Introduction to Molecular Genetics

1

M.M.A.M. Mannens and H.J.M. Smeets

1.1 Introduction

Genetic knowledge is being rapidly introduced into clinical medicine. Knowledge on genes and gene defects, gene expression, and gene products have been gathered as a result of the human genome project and large-scale resequencing projects at a rapid pace.[1,2] The genetic cause of the vast majority of important monogenic disorders is nowadays known, and more rare disorders are being unraveled quickly. Developments in genomics and sequencing technologies enable molecular geneticists with an accelerated and detailed characterization of genetic defects, genetic predisposition, and/or genetic background of individual patients. This also provides extensive knowledge on the molecular pathophysiology.

The introduction of genetic tests for heritable cardiac abnormalities is of a fairly recent nature. Disorders, like hereditary cardiac arrhythmia syndromes or inherited cardiomyopathies, have been genetically unraveled in the last 2 decades, and research is ongoing to improve DNA-diagnostics.[3,4] Genetic testing offers many opportunities, but also a considerable number of risks and uncertainties. Therefore, introduction in the clinic has to be performed with great care. Not every test that can be done should be done. It is evident that genetic testing must be beneficial for the patient and his or her family. If he or she is affected, then the test can be performed to confirm the diagnosis and in some instances to predict prognosis and adjust treatment. It is clear that such genetic tests affect not only the patient involved, but also will be of concern to relatives and future offspring. In asymptomatic relatives of patients, it may be possible to determine their genetic status and predict what the chances will be of developing symptoms in the years to follow. Especially this kind of predictive genetic testing is becoming more and more important in the field of genetic cardiovascular diseases, enabling the discrimination between carriers and noncarriers of a specific genetic risk. It is obvious that in this area of predictive medicine, ethical and social (health insurance) aspects play important roles as well. Therefore, these investigations should be embedded in a multidisciplinary approach of cardiologists, clinical geneticists, laboratory specialists and social workers, psychologists, and ethicists. Society should define the outlines, within which genetic testing should be performed, preventing the social and economical discrimination of individuals based on their genetic burden.

1.1.1 DNA, What Is It, Where Is It, and How Are Proteins Made?

DNA (deoxyribonucleic acid) is an antiparallel dimer of nucleic acid strands (Fig. 1.1a–c). It is composed of nucleotides (base + deoxyribose + phosphate group). These nucleotides are polymerized through a phophodiester linkage. Base pairing of nucleotides are possible between adenine (A) and thymine (T) or cytosine (C) and guanine (G). CG bonds are stronger than AT bonds because the CG base pairs form three hydrogen bonds and the AT base pairs only two. The DNA *double helix* structure was originally published by Watson and

M.M.A.M. Mannens (✉)
Department of Clinical Genetics, Academic Medical Center, P.O. Box 22700 1100 DE, Amsterdam, The Netherlands
e-mail: m.a.mannens@amc.uva.nl

H.F. Baars et al. (eds.), *Clinical Cardiogenetics*,
DOI: 10.1007/978-1-84996-471-5_1,

Crick and is composed of two strands wound around each other in a helical structure and antiparallel, that is, the strands run in opposite directions (5′–3′ versus 3′–5′). DNA does not form a perfect helix since the sugar phosphate backbones are slightly offset from the center of the helix. This creates major (relative open) grooves and minor (relatively closed) grooves (Fig. 1.1c). DNA in eukaryotic cells is located in the nucleus. The mitochondria (mainly involved in the energy supply for the cell) have multiple copies of their own small circular DNA molecules. Since the total DNA content in the human cell is about 2 m long and the nucleus is about 6×6^{-10} m, it is obvious that DNA has to be packed very condensed. DNA is packed in chromosomes, a complex of DNA and histone proteins. The human cell has a diploid genome consisting of 46 *chromosomes*, which means that all chromosomes are present in two copies (homologous pairs) with the exception of the sex chromosomes, which can be present as XX (female) or XY (male). During gametogenesis (forming of oocyte or sperm cell), the chromosomes become haploid (only one copy per cell) and genetic material between both homologous chromosomes is exchanged (*recombination* Fig. 1.2). This means that genetic information (such as disease causing mutations) can switch from one chromosome to the other homologue. This may sometimes be a problem in indirect DNA-diagnostics using markers in the vicinity of the gene defect. Of course *replication* of DNA is needed for cell division. A complex of enzymes unwinds the DNA at many positions and replicates the two strands of the DNA helix resulting in two copies of the original DNA molecule. Since DNA molecules are between 48,000,000 and 240,000,000 nucleotides long, mistakes in replication are likely, leading to potential mutations. The cell, however, has a number of mechanisms (e.g., proofreading activity of enzymes, mismatch repair, or excision repair) that correct such replication errors. However, still one mismatch out of 10^9 nucleotides remains.

The genetic code lies in the genes in the DNA molecules. These genes code for the specific amino acid sequence of proteins. *Messenger RNA* (mRNA) is needed as an intermediate between DNA and the protein since DNA is localized in the nucleus and proteins are synthesized on *ribosomes* outside the nucleus. Eukaryotic genes consist of coding sequences (exons), noncoding sequences (*introns*), and *regulatory sequences*

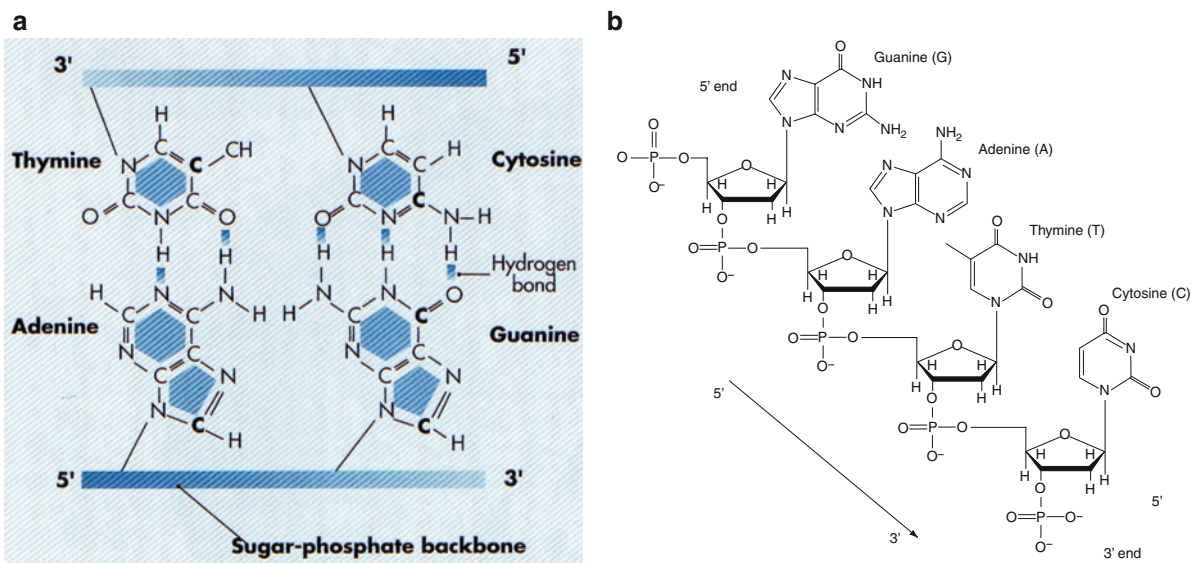

Fig. 1.1 Structure of DNA.a base (C, T, A, or G) combined with a deoxyribose and a phosphate group is called a nucleotide. These nucleotides are polymerized through phophodiester linkage. DNA is read from the 5′ to the 3′ end. (**a**) The four bases that make-up the actual DNA code, Adenine always pairs with Thymine with two hydrogen bonds and Cytosine always pairs with Guanine using three hydrogen bonds()s. (**b**) Chemical structure of DNA, showing the sugar backbone and the polymerization through phosphodiester linkage, as DNA is read in 5′ to 3′ direction, the code of this stretch of DNA would read:GATC. (**c**) DNA double helix istructure, the base pairs in the middle are aligned around the helical axis. The major and minor grooves are indicated. Adapted with permission from Jorde, Carey, Bamshad, White, Medical Genetics third edition, Mosby Elsevier 2006

