Chromosomes 21 and 22: Gene Density

Salvatore Saccone, *University of Catania, Catania, Italy*
Giorgio Bernardi, *Stazione Zoologica 'A. Dohrn', Napoli, Italy*

Isochores are long DNA segments that are fairly homogeneous in base composition. The GC-richest and GC-poorest isochores of the genomes from warm-blooded vertebrates are characterized by the highest and the lowest gene concentrations respectively. Chromosomes 21 and 22 can be used to show some properties of both the GC-richest and GC-poorest isochores of the human genome.

Advanced article

Article contents

doi: 10.1038/npg.els.0005013

Introduction

The recently available DNA sequence of the human genome can be used to directly assess the molar ratio, namely the percentage, of guanine + cytosine in DNA (the GC level) of different genomic regions. This was previously done using a number of different approaches, such as CsCl density gradient and *in situ* hybridization of compositional DNA fractions.

The knowledge of the GC levels of different genomic regions is of special interest because a number of structural and functional properties of the human genome are directly related to the GC levels, namely to the compositional features of the isochores. These properties include gene density replication timing during the S phase of the cell cycle, recombination frequency and the level of transcriptional activity. Other features such as the correlation between isochores and chromosomal bands deserve a detailed analysis because some aspects of the chromosomal bands, such as the different levels of compaction of the DNA, are not yet well understood. (***See*** Chromosomal Bands and Sequence Features; Genome Organization of Vertebrates; Isochores.)

Compositional Properties of the GC-richest Isochores

The two smallest human chromosomes, chromosomes 21 and 22, exhibit very different isochore patterns. While chromosome 21 is made of several compositional regions representing all the isochore families, chromosome 22 is essentially formed by H2 and H3 isochores, with a sizable contribution of H1 isochores (**Figure 1**). Compositional fluctuations are remarkably high in the GC-richest regions compared to the GC-poorest ones. This is evident by comparing the GC level of the long stretch of the L1 isochore of chromosome 21 and the GC-rich regions of chromosome 22 (see, as examples, DNA segments A and B in **Figure 1**), the latter regions being also characterized by a shorter size of isochores (Saccone *et al*., 2001).

The high level of compositional fluctuations observed in the GC-rich chromosomal regions could be due to local dips, associated with the matrix attachment regions, which are short GC-poor regions, in the GC level of a GC-rich isochore.

GC-rich Chromosomal Regions: H3$^+$ Bands

The *in situ* hybridization of the GC-poorest and the GC-richest isochore families led to the identification of

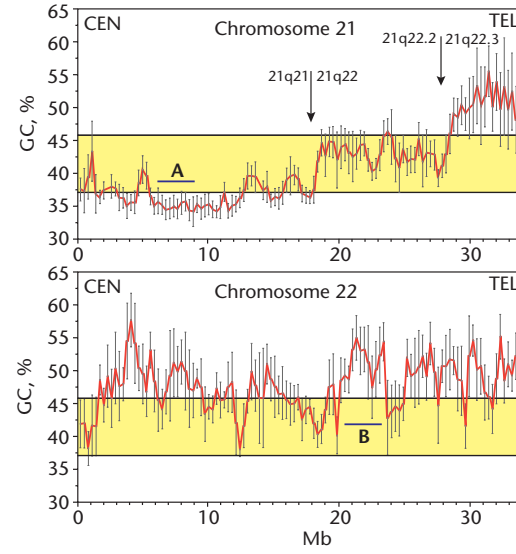

Figure 1 GC level profiles of chromosomes 21 and 22. GC level of 300 kb nonoverlapping segments (average of the GC level from the corresponding 20 kb subwindows) from the long arm of chromosomes 21 and 22. The standard deviation is indicated for each of the 300 kb DNA segments. A and B indicate the chromosomal regions (3 Mb in size) shown in **Figure 4** in more detail. The shaded area represents the intermediate compositional regions, namely the DNA regions, composed by L2 and H1 isochores, separating the GC-richer H2/H3 and the GC-poorer L1 isochores. Two evident band borders are indicated by arrows (see **Figure 2** for details).

