alu-Alu recombination in a subject with familial hypercholesterolemia. Cell 48:827–835
206. Le Marechal C, Masson E, Chen JM, Morel F, Ruszniewski P, Levy P, Ferec C (2006) Hereditary pancreatitis caused by triplication of the trypsinogen locus. Nat Genet 38:1372–1374
207. Lettice LA, Horikoshi T, Heaney SJ, van Baren MJ, van der Linde HC, Breedveld GJ, Joosse M, Akarsu N, Oostra BA, Endo N, Shibata M, Suzuki M, Takahashi E, Shinka T, Nakahori Y, Ayusawa D, Nakabayashi K, Scherer SW, Heutink P, Hill RE, Noji S (2002) Disruption of a long-range cis-acting regulator for Shh causes preaxial polydactyly. Proc Natl Acad Sci USA 99:7548–7553
208. Li X, Scaringe WA, Hill KA, Roberts S, Mengos A, Careri D, Pinto MT, Kasper CK, Sommer SS (2001) Frequency of recent retrotransposition events in the human factor IX gene. Hum Mutat 17:511–519
209. Liebhaber SA, Griese EU, Weiss I, Cash FE, Ayyub H, Higgs DR, Horst J (1990) Inactivation of human alpha-globin gene expression by a de novo deletion located upstream of the alpha-globin gene cluster. Proc Natl Acad Sci USA 87:9431–9435
210. Linton MF, Pierotti V, Young SG (1992) Reading-frame restoration with an apolipoprotein B gene frameshift mutation. Proc Natl Acad Sci USA 89:11431–11435
211. Liquori CL, Ricker K, Moseley ML, Jacobsen JF, Kress W, Naylor SL, Day JW, Ranum LP (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. Science 293:864–867
212. Litt M, Luty JA (1989) A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. Am J Hum Genet 44:397–401
213. Liu HX, Cartegni L, Zhang MQ, Krainer AR (2001) A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. Nat Genet 27:55–58
214. Loeb LA, Kunkel TA (1982) Fidelity of DNA synthesis. Annu Rev Biochem 51:429–457
215. Loots GG, Kneissel M, Keller H, Baptist M, Chang J, Collette NM, Ovcharenko D, Plajzer-Frick I, Rubin EM (2005) Genomic deletion of a long-range bone enhancer misregulates sclerostin in Van Buchem disease. Genome Res 15:928–935
216. Lopez-Bigas N, Ouzounis CA (2004) Genome-wide identification of genes likely to be involved in human genetic disease. Nucleic Acids Res 32:3108–3114
217. Majewski J (2003) Dependence of mutational asymmetry on gene-expression levels in the human genome. Am J Hum Genet 73:688–692
218. Mandel JL (1993) Questions of expansion. Nat Genet 4:8–9
219. Maquat LE (2004) Nonsense-mediated mRNA decay: splicing, translation and mRNP dynamics. Nat Rev Mol Cell Biol 5:89–99
220. Markowitz S, Wang J, Myeroff L, Parsons R, Sun L, Lutterbaugh J, Fan RS, Zborowska E, Kinzler KW, Vogelstein B et al (1995) Inactivation of the type II TGF-beta receptor in colon cancer cells with microsatellite instability. Science 268:1336–1338
221. Matsuura T, Yamagata T, Burgess DL, Rasmussen A, Grewal RP, Watase K, Khajavi M, McCall AE, Davis CF, Zu L, Achari M, Pulst SM, Alonso E, Noebels JL, Nelson DL, Zoghbi HY, Ashizawa T (2000) Large expansion of the ATTCT pentanucleotide repeat in spinocerebellar ataxia type 10. Nat Genet 26:191–194

222. McCarroll SA, Hadnott TN, Perry GH, Sabeti PC, Zody MC, Barrett JC, Dallaire S, Gabriel SB, Lee C, Daly MJ, Altshuler DM (2006) Common deletion polymorphisms in the human genome. Nat Genet 38:86–92
223. McNaughton JC, Cockburn DJ, Hughes G, Jones WA, Laing NG, Ray PN, Stockwell PA, Petersen GB (1998) Is gene deletion in eukaryotes sequence-dependent? A study of nine deletion junctions and nineteen other deletion breakpoints in intron 7 of the human dystrophin gene. Gene 222:41–51
224. McVey JH, Michaelides K, Hansen LP, Ferguson-Smith M, Tilghman S, Krumlauf R, Tuddenham EG (1993) A G–>A substitution in an HNF I binding site in the human alpha-fetoprotein gene is associated with hereditary persistence of alpha-fetoprotein (HPAFP). Hum Mol Genet 2:379–384
225. Mead S, Stumpf MP, Whitfield J, Beck JA, Poulter M, Campbell T, Uphill JB, Goldstein D, Alpers M, Fisher EM, Collinge J (2003) Balancing selection at the prion protein gene consistent with prehistoric kurulike epidemics. Science 300:640–643