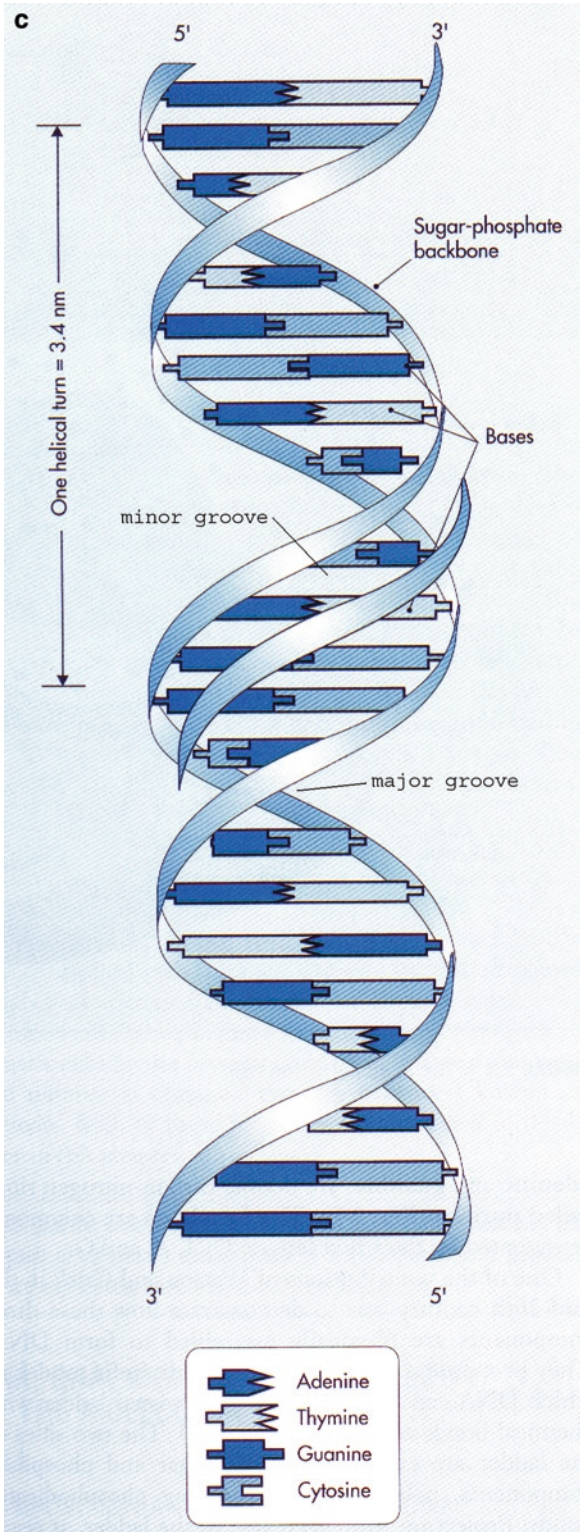

Fig. 1.1 (continued)

(promoter sequence, enhancer, and repressor) that start, enhance, or repress gene expression. Only 2% of the DNA actually codes for the exons in approximately 20,000–25,000 human genes, 25% of the DNA codes for introns. The remainder of the DNA is, for instance, needed for regulation of gene expression, maintenance of chromosomes, syntheses of rRNAs (ribosomal RNAs) or tRNAs (transfer RNAs), or segregation of chromosomes during cell division. Part of the genome might be a coincidental result of evolution and might not have any function at all, but new functions of DNA sequences are still revealed. Recently, microRNAs have been discovered that do not code for a protein but have proven to be regulators of gene expression. Between humans, the genetic variation is less than 0.1%.

The enzyme *RNA polymerase* can use a single-stranded DNA molecule as a template and it copies the genetic code in the genes in a single-stranded mRNA molecule. The chemical structure of mRNA is similar to DNA with the exception that ribose is used instead of deoxyribose for the sugar backbone and that the base *uracil* is used instead of thymine. During a process called *transcription*, a pre-mRNA is (Fig. 1.3) synthesized according to the DNA code. This pre-mRNA consists of introns and exons. The introns are removed by a process called splicing leading to a smaller mRNA that contains only the genetic code for a protein. Transcription is very efficient since sometimes many different mRNA's encoding different proteins can be synthesized from a single gene.

These mRNAs are transported from the nucleus to the *ribosomes* in the cytosol. Ribosomes are cellular complexes consisting of proteins and ribosomal RNAs (rRNAs). In the ribosomes, the mRNA code is used as a template to synthesize proteins (a process called *translation*) (Fig. 1.4). mRNA, ribosomes, transfer RNAs (tRNA), and amino acids are the key components in this process. tRNAs function as amino acid carriers and as recognition molecules that identify the mRNA nucleotide sequence and translate that sequence into the amino acid sequence of proteins. Each tRNA recognizes one of the 20 specific amino acids (Table 1.1). A second tRNA recognition site binds to a specific nucleotide triplet (a combination of three sequential nucleotides in the mRNA [called a codon] that encodes for a specific amino acid, see Fig. 1.4). By repeating this process for all codons and linking subsequent amino acids to the previous one, the sequence of the genetic code is

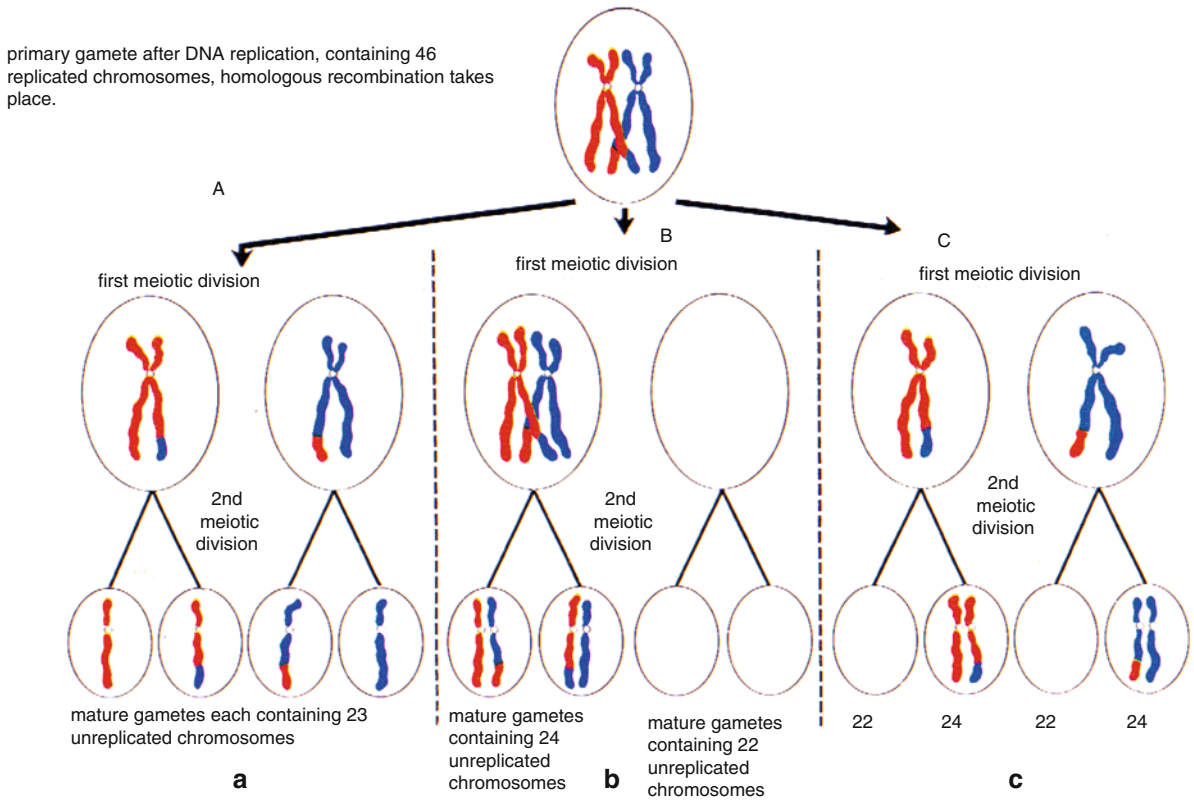

Fig. 1.2 Meiosis: (**a**) demonstrates the normal stages of meiosis(after division one the cell contains 23 replicated chromosomes-22 autosomes and 1 sex-chromosome). (**b**) Demonstrates nondisjunction in meiosis 1(the most frequent cause of for instance Down syndrome). (**c**) Demonstrates nondisjunction in meiosis2. Appreciate the effect of recombination in the mature gamete, in this way each grandparent contributes to both copies of each autosome of his/her grandchild. Adapted from Langman Inleiding tot de embryologie Bohn Scheltema & Holkema 9e revised edition druk 1982

translated into protein. Translation is also a very efficient process since many ribosomes (polysome) translate a single mRNA. Both transcription and translation are initiated and terminated by specific sequences in the genetic code.

