CHAPTER 5

Principles of Clinical Cytogenetics and Genome Analysis

Clinical cytogenetics is the study of chromosomes, their structure, and their inheritance, as applied to the practice of medicine. It has been apparent for over 50 years that chromosome abnormalities—microscopically visible changes in the number or structure of chromosomes—could account for a number of clinical conditions that are thus referred to as **chromosome disorders**. With their focus on the complete set of genetic material, cytogeneticists were the first to bring a genome-wide perspective to the practice of medicine. Today, chromosome analysis—with increasing resolution and precision at both the cytological and genomic levels—is an important diagnostic procedure in numerous areas of clinical medicine. Current genome analyses that use approaches to be explored in this chapter, including **chromosomal microarrays** and **whole-genome sequencing,** represent impressive improvements in capacity and resolution, but ones that are conceptually similar to microscopic methods focusing on chromosomes (Fig. 5-1).

Chromosome disorders form a major category of genetic disease. They account for a large proportion of all reproductive wastage, congenital malformations, and intellectual disability and play an important role in the pathogenesis of cancer. Specific cytogenetic disorders are responsible for hundreds of distinct syndromes that collectively are more common than all the single-gene diseases together. Cytogenetic abnormalities are present in nearly 1% of live births, in approximately 2% of pregnancies in women older than 35 years who undergo prenatal diagnosis, and in fully half of all spontaneous, first-trimester abortions.

The spectrum of analysis from microscopically visible changes in chromosome number and structure to anomalies of genome structure and sequence detectable at the level of whole-genome sequencing encompasses literally the entire field of medical genetics (see Fig. 5-1). In this chapter, we present the general principles of chromosome and genome analysis and focus on the **chromosome mutations** and **regional mutations** introduced in the previous chapter. We restrict our discussion to disorders due to genomic imbalance—either for the hundreds to thousands of genes found on individual chromosomes or for smaller numbers of genes located within a particular chromosome region. Application of these principles to some of the most common and best-known chromosomal and genomic disorders will then be presented in Chapter 6.

INTRODUCTION TO CYTOGENETICS AND GENOME ANALYSIS

The general morphology and organization of human chromosomes, as well as their molecular and genomic composition, were introduced in Chapters 2 and 3. To be examined by chromosome analysis for clinical purposes, cells must be capable of proliferation in culture. The most accessible cells that meet this requirement are white blood cells, specifically T lymphocytes. To prepare a short-term culture that is suitable for cytogenetic analysis of these cells, a sample of peripheral blood is obtained, and the white blood cells are collected, placed in tissue culture medium, and stimulated to divide. After a few days, the dividing cells are arrested in **metaphase** with chemicals that inhibit the mitotic spindle. Cells are treated with a hypotonic solution to release the chromosomes, which are then fixed, spread on slides, and stained by one of several techniques, depending on the particular diagnostic procedure being performed. They are then ready for analysis.

Although ideal for rapid clinical analysis, cell cultures prepared from peripheral blood have the disadvantage of being short-lived (3 to 4 days). Long-term cultures suitable for permanent storage or further studies can be derived from a variety of other tissues. Skin biopsy, a minor surgical procedure, can provide samples of tissue that in culture produce **fibroblasts,** which can be used for a variety of biochemical and molecular studies as well as for chromosome and genome analysis. White blood cells can also be transformed in culture to form **lymphoblastoid** cell lines that are potentially immortal. **Bone marrow** has the advantage of containing a high proportion of dividing cells, so that little if any culturing