226. Mehdi H, Manzi S, Desai P, Chen Q, Nestlerode C, Bontempo F, Strom SC, Zarnegar R, Kamboh MI (2003) A functional polymorphism at the transcriptional initiation site in beta2-glycoprotein I (apolipoprotein H) associated with reduced gene expression and lower plasma levels of beta2-glycoprotein I. Eur J Biochem 270: 230–238
227. Millar DS, Lewis MD, Horan M, Newsway V, Easter TE, Gregory JW, Fryklund L, Norin M, Crowne EC, Davies SJ, Edwards P, Kirk J, Waldron K, Smith PJ, Phillips JA 3rd, Scanlon MF, Krawczak M, Cooper DN, Procter AM (2003) Novel mutations of the growth hormone 1 (GH1) gene disclosed by modulation of the clinical selection criteria for individuals with short stature. Hum Mutat 21:424–440
228. Miller MP, Kumar S (2001) Understanding human disease mutations through the use of interspecific genetic variation. Hum Mol Genet 10:2319–2328
229. Miller MP, Parker JD, Rissing SW, Kumar S (2003) Quantifying the intragenic distribution of human disease mutations. Ann Hum Genet 67:567–579
230. Minegishi Y, Coustan-Smith E, Wang YH, Cooper MD, Campana D, Conley ME (1998) Mutations in the human lambda5/14.1 gene result in B cell deficiency and agammaglobulinemia. J Exp Med 187:71–77
231. Mitchell GA, Labuda D, Fontaine G, Saudubray JM, Bonnefont JP, Lyonnet S, Brody LC, Steel G, Obie C, Valle D (1991) Splice-mediated insertion of an Alu sequence inactivates ornithine delta-aminotransferase: a role for Alu elements in human mutation. Proc Natl Acad Sci USA 88:815–819
232. Mockenhaupt FP, Mandelkow J, Till H, Ehrhardt S, Eggelte TA, Bienzle U (2003) Reduced prevalence of Plasmodium falciparum infection and of concomitant anaemia in pregnant women with heterozygous G6PD deficiency. Trop Med Int Health 8:118–124
233. Moi P, Loudianos G, Lavinha J, Murru S, Cossu P, Casu R, Oggiano L, Longinotti M, Cao A, Pirastu M (1992) Delta-thalassemia due to a mutation in an erythroid-specific binding protein sequence 3' to the delta-globin gene. Blood 79:512–516
234. Mooers BH, Logue JS, Berglund JA (2005) The structural basis of myotonic dystrophy from the crystal structure of CUG repeats. Proc Natl Acad Sci USA 102:16626–16631
235. Morell R, Spritz RA, Ho L, Pierpont J, Guo W, Friedman TB, Asher JH Jr (1997) Apparent digenic inheritance of Waardenburg syndrome type 2 (WS2) and autosomal recessive ocular albinism (AROA). Hum Mol Genet 6:659–664
236. Motulsky AG (1995) Jewish diseases and origins. Nat Genet 9:99–101
237. Muragaki Y, Mundlos S, Upton J, Olsen BR (1996) Altered growth and branching patterns in synpolydactyly caused by mutations in HOXD13. Science 272:548–551
238. Muratani K, Hada T, Yamamoto Y, Kaneko T, Shigeto Y, Ohue T, Furuyama J, Higashino K (1991) Inactivation of the cholinesterase gene by Alu insertion: possible mechanism for human gene transposition. Proc Natl Acad Sci USA 88:11315–11319
239. Nathans J, Piantanida TP, Eddy RL, Shows TB, Hogness DS (1986) Molecular genetics of inherited variation in human color vision. Science 232:203–210
240. Naylor JA, Green PM, Rizza CR, Giannelli F (1993) Analysis of factor VIII mRNA reveals defects in everyone of 28 haemophilia A patients. Hum Mol Genet 2:11–17
241. Neitz M, Neitz J, Grishok A (1995) Polymorphism in the number of genes encoding long-wavelength-sensitive cone pigments among males with normal color vision. Vision Res 35:2395–2407
242. Nicholls RD, Fischel-Ghodsian N, Higgs DR (1987) Recombination at the human alpha-globin gene cluster: sequence features and topological constraints. Cell 49:369–378
243. Nichols WC, Ginsburg D (1997) von Willebrand disease. Medicine (Baltimore) 76:1–20
244. Nickerson DA, Taylor SL, Weiss KM, Clark AG, Hutchinson RG, Stengard J, Salomaa V, Vartiainen E, Boerwinkle E, Sing CF (1998) DNA sequence diversity in a 9.7-kb region of the human lipoprotein lipase gene. Nat Genet 19:233–240