Mutations are changes in the DNA that can lead to disease. Mutations can occur due to chemical modifications such as smoking, UV light (sun), radioactivity, and chemical instability of the DNA. Mutations can also be the result of replication errors. The effect can be a somatic nonhereditary disease (e.g., cancer) or a hereditary disease (e.g., hereditary cancer or for instance a hereditary cardiac condition), provided the mutation is transmitted to the next generation because the mutation is present in or has newly occurred in the gametes. When a single nucleotide is changed (called a *point mutation*) the effect may be alteration of the amino acid code. Besides, such a mutation can affect a regulatory sequence or lead to a premature stop codon. Also one or more nucleotides can be deleted (called a *deletion*) or added (*insertion*). Deletions and insertions of triplets of nucleotides within the reading frame will delete or add the associated amino acids. As nucleotides are read in triplets (called a codon, each codon coding for a specific amino acid) any deletion or insertion that cannot be divided by three will lead to a disturbance of the reading frame of the genetic code (and will therefore cause a frameshift mutation). This leads to a completely different protein (or more often to loss of the protein), as sooner or later a premature stop codon will be introduced. Mutations can cause a loss of function or gain of (an abnormal) function of a protein. A 50% reduced dosage of a protein may not be enough for normal function (this is called *haploinsufficienty*) and may thus cause disease. Some mutations can have a *dominant negative* effect on protein function which means that the mutation leads to an altered protein that interferes with the function of the wild type (=normal) protein, which is produced from the unmutated copy of the gene. Finally, some mutations

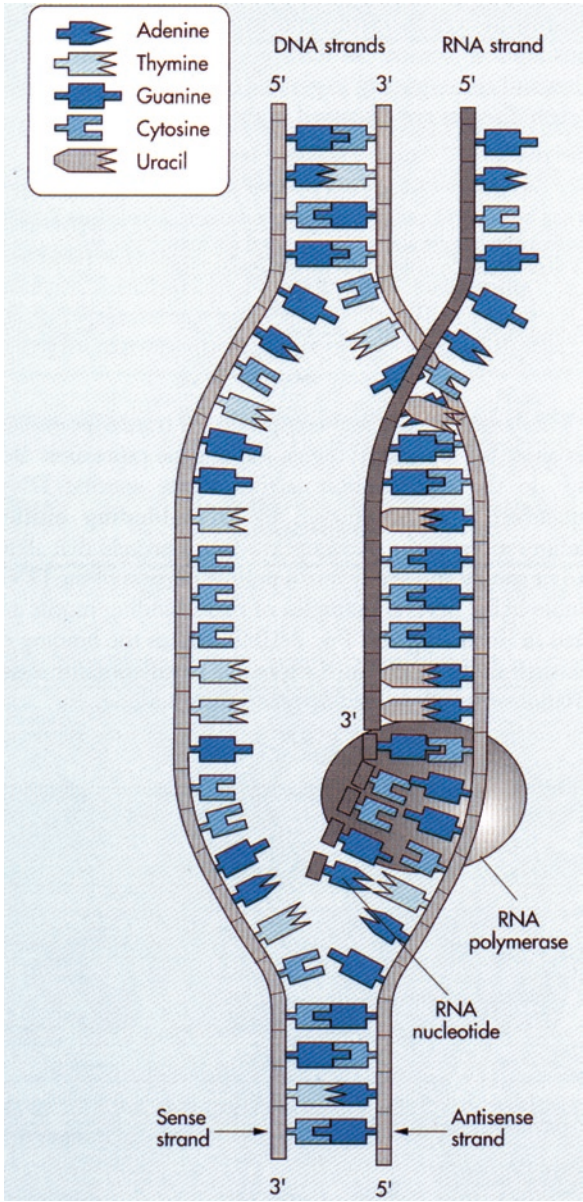

Fig. 1.3 Visualizes the transcription of the DNA code into messenger RNA. This messenger RNA contains both exons and introns. It has to be processed, and the introns (nonprotein coding parts) of the gene have to be removed (a process called splicing), before a mature messenger RNA leaves the nucleus to the cytosol, where it will be translated into protein. Reprinted with permission Jorde, Carey, Bamshad, White, Medical Genetics third edition, Mosby Elsevier 2006

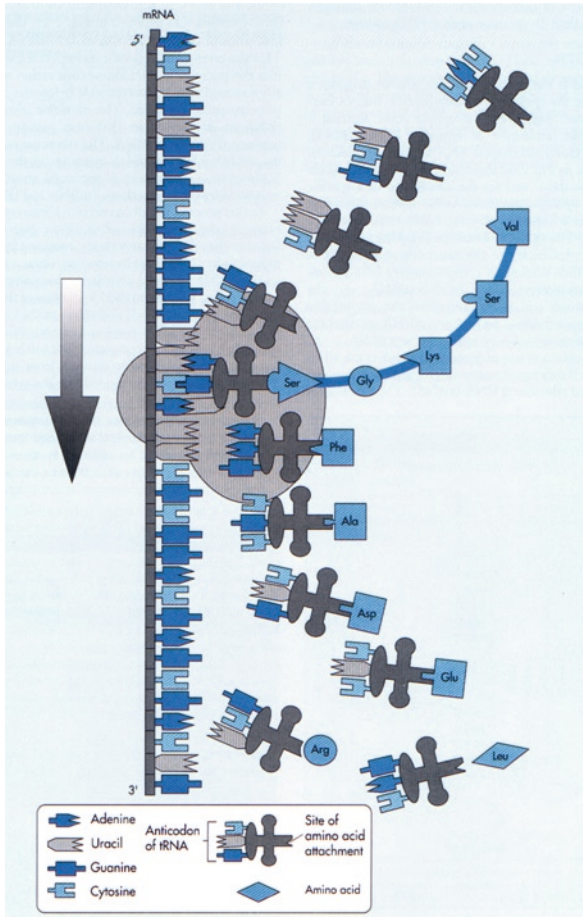

Fig. 1.4 Translation. The genetic code of the mRNA determines the order in which tRNAs are recognized and thus the order of the amino acids in the protein. The light gray oval structure depicts the ribosome, where translation takes place and on which the proteins are formed. Reprinted with permission Jorde, Carey, Bamshad, White, Medical Genetics third edition, Mosby Elsevier 2006

can be lethal leading to a nonviable embryo; such mutations are always de novo, or the transmitting parent has been rescued by the fact that the mutation is not present in all of his or her cells (a situation that is called *mosaicism*). Mutations can vary in size from one nucleotide up to complete extra chromosomes (as for instance in the case of Down's syndrome). Especially for small missense mutations only changing a single amino acid in the protein, the effect of the mutation may be difficult to predict. The location of such a mutation in the protein will to a large extent determine the effect of the mutation on protein function (e.g., is it within an important domain of the protein or is the specific amino acid residue very conserved during evolution, an indication that it may be important for normal function and that changing this residue may be not tolerated). Frameshift mutations at the extreme end of a gene often are less devastating for instance because only a small part of the protein is altered. In addition, a

Table 1.1 key to the genetic code

	T		C		A		G	
T	TTT	F	CTT	L	ATT	I	GTT	V
	TTC	L	CTC	L	ATC	I	GTC	V
	TTA	S	CTA	L	ATA	I	GTA	V
	TTG	L	CTG	L	ATG	M	GTG	V
C	TCT	S	CCT	P	ACT	T	GCT	A
	TCC	S	CCC	P	ATT	T	GCC	A
	TCA	S	CCA	P	ATA	T	GCA	A
	TCG	S	CCG	P	ACG	T	GCG	A
A	TAT	Y	CAT	H	AAT	N	GAT	D
	TAC	Y	CAC	H	AAC	N	GAC	D
	TAA	X	CAA	Q	AAA	K	GAA	E
	TAG	X	CAG	Q	AAG	K	GAG	E
G	TGT	C	CGT	R	AGT	S	GGT	G
	TGC	C	CGC	R	AGC	S	GGC	G
	TGA	X	CGA	R	AGA	R	GGA	G
	TGG	W	CGG	R	AGG	R	GGG	G

Top row depicts the first base of each codon at the DNA level (at the RNA level thymine T is replaced by uracil U
The most left column shows the second base of each codon. Amino acids are indicated by their single letter codes(three-letter codes are also often used): *A* alanine(ala), *C* cysteine(cys), *D* aspartic acid (asp), *E* glutamic acid (glu), *F* phenylalanine(phe), *G* glycine (gly), *H* histidine (his), *I* isoleucine (ile), *K* Lysine (lys), *L* Leucine(leu), *M* Methionine (met), *N* asparagine (asn), *P* proline (pro),*Q* glutamine (gln), *R* arginine (arg), *S* serine (ser), *T* threonine (thr), *V* valine (val), *W* tryptophan (try), *Y* tyrosine (tyr), *X* stop codon
As you can see in the table each amino acid can be coded by different codons

kind of quality control mechanism called *nonsense mediated messenger –RNA decay* largely prevents the expression of truncated erroneous proteins.

1.1.1.1 Genotyping in Mendelian and Non-Mendelian Disorders

Genetic diseases can be divided in *Mendelian* and *non-Mendelian* diseases.[5] The first group is caused by defects in *autosomal* or *X-chromosomal* genes that have a dominant or recessive manifestation. In case of a *dominant* disorder (Fig. 1.5), the disease will already become manifest if only one *allele* (one of both copies of a gene) is affected. In case of *recessive* segregation, both homologous alleles must be affected to develop the disease. Dominant and recessive are no absolute terms. Some carriers of dominant mutations remain healthy, which is called *non-penetrance.*[6] This can be disease, age- and sex-dependent. A *penetrance* of 90% means that 90% of the mutation carriers of a specific gene defect will be affected (at a defined age). Moreover, manifestations of genetic disease often vary among mutation carriers even within a single family. This is called *variable expression*, probably caused by other genetic and environmental factors. One has to realize that not every disease case within a single family has to have a genetic cause. Nongenetic cases mimicking genetic disease are called *phenocopies*, and this can be a problem especially in disorders with a commonly occurring environmentally induced counterpart, like for instance left ventricular hypertrophy due to untreated hypertension in a family with hypertrophic cardiomyopathy caused by a sarcomere gene mutation.