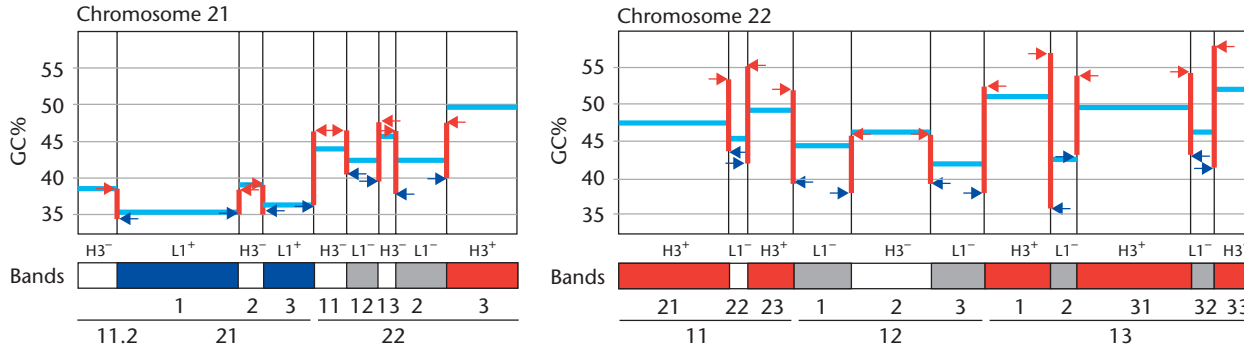

Figure 2 Compositional features of bands from chromosomes 21 and 22. (Bottom to top) Bands, ideograms showing the $H3^+$, $H3^-$, $L1^-$ and $L1^+$ bands. GC%, average GC level of each chromosomal band (horizontal blue lines), and GC levels observed at band borders (red and blue arrows indicate the GC level on the R and G band side respectively; vertical red lines indicate the GC difference over 300 kb regions around band borders). All G bands ($L1^+$ or $L1^-$ bands) show lower GC levels than the adjacent R bands ($H3^+$ or $H3^-$ bands), and all the R bands ($H3^+$ or $H3^-$ bands) show higher GC levels than the adjacent G bands ($L1^+$ or $L1^-$ bands). Note that the sizes of the two chromosomes were scaled according to the cytogenetic ideograms of Francke (1994). (Modified from Saccone *et al.* (2001).)

the $L1^+$ and the $H3^+$ bands (Saccone *et al.*, 1999; Federico *et al.*, 2000). Briefly, $H3^+$ and $H3^-$ bands are the G(iemsa) negative (or Reverse) bands containing or not containing the H3 isochores, whereas $L1^+$ and $L1^-$ bands are the G-positive bands containing or not containing the L1 isochores. Thus, $L1^+$ and $H3^+$ bands represent two chromosomal sets of bands characterized by opposite compositional features. (***See*** Genome Organization of Vertebrates.)

Apart from their very different GC levels, the $H3^+$ and $L1^+$ bands:

- are almost never adjacent and tend to be located in the distal and proximal regions of chromosomes (Federico *et al.*, 2000);
- show different locations also in the interphase nuclei; the GC-rich $H3^+$ band DNA being located in the interior of the nuclei, and the GC-poor $L1^+$ band DNA very close to the nuclear envelope;
- are characterized by more open chromatin in $H3^+$ compared to $L1^+$ bands (Saccone *et al.*, 2002).

These findings are general for warm-blooded vertebrates and have an evolutionary relevance. (***See*** GC-rich Isochores: Origin; Isochores.)

The band borders are associated with compositional jumps (long contiguous stretches of DNA sequences, endowed by a very different average GC level) and identify adjacent bands (**Figures 1** and **2**). Interestingly, while the GC levels of the very GC-rich $H3^+$ bands (that always belong to the R bands) are always higher than those of $L1^+$ bands (that always belong to the G bands), in compositionally intermediate bands (namely the $H3^-$ and the $L1^-$ bands, that always belong to the G-negative and to the G-positive bands respectively), G bands can be endowed with

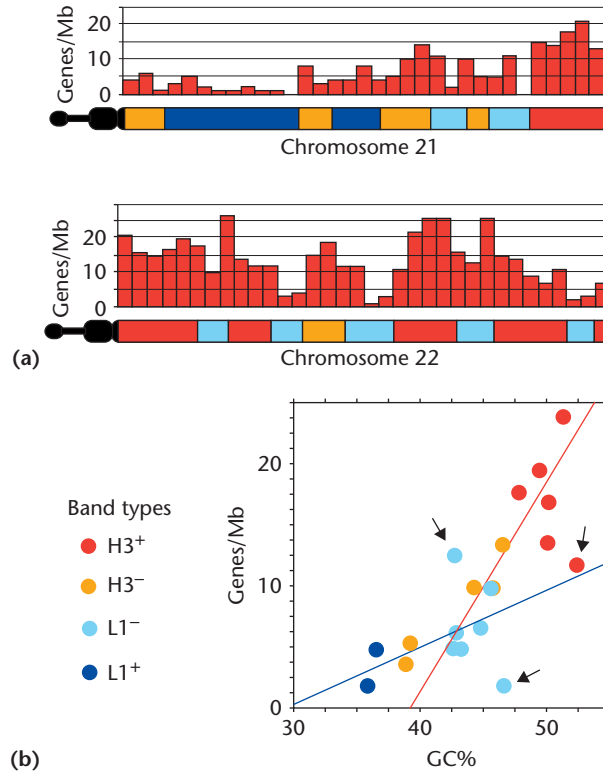

Figure 3 Gene density at the chromosomal band level. (a) Distribution of genes in chromosomes 21 and 22 showing the very different gene density between the $L1^+$ and the $H3^+$ bands. (b) Plot showing the correlation between the average GC level of each band (from chromosomes 21 and 22) and the relative gene density. Three points, indicated by arrows, represent three outliers (two $L1^-$ and one $H3^+$ bands) not taken into consideration when drawing the regression line. Inclusion of these points does not significantly change the lower slope and changes the higher slope only slightly. (Modified from Saccone *et al.* (2001).)