245. Nishihara S, Narimatsu H, Iwasaki H, Yazawa S, Akamatsu S, Ando T, Seno T, Narimatsu I (1994) Molecular genetic analysis of the human Lewis histo-blood group system. J Biol Chem 269:29271–29278
246. Oldenburg J, Rost S, El-Maarri O, Leuer M, Olek K, Muller CR, Schwaab R (2000) De novo factor VIII gene intron 22 inversion in a female carrier presents as a somatic mosaicism. Blood 96:2905–2906
247. Orkin SH, Kazazian HH Jr, Antonarakis SE, Ostrer H, Goff SC, Sexton JP (1982) Abnormal RNA processing due to the exon mutation of beta E-globin gene. Nature 300:768–769
248. Orkin SH, Antonarakis SE, Kazazian HH Jr (1984) Base substitution at position -88 in a beta-thalassemic globin gene. Further evidence for the role of distal promoter element ACACCC. J Biol Chem 259:8679–8681
249. Orkin SH, Cheng TC, Antonarakis SE, Kazazian HH Jr (1985) Thalassemia due to a mutation in the cleavage-polyadenylation signal of the human beta-globin gene. EMBO J 4:453–456
250. Oron-Karni V, Filon D, Rund D, Oppenheim A (1997) A novel mechanism generating short deletion/insertions

following slippage is suggested by a mutation in the human alpha2-globin gene. Hum Mol Genet 6:881–885
251. Ostertag EM, Kazazian HH Jr (2001) Biology of mammalian L1 retrotransposons. Annu Rev Genet 35:501–538
252. Ostrer H (2001) A genetic profile of contemporary Jewish populations. Nat Rev Genet 2:891–898
253. Owen MC, Brennan SO, Lewis JH, Carrell RW (1983) Mutation of antitrypsin to antithrombin. alpha 1-antitrypsin Pittsburgh (358 Met leads to Arg), a fatal bleeding disorder. N Engl J Med 309:694–698
254. Ozisik G, Mantovani G, Achermann JC, Persani L, Spada A, Weiss J, Beck-Peccoz P, Jameson JL (2003) An alternate translation initiation site circumvents an amino-terminal DAX1 nonsense mutation leading to a mild form of X-linked adrenal hypoplasia congenita. J Clin Endocrinol Metab 88:417–423
255. Pagani F, Baralle FE (2004) Genomic variants in exons and introns: identifying the splicing spoilers. Nat Rev Genet 5:389–396
256. Pagani F, Buratti E, Stuani C, Bendix R, Dork T, Baralle FE (2002) A new type of mutation causes a splicing defect in ATM. Nat Genet 30:426–429
257. Pagani F, Raponi M, Baralle FE (2005) Synonymous mutations in CFTR exon 12 affect splicing and are not neutral in evolution. Proc Natl Acad Sci USA 102:6368–6372
258. Pakula AA, Sauer RT (1989) Genetic analysis of protein stability and function. Annu Rev Genet 23:289–310
259. Paoloni-Giacobino A, Rossier C, Papasavvas MP, Antonarakis SE (2001) Frequency of replication/transcription errors in (A)/(T) runs of human genes. Hum Genet 109:40–47
260. Papadopoulos N, Nicolaides NC, Wei YF, Ruben SM, Carter KC, Rosen CA, Haseltine WA, Fleischmann RD, Fraser CM, Adams MD et al (1994) Mutation of a mutL homolog in hereditary colon cancer. Science 263:1625–1629
261. Parchi P, Petersen RB, Chen SG, Autilio-Gambetti L, Capellari S, Monari L, Cortelli P, Montagna P, Lugaresi E, Gambetti P (1998) Molecular pathology of fatal familial insomnia. Brain Pathol 8:539–548
262. Pascoe L, Jeunemaitre X, Lebrethon MC, Curnow KM, Gomez-Sanchez CE, Gasc JM, Saez JM, Corvol P (1995) Glucocorticoid-suppressible hyperaldosteronism and adrenal tumors occurring in a single French pedigree. J Clin Invest 96:2236–2246
263. Patel PI, Roa BB, Welcher AA, Schoener-Scott R, Trask BJ, Pentao L, Snipes GJ, Garcia CA, Francke U, Shooter EM, Lupski JR, Suter U (1992) The gene for the peripheral myelin protein PMP-22 is a candidate for Charcot-Marie-Tooth disease type 1A. Nat Genet 1:159–165
264. Pearson CE, Nichol Edamura K, Cleary JD (2005) Repeat instability: mechanisms of dynamic mutations. Nat Rev Genet 6:729–742
265. Pentao L, Wise CA, Chinault AC, Patel PI, Lupski JR (1992) Charcot-Marie-Tooth type 1A duplication appears to arise from recombination at repeat sequences flanking the 1.5 Mb monomer unit. Nat Genet 2:292–300
266. Perkins AC, Sharpe AH, Orkin SH (1995) Lethal beta-thalassaemia in mice lacking the erythroid CACCC-transcription factor EKLF. Nature 375:318–322