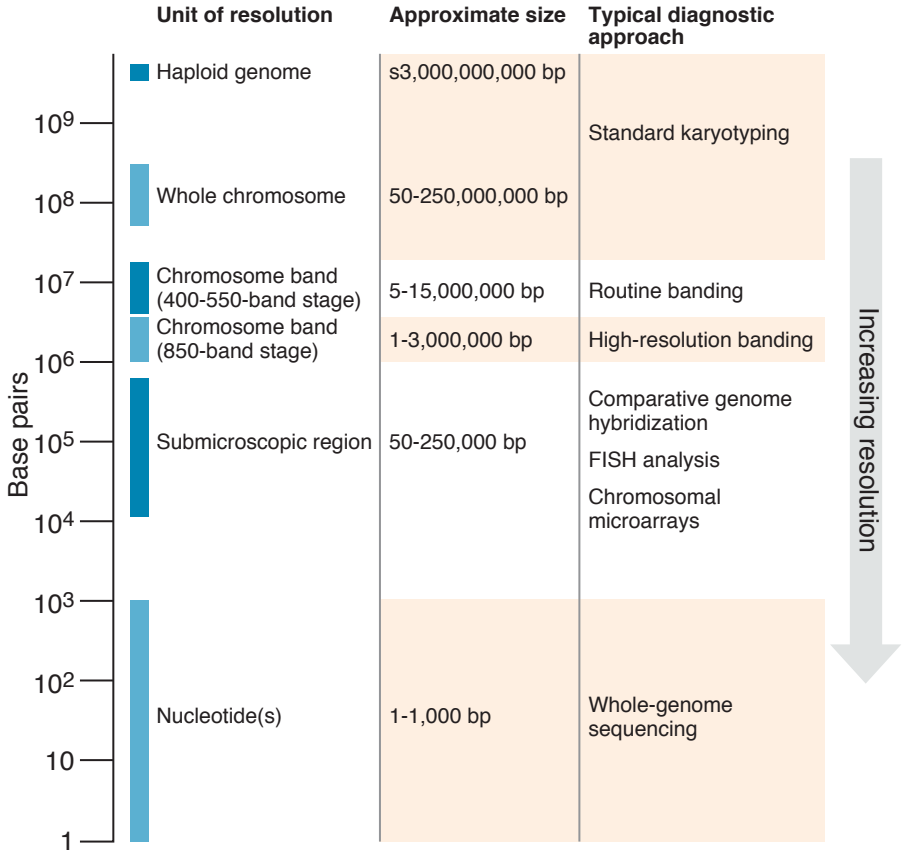

Figure 5-1 **Spectrum of resolution in chromosome and genome analysis.** The typical resolution and range of effectiveness are given for various diagnostic approaches used routinely in chromosome and genome analysis. See text for details and specific examples. FISH, Fluorescence in situ hybridization.

is required; however, it can be obtained only by the relatively invasive procedure of marrow biopsy. Its main use is in the diagnosis of suspected hematological malignancies. **Fetal cells** derived from amniotic fluid (amniocytes) or obtained by chorionic villus biopsy can also be cultured successfully for cytogenetic, genomic, biochemical, or molecular analysis. Chorionic villus cells can also be analyzed directly after biopsy, without the need for culturing. Remarkably, small amounts of **cell-free fetal DNA** are found in the maternal plasma and can be tested by whole-genome sequencing (see Chapter 17 for further discussion).

Molecular analysis of the genome, including whole-genome sequencing, can be carried out on any appropriate clinical material, provided that good-quality DNA can be obtained. Cells need not be dividing for this purpose, and thus it is possible to study DNA from tissue and tumor samples, for example, as well as from peripheral blood. Which approach is most appropriate for a particular diagnostic or research purpose is a rapidly evolving area as the resolution, sensitivity, and ease of chromosome and genome analysis increase (see Box).

Chromosome Identification

The 24 types of chromosome found in the human genome can be readily identified at the cytological level by specific staining procedures. The most common of these, Giemsa banding (**G banding**), was developed in the early 1970s and was the first widely used whole-genome analytical tool for research and clinical diagnosis (see Figs. 2-1 and 2-10). It has been the gold standard for the detection and characterization of structural and numerical genomic abnormalities in clinical diagnostic settings for both constitutional (postnatal or prenatal) and acquired (cancer) disorders.

G-banding and other staining procedures can be used to describe individual chromosomes and their variants or abnormalities, using an internationally accepted system of chromosome classification. Figure 5-2 is an ideogram of the banding pattern of a set of normal human chromosomes at metaphase, illustrating the alternating pattern of light and dark bands used for chromosome identification. The pattern of bands on each chromosome is numbered on each arm from the centromere to the telomere, as shown in detail in Figure 5-3 for several chromosomes. The identity of any particular

CLINICAL INDICATIONS FOR CHROMOSOME AND GENOME ANALYSIS

Chromosome analysis is indicated as a routine diagnostic procedure for a number of specific conditions encountered in clinical medicine. Some general clinical situations indicate a need for cytogenetic and genome analysis:

- ***Problems of early growth and development.*** Failure to thrive, developmental delay, dysmorphic facies, multiple malformations, short stature, ambiguous genitalia, and intellectual disability are frequent findings in children with chromosome abnormalities. Unless there is a definite nonchromosomal diagnosis, chromosome and genome analysis should be performed for patients presenting with any combination of such problems.
- ***Stillbirth and neonatal death.*** The incidence of chromosome abnormalities is much higher among stillbirths (up to approximately 10%) than among live births (approximately 0.7%). It is also elevated among infants who die in the neonatal period (approximately 10%). Chromosome analysis should be performed for all stillbirths and neonatal deaths that that do not have a clear basis to rule out a chromosome abnormality. In such cases, karyotyping (or other comprehensive ways of scanning the genome) is essential for accurate genetic counseling. These analyses may provide important information for prenatal diagnosis in future pregnancies.
- ***Fertility problems.*** Chromosome studies are indicated for women presenting with amenorrhea and for couples with a history of infertility or recurrent miscarriage. A chromosome abnormality is seen in one or the other parent in 3% to 6% of cases in which there is infertility or two or more miscarriages.
- ***Family history.*** A known or suspected chromosome or genome abnormality in a first-degree relative is an indication for chromosome and genome analysis.
- ***Neoplasia.*** Virtually all cancers are associated with one or more chromosome abnormalities (see Chapter 15). Chromosome and genome evaluation in the tumor itself, or in bone marrow in the case of hematological malignant neoplasms, can offer diagnostic or prognostic information.
- ***Pregnancy.*** There is a higher risk for chromosome abnormality in fetuses conceived by women of increased age, typically defined as older than 35 years (see Chapter 17). Fetal chromosome and genome analysis should be offered as a routine part of prenatal care in such pregnancies. As a screening approach for the most common chromosome disorders, noninvasive prenatal testing using whole-genome sequencing is now available to pregnant women of all ages.

band (and thus the DNA sequences and genes within it) can be described precisely and unambiguously by use of this regionally based and hierarchical numbering system.

