Primer on Medical Genetics

2

Medical genetics traditionally deals with clinical conditions purely or largely due to genetic causes such as single-gene (Mendelian) disorders. More than 4500 Mendelian disorders are known but they constitute only 1% of all diseases. Information used by medical geneticists to identify these rare diseases has strong implications for genetic association study design, execution, and interpretation. Most current genetic association studies are typically for **multifactorial** or **complex disorders**, which are determined by the interplay among multiple gene variants and environmental factors, but the principles of medical genetics are important in understanding the nature of multifactorial disorders and setting up informative studies. It should be noted that mutations causing Mendelian disorders were located using **linkage studies**, which are not covered in this book. This chapter will review the main principles of medical genetics and how they connect to the study of multifactorial disorders by association studies.

2.1 The Location of Disease-causing Mutations

Genes with disease-causing mutations can be in any part of the human genome

Disease susceptibility is modified by a mutation that changes the activity of a gene. The terms "mutation" and "polymorphism" are frequently used interchangeably, but the differences are listed in **Table 2.1**. Every polymorphism is a mutation, but it is only referred to as a polymorphism when it occurs in more than 1% of the population. In the context of disease risk, most mutations cause disease in most of their bearers (they have high **penetrance**), but polymorphisms only show a weak effect on predisposition to disease. While disease-causing mutations are usually missense coding-region variants, most polymorphisms change the expression level of a gene; this can be a gene in the vicinity (***cis*-effect**; which is usually defined as within 5 Mb), or at a distance (>5 Mb away), or even on a

Table 2.1 Main differences between mutations and polymorphisms

	Mutation	Polymorphism
Frequency	Rare (≤1%)	Common (>1%)
Role in disease development	Causal; necessary and sufficient (deterministic)	Weak modifying/predisposing effect; neither necessary nor sufficient
Effect size (relative risk)	Very high (>10)	Small to modest (1.1 to 3.0)
Penetrance	Very high (~100%)	Low (~1–2%)
Inheritance pattern	Clear (Mendelian)	Equivocal
Identification	Linkage studies in families	Association studies in unrelated individuals

different chromosome (***trans*-effect**). The gene whose function is changed by a variant is called the **disease gene** if the change is strong enough to be instrumental in disease pathogenesis. The affected gene may be on an autosomal chromosome and the disease it causes is evident in both males and females (autosomal genetic diseases), or it can be on a sex chromosome, in which case one sex is more likely to be affected than the other (sex-linked genetic diseases). The **mitochondrial genome** also contains genes whose variants can cause serious medical conditions.

The majority of disease genes carrying disease-causing mutations are on autosomal chromosomes

Given that 22 of the 23 pairs of chromosomes are autosomal, a great majority of mutations occur in genes on these chromosomes. However, the effects of autosomal genetic variants may be influenced by either the products of genes on a sex chromosome or sex hormones, and may therefore show sex specificity. Examples of sex-specific traits include male pattern baldness and hairy ears, which are encoded by genes in autosomal chromosomes but expressed only in males. This point has to be remembered in genetic association studies: differences in the incidence rates of diseases between the sexes do not necessarily mean the disease gene is on a sex chromosome. Sometimes, the effect of the variant on an autosomal gene depends on the parental origin of the chromosome. This is called a **parental effect** and is discussed in a later section of this chapter.

Only three of the 23 pairs of chromosomes (Chromosomes 21, 13, and 18) can be found in a trisomic state compatible with life. The reason for this is that these chromosomes are the ones with the fewest genes and they can be tolerated in a trisomic state. Fetuses with multiple copies of larger and more gene-dense chromosomes are lost before birth. There is nothing special about the genes on Chromosomes 21, 13, and 18 that cause certain syndromes when present in three copies. Likewise, disease genes are equally distributed in autosomal chromosomes and there is generally no preferential distribution of them on any particular chromosome. Possible exceptions are the high number of disease associations with HLA-region variants and the disproportionately high number of immune system-related genes in the X chromosome.

Disease-causing mutations are also present in genes located on sex chromosomes

The human sex chromosomes are the X and Y chromosomes. The Y chromosome contains around 100 unique protein- and non-protein-coding genes that have no counterparts on the X chromosome, but there are no inherited disorders due to a mutation in a disease gene on the Y chromosome. Microdeletion of the azoospermia factor (AZF) region of the Y chromosome is common in oligospermia/azoospermia, and there are some suggestive studies for associations of Y chromosome variants with cardiovascular disease risk. It should be noted that the Y chromosome is about three times smaller than the X chromosome and a portion at both tips is shared by both chromosomes (**Figure 2.1**). These shared portions are called **pseudoautosomal regions** (PAR1 and PAR2) and act as autosomal genes. Approximately 95% of the Y chromosome is unique and is called the male-specific region. An important feature of this region is that it does not recombine, due to the lack of a corresponding chromosome with which to exchange material. This means that the Y chromosome is transmitted as invariant blocks from male to male. These blocks, consisting of specific variants, are called haplogroups (**Figure 2.2**) and are widely used in

