7

Molecular Correlates of Chromosome Bands

GC- and AT-rich Isochores

Bernardi and his coworkers have demonstrated that the human genome is organized into alternating blocks, each over 300 kb long, of rather homogeneous DNA of quite different GC richness, called *isochores* (Saccone et al., 1996). These can be separated from one another on the basis of their buoyant density in a cesium chloride ultracentrifugal gradient (Fig. 7.1). This reflects the average nucleotide composition of the isochores, because GC base pairs and GC-rich DNA are denser than AT base pairs and AT-rich DNA. There are five classes of isochores, called L1, L2 (pooled in Fig. 7.1), H1, H2, and H3, in order of increasing buoyant density and thus GC-richness. In situ hybridization shows that the H3 isochores are gene-richest (especially in housekeeping genes) and

O. J. Miller et al., *Human Chromosomes*

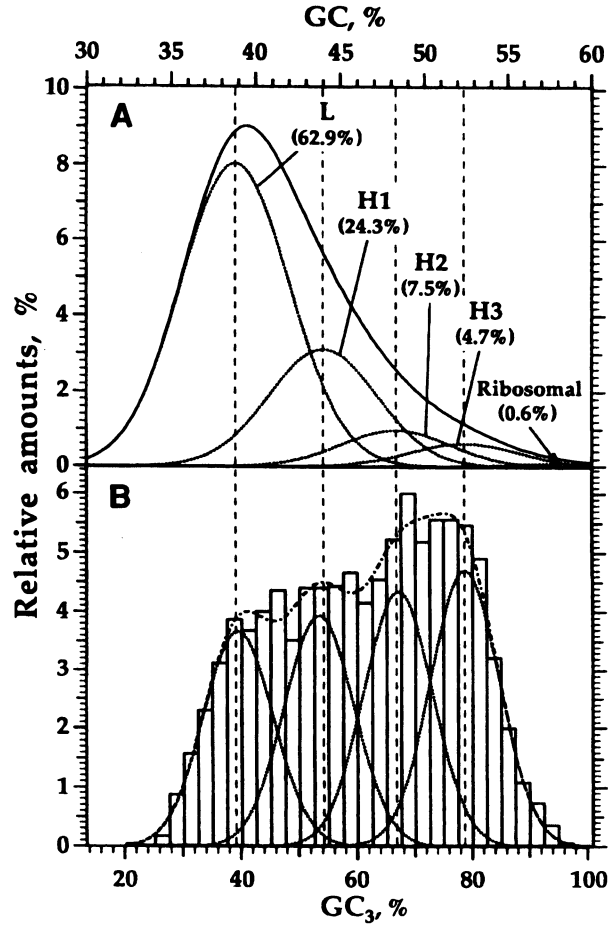

Figure 7.1. (A) The cesium chloride density gradient profile of human DNA can be resolved into four major components and a small ribosomal RNA gene component. (B) A histogram based on the GC richness of the third codon position of 4270 human genes similarly resolved into four components. The GC-riches fractions (H3 and H2) are the richest in genes (reprinted from Gene 174:95–102, 1996, Zoubak et al., The GC-richest (non-ribosomal) component of the human genome, with permission of Elsevier Science).

are present in highest concentration in 28 T-bands and at somewhat lower concentration in 31 additional R-bands (Fig. 7.2; see color insert).

The remaining R-bands at a 400-band resolution contain no H3 isochores but abundant H1 and H2 isochores. They contain more than half the tissue-