Sometimes mutations in the same gene can lead to different disorders, as do for instance mutations in the SCN5A gene involved in Long QT syndrome type 3, Brugada syndrome, or isolated conduction defects.[7–9] Some specific neuromuscular disorders, like myotonic dystrophy (in which about 30% of patients die from cardiac arrythmias), become more severe in subsequent generations. This phenomenon is called *anticipation* and the molecular basis is an unstable stretch of small DNA repeats, which in most cases increase in size when transmitted to the next generation.[5] The size of the unstable DNA fragment is in general related to disease severity and age of onset.

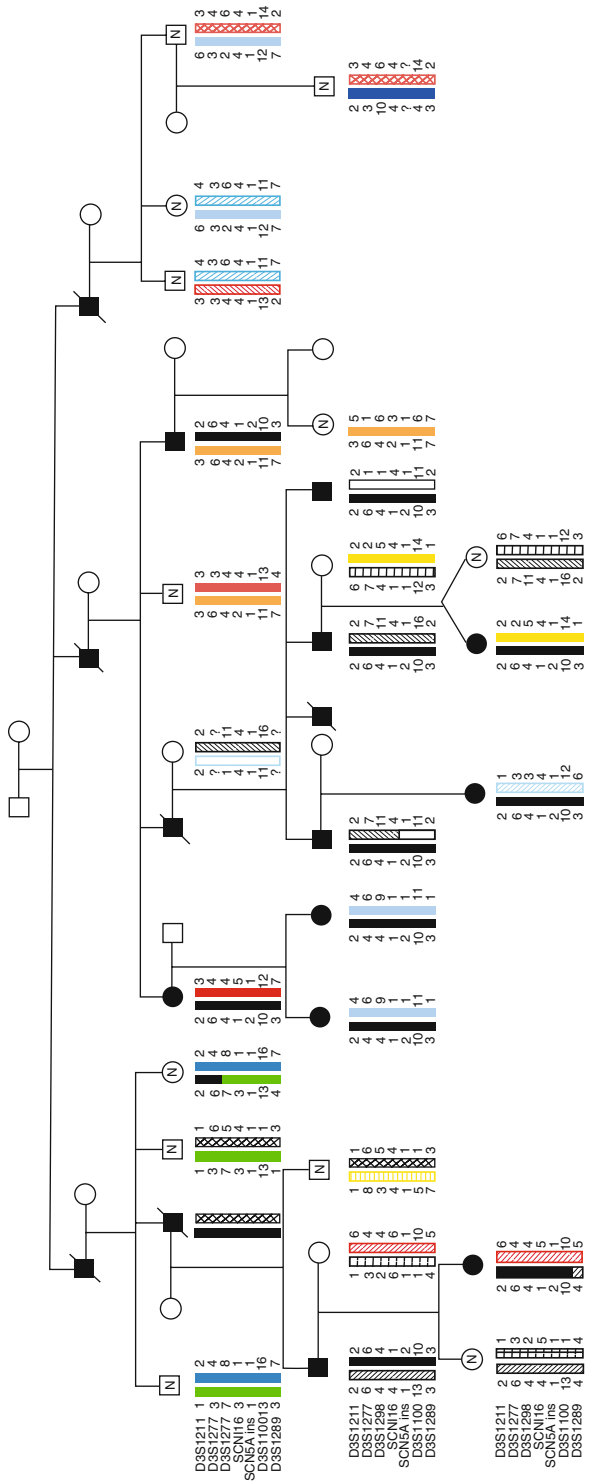

Fig. 1.5 Linkage analyses in a family presenting with "Brugada" syndrome and LQT3. Affected persons are indicated with filled symbols. *N* not affected. Deceased persons are indicated with a crossed line. The disease is showing evident autosomal dominant transmission with an average half of the offspring of affected individuals inheriting the disease. There is male to male transmission, thus excluding X-linked dominant disease. The haplotype (haplotype = a stretch of DNA on a single chromosome as defined by the alleles of a set of linked genetic markers that are usually inherited together, thus without recombination) indicated with a black bar segregates with the disease. Marker alleles are numbered for each tested marker on chromosome 3

Carriers of recessive disorders are usually without symptoms, thus including most females in X-linked recessive disorders. However, in some disorders it is possible to identify them by biochemical testing, reduced enzymatic activity, or by expression of minimal signs (for instance as a marginally increased QT-interval in carriers of the recessive Jervell-Lange-Nielsen syndrome mutations [Jervell-Lange-Nielsen syndrome is a combination of long QT syndromes and deafness]). Finally, some disorders, like inherited cancers can be dominant at the family level and recessive at the gene level. In that case a mutation in a specific gene segregates in the family but a *second hit*, destroying the other wild-type (=normal) allele of the gene, has to occur in order for disease to arise. If the chance of this second hit occurring is very high, than a dominant segregation pattern will be observed. For these reasons terms like dominant and recessive segregation should be carefully defined. This also applies to X-chromosomal diseases, which are predominantly expressed in males, because males only have one X-chromosome, whereas females have two. Dominant and recessive are terms describing genetic diseases and not genes or mutations themselves. For mutations in a specific gene, both dominant and recessive disease can occur, either leading to the same or a different phenotype. Single gene mutations in the KCNQ1 gene lead to dominant Romano-Ward syndrome, whereas mutations affecting both copies of the same gene cause the recessive Jervell-Lange-Nielsen syndrome.[10] The latter may also result from mutations in the KCNE1 gene, demonstrating the *genetic heterogeneity* of this disorder. Often the nomenclature and classification of genetic disorders have to be adjusted at the moment the genes causing the disorders have been identified.

In addition to Mendelian inheritance, other segregation patterns exist for monogenic disorders. Some diseases show a *maternal segregation pattern* and are caused by defects in the *mitochondrial DNA*, which is, for example, the case for a number of isolated or rare syndromic dilated or hypertrophic cardiomyopathies.[11] On rare occasions gene defects only become manifest when transmitted by either the mother or the father. This phenomenon is called *genomic imprinting*[12] and results from the fact that for a limited number of genes only the paternal or maternal copy is expressed. A defect in such an imprinted gene will only be detected, if the mutation is present in the active (expressed) allele of the gene. When looking at phenomena like reduced penetrance and variable expression, it is immediately clear that clinical manifestation of monogenic disorders is not constant within families. No clear discrimination between *monogenic disorders* with one major gene defect and modifying factors, and complex genetic disorders with more than one minor gene defects and contributing environmental factors, exists. *Polygenic diseases* are caused by more than one gene, which often contribute in a quantitative manner to the clinical manifestations and are considered risk factors, as the manifestations may vary, depending on the genetic background. Complex genetic disorders are polygenic disorders, which can often be influenced by the environment to a relatively large extent. These disorders are much more frequent than monogenic disorders and the role of the environment is often stronger, which means that a disease-state does not always have a major genetic cause.[13] The underlying genetic risk factors can be identified in large-scale population studies in which common genetic variants (frequency > 1%) are being tested across the entire genome (*whole genome association (WGA) studies*) for being more or less frequent in the patient population compared to a matched control population. These variants can be the functional causative factor or they can be merely markers that such a contributing factor is located in their immediate vicinity. The success of these studies largely depends on the size and composition of the population and the number of genetic factors involved. If the size of the population is too small, the number of genetic factors too large, and the effect of the genetic factor too small, then it is unlikely that any meaningful association can be found. However, many loci and factors are currently being identified, which usually have only a small to moderate effect on the manifestation of the complex genetic disorder. This implies that these factors may allow for the identification of the underlying pathophysiological processes, but by themselves only have a limited value in predicting the risk of developing the disorder under investigation in a specific individual. Apart from these common variants, it is anticipated that many different, rare variants in candidate genes can also explain part of the genetic risk of complex diseases. These cannot be assessed by WGA, but possibly by sequencing all or large numbers of candidate genes or complete exomes (all exons in the genome), using massive parallel or next-generation sequencing approaches. These rare variants contributing to complex genetic diseases will be found in the near future.

1.1.1.2 Genetic Heterogeneity of Monogenic Disorders

Diseases like familial hypertrophic cardiomyopathy or congenital Long QT syndrome are genetically heterogeneous monogenic disorders. The diseases are caused by single gene defects in patients, but defects in a large number of different genes can lead to the same or very similar disease and the severity of clinical manifestation may depend on additional mutations in other genes and/or environmental factors. Familial hypertrophic cardiomyopathy is mainly a disease of the cardiac sarcomere and many mutations in many sarcomeric genes have been detected in patients.[4,14] At this point, a complete screening of all sarcomere genes would be technically possible, but very difficult to implement in a routine diagnostic setting. It is also not necessary for those families in which hypertrophy can be clearly demonstrated and progression of disease is relatively benign. However, testing can be essential in those families in which sudden cardiac death (SCD) occurs, especially in those cases where hypertrophy is mild, and individuals at risk for dangerous arrhythmias may not be so easy to detect by nongenetic means. At this moment candidate genes are routinely analyzed sequentially, which is laborious and time-consuming, but the pace of technical developments in this field will allow for more optimal diagnostic strategies in the near future, making analysis of a large number of involved genes in parallel feasible. Such protocols will become available to diagnostic laboratories in the years to come, and can be considered a first step toward the personal genome (see also section alternative technologies).