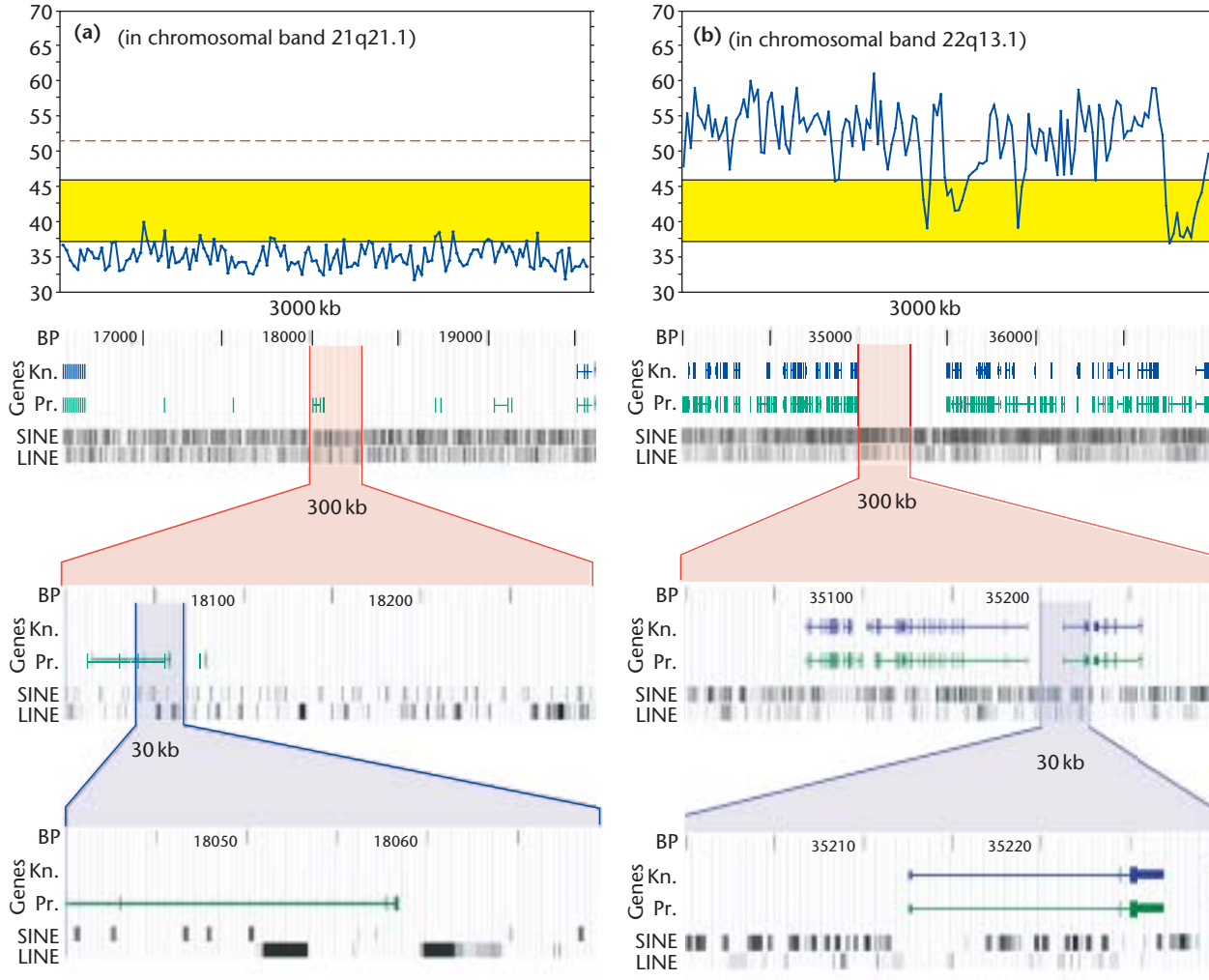

Figure 4 GC level-related properties in two compositionally different genomic regions. (a) and (b) show, at higher resolution, the DNA regions indicated in **Figure 1**. Upper panels: Average GC-level profiles of the 20 kb nonoverlapping windows that form each DNA region. The yellow areas indicate the intermediate compositional regions. The horizontal broken line is the average GC level of the GC-richest H3 isochores. Bottom panels: At different levels of resolution, gene and repeated sequence contents are shown for each chromosomal region (from the UCSC Human Genome Browser http://genome.ucsc.edu/). Each gene is indicated by vertical (exons) and horizontal (introns) lines. BP: base position, indicating nucleotides from the short arm telomere. Genes Kn.: indicate the known protein-coding genes; Pr.: indicate the Fgenesh++ prediction based on Softberry's gene-finding software (see UCSC Human Genome Browser http://genome.ucsc.edu/); SINEs and LINEs: location of these repeats in the sequence.