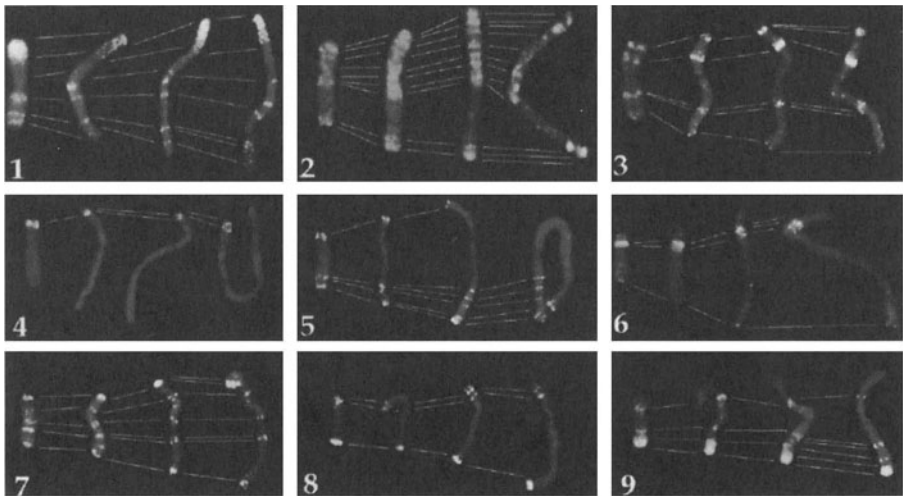

Figure 7.2. Hybridization of GC-rich H3 isochore DNA to chromosomes 1–9 at different stages of contraction. Biotinylated H3 isochore DNA was detected with avidin-FITC (yellow) on propidium iodide-stained chromosomes (red) (Saccone et al., Identification of the gene-richest bands in human prometaphase chromosomes, Chrom Res 7: figure 1, p 382, copyright 1999, with kind permission from Kluwer Academic Publishers) (See color insert).

restricted (tissue-specific) genes, as well as many housekeeping genes. The G-bands contain mainly the L1 and L2 (AT-rich) isochores and nearly half of the tissue-specific genes. The GC-richest component of the genome is at least 17 times as rich in genes as the GC-poor regions are. There is about one gene per 9kb in H3 isochore DNA, one gene per 15kb in H2 isochore DNA, one gene per 54kb in H1 isochore DNA, and one gene per 150kb, on average, in the combined L1 and L2 isochore DNAs (Zoubak et al., 1996). Adjacent isochores need not be functionally distinct. Thus, the *SOX9* gene belongs to a GC-rich H2 isochore, but almost all of the 1063kb of DNA upstream (5′) of the gene belongs to the GC-poor L1 and L2 isochore families. However, when this region is disrupted at any point by a translocation, the expression of the *SOX9* gene is impaired (Pfeifer et al., 1999).

How sharp are the boundaries between adjacent isochores, and what is their structure and functional significance? There is growing evidence that chromosome bands may have specific boundary sequences that separate a

GC-rich band from a less GC-rich band. The still-homologous pseudoautosomal region on Xp and Yp (Chapter 17) has a boundary sequence, and pseudoautosomal boundary-like sequences (PABLs) have been found throughout the genome. They have a highly conserved 650-bp consensus sequence. One example is provided by the major histocompatibility complex (*MHC*) gene locus at 6p21.3 (Fig. 29.3). This is a wide R-band that includes a very GC-rich T-band. There is also an intervening G-band only 200 kb in length at 6p21.32, which separates the *MHC* class II genes from class I and class III genes. This tiny region contains a PABL1 element and a cluster of LINE1 elements (Fukagawa et al., 1996).

Unmethylated CpG Clusters and Housekeeping and Tissue-Specific Genes

The restriction endonuclease *Msp*I cleaves double-stranded DNA at CCGG sites in DNA whether or not the internal C is methylated, whereas another restriction enzyme, *Hpa*II, cleaves CCGG sites in DNA only if the internal C is not methylated. Bird and his collaborators have shown that a tiny fraction of the genome, about 1%, is highly enriched in methylatable sites (CGs, usually referred to as CpGs) that are not methylated. These were first detected as clusters of *Hpa*II-generated tiny fragments, or HTF islands. Analysis of 375 sequenced genes indicated that all housekeeping genes and about 40% of tissue-restricted genes are associated with CpG clusters. The majority of these clusters, which are generally 200–1400 bp long, are at the 5′ ends of housekeeping genes, but there is no such bias towards the 5′ end of tissue-restricted genes (Larsen et al., 1992).

Antiqua and Bird (1995) showed that there are about 45,000 CpG islands in the haploid human genome. Combining this with the data of Larsen et al. (1992) and data indicating there is one CpG island per 36 kb in R-(including T-) bands, Antiqua and Bird estimated the total number of genes in the human genome to be about 80,000, made up of 22,000 housekeeping and 58,000 tissue-restricted genes. The housekeeping genes and about half of the tissue-specific genes are in the T- and R-bands, with the remainder of the tissue-specific genes situated in the Q- or G-bands (Holmquist, 1992; Saccone et al., 1996; Zoubak et al., 1996).

