have been discovered falls well short of estimates of genetic inheritance predicted by twin studies, for example it may be that the missing risk factors will prove to be rare genetic variants that are more difficult to study, and have yet to be identified.

Epigenetic Alterations

Epigenetics is defined as the study of heritable chemical modification of DNA or chromatin that does not alter the DNA sequence itself. Examples of such modification include the methylation of DNA, and the methylation and acetylation of histones (Chapter 1). Our understanding of these types of molecular alterations is rapidly growing, and it is clear that epigenetic modifications are critical for normal human development—including the regulation of tissue-specific gene expression, X chromosome inactivation, and imprinting, as well as for understanding of the cellular perturbations in the aging and cancer.

Gene expression frequently correlates with the level of methylation of DNA, usually of cytosines specifically in CG dinucleotide-rich promoter regions known as CpG islands. As discussed earlier in the section on genomic imprinting, increased methylation of these loci is associated with decreased gene expression and is accompanied by concomitant specific patterns of histone methylation and acetylation. An ever increasing number of disease states warrant analysis of promoter methylation—for example, in the diagnosis of fragile X syndrome, in which hypermethylation results in *FMR1* silencing. Methylation analysis is also essential in the diagnosis of Prader-Willi and Angelman syndromes.

Since traditional Sanger sequencing alone cannot detect DNA methylation, other techniques have been developed to uncover these chemical modifications. One common approach is to treat genomic DNA with sodium bisulfite, a chemical that converts unmethylated cytosine to uracil, which acts like thymine in downstream reactions. Methylated cytosines are protected from modification and remain unchanged. After treatment, it is then straightforward to discriminate the unmethylated (modified) DNA from the methylated (unmodified) DNA on the basis of sequence analysis.

RNA Analysis

Because DNA exerts its effects on the cell through RNA expression and mature mRNA contains the coding sequences of all expressed genes, RNA can substitute for DNA in a wide range of diagnostic applications. From a practical standpoint, however, DNA-based diagnosis is usually preferred, since DNA is much more stable. Nonetheless, RNA analysis is critical in several areas of molecular diagnostics. The most important application is the detection and quantification of RNA viruses such as HIV and hepatitis C virus. Furthermore, mRNA expression profiling (described for breast cancer in Chapter 23) is emerging as an important tool for molecular stratification of tumors. In some instances cancer cells bearing particular chromosomal translocations are detected with greater sensitivity by analyzing mRNA (e.g., the *BCR-ABL* fusion transcript in CML). The principal reason for this is that most translocations occur in scattered locations within particular introns, which can be very large, beyond the capacity of conventional PCR amplification. Since introns are removed by splicing during the formation of mRNA, PCR analysis is possible if RNA is first converted to cDNA by reverse transcriptase. Real-time PCR performed on cDNA is the method of choice for monitoring residual disease in patients with CML and certain other hematologic malignancies (Chapter 13).

Next-Generation Sequencing

***Next-generation sequencing* (NGS) is a term used to describe several newer DNA sequencing technologies that are capable of producing large amounts of sequence data in a massively parallel manner.** These technologies, developed over the past decade, have already revolutionized biomedical research, and they are now beginning to have a similar impact on molecular diagnostics. The factors propelling rapid adoption of NGS are both price and performance: NGS allows us to perform previously impossible analyses at extremely low relative cost.

The fundamental factor that sets NGS apart from traditional Sanger sequencing is its input sample requirements. Whereas Sanger sequencing requires a single, simple, homogenous template DNA (usually either a specific PCR product or prepared plasmid), NGS has no such requirement: any DNA from almost any source can be used. Because Sanger sequencing essentially provides an "average" result for a DNA sample, samples with extreme sequence heterogeneity amongst input DNA molecules produce uninterpretable results. NGS instruments, in contrast, are well suited to heterogeneous DNA samples due to the application of these common basic processes (Fig. 5-33):

- *Spatial separation.* At the beginning of the procedure, individual input DNA molecules are physically isolated from each other in space. The specifics of this process are platform-dependent.
- *Local amplification.* After separation, the individual DNA molecules are amplified in situ using a limited number of PCR cycles. Amplification is necessary so that sufficient signal can be generated to ensure detection and accuracy.
- *Parallel sequencing.* The amplified DNA molecules are simultaneously sequenced by the addition of polymerases and other reagents, with each spatially separated and amplified original molecule yielding a "read" corresponding to its sequence. Sequence reads from NGS instruments are generally short, approximately less than 500 bp.

Bioinformatics

NGS instruments can generate a staggering amount of sequence data. For example one newer instrument is capable of analyzing over 500 million individual DNA clusters and producing 180 billion base pairs or more of sequence in a little more than a day. This is enough to produce a high-quality sequence spanning an entire human genome. The downstream analysis necessary to make sense of these enormous data sets is sufficiently complex that specialized training in bioinformatics is frequently needed to ensure its proper interpretation.

Bioinformatic computational pipelines can vary tremendously based on particular applications and sample types,