When no good phenotypic selection criteria for which genes to test first exist, the frequency with which causative mutations in the different genes are found can be used to establish *a mutation screening strategy*, starting with the most commonly mutated gene. For familial hypertrophic cardiomyopathy, one generally starts with the beta-MHC gene (MYH7) (±35% of the mutations), followed by the myosin binding protein-gene (MYBPC3) (±15%) and the Troponin T gene (TNNT2) (±3–5%). A bias exists in these percentages as some genes have been more frequently screened in patients than others. Moreover, these percentages can be different in different populations. For instance in the Netherlands a single mutation in the myosin binding protein gene accounts for almost 25% of hypertrophic cardiomyopathy families. It should be stressed that the chances of rapidly identifying a mutation in a patient increases if many affected family members are available for testing, and genetic markers can be used to identify the locus involved (many candidate loci can then be easily excluded using polymorphic genetic markers). In some families it is also possible to perform DNA-diagnostics, if the underlying genetic defect is not located in any of the known genes. A prerequisite is that the size of the family is such that segregation of the disorder with a specific locus can be proven with high likelihood (as a general rule of thumb one would need >10 patients with unambiguous genetic status consenting to participate in case of an autosomal dominant disorder). It should be noted that careful clinical examination is a prerequisite as any false diagnosis will lead to a wrong conclusion with respect to the gene or genetic locus involved. Also carriers of the mutation that are not yet expressing the disease as a result of their young age may confuse such a genetic undertaking. The combination of clinical and genetic data can not only be of use in care for individual patients and their families, but is also important for extending the knowledge on the disease (frequency of non-penetrant carriers, number of de novo *(=new or spontaneous)* mutations), and for instance for the development of clinical selection criteria for the gene involved, like in the congenital Long QT syndrome, and for establishing a more exact genotype-based prognosis. *Genotype–phenotype correlations* are important for counseling but as long as not all contributing factors to the phenotype are known such knowledge should be used with great care in individual cases. Congenital Long QT syndrome is most often caused by mutations in cardiac potassium or sodium channel genes.[3,15] Based on genotype–phenotype correlations, it is possible to select the most likely gene involved using clinical information such as the trigger of cardiac events (physical stress and diving as a trigger for arrythmia point to the KCNQ1 gene and acoustic stimuli to KCNH2 gene). The characteristic repolarization pattern on the ECG also gives information on which gene is most likely to be involved.[16] The identification of the causative mutation has to confirm this. This is specifically important, because for these disorders gene-specific prophylactic therapies exists and can save lives, if carriers are identified early. It can be expected that gene-specific therapy may evolve to mutation specific therapy in the future. Genetic testing can provide for these disorders a clear contribution to patient management, either by

directing therapy/follow-up or to give advice concerning lifestyle. This may be very important for asymptomatic carriers, who can only be identified by screening family members at risk for the gene defect (either by cardiac screening or genetic screening, once the family specific mutation has been established).

The discovery of genes and (possible) pathogenic mutations in a genetic disorder does not solve all diagnostic or prognostic problems. Mutation carriers and segregation patterns can be determined with great certainty, but in case of unique mutations, or DNA variants with an unclear pathogenic significance (not every unknown DNA change is a disease-causing mutation. It may also be a rare polymorphism with no influence on protein function whatsoever). Especially, when such an unknown variant is detected in a small family, where the segregation of the variant cannot be compared to segregation of the disease, it can be impossible to give an accurate risk estimate for the development of clinical features for a person that carries the *uncertain genetic variant.* It should be stressed that it is quite damaging to over-interpret such an unknown variant and mistake it for a real mutation. On the one hand, carriers of the variant may be unnecessarily frightened and surveyed, while on the other hand the real causative mutation may remain undetected. Relatives not carrying the variant may be released from surveillance, based on inaccurate genetic findings. It should also be noted that in clinical practice the finding, of such an *unclassified variant* (UV) or even a genetic risk factor of small effect size, may trigger a direct request from clinicians or family members to perform genetic testing in relatives, at a moment in time that the clinical implications of the genetic findings still have to be determined, which of course is an undesirable situation. Therefore, the importance of close cooperation between organ specialists (cardiologists) and genetic specialists cannot easily be overrated. In all predictive genetic testing the patient should be informed and guided through the process by a multidisciplinary team of experts in the field of cardiogenetics to determine what is in his or her best interest. Although DNA-diagnostics are becoming available for an increasing number of cardiovascular disorders, this does not mean that in all cases a molecular diagnosis can be achieved. To improve the success rate, a good collaboration between the laboratory and the referring cardiologist and clinical geneticist is necessary. Often, the clinical data such as the patients' history or an ECG can be extremely helpful to identify the most likely causative gene.

1.1.1.3 DNA-Diagnostics: Basic Strategies

DNA-tests are usually performed on DNA isolated from blood. As long as only DNA-analysis is performed specific tissues are not required (genes do not have to be expressed in order to be analyzed). An exception should be made for mitochondrial DNA mutations, which can vary in mutation load between tissues. DNA-isolation is not very critical and EDTA or Heparin blood samples can be kept at room temperature for several days without interfering with the analysis. If RNA-analysis is required or immortalized cell lines need to be prepared for gene expression or protein studies, then the blood needs to be delivered within a single day at the DNA-laboratory. Recently, specific tubes have been developed, which stabilize the RNA and enable more standard logistics.

DNA-diagnostics falls apart in strategies to identify already known family specific mutations and strategies to screen for unknown genes or mutations. The identification of an already identified mutation in a specific family is straightforward and can be completed within days, which is a big advantage for predictive testing. Dependent on the specific mutation, either p*olymerase chain reaction (PCR)* or Southern blot technology can be applied.[5] In most cases, the amplification by PCR of the DNA fragment to be investigated (genes, exons) is required. For PCR, two synthetic oligonucleotides (small single-stranded DNA fragments called primers), are added to single-stranded DNA. These primers flank the DNA fragment that has to be analyzed. Each primer hybridizes to one of the DNA strands and serves as a starting point for the synthesis of new DNA by the enzyme Taq DNA-polymerase (Taq stands for Thermus aquaticus, the bacteria from which it was isolated). After the first round of amplification, the DNA is denatured again and the next round of amplification follows (Fig. 1.6). After about 20–30 rounds of amplification, the DNA fragment to be analyzed has increased a million fold and can be visualized by gel electrophoresis and ethidium bromide staining. The PCR fragments can be analyzed for the specific mutations by a variety of techniques, such as Direct Sequencing, allele-specific oligonucleotides (ASO), allele-specific primer extension (APEX), oligonucleotide ligation assay (OLA), and mutation-specific restriction digestion.[17] For large rearrangements, deletions, or duplications in genes, Southern blot analysis is usually performed. DNA can be size separated directly on a gel after being

cut by restriction enzymes. Numerous enzymes are available that cut at specific places in the DNA, generating DNA fragments of various sizes. These size-separated DNA fragments can be transferred to a membrane to which the DNA is bound. The relevant DNA fragments on this membrane can be visualized by hybridizing the membrane to labeled (radioactive or fluorescent) DNA fragments of the genes/exons that are investigated (see Fig. 1.7b). An alternative can be long-range PCR, by which fragments of more than 10 kb can be achieved (Fig. 1.7a). Multiplex ligation-dependent probe amplification (MLPA),a quantitative PCR based technology, is another option.[18]

In case the genetic defect in an affected family is unknown, then basically two strategies could be applied in DNA-diagnostics to identify the causative genetic locus or defect.

1. When a limited number of candidate genes is associated with the phenotype, these candidates can be analyzed sequentially (prioritized based on clinical information or genetic epidemiology). Of course, it is important that an index patient that is most certainly affected is available for this analysis. If the number of candidates is still quite large and if multiple affected individuals are available within the family, it can be helpful to do a risk haplotype analysis (haplotype = a stretch of DNA on a single chromosome as defined by the alleles of a set of linked genetic markers that are usually inherited together, thus without recombination). Polymorphic markers associated with the various candidate loci are tested in all affected individuals, only a locus with shared polymorphic marker alleles in all affected individuals remains a plausible candidate locus for the phenotype in this family. Subsequently, the candidate gene at the shared locus is analyzed in more detail.
2. Whenever the gene and locus are entirely unknown (no plausible candidates available) and a large and very well-phenotyped family is available with