Human chromosomes are often classified into three types that can be easily distinguished at metaphase by the position of the **centromere,** the primary constriction visible at metaphase (see Fig. 5-2): **metacentric** chromosomes, with a more or less central centromere and arms of approximately equal length; **submetacentric** chromosomes, with an off-center centromere and arms of clearly different lengths; and **acrocentric** chromosomes, with the centromere near one end. A potential fourth type of chromosome, **telocentric,** with the centromere at one end and only a single arm, does not occur in the normal human karyotype, but it is occasionally observed in chromosome rearrangements. The human acrocentric chromosomes (chromosomes 13, 14, 15, 21, and 22) have small, distinctive masses of chromatin known as **satellites** attached to their short arms by narrow stalks (called secondary constrictions). The stalks of these five chromosome pairs contain hundreds of copies of genes for ribosomal RNA (the major component of ribosomes; see Chapter 3) as well as a variety of repetitive sequences.

In addition to changes in banding pattern, nonstaining gaps—called **fragile sites**—are occasionally observed at particular sites on several chromosomes that are prone to regional genomic instability. Over 80 common fragile sites are known, many of which are heritable variants. A small proportion of fragile sites are associated with specific clinical disorders; the fragile site most clearly shown to be clinically significant is seen near the end of the long arm of the X chromosome in males with a specific and common form of X-linked intellectual disability, **fragile X syndrome** (Case 17), as well as in some female carriers of the same genetic defect.

High-Resolution Chromosome Analysis

The standard G-banded karyotype at a 400- to 550-band stage of resolution, as seen in a typical metaphase preparation, allows detection of deletions and duplications of greater than approximately 5 to 10 Mb anywhere in the genome (see Fig. 5-1). However, the sensitivity of G-banding at this resolution may be lower in regions of the genome in which the banding patterns are less specific.

To increase the sensitivity of chromosome analysis, high-resolution banding (also called **prometaphase banding**) can be achieved by staining chromosomes that have been obtained at an early stage of mitosis (prophase or prometaphase), when they are still in a relatively uncondensed state (see Chapter 2). High-resolution banding is especially useful when a subtle structural abnormality of a chromosome is suspected. Staining of prometaphase chromosomes can reveal up to 850 bands or even more in a haploid set, although this method is frequently replaced now by microarray analysis (see later). A comparison of the banding patterns at three different stages of resolution is shown for one chromosome in Figure 5-4, demonstrating the increase in diagnostic precision that one obtains with these longer chromosomes. Development of high-resolution chromosome analysis in the early 1980s allowed the discovery of a number of new so-called **microdeletion syndromes**