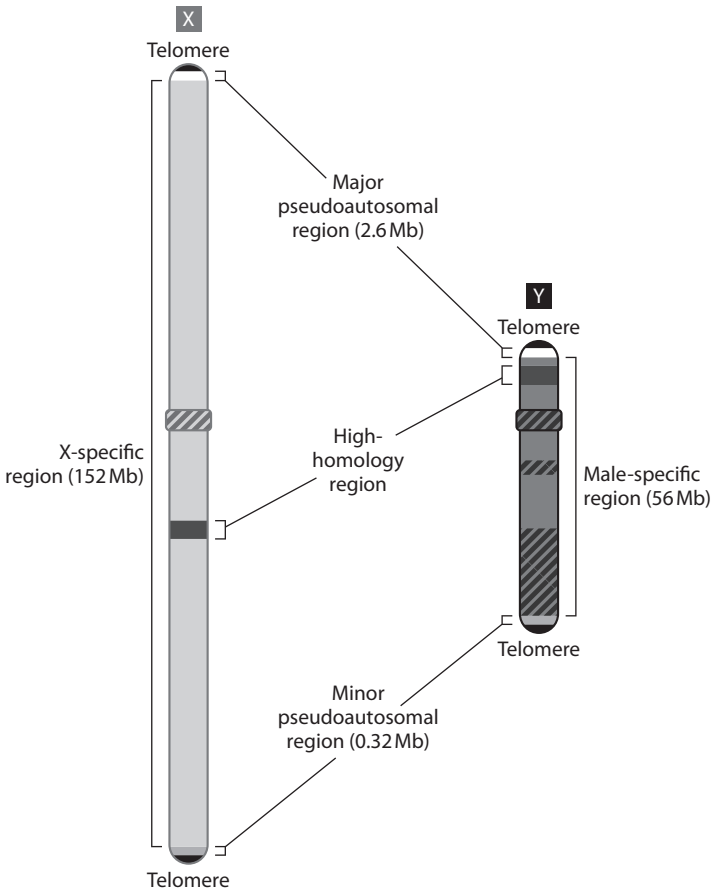

Figure 2.1 Shared pseudoautosomal regions at the tips of X and Y chromosomes. These regions behave like autosomal genes rather than as sex chromosomes in inheritance patterns. (From Strachan T, Goodship J & Chinnery P [2015] Genetics and Genomics in Medicine. Garland Science.)

phylogenetic studies. Genetic association studies examining why there may be sex disparity in certain diseases have recently started to use the Y chromosome haplogroups. The microarray chips used in genome-wide association studies have only recently included SNPs from the Y chromosome and more data on Y chromosome polymorphisms and disease susceptibility should become available soon. There is no reason to exclude these

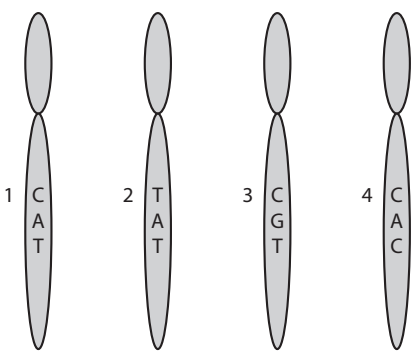

Figure 2.2 Haplogroups are formed by different constellations of multiple polymorphisms on the same chromosome. In this example, three SNPs (C > T; A > G; T > C) form multiple haplotypes: haplogroup 1 is represented by C-A-T; haplogroup 2 by T-A-T; haplogroup 3 by C-G-T; and haplogroup 4 by C-A-C.

SNPs from any study. However, data from the Y chromosome should be analyzed separately as allelic associations and, obviously, only in males.

While any Y chromosome-linked trait is transmitted from male to male, traits encoded in the X chromosome can be transmitted from either parent but are usually expressed in males. The preferential expression in males is due to the presence of only one copy of the X chromosome in males (females have two copies). While both males and females express dominant mutations, which require only one copy, females need to have two copies of a recessive mutation to express the trait. Males express single-copy recessive mutations because there is no second copy of the X chromosome to mask the effect of the recessive mutation. The imbalance in the number of X chromosomes in males and females creates some difficulty in the statistical analysis of data from X chromosome SNPs. Early association studies either did not include the X chromosome in the study or did not analyze the data. With the advent of statistical methods that make these analyses feasible, X chromosome associations are now being examined. Such analysis should not be restricted to disorders with sex effects because any SNP on the X chromosome, especially in PARs, may be involved in any disease with or without a known sex-specific presentation.

Mitochondrial disease genes can only be inherited from the mother

While Y chromosomes are only transmitted to the offspring from the father, the mitochondrial genome is transmitted only by the mother (to both male and female offspring). Depending on the cell type, each cell contains from a few to hundreds of mitochondria, and each mitochondrion has several copies of the mitochondrial genome. The result is that each cell contains an average of 500 copies of the mitochondrial genome. Mitochondrial DNA predominantly encodes proteins that are components of the mitochondrial respiratory chain. Some nuclear genes influence mitochondrial functions, so the source of a mitochondrial disorder may not exclusively be in the mitochondrial genome. Due to the variable number of copies of the mitochondrial genome in each cell, cells may contain a mixture of mutant and wild-type mitochondrial genomes, a situation called **heteroplasmy**, which may lead to phenotypic variability among the offspring of the same mother. The ratio of wild-type and mutant chromosomes may also shift during each cell replication. Because redox reactions take place in mitochondria, the production of reactive oxygen species is high and the ***de novo* mutation** rate is around 1000 times higher in mitochondrial DNA compared with nuclear DNA. This results in sporadic disorders caused by mitochondrial DNA mutations. The high mutation rate also generates a lot of polymorphisms that are useful markers for disease association studies and, in the case of mitochondrial haplogroups, for phylogenetic studies. Comprehensive genetic association studies should include mitochondrial DNA SNPs in their genotyping schemes.