higher GC levels than those of (nonadjacent) R bands. Indeed, $L1^-$ bands of chromosome 22 are GC-richer than most $H1^-$ bands of chromosome 21. The average GC level of a G band is, however, always lower than those of adjacent R bands and, conversely, the average GC level of an R band is always higher than those of adjacent G bands. Moreover, the GC levels at each band border (over regions of about 300 kb) are always higher on the R side compared to the G side. Incidentally, this stresses the generality of sharp discontinuities at isochore/band borders first observed by Fukagawa *et al.* (1995). One should, therefore, conclude that the G or R bands are not simply associated with the GC level of given chromosomal regions, but also with the composition of the flanking regions (**Figure 2**).

Gene Density

In the human genome (and more generally in the genomes of warm-blooded vertebrates), gene concentration increases with increasing GC levels of isochores (Bernardi *et al.*, 1985; Zoubak *et al.*, 1996) reaching a 20-fold higher level in H3 compared to L1 isochores. Thus, at the chromosomal band level, gene density is very low in $L1^+$ bands and very high in $H3^+$ bands, a

correlation best described by two regression lines with different slopes (**Figure 3**), in agreement with previous results obtained on DNA molecules in the 100 kb size range using an independent approach (Zoubak *et al*., 1996). **Figure 4** shows, at higher resolution, the very high gene density of a GC-rich region compared to a GC-poor one, the latter only containing two identified genes, and 10 predicted genes. (***See*** Gene Distribution on Human Chromosomes.)

Interspersed Repeated Sequence Density

SINEs and LINEs are widespread classes of repeats. Briefly, SINEs are short (about 300 bp) GC-rich, non-autonomous elements, which are derived from 7SL RNA. Alus (the more common SINEs in the human genome) make up 10% of the human genome. Instead, full-length LINEs are long (6–8 kb) GC-poor sequences encoding an RNA-binding protein and a reverse transcriptase/endonuclease. LINE1 elements represent the most abundant group of LINEs and correspond to 15% of the human genome (Smit, 1996). (***See*** Long Interspersed Nuclear Elements (LINEs), Short Interspersed Elements (SINEs).)

The majority of the GC-rich SINEs and of the GC-poor LINEs are very unevenly distributed in the human genome (Soriano *et al*., 1983) being prevalently located in the GC-rich and in the GC-poor isochores respectively (**Figure 4**).

See also

Chromosome 21
Chromosome 22
Chromosomes 21 and 22: Comparisons
Isochores
GC-rich Isochores: Origin

References

Bernardi G, Olofsson B, Filipski J, *et al*. (1985) The mosaic genome of warm-blooded vertebrates. *Science* **228**: 953–958.

Federico C, Andreozzi L, Saccone S and Bernardi G (2000) Gene density in the Giemsa bands of human chromosomes. *Chromosome Research* **8**: 737–746.

Francke U (1994) Digitized and differentially shaded human chromosome ideograms for genomic applications. *Cytogenetics and Cell Genetics* **6**: 206–219.

Fukagawa T, Sugaya K, Matsumoto K, *et al*. (1995) A boundary of long-range G + C% mosaic domains in the human MHC locus: pseudoautosomal boundary-like sequence exists near the boundary. *Genomics* **25**: 184–191.

Saccone S, Federico C and Bernardi G (2002) Localization of the gene-richest and the gene-poorest isochores in the interphase nuclei of mammals and birds. *Gene* **300**: 169–178.

Saccone S, Federico C, Solovei I, *et al*. (1999) Identification of the gene-richest bands in human prometaphase chromosomes. *Chromosome Research* **7**: 379–386.

Saccone S, Pavlicek A, Federico C, Paces J and Bernardi G (2001) Genes, isochores and bands in human chromosomes 21 and 22. *Chromosome Research* **9**: 533–539.

Smit AF (1996) The origin of interspersed repeats in the human genome. *Current Opinions in Genetics and Development* **6**: 743–748.

Soriano P, Meunier-Rotival M and Bernardi G (1983) The distribution of interspersed repeats is non-uniform and conserved in the mouse and human genomes. *Proceedings of the National Academy of Sciences of the United States of America* **80**: 1816–1820.

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