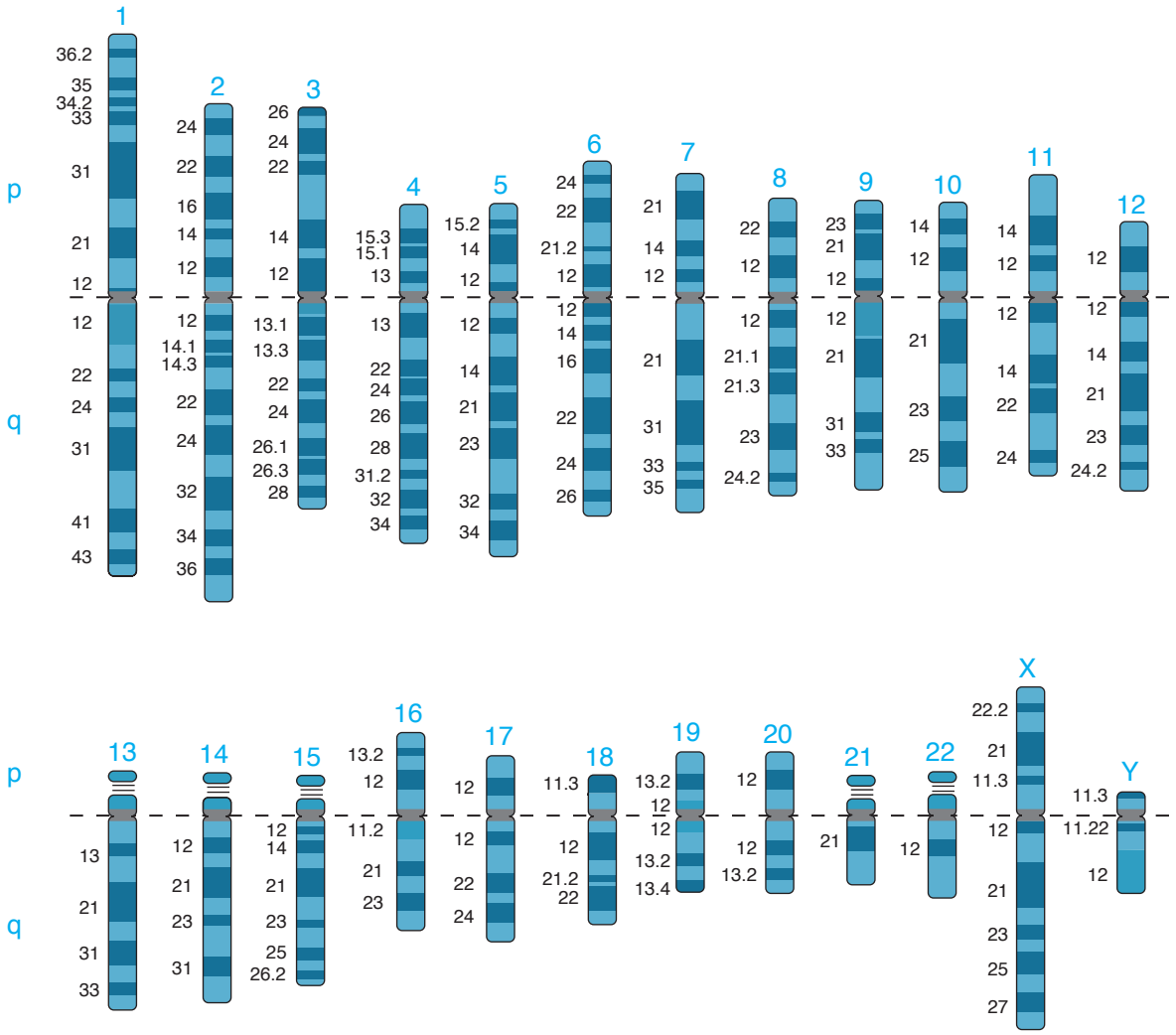

Figure 5-2 Ideogram showing G-banding patterns for human chromosomes at metaphase, with approximately 400 bands per haploid karyotype. As drawn, chromosomes are typically represented with the sister chromatids so closely aligned that they are not recognized as distinct entities. Centromeres are indicated by the primary constriction and narrow *dark gray* regions separating the p and q arms. For convenience and clarity, only the G-dark bands are numbered. For examples of full numbering scheme, see Figure 5-3. *See Sources & Acknowledgments.*

caused by smaller genomic deletions or duplications in the 2- to 3-Mb size range (see Fig. 5-1). However, the time-consuming and technically difficult nature of this method precludes its routine use for whole-genome analysis.

Fluorescence In Situ Hybridization

Targeted high-resolution chromosome banding was largely replaced in the early 1990s by **fluorescence in situ hybridization (FISH)**, a method for detecting the presence or absence of a particular DNA sequence or for evaluating the number or organization of a chromosome or chromosomal region in situ (literally, "in place") in the cell. This convergence of genomic and cytogenetic approaches—variously termed *molecular cytogenetics, cytogenomics,* or *chromonomics*—dramatically expanded both the scope and precision of chromosome analysis in routine clinical practice.

FISH technology takes advantage of the availability of ordered collections of recombinant DNA clones containing DNA from around the entire genome, generated originally as part of the Human Genome Project. Clones containing specific human DNA sequences can be used as probes to detect the corresponding region of the genome in chromosome preparations or in interphase nuclei for a variety of research and diagnostic purposes, as illustrated in Figure 5-5:

- DNA probes specific for individual chromosomes, chromosomal regions, or genes can be labeled with different fluorochromes and used to identify particular chromosomal rearrangements or to rapidly diagnose the existence of an abnormal chromosome number in clinical material.

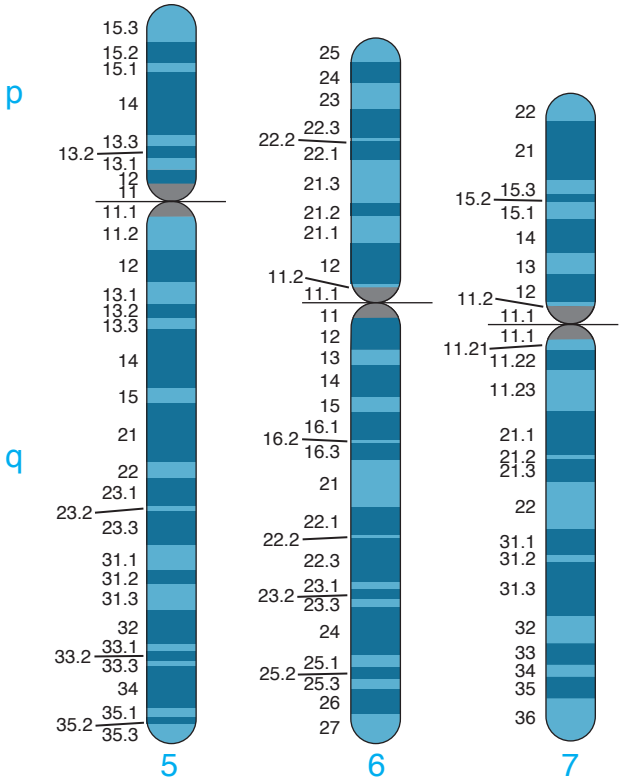

Figure 5-3 **Examples of G-banding patterns for chromosomes 5, 6, and 7 at the 550-band stage of condensation.** Band numbers permit unambiguous identification of each G-dark or G-light band, for example, chromosome 5p15.2 or chromosome 7q21.2. *See Sources & Acknowledgments.*