Interspersed Repetitive DNA Sequences and Transposable Elements

The human genome contains an enormous number of interspersed repetitive elements that have spread through the genome by some sort of transposition process. They make up almost 35% of the genome (Kazazian and Moran, 1998). Short interspersed elements (SINEs), such as the 340-bp *Alu* sequences, whose name reflects the presence of an *Alu*I restriction endonuclease cutting site in each one, are especially abundant; they make up more than 10% of the genome. They are located mainly in the R-bands (Korenberg and Rykowski, 1988). A small subset of the *Alu* repeats can be transcribed and a DNA copy made by a special DNA polymerase called *reverse transcriptase*, because it makes DNA from an RNA template. This DNA copy can be transposed into a new genomic site. If this disrupts a gene (insertion mutagenesis), the latter may be inactivated, as exemplified by a cholinesterase gene mutation (Muratani et al., 1991).

Retroelements tend to integrate into actively transcribed DNA, in part because its more open conformation makes its DNA more accessible. *Alu* sequences are thus most abundant in the gene-rich T- and R-bands. For example, the *RCC1* gene, a regulator of chromosome condensation (Chapter 2), contains 37 *Alu* sequences within its 35 kb of DNA (Furuno et al., 1991). The consensus sequence of one *Alu* subfamily contains a binding site for the retinoic acid receptor that is present as part of the 5′ promoter region of many genes. This provides a mechanism by which retrotransposition of *Alu* sequences can alter the expression of many genes and contribute to evolutionary potential (Vansant and Reynolds, 1995).

Long interspersed elements (LINEs), with about 50,000–100,000 copies, make up nearly 15% of the genome. They are mainly located in the Q- (G-)bands and are particularly abundant on the X chromosome, especially in the pericentromeric region (Korenberg and Rykowski, 1988). Full length LINE sequences are over 6 kb long and, like similar transposable elements, contain several genes; however, most copies are truncated to a variable extent at the 5′ end and are much shorter. A significant number, estimated at 30–60, of the 3,000–4,000 full-length LINE L1 elements are able to initiate transposition (Sassaman et al., 1997). These encode an endonuclease that can nick DNA and a reverse transcriptase that can make a usually truncated DNA copy of an RNA transcript.

This can integrate into a new chromosomal site. Insertion of a truncated LINE L1 element into the X-linked factor VIII gene has been seen in a number of patients with hemophilia A (Dombroski et al., 1991).

Sometimes DNA copies of unrelated mRNAs are generated using reverse transcriptase and are integrated into the genome. These *processed pseudogenes* have no introns, because these are removed during the processing of RNA transcripts into mature mRNA. Usually, there are only a few copies of these pseudogenes, which are thus a minor component of interspersed repeats. Transposable elements are much more abundant. One, called *mariner*, or MITE (*mariner* insert-like transposable element), is present in about 1000 copies, and so are several other such elements. The complete *mariner* transposon contains inverted terminal repeats that enable it to bind a transposase and thus to move. There are more than 100 of these complete *mariner* elements scattered throughout the genome, except on the Y, as shown by primed in situ labeling (PRINS) with primers matching the right and left inverted repeats. Many of them are at sites where a disease-producing deletion, duplication, or inversion has occurred, suggesting that they are involved in initiating homologous recombination events (Reiter et al., 1999).

An updated list of retroelements can be obtained from the repetitive element data base (Repbase.http://www.girinst.org), or those on chromosome 6 from the chromosome 6 database (http://www.sanger.ac.uk/chr6). Their distribution in relation to bands is not well established. Given the frequency and amount of transposition characteristic of the human and other genomes, how have such striking differences in GC-richness of isochores been achieved, and maintained? One exciting hypothesis is that integration of exogenous DNA tends to be isopycnic, that is, the GC level of the incoming DNA sequence matches that of the host sequences around the integration site. This has been demonstrated for multiple integration sites of hepatitis B (HBV), AIDS (HIV-1), and human T-cell leukemia, type 1 (HTLV-1) viruses (Glukhova et al., 1999).