267. Petrov A, Pirozhkova I, Carnac G, Laoudj D, Lipinski M, Vassetzky YS (2006) Chromatin loop domain organization within the 4q35 locus in facioscapulohumeral dystrophy patients versus normal human myoblasts. Proc Natl Acad Sci USA 103:6982–6987
268. Pfeifer D, Kist R, Dewar K, Devon K, Lander ES, Birren B, Korniszewski L, Back E, Scherer G (1999) Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to SOX9: evidence for an extended control region. Am J Hum Genet 65:111–124
269. Pier GB (2000) Role of the cystic fibrosis transmembrane conductance regulator in innate immunity to Pseudomonas aeruginosa infections. Proc Natl Acad Sci USA 97: 8822–8828
270. Pirastu M, Saglio G, Chang JC, Cao A, Kan YW (1984) Initiation codon mutation as a cause of alpha thalassemia. J Biol Chem 259:12315–12317
271. Poort SR, Rosendaal FR, Reitsma PH, Bertina RM (1996) A common genetic variation in the 3'-untranslated region of the prothrombin gene is associated with elevated plasma prothrombin levels and an increase in venous thrombosis. Blood 88:3698–3703
272. Rabbitts TH (1994) Chromosomal translocations in human cancer. Nature 372:143–149
273. Rebbeck TR (1997) Molecular epidemiology of the human glutathione S-transferase genotypes GSTM1 and GSTT1 in cancer susceptibility. Cancer Epidemiol Biomarkers Prev 6:733–743
274. Redon R, Ishikawa S, Fitch KR, Feuk L, Perry GH, Andrews TD, Fiegler H et al (2006) Global variation in copy number in the human genome. Nature 444:444–454
275. Reich DE, Schaffner SF, Daly MJ, McVean G, Mullikin JC, Higgins JM, Richter DJ, Lander ES, Altshuler D (2002) Human genome sequence variation and the influence of gene history, mutation and recombination. Nat Genet 32:135–142
276. Reiter LT, Hastings PJ, Nelis E, De Jonghe P, Van Broeckhoven C, Lupski JR (1998) Human meiotic recombination products revealed by sequencing a hotspot for homologous strand exchange in multiple HNPP deletion patients. Am J Hum Genet 62:1023–1033
277. Rigat B, Hubert C, Alhenc-Gelas F, Cambien F, Corvol P, Soubrier F (1990) An insertion/deletion polymorphism in the angiotensin I-converting enzyme gene accounting for half the variance of serum enzyme levels. J Clin Invest 86:1343–1346
278. Risch N, Tang H, Katzenstein H, Ekstein J (2003) Geographic distribution of disease mutations in the Ashkenazi Jewish population supports genetic drift over selection. Am J Hum Genet 72:812–822
279. Rosenthal A, Jouet M, Kenwrick S (1992) Aberrant splicing of neural cell adhesion molecule L1 mRNA in a family with X-linked hydrocephalus. Nat Genet 2:107–112
280. Rossiter JP, Young M, Kimberland ML, Hutter P, Ketterling RP, Gitschier J, Horst J, Morris MA, Schaid DJ, de Moerloose P, Sommer SS, Kazazian HH, Antonarakis SE (1994) Factor VIII gene inversions causing severe haemophilia A originate almost exclusively in male germ cells. Hum Mol Genet 3:1035–1039
281. Roth DB, Wilson JH (1986) Nonhomologous recombination in mammalian cells: role for short sequence homologies in the joining reaction. Mol Cell Biol 6: 4295–4304

282. Rousseau F, Heitz D, Mandel JL (1992) The unstable and methylatable mutations causing the fragile X syndrome. Hum Mutat 1:91–96
283. Rovelet-Lecrux A, Hannequin D, Raux G, Le Meur N, Laquerriere A, Vital A, Dumanchin C, Feuillette S, Brice A, Vercelletto M, Dubas F, Frebourg T, Campion D (2006) APP locus duplication causes autosomal dominant early-onset Alzheimer disease with cerebral amyloid angiopathy. Nat Genet 38:24–26
284. Ruwende C, Khoo SC, Snow RW, Yates SN, Kwiatkowski D, Gupta S, Warn P, Allsopp CE, Gilbert SC, Peschu N et al (1995) Natural selection of hemi- and heterozygotes for G6PD deficiency in Africa by resistance to severe malaria. Nature 376:246–249
285. Sachidanandam R, Weissman D, Schmidt SC, Kakol JM, Stein LD, Marth G, Sherry S et al (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409:928–933