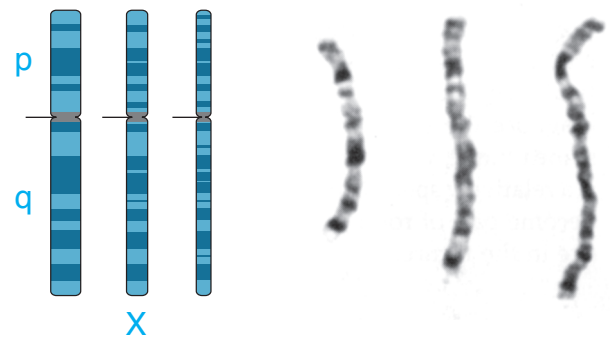

Figure 5-4 The X chromosome: ideograms and photomicrographs at metaphase, prometaphase, and prophase (*left to right*). *See Sources & Acknowledgments.*

- Repetitive DNA probes allow detection of satellite DNA or other repeated DNA elements localized to specific chromosomal regions. Satellite DNA probes, especially those belonging to the α-satellite family of centromere repeats (see Chapter 2), are widely used for determining the number of copies of a particular chromosome.

Although FISH technology provides much higher resolution and specificity than G-banded chromosome analysis, it does not allow for efficient analysis of the entire genome, and thus its use is limited by the need to target a specific genomic region based on a clinical diagnosis or suspicion.

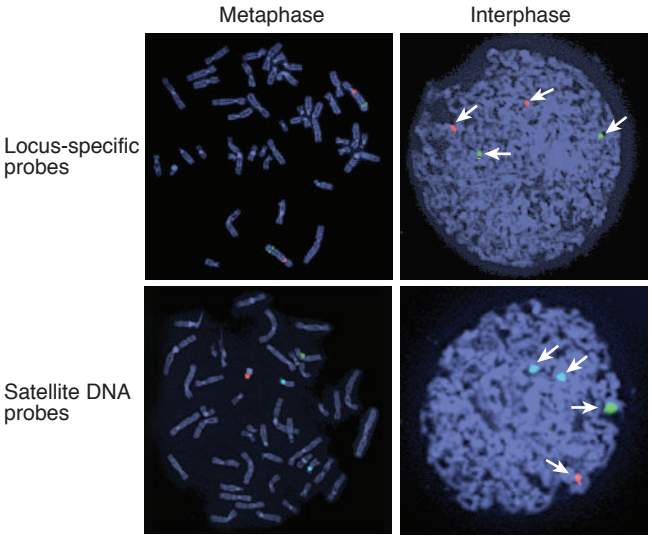

Figure 5-5 **Fluorescence in situ hybridization to human chromosomes at metaphase and interphase, with different types of DNA probe.** *Top*, Single-copy DNA probes specific for sequences within bands 4q12 (*red* fluorescence) and 4q31.1 (*green* fluorescence). *Bottom*, Repetitive α-satellite DNA probes specific for the centromeres of chromosomes 18 *(aqua)*, X *(green)*, and Y *(red)*. *See Sources & Acknowledgments.*

Genome Analysis Using Microarrays

Although the G-banded karyotype remains the frontline diagnostic test for most clinical applications, it has been complemented or even replaced by genome-wide approaches for detecting copy number imbalances at higher resolution (see Fig. 5-1), extending the concept of targeted FISH analysis to test the entire genome. Instead of examining cells and chromosomes in situ one probe at a time, chromosomal microarray techniques simultaneously query the whole genome represented as an ordered array of genomic segments on a microscope slide containing overlapping or regularly spaced DNA segments that represent the entire genome. In one approach based on **comparative genome hybridization (CGH),** one detects relative copy number gains and losses in a genome-wide manner by hybridizing two samples—one a control genome and one from a patient—to such microarrays. An excess of sequences from one or the other genome indicates an overrepresentation or underrepresentation of those sequences in the patient genome relative to the control (Fig. 5-6). An alternative approach uses "single nucleotide polymorphism (SNP) arrays" that contain versions of sequences corresponding to the two alleles of various SNPs around the genome (as introduced in Chapter 4). In this case, the relative representation and intensity of alleles in different regions of the genome indicate if a chromosome or chromosomal region is present at the appropriate dosage (see Fig. 5-6).