Tandemly Repetitive Elements: Telomeres, Centromeres, and Satellite DNAs

At the molecular level, chromosomes consist of highly repeated sequences (most of 100,000 copies or more), middle repeated sequences (most of 100–10,000 copies) and unique sequences (one or several copies). The boundaries between

these classes are somewhat arbitrary. In the highly repetitive sequences, the repeating units vary in length from 2 to 2000 base pairs. In simple-sequence DNA, such as that of constitutive heterochromatin, short sequences are repeated over and over again to make long, meaningless stretches of DNA. These simple DNAs show extensive sequence variation. Simple-sequence DNA may have a buoyant density distinctly different from that of the bulk DNA of the organism, forming satellite bands in density gradient centrifugation; hence the name *satellite* DNA. There are three types of satellite DNA. Megabase pair satellite DNAs are limited mainly to centromeric heterochromatin and distal Yq. Kilobase-pair *minisatellite* DNAs are preferentially (90%) located in subtelomeric T-bands. The roughly 10 to 40-base-pair *microsatellite* DNAs are widely distributed throughout the genome.

Multigene families, which encode ribosomal RNA, histones, and a few other classes of genes, represent middle repetitive DNA. These genes, which encode RNAs or proteins needed in large quantities in the cell, are repeated hundreds of times in the genome. Many of them are located in T-bands and some in R- or G-bands. The clusters of ribosomal RNA genes are located within satellite DNA and are virtually the only type of gene able to avoid being inactivated by such close contact with heterochromatin. Transposable elements, which do not have a fixed location on the chromosomes, make up a considerable proportion of middle repetitive DNA. Other multigene families, of which globin and immunoglobulin genes are typical examples, bridge the gap between unique sequences and middle repeated sequences. Characteristic of genes in such multigene families are multiplicity, close linkage, sequence homology, and similar or overlapping functions. Unique sequences can be divided into transcribed genes and noncoding sequences, including spacers and nonfunctional pseudogenes.

Chromosomal Proteins and Chromatin Conformation

The usual technique for labeling DNA for molecular hybridization is called *nick translation*. It involves nicking DNA with DNAase I in the presence of DNA polymerase and radioactive nucleotides. The polymerase fills in the gaps produced by the DNAase I with nucleotides, producing a labeled probe. Kerem et al. (1984) applied nick translation in situ to fixed chromosome spreads, using a limiting concentration of DNAase I, and found that DNAase I–sensitive regions

were concentrated on the active X chromosome in preference to the inactive X and in R-bands in preference to G-bands. This confirms the more open conformation of the chromatin on the active X and in the R-bands.

This difference in chromatin structure is associated with striking differences in the proteins of the various types of bands. Histone H4 is hyperacetylated in both T- and R-bands but not in G- or C-bands, as shown by immunofluorescence studies (Jeppesen and Turner, 1993). Nucleosomal core histones are also hyperacetylated in interphase chromatin fractions that contain unmethylated CpG islands, which come from T- and R-bands. Histone H1, an inhibitor of gene activity, is depleted in these fractions (Tazi and Bird, 1990). In contrast, histone H1 is enriched in nucleosomes that contain methylated DNA, is more abundant in G-bands than in R-bands, and can be depleted from nucleosome arrays by hyperacetylation. The high-mobility-group protein 1 (HMG1) localizes to G- and C-bands (Disney et al., 1989). The chromosome scaffold proteins topoisomerase II and SCII are especially abundant in centromeric heterochromatin, and topoisomerase II is also found in G-bands (Saitoh and Laemmli, 1994; Sumner, 1996). Their role in organizing the chromosome scaffold that anchors

Table 7.1. Characteristics of the Different Types of Chromosome Bands

Feature	T-bands	R-bands	G-bands	C-bands
Quinacrine or DAPI	Dull	Dull	Bright	Either
Acridine orange/heat	Bright	Bright	Dull	Dull
Chromomycin A3	Bright	Bright	Dull	Dull
Replication time	Very early	Early	Late	Very late
DNAase I sensitivity	High	High	Low	Low
In or between chromomeres	Between	Between	In	In
Chiasma frequency	Very high	High	Low	Absent
Acetylation of histone H4	High	High	Low	Low
Unmethylated CpG islands	Very many	Many	Few	None
GC-richest (H3) isochores	Many	Some	None	None
GC-rich (H1, H2) isochores	Few	Many	Rare	None
GC-poor (L1, L2) isochores	None	None	Many	Some
Interspersed repeats	SINEs	SINEs	LINEs	Rare
Satellite DNAs	Mini-satellites	Micro-satellites	Micro-satellites	Satellites

each loop domain (Chapter 5) is clearly reflected in their lower concentrations in T- and R-bands, with their longer loops and looser packaging.