286. Saiki RK, Scharf S, Faloona F, Mullis KB, Horn GT, Erlich HA, Arnheim N (1985) Enzymatic amplification of beta-globin genomic sequences and restriction site analysis for diagnosis of sickle cell anemia. Science 230:1350–1354
287. Santisteban I, Arredondo-Vega FX, Kelly S, Loubser M, Meydan N, Roifman C, Howell PL, Bowen T, Weinberg KI, Schroeder ML et al (1995) Three new adenosine deaminase mutations that define a splicing enhancer and cause severe and partial phenotypes: implications for evolution of a CpG hotspot and expression of a transduced ADA cDNA. Hum Mol Genet 4:2081–2087
288. Savkur RS, Philips AV, Cooper TA, Dalton JC, Moseley ML, Ranum LP, Day JW (2004) Insulin receptor splicing alteration in myotonic dystrophy type 2. Am J Hum Genet 74:1309–1313
289. Schmucker B, Krawczak M (1997) Meiotic microdeletion breakpoints in the BRCA1 gene are significantly associated with symmetric DNA-sequence elements. Am J Hum Genet 61:1454–1456
290. Schmutte C, Jones PA (1998) Involvement of DNA methylation in human carcinogenesis. Biol Chem 379:377–388
291. Schollen E, Pardon E, Heykants L, Renard J, Doggett NA, Callen DF, Cassiman JJ, Matthijs G (1998) Comparative analysis of the phosphomannomutase genes PMM1, PMM2 and PMM2psi: the sequence variation in the processed pseudogene is a reflection of the mutations found in the functional gene. Hum Mol Genet 7:157–164
292. Schroeder SA, Gaughan DM, Swift M (1995) Protection against bronchial asthma by CFTR delta F508 mutation: a heterozygote advantage in cystic fibrosis. Nat Med 1:703–705
293. Scott HS, Kudoh J, Wattenhofer M, Shibuya K, Berry A, Chrast R, Guipponi M, Wang J, Kawasaki K, Asakawa S, Minoshima S, Younus F, Mehdi SQ, Radhakrishna U, Papasavvas MP, Gehrig C, Rossier C, Korostishevsky M, Gal A, Shimizu N, Bonne-Tamir B, Antonarakis SE (2001) Insertion of beta-satellite repeats identifies a transmembrane protease causing both congenital and childhood onset autosomal recessive deafness. Nat Genet 27:59–63
294. Sebat J, Lakshmi B, Troge J, Alexander J, Young J, Lundin P, Maner S, Massa H, Walker M, Chi M, Navin N, Lucito R, Healy J, Hicks J, Ye K, Reiner A, Gilliam TC, Trask B, Patterson N, Zetterberg A, Wigler M (2004) Large-scale copy number polymorphism in the human genome. Science 305:525–528
295. Semenza GL (1994) Transcriptional regulation of gene expression: mechanisms and pathophysiology. Hum Mutat 3:180–199
296. Sethupathy P, Borel C, Gagnebin M, Grant GR, Deutsch S, Elton TS, Hatzigeorgiou AG, Antonarakis SE (2007) Human microRNA-155 on chromosome 21 differentially interacts with its polymorphic target in the AGTR1 3' untranslated region: a mechanism for functional single-nucleotide polymorphisms related to phenotypes. Am J Hum Genet 81(2):405–413
297. Sgourou A, Routledge S, Antoniou M, Papachatzopoulou A, Psiouri L, Athanassiadou A (2004) Thalassaemia mutations within the 5'UTR of the human beta-globin gene disrupt transcription. Br J Haematol 124:828–835
298. Shaikh TH, Kurahashi H, Saitta SC, O'Hare AM, Hu P, Roe BA, Driscoll DA, McDonald-McGinn DM, Zackai EH, Budarf ML, Emanuel BS (2000) Chromosome 22-specific low copy repeats and the 22q11.2 deletion syndrome: genomic organization and deletion endpoint analysis. Hum Mol Genet 9:489–501
299. Shapiro LJ, Yen P, Pomerantz D, Martin E, Rolewic L, Mohandas T (1989) Molecular studies of deletions at the human steroid sulfatase locus. Proc Natl Acad Sci USA 86:8477–8481
300. Sharp AJ, Locke DP, McGrath SD, Cheng Z, Bailey JA, Vallente RU, Pertz LM, Clark RA, Schwartz S, Segraves R, Oseroff VV, Albertson DG, Pinkel D, Eichler EE (2005) Segmental duplications and copy-number variation in the human genome. Am J Hum Genet 77:78–88
301. Sharp AJ, Hansen S, Selzer RR, Cheng Z, Regan R, Hurst JA, Stewart H, Price SM, Blair E, Hennekam RC, Fitzpatrick CA, Segraves R, Richmond TA, Guiver C, Albertson DG, Pinkel D, Eis PS, Schwartz S, Knight SJ, Eichler EE (2006) Discovery of previously unidentified genomic disorders from the duplication architecture of the human genome. Nat Genet 38:1038–1042