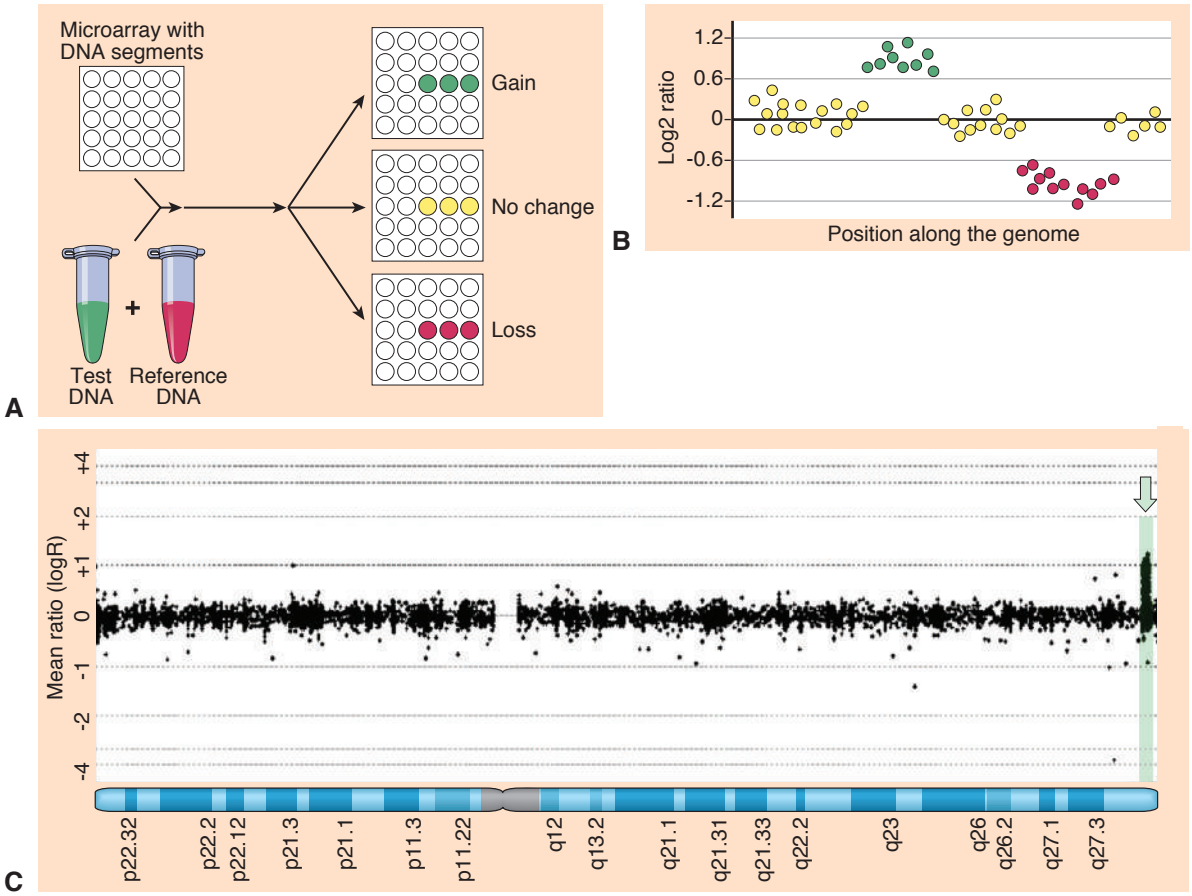

Figure 5-6 **Chromosomal microarray to detect chromosome and genomic dosage. A,** Schematic of an array assay based on comparative genome hybridization (CGH), where a patient's genome (denoted in *green*) is cohybridized to the array with a control reference genome (denoted in *red*). The probes are mixed and allowed to hybridize to their complementary sequences on the array. Relative intensities of hybridization of the two probes are measured, indicating equivalent dosage between the two genomes *(yellow)* or a relative gain *(green)* or loss *(red)* in the patient sample. **B,** A typical output plots the logarithm of the fluorescence ratios as a function of the position along the genome. **C,** Array CGH result for a patient with Rett syndrome (Case 40), indicating a duplication of approximately 800 kb in band Xq28 containing the *MECP2* gene. LogR of fluorescence ratios are plotted along the length of the X chromosome. Each dot represents the ratio for an individual sequence on the array. Sequences corresponding to the *MECP2* gene and its surrounding region are duplicated in the patient's genome, leading to an increased ratio, indicated by the *green arrow* and *shaded box* in that region of the chromosome. *See Sources & Acknowledgments.*

For routine clinical testing of suspected chromosome disorders, probe spacing on the array provides a resolution as high as 250 kb over the entire unique portion of the human genome. A higher density of probes can be used to achieve even higher resolution (<25-50 kb) over regions of particular clinical interest, such as those associated with known developmental disorders or congenital anomalies (see Fig. 5-6; for other examples, see Chapter 6). This approach, which is being used in an increasing number of clinical laboratories, complements conventional karyotyping and provides a much more sensitive, high-resolution assessment of the genome. Microarrays have been used successfully to identify chromosome and genome abnormalities in children with unexplained developmental delay, intellectual disability, or birth defects, revealing a number of pathogenic genomic alterations that were not detectable by conventional G banding. Based on this significantly increased yield, genome-wide arrays are replacing the G-banded karyotype as the routine frontline test for certain patient populations.