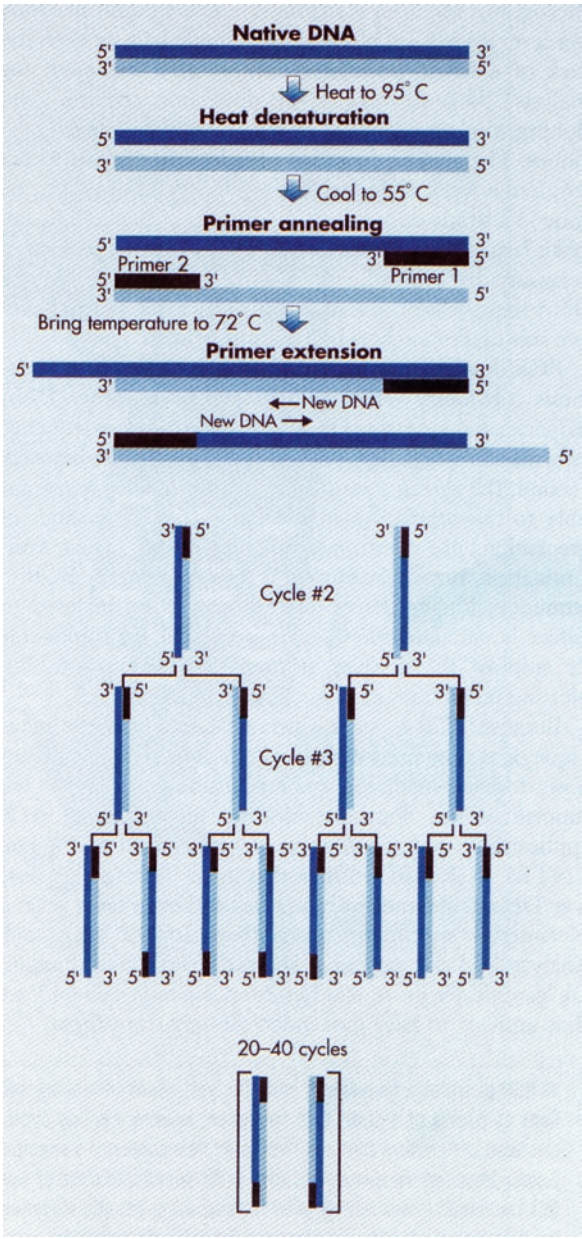

Fig. 1.6 Illustration of the polymerase chain reaction (PCR). This reaction is used in order to amplify a specific stretch of DNA (usually an exon or part of an exon of a disease gene) by several million folds. The reaction requires the presence of patient/subject DNA, two primers (called forward and reverse primer), a thermally stable form of the enzyme DNA polymerase, and a large quantity of free DNA nucleotides. These primers are oligonucleotides (15–20 bases in length) that are designed so that they are complementary to the sequences immediately flanking the region that has to be amplified. During the first step patient DNA is heated to denature the DNA (so that it becomes single stranded). Then the mixture is cooled down, allowing the primers to anneal to the appropriate sequences. In the third step the mixture is heated again to intermediate temperature, to allow for DNA replication in 5′>3′direction (as a result of the added DNA polymerase).This process is called primer extension. After 1 cycle the copy number of the region of interest has doubled. This happens during each cycle (i.e., 2, 4, 8, 16, 32 copies …), hence the name chain reaction. Typically, the cycles are repeated 20–30 times. Reprinted with permission Jorde, Carey, Bamshad, White, Medical Genetics third edition, Mosby Elsevier 2006

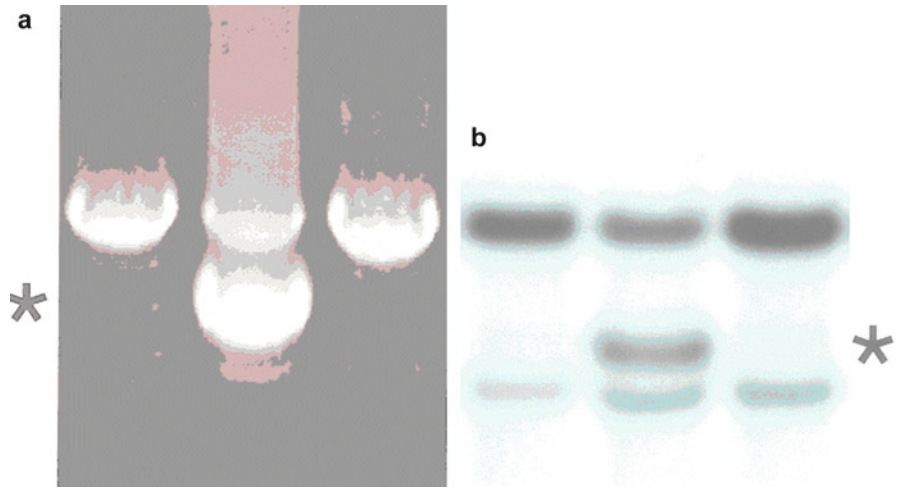

Fig. 1.7 (**a**) Gel electrophoresis of a long-segment PCR fragment containing exons 6–11 of the LDL gene. In lane 2, a 2.5 kb deletion containing exons 7 and 8 is visible (asterisk). (**b**) Autoradiograph of the same 2.5 kb deletion (lane 2, asterisk) as analyzed by Southern blotting and hybridization with a radioactive probe containing exons 2–11

multiple affected persons (≥10 individuals with unambiguous genetic status in case of autosomal dominant disease), a genome wide *linkage analysis* could be performed (actually this is more of a research project than routine DNA-diagnostics). The segregation of the disease in the family is compared to the segregation of a panel of polymorphic markers covering the whole genome and usually it will be possible to identify a set of closely linked marker alleles that are inherited by all affected individuals and not by the unaffected individuals (as in Fig. 1.5).It can then be assumed that the gene harboring the causative mutation is somewhere in the immediate vicinity of these linked markers. When the odds of the observed segregation occurring as a result of actual linkage as opposed to chance are 1,000:1, this is usually taken as evidence for linkage. At this point a locus has become known, that must contain the causative mutated gene. However, the identified region may still be quite large and may contain a large number of possible candidate genes, which would have to be analyzed one by one to finally find the culprit in this specific family. Instead, once the responsible locus has been identified without doubt, the risk allele can be followed with closely linked, preferably disease locus flanking, polymorphic DNA markers (microsatellites, Single Nucleotide Polymorphisms (SNPs)) that can identify the two homologous chromosome loci. Without knowing the actual defect (gene mutation), now a prediction can be made about the mutation carrier status of an individual based on the marker haplotype (established in the linkage analysis). In most cases, the accuracy will be >>95%.

If the family is too small, which is most often the case in clinical practice nowadays, linkage analysis is no option. In addition, meiotic crossover events (naturally occurring exchange of DNA between homologous chromosomes during gametogenesis) and lack of informativity of the polymorphic markers used can obstruct linkage analysis. *Risk haplotype* analysis for known disease loci as described above can usually be rapidly performed. In case of recessive diseases and (distant) consanguinity, homozygosity mapping can be applied. In such an analysis, the gene defect is supposed to be identical in both copies of the gene in a patient and located in a region in which all marker alleles are identical (homozygous). Although homozygosity mapping has been very successful in finding disease genes for autosomal recessive disorders, it is a tool which is restricted to only a limited part of clinical DNA diagnostics.

To avoid problems associated with linkage analysis, mutation screening of known disease genes or functional candidate genes is preferred in most cases. This can be costly and time-consuming, dependent on the size and number of potential, sequentially analyzed candidate genes. The coding and, wherever applicable, regulatory sequences are scanned in patient DNA and compared to the known, un-mutated (wild-type) sequence. The technology used is usually PCR-based and involves either sequencing or gene scanning (see below), followed by sequencing of a specific part of the gene in case of an aberration. Many methods, all with specific advantages and disadvantages to quickly scan through a gene exist. Apart from HPLC, mass spectrometry, and high-resolution melting curve analysis, most of this scanning technology (SSCP, DGGE, PCR or Southern blotting) is

based on *electrophoresis*. In electrophoresis DNA fragments are separated on a gel where the mutant fragment shows an aberrant mobility through the gel. To identify large gene rearrangements, duplications, or deletions, *Southern blot* analysis is applied. Such mutations will be missed if nonquantitative PCR-based strategies are used. MLPA is a quantitative PCR-based technology that is frequently used to identify large intragenic deletions, for instance deletion of an entire exon. Currently, high-density SNP-assays are filling the gap between large deletions ($>3*10^6$ basepairs), which can be visualized by cytogenetic techniques, and the small deletions in genes, which are detected by the quantitative PCR techniques or Southern blot assays as mentioned above. This implies that theoretically any kind of copy number variation (deletion or duplication) is now detectable using molecular technology.

Besides huge advantages, also pitfalls exist in DNA-diagnostics. Identification of a sequence variant does not always mean that this variant is causative for the disease. Apart from obvious mutations such as deletions, frameshifts, and premature stops, there are many DNA-variants (p*olymorphisms*) in the general population that are not pathogenic. In most cases, mutations leading to single amino acid changes need to be further investigated to rule out the possibility of a sequence polymorphism with no functional significance whatsoever. Population studies have to be performed in appropriate controls and functional analyses would also be desired, but one should realize that this is often not possible in a diagnostic setting. Literature and mutation database searches are necessary and the position of a mutation within important domains of a gene or conservation of the altered amino acid throughout evolution can provide some evidence that the variant is indeed a deleterious mutation, although this type of evidence usually does not provide absolute certainty. It can also be helpful to demonstrate cosegregation of the variant with the disorder in a family (provided that the family is cooperative and large enough to draw conclusions from). Lack of cosegregation of a variant with the disease is very strong evidence against it being the causal mutation, and this can sometimes be proven even if the family is quite small. Besides, since the mutation frequency of individual genes is extremely low, the demonstration that the variant is absent in the DNA of both healthy parents and thus must have occurred de novo, is considered convincing evidence in favor of it being a real mutation The other way round, if the variant is present in one of the healthy parents, this cannot be automatically considered as proof of it being a harmless variant. The parent having the variant, may be an as yet non-penetrant carrier of a deleterious mutation, therefore this type of evidence has to be weighted carefully (what age does the parent have and what investigations have been performed in order to rule out minimal manifestations of the disease in the parent?). Finally, one should realize that it is virtually impossible to exclude the involvement of a specific gene, based on the fact that no mutations have been found. This is due to the fact that causative mutations can be present in noncoding regions (introns or intergenic regulating regions), which are not and cannot be fully included in the current diagnostic approaches.