2.2 Single-Gene, Oligogenic, and Multigenic Disorders

Medical geneticists traditionally deal with genetically well-characterized conditions called **single-gene disorders**, where a gene becomes nonfunctional through a high-penetrance mutation. Most of the well-documented medical genetic conditions—such as cystic fibrosis, phenylketonuria, congenital adrenal hyperplasia, and rare inherited forms of common diseases such as Parkinson's disease and breast cancer—are single-gene disorders. They are called single-gene disorders but they do not have to be single-mutation disorders as the disease gene is commonly incapacitated by different mutations. Inheritance of these diseases follows Mendelian patterns (**Table 2.2**). Some diseases, however,

Table 2.2 Main features of Mendelian inheritance patterns

	Autosomal dominant	Autosomal recessive	X-linked dominant	X-linked recessive	Y-linked	Mitochondrial
Example	Huntington's disease	Cystic fibrosis	Vitamin D-resistant rickets	Hemophilia	Male infertility	Leber's hereditary optic neuropathy
Multiple generations affected?	YES	**NO** (skips generation)	YES	**NO** (skips generation)	YES	YES
Is a parent always affected?	YES (unless it is a new mutation)	NO	YES	NO	YES	YES
Are both sexes affected?	YES	YES	YES (F > M)	YES (M > F)	**NO**[a]	YES
Are all affected individuals male?	NO	NO	NO	**YES**[b]	**YES**[a]	NO
Male-to-male transmission?	POSSIBLE	POSSIBLE	**IMPOSSIBLE**[c]	**IMPOSSIBLE**[c]	YES	IMPOSSIBLE
Is it always transmitted from the mother?	NO	NO	NO	NO	NO	**YES**[d]
Is it always transmitted from the father?	NO	NO	NO	NO	**YES**	NO
	50% of the children of an affected parent will be affected. Each affected child has an affected parent	One in four children of healthy (carrier) parents will be affected. An affected child usually has unaffected parents	All female children of an affected father are affected	No male children of an affected father are affected	Only males are affected	Transmitted only from an affected mother (no transmission from an affected father)

[a]A father's Y chromosome is always and only inherited by sons (never by daughters); [b]a son receives an X chromosome from his mother and does not inherit his father's X chromosome; [c]a father's X chromosome is always inherited only by a daughter; [d]mitochondrial DNA is only transmitted from the mother.

are caused or modulated by more than one gene. It is now clear that even single-gene disorders are strongly modulated by additional (modifier) genes. Diseases caused by interactions of mutations in several genes are called **oligogenic disorders**. One example is the digenic form of retinitis pigmentosa, which results from simultaneous heterozygous mutations in two genes, *ROM1* and *RDS*. Another disease that has a digenic form is Hirschsprung disease, where the *RET* and *GDNF* genes are mutated. In Bardet-Biedel syndrome, three mutations in two genes are needed for disease development (tri-allelic inheritance). **Multigenic disorders** are more common than single-gene or oligogenic disorders and are caused by mutations or functional variants in many genes. The genetic basis of multigenic disorders is explored by most current genetic association studies. Even in single-gene disorders the environment also plays a role, interacting with genetic variants. For example, phenylketonuria cannot result from the gene mutation only; if the environment (dietary phenylalanine intake) is controlled, the patient will never develop the disease. Thus, almost all diseases can now be called multifactorial.

Different mutations can cause the same phenotype

With increasing knowledge of the human genome, our view of the genetic basis of even the simplest genetic disorders is evolving rapidly. Most single-gene disorders are first attributed to a single mutation. Additional mutations are then recognized either in the same gene and with similarly strong effects on the phenotype or in additional genes that result in a similar phenotype. These two situations are known as **allelic heterogeneity** and **genetic heterogeneity**, respectively (**Figure 2.3**). Genetic heterogeneity is also known as **locus heterogeneity**. These concepts are highly relevant in complex disease genetics as they are always caused by multiple genetic variants. Recently, epigenetic changes that modify gene activity have been added to the mix of genomic changes that cause or modify single-gene disorders.

One condition with tremendous heterogeneity is retinitis pigmentosa. It has monogenic and digenic forms, and a large proportion of cases present with an unknown genetic basis. Another feature of retinitis pigmentosa is that even a certain gene mutation may be inherited via different modes, causing an autosomal recessive inheritance pattern in one family and an autosomal dominant pattern in another. Allelic and genetic heterogeneity both have important implications in genetic association studies. In two different studies, it is not uncommon to find associations with different variants of the same gene. Likewise, failing to replicate an association finding in a second study may be due to genetic heterogeneity and the predominant effect of a different gene in a different population due to different modifiers. Therefore, even negative results may have a plausible biological explanation in light of allelic and genetic heterogeneity.