Two important limitations of this technology bear mentioning, however. First, array-based methods measure only the relative copy number of DNA sequences but not whether they have been translocated or rearranged from their normal position(s) in the genome. Thus confirmation of suspected chromosome or genome abnormalities by karyotyping or FISH is important to determine the nature of an abnormality and thus its risk for recurrence, either for the individual or for other family members. And second, high-resolution genome analysis can reveal variants, in particular small differences in copy number, that are of uncertain clinical significance. An increasing number of such variants are

being documented and catalogued even within the general population. As we saw in Chapter 4, many are likely to be benign **copy number variants**. Their existence underscores the unique nature of each individual's genome and emphasizes the diagnostic challenge of assessing what is considered a "normal" karyotype and what is likely to be pathogenic.

Genome Analysis by Whole-Genome Sequencing

At the extreme end but on the same spectrum as cytogenetic analysis and microarray analysis, the ultimate resolution for clinical tests to detect chromosomal and genomic disorders would be to sequence patient genomes in their entirety. Indeed, as the efficiency of whole-genome sequencing has increased and its costs have fallen, it is becoming increasingly practical to consider sequencing patient samples in a clinical setting (see Fig. 5-1).

The principles underlying such an approach are straightforward, because the number and composition of any particular segment of an individual's genome will be reflected in the DNA sequences generated from that genome. Although the sequences routinely obtained with today's technology are generally short (approximately 50 to 500 bp) compared to the size of a chromosome or even a single gene, a genome with an abnormally low or high representation of those sequences from a particular chromosome or segment of a chromosome is likely to have a numerical or structural abnormality of that chromosome. To detect numerical abnormalities of an entire chromosome, it is generally not necessary to sequence a genome to completion; even a limited number of sequences that align to a particular chromosome of interest should reveal whether those sequences are found in the expected number (e.g., equivalent to two copies per diploid genome for an autosome) or whether they are significantly overrepresented or underrepresented (Fig. 5-7). This concept is now being applied to the prenatal diagnosis of fetal chromosome imbalance (see Chapter 17).

To detect balanced rearrangements of the genome, however, in which no DNA in the genome is either gained or lost, a more complete genome sequence is required. Here, instead of sequences that align perfectly to the reference human genome sequence, one finds rare sequences that align to two *different* and *normally noncontiguous* regions in the reference sequence (whether on the same chromosome or on different chromosomes) (see Fig. 5-7). This approach has been used to identify the specific genes involved in some cancers, and in children with various congenital defects due to translocations, involving the juxtaposition of sequences that are

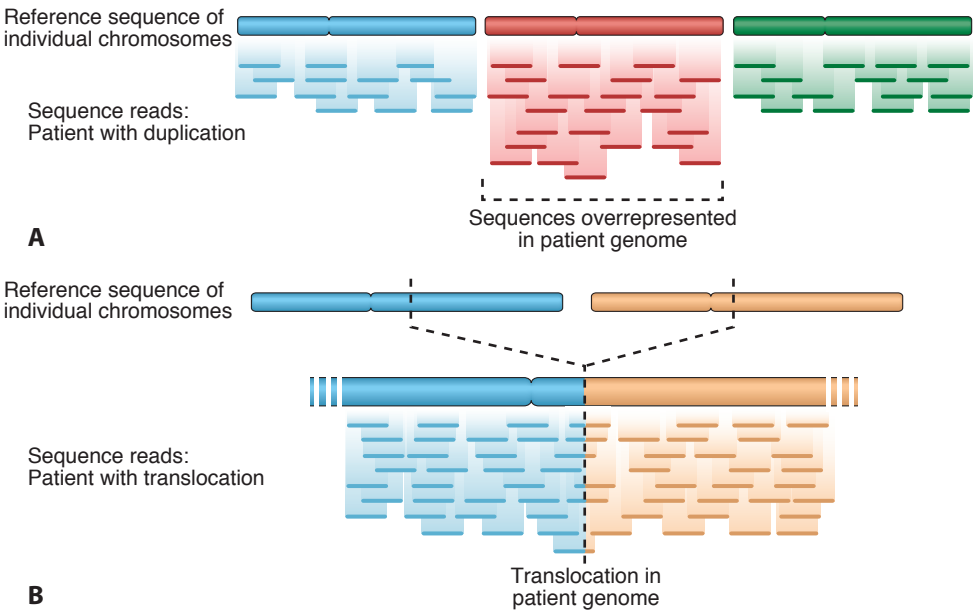

Figure 5-7 Strategies for detection of numerical and structural chromosome abnormalities by whole-genome sequence analysis. Although only a small number of reads are illustrated schematically here, in practice many millions of sequence reads are analyzed and aligned to the reference genome to obtain statistically significant support for a diagnosis of aneuploidy or a structural chromosome abnormality. **A**, Alignment of sequence reads from a patient's genome to the reference sequence of three individual chromosomes. Overrepresentation of sequences from the *red* chromosome indicates that the patient is aneuploid for this chromosome. **B**, Alignment of sequence reads from a patient's genome to the reference sequence of two chromosomes reveals a number of reads that contain contiguous sequences from *both* chromosomes. This indicates a translocation in the patient's genome involving the *blue* and *orange* chromosomes at the positions designated by the *dotted lines*.

normally located on different chromosomes (see Chapters 6 and 15).