1.1.1.4 Indirect DNA-Diagnostics: Linkage Analysis and Risk Haplotype Analysis

With polymorphic DNA markers a disease locus can be followed within a family (see Fig. 1.5).[5,13] In this way, the carrier status of a patient can be determined indirectly once a locus for the gene harboring the causative mutation has been established. If a polymorphic marker is close to the disease locus (1–5 cM) (cM stands for centi morgan, which is a measure of genetic distance. If two loci are 1 cM apart, the chance that a crossover will occur between the two loci during meiosis is 1%) the marker will segregate in the vast majority of cases with the disease locus as the occurrence of chromosomal crossover events between the marker and the actual mutation during meiosis is unlikely (1–5%). If such markers on both sides of the mutation are available, the chance of double crossovers that would lead to false interpretation of the test is even smaller (1–5% × 1–5% = 1/10,000–25/10,000), whereas single crossovers that make the test impossible to interpret will be easily detected with flanking markers. The polymorphic markers are often PCR fragments or large numbers of SNPs that, due to the variations in DNA, differ in length or sequence and can be separated with electrophoresis or alternative techniques. In general, risk haplotype analyses can only be used in cases where the disease locus and carrier status of at least two individuals within the family are known. For genome wide linkage analyses (completely unknown locus) the family has to be large enough (≥10 individuals with unambiguous genetic status in case of autosomal dominant disease to exclude segregation of a marker with the disease locus by chance). Computer

analyses are necessary but linkage of a marker with a disease locus is considered to be proven if the chance of linkage as compared to chance segregation is 1,000:1. Linkage is excluded at odds of chance to linkage of 100:1. These likelihoods are given in logarithms or LOD scores (log of the odds). A lod score of 3 means a chance of 1,000 to 1 odds for linkage as compared to chance. Lod scores of 3 can only be achieved in large families (roughly more than ten informative meioses, i.e., heterozygotes for whom the segregation of the homologous chromosomes in relation to the disease can be determined). This means that the genetic status of at least 11 individuals should be known without doubt. Smaller families are often studied to give an indication of the gene involved or exclude genes not involved, but conclusive DNA-diagnosis by linkage is not possible in such smaller families. Figure 1.5 gives an example of a linkage study. Polymorphic markers are numbered according to their length. The "black bar" combination of polymorphic markers (haplotype) is segregating with the disease locus, a combination of LQT3 and "Brugada" syndrome segregating within a single family. These markers on chromosome 3 segregate with the disease giving a lod score of 6.5. Analysis of the relevant gene (a sodium channel SCN5A) revealed a three nucleotide insertion (TGA) at position 5537, causing the disease.

1.1.1.5 Direct DNA-Diagnostics: Screening for Small Mutations

Sanger Sequencing

The sequence of a particular DNA fragment can be determined by using a primer sequence (short oligonucleotide complementary to a small DNA fragment adjacent to the DNA fragment to be sequenced).[5] Using this primer as a starting point, the sequence of a single-stranded DNA fragment can be artificially copied by an enzyme called DNA-polymerase and the addition of the chemical components of DNA (the nucleotides adenine, cytosine, thymine, and guanine). The newly synthesized strands are chemically modified by the addition of small amounts of dideoxynucleotides that are labeled with fluorochromes or radioactive isotopes. The dideoxynucleotides represent all four existing nucleotides (A, T, C, or G) that are added in four distinct reactions or are labeled differently so they can be recognized as such. Incorporation of a dideoxynucleotide stops the synthesis of DNA. This occurs randomly, but with a predetermined frequency. Addition of modified adenine for instance reveals the position of all adenines in the genetic code by creating partially synthesized DNA fragments that stop at all positions of adinine. By doing so with all four modified nucleotides and by size separating fragments with a resolution of one nucleotide by high voltage electrophoresis, the genetic code can be read simply by determining the length of the all fragments. The nucleotide present in a particular fragment comes before that of a longer fragment. Figure 1.8 shows an example of such a sequence reaction. The procedure is nowadays fully automated and the dideoxynucleotides are labeled with fluorochromes of four different colors. For capillary sequencers, the separated fragments are detected with laser technology (colored peaks). By comparing the known sequence with the sequence in patients, a mutation can be found. Although the sequencing is fully automated, the quality control of the sequence and comparison to the wild type (normal) sequence is not completely automated; therefore, sequencing an entire gene for mutations is still quite laborious. In the case shown (Fig. 1.8), a DNA mutation was found leading to a change from amino acid glutamine to a stop at position 356. The result is an incomplete nonfunctional ion channel protein (KCNQ1) that causes the long QT syndrome type 1 (Romano-Ward syndrome) in this patient.

Gene Scanning Methods: Alternatives to Sequencing Entire Candidate Genes

As an alternative to the laborious sequencing of entire genes many clinical laboratories use a *gene scanning method* to quickly go through a gene for possible aberrations; subsequently they only sequence those parts of the gene that demonstrated an abnormal pattern using the scanning technique. Many different techniques exist, each with their pros and cons, and differing sensitivity and specificity. Most of these techniques are either based on the fact that most mutations will cause a conformational change of the single-stranded DNA and therefore altered characteristics at gel electrophoresis (SSCP single-stranded conformation polymorphism), or on the difference in stability between heteroduplexes and homoduplexes (DGGE denaturing

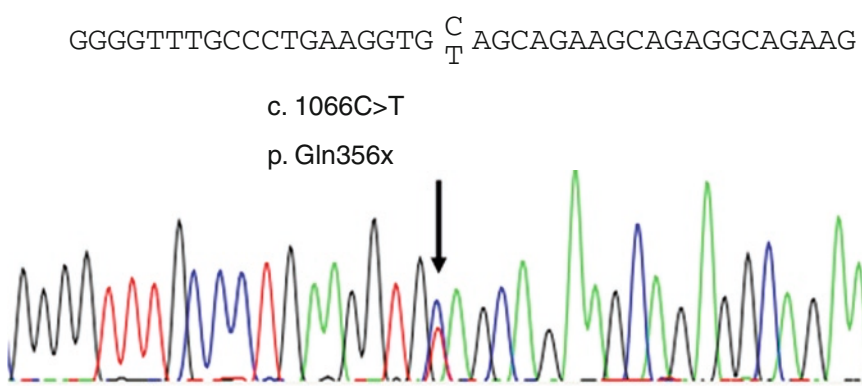

Fig. 1.8 Sequence analysis. A DNA heterozygous C to T mutation was found at position 1066, leading to a change from a codon (CAG) for amino acid glutamine at position 356 in the protein to a stop codon (TAG) at position 356. The result is an incomplete nonfunctional ion channel protein (KCNQ1) that causes the long QT syndrome type 1 (Romano-Ward syndrome) in this patient

gradient gel electrophoresis, DHPLC denaturing high-performance liquid chromatography, HRMC high resolution melting curve analysis).

These latter methods detect DNA consisting of mismatched mutant and wild-type DNA strands.[5] These so-called heteroduplexes are less stable than homoduplexes (consisting of either two mutated or two wild type strands)and can be, respectively, detected by an altered melting profile, using denaturing agents and HPLC, saturating dyes and heating or denaturing agents and electrophoresis. The methods are relatively easy to perform and detect up to 100% of mutations. These approaches are illustrated by an example of DGGE (Fig. 1.9). In this technique, the DNA is denatured (made single stranded) and then allowed to reanneal (become double stranded again), in case of a heterozygous mutation different types of molecules will be formed: homoduplexes when two wild type (normal strands) reanneal or when two mutated strands reanneal, and heteroduplexes when a wild type strand reanneals with a mutated strand.

The newly formed double-stranded DNA molecules demonstrate changes in electrophoretic mobility on a gel with a gradient of denaturing agents. Because of the mutation, DNA fragments will melt at different positions in the denaturing gel and thus get stuck at different positions (Fig. 1.9). Formally, these techniques have been designed to detect heterozygous mutations as heteroduplexes are easiest to identify, but often homoduplexes consisting of two mutated strands show a pattern that can also be discerned from the wild-type homoduplex (see Fig. 1.9), thus allowing for the detection of homozygous mutations. Figure 1.9b. shows a N543H heterozygous mutation in the LDL receptor gene causing familial hypercholesterolemia compared to control DNA. The aberrant homo- and heteroduplexes are clearly visible.