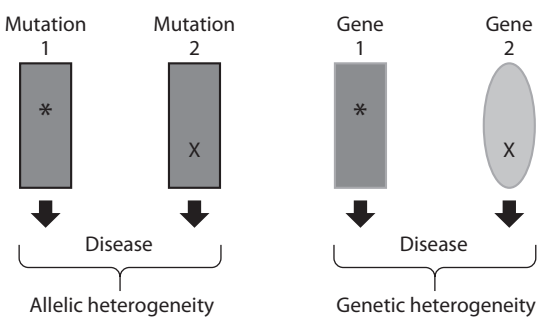

Figure 2.3 Allelic and genetic heterogeneity. In allelic heterogeneity, different variants of the same gene increase the risk for the same disease. In genetic heterogeneity, variants in different genes independently increase the risk for the same disease.

A combination of multiple genetic and nongenetic factors leads to multifactorial inheritance

As opposed to single-gene disorders, disorders such as diabetes, rheumatoid arthritis, schizophrenia, and most cancers are determined by a large, but unknown, number of genetic and nongenetic variables, making them multifactorial traits. The genetic variables participating in susceptibility to multifactorial diseases are generally low-penetrance variants that would yield very small, if any, **effect sizes** on their own. It is their interaction with one another that generates the **genetic load** for increased susceptibility. Susceptibility to a multifactorial disease can be seen as a quantitative trait, and when the accumulation of genetic and environmental factors exceeds a critical threshold, the disease occurs. This concept is known as **Falconer's polygenic threshold model** for dichotomous non-Mendelian characters (**Figure 2.4**). There are several implications of this model:

- The greater the number of predisposing risk genes possessed by the parents, the greater the probability that they will have an affected offspring, depending on the contribution of the environmental factors.

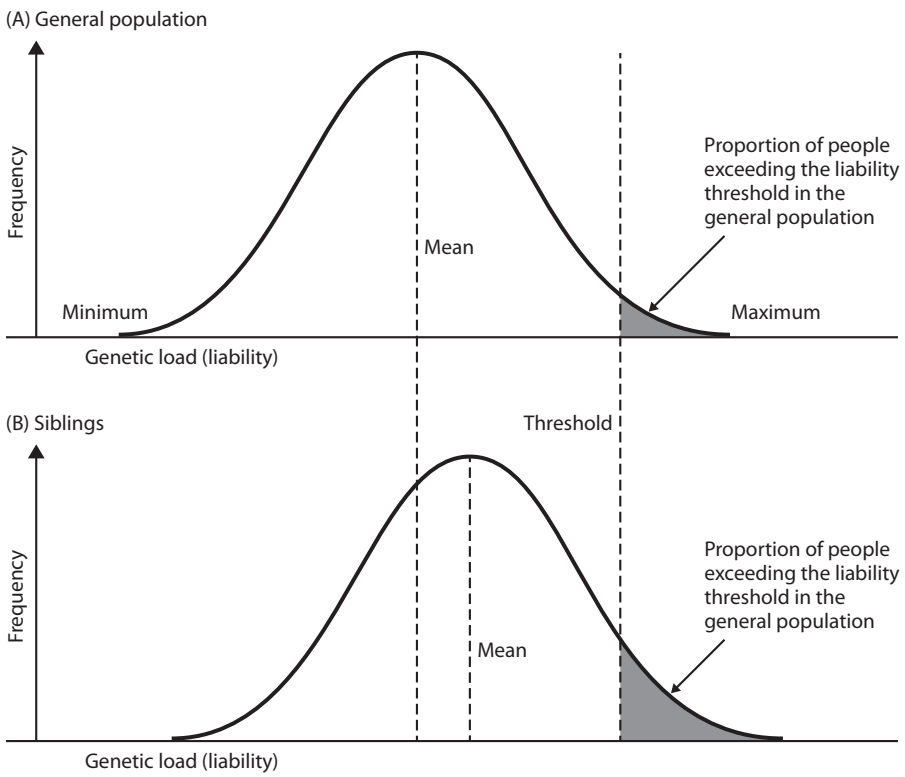

Figure 2.4 Falconer's polygenic threshold model. (A) The distribution of liability or the amount of genetic contributors to risk in the general population shows a normal distribution, with most people having average liability. Only those whose liability exceeds the polygenic threshold develop the disease. (B) Siblings of diseased individuals have higher liability on average, and a higher proportion of siblings have a liability exceeding the threshold. Thus, a higher proportion of siblings develop the disease than the proportion of the general population.

- The disease may show a familial aggregation but without a discernible Mendelian segregation pattern.
- Familial risk declines with increasingly remote degrees of relationship (from first-degree relatives toward third- and lower-degree relatives).
- The greater the number of affected family members, the higher the recurrence risk in other family members.
- Recurrence risk increases with severity of the disorder in the index case.
- When there is a marked difference in incidence between sexes, it is because of differences in risk threshold between males and females. The less frequently affected sex has a higher risk threshold and, simply because of greater load, transmits the condition more often to the more frequently affected sex, which has a lower threshold.

Single-gene disorders are now being considered more as multifactorial disorders with one major mutation and multiple modifiers. The distinction between single-gene disorders and multifactorial or complex disorders is therefore becoming blurred. However, in a complex trait, there is no main determinant and the interaction of multiple modifier genes results in the disease. The main implication for genetic association studies relates to study design and the need to have comprehensive coverage of genetic and environmental variables.