Functional Significance of Chromosome Bands

Chromosome bands reflect the functional organization of the genome that is necessary for regulating DNA replication and repair, transcription, genetic recombination, and transposition (Bickmore and Craig, 1997). Table 7.1 summarizes some of the characteristic differences among T-, R-, G-, and C-bands. There are, of course, additional differences already known and possibly many yet to be discovered. T-bands contain only 15% of the DNA but 65% of the mapped genes, most of the meiotic chiasmata, and most of the cancer-associated or X-ray-induced rearrangements (Holmquist, 1992). Genes in T- and R-bands are at or near the periphery of interphase chromosome domains, while genes in G-bands tend to be buried in the interior of the domain (Chapter 5). C-bands, in contrast, tend to clump together at the nuclear envelope or around a nucleolus.

References

Antiqua F, Bird A (1995) Number of CpG islands and genes in human and mouse. Proc Natl Acad Sci USA 90:11995–11999

Bickmore WA, Craig J (1997) Chromosome bands: patterns in the genome. Chapman & Hall, New York

Disney JE, Johnson KR, Magnuson NS, et al. (1989) High-mobility group protein HMG-1 localizes to G/Q- and C-bands of human and murine chromosomes. J Cell Biol 109:1975–1982

Dombroski B, Mathias SL, Nanthakumar E, et al. (1991) Isolation of an active human transposable element. Science 254:1805–1808

Fukagawa T, Nakamura Y, Okumura K, et al. (1996) Human pseudoautosomal boundary-like sequences: expression and involvement in evolutionary formation of the present-day pseudoautosomal boundary of human sex chromosomes. Hum Mol Genet 5:23–32

Furuno N, Nakagawa K, Eguchi U, et al. (1991) Complete nucleotide sequence of the human RCC1 gene involved in coupling between DNA replication and mitosis. Genomics 11:459–461

Glukhova LA, Zoubak SV, Rynditch AV, et al. (1999) Localization of HTLV-1 and HIV-1 proviral sequences in chromosomes of persistently infected cells. Chrom Res 7:177–183

Holmquist GP (1992) Review article: chromosome bands, their chromatin flavors, and their functional features. Am J Hum Genet 51:17–37

Jeppesen P, Turner BM (1993) The inactive X chromosome in female mammals is distinguished by a lack of histone H4 acetylation, a cytogenetic marker for gene expression. Cell 74:281–289

Kazazian HH Jr, Moran JV (1998) The impact of retrotransposons on the human genome. Nat Genet 19:19–24

Kerem B-S, Goitein R, Diamond G, et al. (1984) Mapping of DNAase I sensitive regions in mitotic chromosomes. Cell 38:493–499

Korenberg JR, Rykowski MC (1988) Human genome organization: Alu, Lines, and the molecular structure of metaphase chromosome bands. Cell 53: 391–400

Larsen F, Gunderson G, Lopez R, et al. (1992) CpG islands as gene markers in the human genome. Genomics 13:1095–1107

Muratani K, Hada T, Yamamoto Y, et al. (1991) Inactivation of the cholinesterase gene by *Alu* insertion: Possible mechanism for human gene transposition. Proc Natl Acad Sci USA 88:11315–11319

Pfeifer D, Kist R, Dewar K, et al. (1999) Campomelic dysplasia translocation breakpoints are scattered over 1 Mb proximal to *SOX9*: evidence for an extended control region. Am J Hum Genet 65:111–124

Reiter LT, Liehr T, Rautenstrauss B, et al. (1999) Localization of *mariner* DNA transposons in the human genome by PRINS. Genome Res 9:839–843

Saccone S, Caccio S, Kasuda J, et al. (1996) Identification of the gene-richest bands in human chromosomes. Gene 174:85–94

Saitoh Y, Laemmli UK (1994) Metaphase chromosome structure: bands arise from a differential folding path of the highly AT-rich scaffold. Cell 76: 609–622

Sassaman DM, Dombroski BA, Moran JV, et al. (1997) Many human L1 elements are capable of retrotransposition. Nat Genet 16:37–43

Sumner AT (1996) The distribution of topoisomerase II on mammalian chromosomes. Chrom Res 4:5–14

Tazi J, Bird A (1990) Alternative chromatin structure at CpG islands. Cell 60: 909–920

Vansant G, Reynolds WF (1995) The consensus sequence of a major *Alu* subfamily contains a functional retinoic acid response element. Proc Natl Acad Sci USA 92:8229–8233

Zoubak S, Clay O, Bernardi G (1996) The gene distribution of the human genome. Gene 174:95–102