An alternative technology to DGGE is denaturing high-performance liquid chromatography (DHPLC).[19] In contrast to DGGE, DHPLC uses chromatography instead of electrophoresis to detect aberrant DNA fragments such as insertions, deletions, and also single base substitutions. The sensitivity of the technology is comparable to sequencing, although a sequence reaction will still be needed to determine the actual sequence change. The detection level of DHPLC is even higher than in sequencing (1% of a mutation in a pool of wild type (=normal) molecules is detectable[20]). This may be of importance when looking for mosaic mutations that are present in only a fraction of the cells.

New Technologies Emerging on the Horizon

Many alternatives techniques exist or emerge that fall outside the scope of this book. Most promising are those techniques that allow screening of multiple genes in parallel for each patient, allowing a more rapid identification of the underlying gene defect, especially for those genes in which mutations are rarely found, and yielding a more complete picture of all mutations and variants involved. This is of importance for both

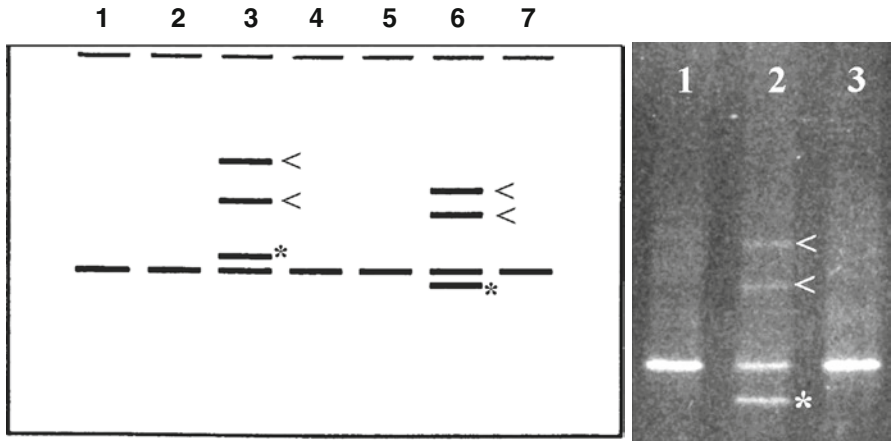

Fig. 1.9 The left-hand part shows a theoretical DGGE analysis. Mutants are seen in lane 3 and 6, revealing homo- and heteroduplexes (heteroduplexes indicated with < and mutated homoduplexes indicated with *). The heteroduplexes are less stable as a result of the mismatch so they denature more easily and get stuck higher in the gel at a lower concentration of denaturing agents than the homoduplexes. Besides heteroduplexes, both gels show an abnormal homoduplex different from the homoduplex present in the loaded controls, corresponding to the homoduplex formed by two mutated strands. The right-hand part shows a N543H heterozygous mutation in the LDL receptor gene (lane 2) causing familial hypercholesterolemia compared to control DNA (lane 1 and 3). The aberrant homo- and heteroduplexes are clearly visible on this DGGE gel. (*Reprinted with permission Jorde, Carey, Bamshad, White, Medical Genetics third edition, Mosby Elsevier 2006)

diagnostics and prognostics. Mass spectrometry is an example that can be used to quickly screen large numbers of samples for known mutations. By using microarray or DNA-microchip approaches the entire sequence of multiple genes can be rebuilt by overlapping oligonucleotides and the hybridization pattern of the DNA of the patient resolves its sequence rapidly.[21] Next generation sequencing (also called massive parallel sequencing) will enable us to screen multiple genes in multiple patients in one short experiment (400–600 Mb per run). Both approaches still require prior amplification or capture of the candidate genes involved for which a variety of enrichment strategies are available. So far these non-PCR-based enrichment strategies have not reach the required coverage for diagnostic applications. However, developments in sequencing technology proceed in a very rapid pace and it can be expected that within 3–5 years a complete human genome can be determined within a week for less than $1,000 without any prior handling or enrichment. However, in general practice, many of these techniques are not yet or only just available or applicable. They have been successful in research, but they have to be validated for diagnostic purposes. These approaches will yield massive numbers of data and variants in patients, which have to be filtered and broken down to the pathological variants, which are responsible for or involved in the disease manifestations. It will require a major effort from bioinformatics and functional approaches to cope with and interpret these huge amounts of sequence variants. ICT solutions are required to deal with the immense data sets generated by these technologies.

In case of quantitative differences, like deletions and duplications, real-time PCR or MLPA has been proven very valuable since these techniques can provide information on gene (or chromosome) copy number within hours. Sometimes, analyses at the RNA or protein level are necessary to predict the result of a DNA chance or as alternative in cases where no DNA alteration can be found. A major breakthrough is also expected in the field of copy number variations from the microarray or DNA-microchip and next generation sequencing approaches mentioned above. These technologies can be used to screen the complete sequence of multiple genes for small mutations or to analyse large numbers of clones (cDNA or genomic clones) for major rearrangements. In a different application oligonucleotide clones are attached on the array and the expression level of those genes can be determined in RNA samples

from normal or affected tissue.[22] Specific expression profiles may exist for specific gene defects or pathogenic states, allowing a refined characterization of the underlying genetic cause or disease state in the patient. As the microarray and next generation sequencing technology approach can also be used to determine large numbers of risk factors, it is clear that the time where a more complete knowledge of genetic defects and predisposition of patients emerge is nearby. It is, however, also clear that interpreting this information and explaining it to patients will be the major bottleneck in the years to come. It will be important to focus on relevant information and established knowledge in health care. Scientists, clinicians, and society will have to collaborate closely to make the genomic revolution a success, and that will be of benefit to mankind.

References

1. International Human Genome Consortium. Initial sequencing and analysis of the human genome. *Nature*. 2001;409: 860-921.
2. Venter JC, Adams MD, Myers EM, et al. The sequence of the human genome. *Science*. 2001;291:1304-1351.
3. Van Langen IM, Hofman N, Wilde AAM. Family and population strategies for screening and counselling of inherited cardiac arrhythmias. *Ann Med*. 2004;36(Suppl 1):116-124.
4. Morita H, Seidman J, Seidman CE. Genetic causes of human heart failure. *J Clin Invest*. 2005;115:518-526.
5. Strachan T, Read AP. *Human Molecular Genetics*. 2nd ed. Oxford: Bios Scientific Publishers; 1999.
6. Priori SG, Napolitano C, Schwartz PJ. Low penetrance in the Long-QT syndrome. Clinical impact. *Circulation*. 1999;99: 529-533.
7. Priori SG. Long QT and Brugada syndromes: from genetics to clinical management. *J Cardiovasc Electrophysiol*. 2000; 11:1174-1178.
8. Bezzina C, Veldkamp MW, Van den Berg MP, et al. A single Na^+ channel mutation causing both long-QT and Brugada syndromes. *Circ Res*. 1999;85:1206-1213.
9. Bezzina CR, Rook MB, Wilde AAM. Cardiac sodium channel and inherited arrhythmia syndromes. *Cardiovasc Res*. 2001;49:257-271.
10. Neyroud N, Tesson F, Denjoy I, et al. A novel mutation in the potassium channel gene KVLQT1 causes the Jervell and Lange-Nielsen cardioauditory syndrome. *Nat Genet*. 1997; 15:186-189.
11. Towbin JA, Lipshultz SE. Genetics of neonatal cardiomyopathy. *Curr Opin Cardiol*. 1999;14:250-262.
12. Reik W, Walter J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet*. 2001;2:21-32.
13. Haines Jl, Pericak-Vance MA, eds. *Approaches to Gene Mapping in Complex Human Diseases*. New York: Wiley-Liss; 1998.
14. Bonne G, Carrier L, Richard P, Hainque B, Schwartz K. Familial hypertrophic cardiomyopathy. From mutations to functional defects. *Circ Res*. 1998;83:580-593.
15. Nicol RL, Frey N, Olson EN. From the sarcomere to the nucleus: role of genetics and signaling in structural heart disease. *Annu Rev Genomics Hum Genet*. 2000;01:179-223.
16. Wilde AAM, Roden DM. Predicting the long-QT genotype from clinical data. From sense to science. *Circulation*. 2000;102:2796-2798.
17. Hawkins JR. *Finding Mutations, the Basics*. Oxford: Oxford University Press; 1997.
18. DeCoo IFM, Gussinklo T, Arts PJW, van Oost BA, Smeets HJM. A PCR test for progressive external ophthalmoplegia and Kearns-Sayre syndrome on DNA from blood samples. *J Neur Sci*. 1997;149:37-40.
19. Underhill PA, Jin L, Lin AA, et al. Detection of numerous Y chromosome biallelic polymorphisms by denaturing high-performance liquid chromatography. *Genome Res*. 1997; 10:996-1005.
20. van den Bosch BJC, de Coo RFM, Scholte HR, et al. Mutation analysis of the entire mitochondrial genome using denaturing high performance liquid chromatography. *Nucl Acids Res*. 2000;28:e89-96.
21. Lipshutz RJ, Fodor SPA, Gingeras TR, Lockhart DJ. High density synthetic oligonucleotide arrays. *Nat Genet*. 1992; 1(Suppl):20-25.
22. Young RA. Biomedical discovery with DNA arrays. *Cell*. 2000;102:9-15.