302. Sharp PA (1987) Splicing of messenger RNA precursors. Science 235:766–771
303. Shaw CJ, Lupski JR (2004) Implications of human genome architecture for rearrangement-based disorders: the genomic basis of disease. Hum Mol Genet 13(Spec No 1): R57–64
304. Shaw MA, Brunetti-Pierri N, Kadasi L, Kovacova V, Van Maldergem L, De Brasi D, Salerno M, Gecz J (2003) Identification of three novel SEDL mutations, including mutation in the rare, non-canonical splice site of exon 4. Clin Genet 64:235–242
305. Shaw-Smith C, Pittman AM, Willatt L, Martin H, Rickman L, Gribble S, Curley R, Cumming S, Dunn C, Kalaitzopoulos D, Porter K, Prigmore E, Krepischi-Santos AC, Varela MC, Koiffmann CP, Lees AJ, Rosenberg C, Firth HV, de Silva R, Carter NP (2006) Microdeletion encompassing MAPT at chromosome 17q21.3 is associated with developmental delay and learning disability. Nat Genet 38:1032–1037
306. Singleton AB, Farrer M, Johnson J, Singleton A, Hague S, Kachergus J, Hulihan M, Peuralinna T, Dutra A, Nussbaum R, Lincoln S, Crawley A, Hanson M, Maraganore D, Adler C, Cookson MR, Muenter M, Baptista M, Miller D,

Blancato J, Hardy J, Gwinn-Hardy K (2003) Alpha-Synuclein locus triplication causes Parkinson's disease. Science 302:841
307. Small K, Iber J, Warren ST (1997) Emerin deletion reveals a common X-chromosome inversion mediated by inverted repeats. Nat Genet 16:96–99
308. Spitz F, Montavon T, Monso-Hinard C, Morris M, Ventruto ML, Antonarakis S, Ventruto V, Duboule D (2002) A t(2;8) balanced translocation with breakpoints near the human HOXD complex causes mesomelic dysplasia and vertebral defects. Genomics 79:493–498
309. Spritz RA, Giebel LB, Holmes SA (1992) Dominant negative and loss of function mutations of the c-kit (mast/stem cell growth factor receptor) proto-oncogene in human piebaldism. Am J Hum Genet 50:261–269
310. Stamatoyannopoulos G (1991) Human hemoglobin switching. Science 252:383
311. Stefansson H, Helgason A, Thorleifsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A et al (2005) A common inversion under selection in Europeans. Nat Genet 37:129–137
312. Stenson PD, Ball EV, Mort M, Phillips AD, Shiel JA, Thomas NS, Abeysinghe S, Krawczak M, Cooper DN (2003) Human Gene Mutation Database (HGMD): 2003 update. Hum Mutat 21:577–581
313. Stoppa-Lyonnet D, Duponchel C, Meo T, Laurent J, Carter PE, Arala-Chaves M, Cohen JH, Dewald G, Goetz J, Hauptmann G et al (1991) Recombinational biases in the rearranged C1-inhibitor genes of hereditary angioedema patients. Am J Hum Genet 49:1055–1062
314. Subramanian S, Kumar S (2006) Evolutionary anatomies of positions and types of disease-associated and neutral amino acid mutations in the human genome. BMC Genomics 7:306
315. Sunyaev S, Ramensky V, Koch I, Lathe W 3rd, Kondrashov AS, Bork P (2001) Prediction of deleterious human alleles. Hum Mol Genet 10:591–597
316. Suter CM, Martin DI, Ward RL (2004) Germline epimutation of MLH1 in individuals with multiple cancers. Nat Genet 36:497–501
317. Terp BN, Cooper DN, Christensen IT, Jorgensen FS, Bross P, Gregersen N, Krawczak M (2002) Assessing the relative importance of the biophysical properties of amino acid substitutions associated with human genetic disease. Hum Mutat 20:98–109
318. Thein SL (2004) Genetic insights into the clinical diversity of beta thalassaemia. Br J Haematol 124:264–274
319. Thomas JW, Touchman JW, Blakesley RW, Bouffard GG, Beckstrom-Sternberg SM, Margulies EH, Blanchette M et al (2003) Comparative analyses of multi-species sequences from targeted genomic regions. Nature 424:788–793
320. Tishkoff SA, Verrelli BC (2003) Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet 4:293–340
321. Touchon M, Nicolay S, Audit B, BrodieofBrodie EB, d'Aubenton-Carafa Y, Arneodo A, Thermes C (2005) Replication-associated strand asymmetries in mammalian genomes: toward detection of replication origins. Proc Natl Acad Sci USA 102:9836–9841
322.