2.3 Copy Number of Mutant Genes

The term "disease gene" is often used, but what is meant by this is a gene that is rendered less functional by mutation. Deactivation of a gene may result from both copies being deleted or inactivated due to deleterious mutations (or epigenetic changes), or one copy being inactivated and the remaining copy being either insufficient to maintain normal function (haploinsufficiency) or rendered inactive by the mutant copy (dominant-negative mutation), usually by physical interaction (**Figure 2.5**). Mutations on one or both copies of a gene can therefore cause disease. The nature of the mutation and whether one or two copies of it cause the disease determines the inheritance pattern (dominant or recessive).

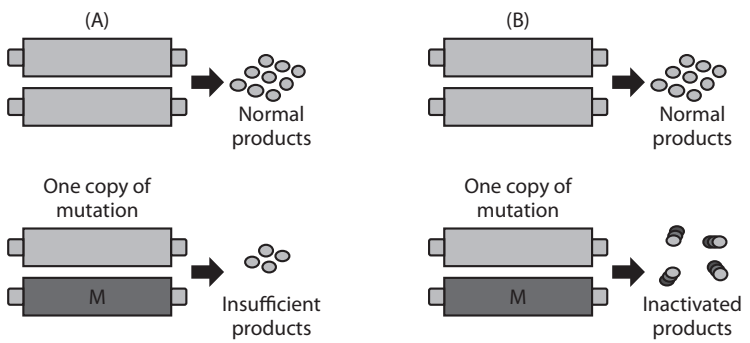

Figure 2.5 Two mechanisms for how a single copy of a mutation renders the gene inactive. (A) In haploinsufficiency, one intact copy of the gene does not produce a sufficient amount of the protein product. (B) A dominant-negative mutation results in the production of a mutant peptide that inactivates the normal product. In both instances, not enough normal product is available to carry out the function.

The location of the gene further determines the autosomal or sex chromosome inheritance pattern. If there is a discernible pattern in disease transmission for a complex disease, determining this pattern may prove useful in the design and analysis of genetic association studies. The main features of inheritance patterns (including mitochondrial transmission) are summarized in Table 2.2. In a complex disease, the mode of action of an individual SNP is usually unknown. It may therefore be necessary to examine different genetic risk models for different inheritance patterns for SNPs in genetic association studies rather than assuming that one model is powerful enough to yield associations for any model.

Disease-causing mutations are propagated by a variety of mechanisms

It is counterintuitive that a mutation that kills its bearers, even before their reproductive age, may continue to be present in the population. It is, however, possible for mutations, especially those that cause recessive diseases, to propagate themselves mainly via unaffected heterozygotes. Although those individuals who develop the disease may not be able to pass the mutant gene to the next generation, there will be a few individuals who have not yet developed the disease or who have developed a milder form and they will transmit the mutation. Furthermore, it has been documented for a number of recessive disease-causing mutations that they may even confer an advantage to the carriers of one copy of the mutation (in heterozygous form). This phenomenon is one form of **heterozygote advantage** and most commonly occurs by conferring resistance to an infectious agent; such mutations remain in the population. Some examples of advantages conferred by disease-causing mutations are shown in **Table 2.3**. For any homozygous offspring who has inherited two copies of the mutant gene, there are two heterozygous siblings who have inherited only one copy, and those individuals will

Table 2.3 Suggested mechanisms for the maintenance of deleterious mutations in the population

Disease	Mutant gene	Advantage for heterozygotes
Sickle-cell anemia	*HBB*	Resistance to *Plasmodium falciparum*
α-Thalassemia	*HBA1/2*	Resistance to *Plasmodium falciparum*
Hemoglobin C	*HBB*	Resistance to *Plasmodium falciparum*
Glucose 6-phosphate dehydrogenase deficiency	*G6PD*	Resistance to *Plasmodium falciparum*
Cystic fibrosis	*CFTR*	Resistance to cholera toxin and *Salmonella enterica* serovar Typhi; increased fertility
Congenital adrenal hyperplasia	*CYP21A2*	Increased fertility; brisk cortisol response in stress
Hereditary hemochromatosis	*HFE*	Protection from iron deficiency; protection from *Salmonella enterica* serovar Typhimurium infection in mice
Tay-Sachs disease	*HEXA*	Resistance to *Mycobacterium tuberculosis*
Phenylketonuria	*PAH*	Protection from spontaneous miscarriage (via reversal of toxicity of ochratoxin A)
Huntington's disease	*HTT*	Increased fertility, decreased risk of cancer
α_1-Antitrypsin deficiency	*SERPINA1*	Increased fertility and twinning

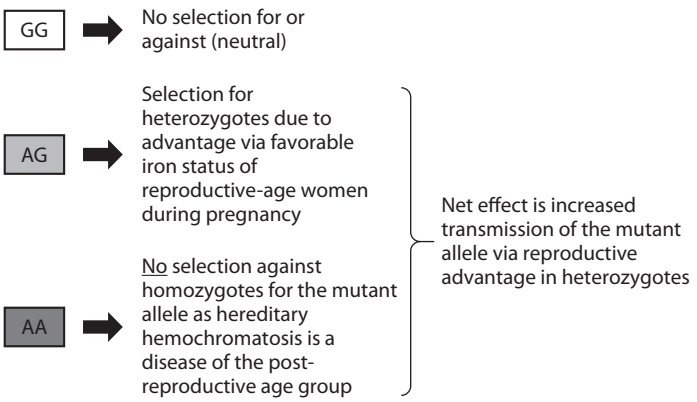

Figure 2.6 The *HFE* mutation C282Y that causes the autosomal recessive disease hereditary hemochromatosis is maintained in the population via heterozygote advantage. The C282Y mutation is due to a nucleotide substitution, G to A. The common genotype GG confers no advantage or disadvantage to its carriers. Heterozygotes (AG) are at an advantage due to increased iron absorption. Because the disease occurs later in life, selection against mutation homozygotes (AA) does not occur. Selection for heterozygous women of reproductive age maintains the frequency of the mutation in the population.

continue transmitting the disease mutation. An example of propagation of a disease-causing mutation is given in **Figure 2.6**.

