Another surprising revelation from the recent genomic studies is that, on average, any two individuals share greater than 99.5% of their DNA sequences. It is perhaps more remarkable that we are 99% identical with chimpanzees! Thus, person-to-person variation, including differential susceptibility to diseases and in response to environmental agents and drugs, is encoded in less than 0.5% of our DNA. Though small when compared to the total nucleotide sequences, this 0.5% represents about 15 million base pairs. **The two most common forms of DNA variation in the human genome are *single-nucleotide polymorphisms (SNPs) and copy number variations (CNVs).*** SNPs are variants at single nucleotide positions and are almost always biallelic (i.e., only two choices exist at a given site within the population, such as A or T). Much effort has been devoted to mapping common SNPs in human populations. Over 6 million human SNPs have been identified, many of which show wide variation in frequency in different populations. SNPs occur across the genome—within exons, introns, intergenic regions, and coding regions. Overall, about 1% of SNPs occur in coding regions, which is about what would be expected by chance, since coding regions comprise about 1.5% of the genome. SNPs located in non-coding regions may fall in regulatory elements in the genome, thereby altering gene expression; in such instances the SNP may have a direct influence on disease susceptibility. In other instances, the SNP may be a "neutral" variant that has no effect on gene function or carrier phenotype. However, even "neutral" SNPs may be useful markers if they happen to be co-inherited with a disease-associated gene as a result of physical proximity. In other words, the SNP and the causative genetic factor are in *linkage disequilibrium.* There is hope that groups of SNPs may serve as markers of risk for multigenic complex diseases such as type II diabetes and hypertension. However, the effect of most SNPs on disease susceptibility is weak, and it remains to be seen if identification of such variants, alone or in combination, can be used to develop effective strategies for disease prevention.

CNVs are a more recently identified form of genetic variation consisting of different numbers of large contiguous stretches of DNA from 1000 base pairs to millions of base pairs. In some instances these loci are, like SNPs, biallelic and simply duplicated or deleted in a subset of the population. In other instances there are complex rearrangements of genomic material, with multiple alleles in the human population. Current estimates are that CNVs are responsible for between 5 and 24 million base pairs of sequence difference between any two individuals. Approximately 50% of CNVs involve gene-coding sequences; thus, CNVs may underlie a large portion of human phenotypic diversity. We currently know much less about CNVs than SNPs, therefore their influence on disease susceptibility is less established.

It should be pointed out that despite all these advances in the understanding of human genetic variation, it is clear that alterations in DNA sequence cannot by themselves explain the diversity of phenotypes in human populations. Nor can classic genetics explain how monozygotic twins can have differing phenotypes. The answer may lie in *epigenetics,* which is defined as heritable changes in gene expression that are not caused by alterations in DNA sequence. The molecular basis of epigenetic changes will be discussed next.

Histone Organization

Even though virtually all cells in the body contain the same genetic material, terminally differentiated cells have distinct structures and functions. Clearly, different cell types are distinguished by lineage-specific programs of gene expression. Such cell type-specific differences in DNA transcription and translation depend on *epigenetic factors* (literally, factors that are "above genetics") that can be conceptualized as follows (Fig. 1-2):

- *Histones and histone modifying factors. Nucleosomes* consist of DNA segments 147 base pairs long that are wrapped around a central core structure of highly conserved low molecular weight proteins called *histones*. The resulting DNA-histone complex resembles a series of beads joined by short DNA linkers and is generically called *chromatin*. The naked DNA of a human cell is about 1.8 meters long, but wound around histones like spools, DNA can be packed into a nucleus as small as 7 to 8 micrometers in diameter—the width of a resting lymphocyte. In most cases, DNA is not *uniformly* and compactly wound. Thus, at the light microscopic level, nuclear chromatin exists in two basic forms: (1) cytochemically dense and transcriptionally inactive *heterochromatin* and (2) cytochemically dispersed and transcriptionally active *euchromatin* (Fig. 1-1). Moreover, which portion of the nuclear chromatin is "unwound" regulates gene expression and thereby dictates cellular identity and activity.

  Histones are not static, but rather are highly dynamic structures regulated by a host of nuclear proteins and chemical modifications. Thus, *chromatin remodeling complexes* can reposition nucleosomes on DNA, exposing (or obscuring) gene regulatory elements such as promoters. "*Chromatin writer" complexes*, on the other hand, carry out more than 70 different histone modifications generically denoted as *marks*. Such covalent alterations include methylation, acetylation, or phosphorylation of specific amino acid residues on the histones.

  Actively transcribed genes in euchromatin are associated with histone marks that make the DNA accessible to RNA polymerases. In contrast, inactive genes have histone marks that enable DNA compaction into heterochromatin. Histone marks are reversible through the activity of "*chromatin erasers.*" Still other proteins function as *"chromatin readers,"* binding histones that bear particular marks and thereby regulating gene expression.
- *Histone methylation*. Both lysines and arginines can be methylated by specific writer enzymes; in particular, methylation of lysine residues in histones may be associated with either transcriptional activation or repression, depending on the histone residue that is "marked".
- *Histone acetylation.* Lysine residues are acetylated by histone acetyl transferases (HAT), whose modifications tend to open up the chromatin and increase transcription. In turn, these changes can be reversed by histone deacetylases (HDAC), leading to chromatin condensation.
- *Histone phosphorylation*. Serine residues can be modified by phosphorylation; depending on the specific residue,