323. Treisman R, Orkin SH, Maniatis T (1983) Specific transcription and RNA splicing defects in five cloned beta-thalassaemia genes. Nature 302:591–596
324. Tufarelli C, Stanley JA, Garrick D, Sharpe JA, Ayyub H, Wood WG, Higgs DR (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. Nat Genet 34:157–165
325. Tuffery-Giraud S, Saquet C, Chambert S, Claustres M (2003) Pseudoexon activation in the DMD gene as a novel mechanism for Becker muscular dystrophy. Hum Mutat 21:608–614
326. Turner C, Killoran C, Thomas NS, Rosenberg M, Chuzhanova NA, Johnston J, Kemel Y, Cooper DN, Biesecker LG (2003) Human genetic disease caused by de novo mitochondrial-nuclear DNA transfer. Hum Genet 112:303–309
327. Tusie-Luna MT, White PC (1995) Gene conversions and unequal crossovers between CYP21 (steroid 21-hydroxylase gene) and CYP21P involve different mechanisms. Proc Natl Acad Sci USA 92:10796–10800
328. Upadhyaya M, Han S, Consoli C, Majounie E, Horan M, Thomas NS, Potts C, Griffiths S, Ruggieri M, von Deimling A, Cooper DN (2004) Characterization of the somatic mutational spectrum of the neurofibromatosis type 1 (NF1) gene in neurofibromatosis patients with benign and malignant tumors. Hum Mutat 23:134–146
329. Van Esch H (2006) The Fragile X premutation: new insights and clinical consequences. Eur J Med Genet 49:1–8
330. van Leeuwen FW, de Kleijn DP, van den Hurk HH, Neubauer A, Sonnemans MA, Sluijs JA, Koycu S, Ramdjielal RD, Salehi A, Martens GJ, Grosveld FG, Peter J, Burbach H, Hol EM (1998) Frameshift mutants of beta amyloid precursor protein and ubiquitin-B in Alzheimer's and Down patients. Science 279:242–247
331. Velagaleti GV, Bien-Willner GA, Northup JK, Lockhart LH, Hawkins JC, Jalal SM, Withers M, Lupski JR, Stankiewicz P (2005) Position effects due to chromosome breakpoints that map approximately 900 Kb upstream and approximately 1.3 Mb downstream of SOX9 in two patients with campomelic dysplasia. Am J Hum Genet 76:652–662
332. Verpy E, Leibovici M, Zwaenepoel I, Liu XZ, Gal A, Salem N, Mansour A, Blanchard S, Kobayashi I, Keats BJ, Slim R, Petit C (2000) A defect in harmonin, a PDZ domain-containing protein expressed in the inner ear sensory hair cells, underlies Usher syndrome type 1C. Nat Genet 26:51–55
333. Vidaud D, Vidaud M, Bahnak BR, Siguret V, Gispert Sanchez S, Laurian Y, Meyer D, Goossens M, Lavergne JM (1993) Haemophilia B due to a de novo insertion of a human-specific Alu subfamily member within the coding region of the factor IX gene. Eur J Hum Genet 1:30–36
334. Vitkup D, Sander C, Church GM (2003) The amino-acid mutational spectrum of human genetic disease. Genome Biol 4:R72
335. Vnencak-Jones CL, Phillips JA 3rd (1990) Hot spots for growth hormone gene deletions in homologous regions outside of Alu repeats. Science 250:1745–1748
336. Vogel F, Motulsky A (1986) Human genetics. Springer, Berlin

337. Vogt G, Chapgier A, Yang K, Chuzhanova N, Feinberg J, Fieschi C, Boisson-Dupuis S et al (2005) Gains of glycosylation comprise an unexpectedly large group of pathogenic mutations. Nat Genet 37:692–700
338. von Ahsen N, Oellerich M (2004) The intronic prothrombin 19911A>G polymorphism influences splicing efficiency and modulates effects of the 20210G>A polymorphism on mRNA amount and expression in a stable reporter gene assay system. Blood 103:586–593
339. Vyas P, Vickers MA, Simmons DL, Ayyub H, Craddock CF, Higgs DR (1992) Cis-acting sequences regulating expression of the human alpha-globin cluster lie within constitutively open chromatin. Cell 69:781–793
340. Wacey AI, Cooper DN, Liney D, Hovig E, Krawczak M (1999) Disentangling the perturbational effects of amino acid substitutions in the DNA-binding domain of p53. Hum Genet 104:15–22
341. Wallace MR, Andersen LB, Saulino AM, Gregory PE, Glover TW, Collins FS (1991) A de novo Alu insertion results in neurofibromatosis type 1. Nature 353:864–866
342. Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G et al (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science 280:1077–1082
343. Wang Z, Moult J (2001) SNPs, protein structure, and disease. Hum Mutat 17:263–270
344. Warren ST, Nelson DL (1994) Advances in molecular analysis of fragile X syndrome. JAMA 271:536–542