These observations have implications for genetic association studies. Heterozygote advantage is real and should be considered in the analysis phase of a genetic association study. Heterozygote advantage is also highly relevant in infectious disease susceptibility determined by HLA antigens. The observations listed in Table 2.3 are relevant to SNP selection or prioritization in genetic association studies. If there are indications that a SNP or a genomic region has been subject to selective pressures, those SNPs should be given priority over others since this is evidence for their functionality.

Heterozygosity for mutations causing recessive diseases may have functional effects

Recessive traits are caused by mutations in both copies of a gene, although the mutations in each copy may be different, a situation known as **compound heterozygosity**. Hereditary hemochromatosis shows compound heterozygosity for *HFE* mutations C282Y and H63D, and congenital adrenal hyperplasia shows compound heterozygosity for a number of *CYP21A2* mutations. While such allelic heterogeneity is well recognized, there has been less interest in the effects of heterozygosity for recessive mutations. Heterozygosity does not cause the disease—or even a milder form of clinically detectable disease—in carriers, but, at least for biochemical traits, there is evidence that biochemical changes are detectable. For example, in heterozygotes for *HFE* mutations, serum iron parameters are often mildly changed but do not come close to the iron overload levels seen in some homozygotes. Likewise, heterozygotes for *CYP21A2* mutations have a lesser degree of the biochemical signs of hormonal changes that are seen in homozygotes with full-blown disease. Although there is considerable overlap of the biochemical values observed in heterozygotes and mutation-free healthy people, a proportion of heterozygotes have hormone levels slightly above normal values. Such people are called **biochemically manifesting heterozygotes** or just manifesting carriers. Heterozygosity for mutations causing rare

recessive disorders is not as rare as homozygosity. In certain populations, up to 50% of individuals may be heterozygous for *HFE* mutations. It is estimated that in Europeans, carrier frequencies for cystic fibrosis and factor V Leiden deficiency are 1/29 and 1/14, respectively, but there has been no long-term follow-up of the carriers of recessive disease mutations for health outcomes. Recessive mutations should not be dismissed when they exist in one copy and should be included in genetic association studies whenever they are relevant to the phenotype under study.

Polymorphisms in disease genes may also modify disease risk

Disease-causing mutations are generally highly functional mutations that cause a severe loss of gene function. The *HFE* C282Y mutation, for example, causes a total loss of the cell surface expression of HFE protein. *CYP21A2* mutations similarly decrease the enzymatic activity of the gene product to zero or near zero. The tumor suppressor gene *TP53* may be totally inactivated by its cancer-causing mutations. However, disease-causing mutations are not the only changes in the respective genes. Like all genes, genes involved in single-gene disorders also have low-penetrance polymorphisms, mainly in the form of SNPs, and some of these will be functional. The effects on gene function of low-penetrance SNPs are much less than those of disease-causing, high-penetrance mutations, but they may be detectable in a genetic association study. For *HFE*, *CYP21A2*, *TP53*, *BRCA1*, and *BRCA2*, such functional polymorphisms have been identified and associations with various traits have been reported. Thus, polymorphisms of disease genes may still contribute to disease susceptibility even when there are no mutations, and they should be considered in genetic association studies when relevant.

2.4 Inherited Mutations

The mutations that cause inherited disorders are present in germ-line DNA in either, or both, the sperm or egg, so that the offspring inherits the mutation from either or both parents and the mutation is present in all cells of the offspring derived from the zygote. Polymorphisms that modify disease risk are also present in the germ line. Rarely, a mutation may occur during intrauterine development and it will then only be present in the lineage of the mutant cell. This is known as **mosaicism** and describes the co-presence of mutant and nonmutant cells in the body. A well-known example of mosaicism is the Dalmatian dogs' coat, which contains a mixture of mutant and nonmutant pigment cells that generate the typical appearance of spots. Mosaicism is also possible in the germ-cell population of either parent. Mosaicism is frequently sought either in parents or in offspring when an inheritance pattern is difficult to explain. One example is the appearance of an autosomal dominant disorder in multiple offspring when both parents are unaffected. This can happen if either one of the parents has germ-cell mosaicism where the mutation is present in their germ cells but not in their somatic cells. The situation could also arise if all affected offspring had *de novo* mutations that occurred during their development, but this is very unlikely. A mosaic parent with a mild form of a single-gene disorder can have offspring that are more severely affected. In this case, the offspring will have the mutant gene in every cell of their body while the parent only has the mutation in a proportion of their somatic cells.

Undetected mosaicism is probably more common than detected mosaicism but does not cause great concern in a genetic association study. In other words, one should expect to find the same genotype in any biospecimen collected from the same individual.