CHROMOSOME ABNORMALITIES

Abnormalities of chromosomes may be either numerical or structural and may involve one or more autosomes, sex chromosomes, or both simultaneously. The overall incidence of chromosome abnormalities is approximately 1 in 154 live births (Fig. 5-8), and their impact is therefore substantial, both in clinical medicine and for society. By far the most common type of clinically significant chromosome abnormality is **aneuploidy,** an abnormal chromosome number due to an extra or missing chromosome. An aneuploid karyotype is always associated with physical or mental abnormalities or both. **Structural abnormalities** (rearrangements involving one or more chromosomes) are also relatively common (see Fig. 5-8). Depending on whether or not a structural rearrangement leads to an imbalance of genomic content, these may or may not have a phenotypic effect. However, as explained later in this chapter, even balanced chromosome abnormalities may be at an increased risk for abnormal offspring in the subsequent generation.

Chromosome abnormalities are described by a standard set of abbreviations and nomenclature that indicate the nature of the abnormality and (in the case of analyses performed by FISH or microarrays) the technology used. Some of the more common abbreviations and examples of abnormal karyotypes and abnormalities are listed in Table 5-1.

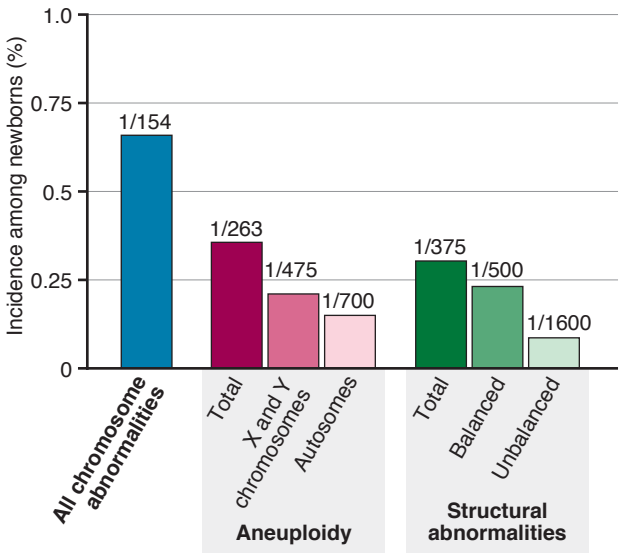

Figure 5-8 Incidence of chromosome abnormalities in newborn surveys, based on chromosome analysis of over 68,000 newborns. *See Sources & Acknowledgments.*

TABLE 5-1 Some Abbreviations Used for Description of Chromosomes and Their Abnormalities, with Representative Examples

Abbreviation	Meaning	Example	Condition
		46,XX	Normal female karyotype
		46,XY	Normal male karyotype
cen	Centromere		
del	Deletion	46,XX,del(5)(q13)	Female with terminal deletion of one chromosome 5 distal to band 5q13
der	Derivative chromosome	der(1)	Translocation chromosome derived from chromosome 1 and containing the centromere of chromosome 1
dic	Dicentric chromosome	dic(X;Y)	Translocation chromosome containing the centromeres of both the X and Y chromosomes
dup	Duplication		
inv	Inversion	inv(3)(p25q21)	Pericentric inversion of chromosome 3
mar	Marker chromosome	47,XX,+mar	Female with an extra, unidentified chromosome
mat	Maternal origin	47,XY,+der(1)mat	Male with an extra der(1) chromosome inherited from his mother
p	Short arm of chromosome		
pat	Paternal origin		
q	Long arm of chromosome		
r	Ring chromosome	46,X,r(X)	Female with ring X chromosome
rob	Robertsonian translocation	rob(14;21)(q10;q10)	Breakage and reunion have occurred at band 14q10 and band 21q10 in the centromeric regions of chromosomes 14 and 21
t	Translocation	46,XX,t(2;8)(q22;p21)	Female with balanced translocation between chromosomes 2 and 8, with breaks in bands 2q22 and 8p21
+	Gain of	47,XX,+21	Female with trisomy 21
–	Loss of	45,XY,–22	Male with monosomy 22
/	Mosaicism	46,XX/47,XX,+21	Female with two populations of cells, one with a normal karyotype and one with trisomy 21

Abbreviations from Shaffer LG, McGowan-Jordan J, Schmid M, editors: *ISCN 2013: an international system for human cytogenetic nomenclature*, Basel, 2013, Karger.