345. Watnick TJ, Gandolph MA, Weber H, Neumann HP, Germino GG (1998) Gene conversion is a likely cause of mutation in PKD1. Hum Mol Genet 7:1239–1243
346. Weatherall DJ, Clegg JB, Higgs DR, Wood WG (2001) The hemoglobinopathies. In: Scriver CR, Beaudet AL, Valle D et al (eds) The metabolic and molecular bases of inherited disease. McGraw-Hill, New York, pp 4571–4636
347. Weber JL, May PE (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. Am J Hum Genet 44:388–396
348. Wells RD, Warren AC (1998) Genetic instabilities and hereditary neurological disorders. Academic, San Diego
349. Wexler NS, Young AB, Tanzi RE, Travers H, Starosta-Rubinstein S, Penney JB, Snodgrass SR, Shoulson I, Gomez F, Ramos Arroyo MA et al (1987) Homozygotes for Huntington's disease. Nature 326:194–197
350. Williams TN, Mwangi TW, Wambua S, Peto TE, Weatherall DJ, Gupta S, Recker M, Penman BS, Uyoga S, Macharia A, Mwacharo JK, Snow RW, Marsh K (2005) Negative epistasis between the malaria-protective effects of alpha+-thalassemia and the sickle cell trait. Nat Genet 37:1253–1257
351. Williams TN, Wambua S, Uyoga S, Macharia A, Mwacharo JK, Newton CR, Maitland K (2005) Both heterozygous and homozygous alpha+ thalassemias protect against severe and fatal Plasmodium falciparum malaria on the coast of Kenya. Blood 106:368–371
352. Witchel SF, Lee PA, Suda-Hartman M, Trucco M, Hoffman EP (1997) Evidence for a heterozygote advantage in congenital adrenal hyperplasia due to 21-hydroxylase deficiency. J Clin Endocrinol Metab 82:2097–2101
353. Wolf U (1997) Identical mutations and phenotypic variation. Hum Genet 100:305–321
354. Wong C, Dowling CE, Saiki RK, Higuchi RG, Erlich HA, Kazazian HH Jr (1987) Characterization of beta-thalassaemia mutations using direct genomic sequencing of amplified single copy DNA. Nature 330:384–386
355. Woods-Samuels P, Wong C, Mathias SL, Scott AF, Kazazian HH Jr, Antonarakis SE (1989) Characterization of a nondeleterious L1 insertion in an intron of the human factor VIII gene and further evidence of open reading frames in functional L1 elements. Genomics 4:290–296
356. Woods-Samuels P, Kazazian HH Jr, Antonarakis SE (1991) Nonhomologous recombination in the human genome: deletions in the human factor VIII gene. Genomics 10:94–101
357. Woodward K, Kendall E, Vetrie D, Malcolm S (1998) Pelizaeus-Merzbacher disease: identification of Xq22 proteolipid-protein duplications and characterization of breakpoints by interphase FISH. Am J Hum Genet 63:207–217
358. Wyman AR, White R (1980) A highly polymorphic locus in human DNA. Proc Natl Acad Sci USA 77: 6754–6758
359. Yang WS, Nevin DN, Peng R, Brunzell JD, Deeb SS (1995) A mutation in the promoter of the lipoprotein lipase (LPL) gene in a patient with familial combined hyperlipidemia and low LPL activity. Proc Natl Acad Sci USA 92:4462–4466
360. Young M, Inaba H, Hoyer LW, Higuchi M, Kazazian HH Jr, Antonarakis SE (1997) Partial correction of a severe molecular defect in hemophilia A, because of errors during expression of the factor VIII gene. Am J Hum Genet 60:565–573
361. Youssoufian H, Kazazian HH Jr, Phillips DG, Aronis S, Tsiftis G, Brown VA, Antonarakis SE (1986) Recurrent mutations in haemophilia A give evidence for CpG mutation hotspots. Nature 324:380–382
362. Youssoufian H, Antonarakis SE, Aronis S, Tsiftis G, Phillips DG, Kazazian HH Jr (1987) Characterization of five partial deletions of the factor VIII gene. Proc Natl Acad Sci USA 84:3772–3776
363. Youssoufian H, Antonarakis SE, Bell W, Griffin AM, Kazazian HH Jr (1988) Nonsense and missense mutations in hemophilia A: estimate of the relative mutation rate at CG dinucleotides. Am J Hum Genet 42:718–725
364. Yu S, Mangelsdorf M, Hewett D, Hobson L, Baker E, Eyre HJ, Lapsys N, Le Paslier D, Doggett NA, Sutherland GR, Richards RI (1997) Human chromosomal fragile site FRA16B is an amplified AT-rich minisatellite repeat. Cell 88:367–374
365. Zlotogora J (1994) High frequencies of human genetic diseases: founder effect with genetic drift or selection? Am J Med Genet 49:10–13
366. Zlotogora J (1997) Dominance and homozygosity. Am J Med Genet 68:412–416
367. Zlotogora J (1998) Germ line mosaicism. Hum Genet 102:381–386
368. Zschocke J (2003) Phenylketonuria mutations in Europe. Hum Mutat 21:345–356