Therefore, it is acceptable, but not ideal, to have DNA extracted from peripheral blood cells in patients and from buccal cells in healthy controls, as is frequently the case in studies of childhood disorders. In a survey of 100 families with children with genetic disorders, 4% of the healthy parents were found to have indications of somatic mosaicism. These individuals with somatic mosaicism have a mixture of mutant and nonmutant cells.

Certain environmental exposures cause mutations only in a certain tissue, for example in the skin following ultraviolet (UV) radiation exposure and in liver cells after aflatoxin exposure. These mutations may cause local effects (skin cancer or liver cancer in these examples) but would not be detectable in other tissues such as peripheral blood cells. Cancer-related mutations detected in genomic studies of tumor cells, such as the ongoing The Cancer Genome Atlas (TCGA) study by the NIH, do not exist in other somatic cells of affected individuals. The implications of these findings in genomic association studies are not clear but are probably minimal. Likewise, such somatic mutations may not have strong implications for understanding what has caused that cell to turn into a cancer cell. What is important in genetic association studies is to identify the variations that predispose individuals (hosts) to cancer development; these are likely to be very different from those detected in an end-stage tumor cell, which has gone through multiple rounds of genomic changes including structural, mutational, and epigenetic alterations. Thus, designing a genetic association study in cancer to examine the variants of a gene that has been found to be highly expressed in a tumor cell, but not in a normal host cell, would not have a strong justification.

Mutations do not always produce equal effects in populations

Two populations with the same frequency of a disease-causing mutation may have different frequencies of the disease, and diseased individuals in each population may have different average levels of disease severity. These two situations are determined by the penetrance and **expressivity** of the mutation. If only a proportion of people with the mutation develop the disease, it is called reduced penetrance**.** If all affected people do not suffer from the disease at the same severity, it is called variable expressivity (**Figure 2.7**). Reduced penetrance may even be evident in a single family, for example when an individual with the mutation of an autosomal dominant disorder such as retinoblastoma does not have the disease despite having a diseased parent and a diseased child. This is a case of reduced penetrance in an individual who has the mutation but is missing a crucial contributor to disease development. In addition to genetic and allelic heterogeneity (see Figure 2.3), there are often unknown genetic and environmental modifiers that determine whether a mutation causes a disease and the severity of the disease caused.

The iron overload disease hereditary hemochromatosis, caused by the *HFE* C282Y mutation, is a well-known example of an inherited disease that is strongly modified by other genetic variants and lifestyle factors. Less than 5% of people homozygous for the C282Y mutation develop the clinical disease (although more may have biochemical changes) and not all of them have a severe form of disease. Both the penetrance and expressivity of the C282Y mutation are modified by other factors such as genetic variants of iron regulatory genes that are also influential on iron homeostasis, behavioral factors like alcohol consumption, or regular blood loss via menses or blood donation. Age is also a modifier since the iron accumulation gets stronger with age. The subset of breast cancer caused by inherited *BRCA1* or *BRCA2* mutations is another example of reduced penetrance. These mutations increase the lifetime risk for breast cancer considerably, up to eight times, but not to 100%. Genetic modifiers of these mutations are being identified.

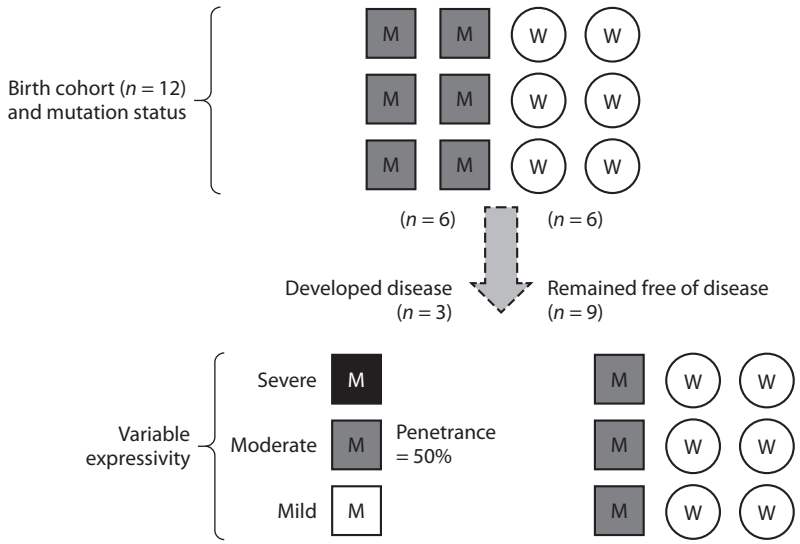

Figure 2.7 Schematic explanation of penetrance and expressivity. M denotes individuals positive for the mutation and W denotes those without the mutation (wild-type genotype). Half of the mutation-positive individuals (3 of 6) developed the disease (thus, penetrance is 50%) with variable severity (expressivity), and all those with the wild-type genotype remained free of disease.

Both penetrance and expressivity (see Figure 2.7) have implications for the design and interpretation of genetic association studies. Since even single-gene diseases can be modified by factors other than the disease mutation, it should be expected that a similar situation exists for genetic variants that modify disease predisposition. It is therefore crucial to collect as much data as possible for genetic and environmental variables in order to have reliable information to help with data analysis and interpretation of results. Such data will be very useful at the analysis stage to identify confounding factors as well as statistical interactions that identify effect modifiers. The current common practice of analyzing SNP data as an individual data point is likely to miss associations that would have been unraveled if the SNP data were analyzed in conjunction with other variables.

Another phenomenon well known to medical geneticists is the occurrence of a disease at younger ages in successive generations. Trinucleotide (triplet) repeat disorders such as Huntington's disease, spinocerebellar ataxia, myotonic dystrophy, and fragile X syndrome are characterized by **genetic anticipation**. With an increasing number of repeats at each generation, called expansion, the age at disease onset gets younger. For some diseases, larger expansions in the trinucleotide repeats are caused by paternal transmission, but for other diseases it is maternal transmission. Such a parental effect is not unique to this group of disorders and is known to exist in other inherited and complex disorders. In diseases either caused or modified by DNA sequence variants, where heterozygosity is sufficient to increase the disease risk, the parental origin of the variant may have to be taken into account to unravel an association. Another situation where a parental effect is observed is in a multifactorial disease such as congenital pyloric stenosis. While the disease is more common in male newborns, children of affected females (the less frequently affected sex) are more likely to be born with the pyloric stenosis than the children of affected males. This is because females need a higher genetic load to express the disease themselves and transmit more of that load to the offspring. Among the offspring of an

affected female, males are more likely to express the disease because of the requirement for a lower genetic load. The end result is that the less frequently affected sex transmits the disease more easily to the more frequently affected sex. While family-based association study designs may be able to unravel parental effects, in a **population-based association study** such effects cannot be examined, resulting in a lower chance of identifying an association.

Key Points

- Even in single-gene disorders there are strong modifiers that can be genetic or environmental. This is why, in a genetic association study of a complex multifactorial disease, a single variant will yield a very small effect size.
- A disease-causing mutation is rare and highly deleterious for gene function. A disease-associated polymorphism (like a SNP) is usually common and alters gene function but is not very deleterious.
- Disease-causing mutations can be on autosomal or sex chromosomes, or in the mitochondrial genome.
- Mutations for diseases that show higher frequency in one sex do not have to be in sex chromosomes.
- A Mendelian pattern of inheritance is a feature of very rare diseases; the lack of such a pattern does not rule out a genetic basis for disease.
- Common human diseases are multifactorial, result from complex interactions between genetic and environmental factors, and do not show a straightforward inheritance pattern.
- Heterozygosity for a mutation causing a recessive disease is generally tolerable, and may confer some advantage to its carriers. This is the reason for the maintenance of some disease-causing mutations in the population.
- Even if a variant is strongly associated with disease susceptibility, the mere presence of it is not generally sufficient to determine with certainty that the associated disease will develop.
- Diseases with a strong genetic basis are still modified heavily by other factors, and genetic association studies should therefore be as comprehensive as possible to unravel the joint effects of multiple factors.

URL List

Genetics for Epidemiologists: Application of Human Genomics to Population Sciences. National Human Genome Research Institute. http://www.genome.gov/27026645

Strachan T & Read AP (1991) Human Molecular Genetics, 2nd ed. Bookshelf. National Center for Biotechnology Information. http://www.ncbi.nlm.nih.gov/books/NBK7580

Further Reading

Disease genes (autosomal and sex chromosomes)

Bodmer W & Bonilla C (2008) Common and rare variants in multifactorial susceptibility to common diseases. *Nat Genet* 40, 695–701 (doi: 10.1038/ng.f.136).

Charchar FJ, Bloomer LD, Barnes TA et al. (2012) Inheritance of coronary artery disease in men: an analysis of the role of the Y chromosome. *Lancet* 379, 915–922 (doi: 10.1016/S0140-6736(11)61453-0).

Clayton DG (2009) Sex chromosomes and genetic association studies. *Genome Med* 1, 110 (doi: 10.1186/gm110).

Jimenez-Sanchez G, Childs B & Valle D (2001) Human disease genes. *Nature* 409, 853–855 (doi: 10.1038/35057050).

Inheritance patterns (monogenic versus polygenic)

Badano JL & Katsanis N (2002) Beyond Mendel: an evolving view of human genetic disease transmission. *Nat Rev Genet* 3, 779–789 (doi: 10.1038/nrg910).

Dipple KM & McCabe ER (2000) Modifier genes convert "simple" Mendelian disorders to complex traits. *Mol Genet Metab* 71, 43–50 (doi: 10.1006/mgme.2000.3052).

Wang DY, Chan WM, Tam PO et al. (2005) Gene mutations in retinitis pigmentosa and their clinical implications. *Clin Chim Acta* 351, 5–16 (doi: 10.1016/j.cccn.2004.08.004).

Homozygosity versus heterozygosity

Rotter JI & Diamond JM (1987) What maintains the frequencies of human genetic diseases? *Nature* 329, 289–290 (doi: 10.1038/329289a0).

Sidransky E (2006) Heterozygosity for a Mendelian disorder as a risk factor for complex disease. *Clin Genet* 70, 275–282 (doi: 10.1111/j.1399-0004.2006.00688.x).

Germ-line mutations

Rochette J, Le Gac G, Lassoued K et al. (2010) Factors influencing disease phenotype and penetrance in *HFE* haemochromatosis. *Hum Genet* 128, 233–248 (doi: 10.1007/s00439-010-0852-1).

Stein CM & Elston RC (2009) Finding genes underlying human disease. *Clin Genet* 75, 101–106 (doi: 10.1111/j.1399-0004.2008.01083.x).