Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 4913–4917, June 1992
Genetics

The highest gene concentrations in the human genome are in telomeric bands of metaphase chromosomes

(isochores/telomeres/chromosome *in situ* suppression hybridization)

SALVATORE SACCONE*, ALBERTINA DE SARIO†, GIULIANO DELLA VALLE*, AND GIORGIO BERNARDI†‡

*Dipartimento di Genetica e Microbiologia A. Buzzati-Traverso, University of Pavia, via Abbiategrasso 207, 27100 Pavia, Italy; and †Laboratoire de Génétique Moléculaire, Institut Jacques Monod, 2, Place Jussieu, 75005 Paris, France

Communicated by Gary Felsenfeld, February 25, 1992 (received for review October 5, 1991)

ABSTRACT **Chromosome *in situ* suppression hybridization has been carried out on human metaphase chromosomes to localize the G+C-richest human DNA fraction (which only represents 3.5% of the genome), as isolated by preparative equilibrium centrifugation in Cs_2SO_4/3,6-bis(acetatomercurimethyl)-1,4-dioxane density gradient. This fraction essentially corresponds to isochore family H3. The rationale for carrying out this experiment is that this isochore family has, by far, the highest gene concentration, the highest concentration in CpG islands, the highest transcriptional and recombinational activity, and a distinct chromatin structure. The *in situ* hybridization results obtained show that the H3 isochore family is localized in two coincident sets of bands of human metaphase chromosomes: telomeric bands and chromomycin A3-positive 4′,6-diamidino-2-phenylindole-negative bands. This result is the first step toward a complete compositional map of the human karyotype. Because the G+C gradient across isochore families is paralleled by a gene concentration gradient, such a map has structural, functional, and evolutionary relevance.**

The human genome is a mosaic of isochores, namely, of large DNA regions (>300 kilobases, on the average) that are compositionally homogeneous (above a size of 3 kilobases) and belong to a small number of families characterized by different G+C levels (1, 2). The distribution of genes in isochore families is strikingly nonuniform (1). Indeed, as shown by recent investigations (3), gene concentration in the G+C-richest 3% of the genome (the isochore family H3) is at least 8 times higher than in the G+C-rich 31% of the genome (the isochore families H1 and H2) and at least 16 times higher than in the G+C-poor 62% of the genome (the isochore families L1 and L2). It should be stressed, however, that the ratios indicated are minimal estimates (see ref. 3).

The G+C-richest H3 family is endowed with exceptionally interesting structural and functional features, in that it has not only the highest concentration of genes and CpG islands but also the highest transcriptional and recombinational activities and a distinct chromatin structure (2–7). In view of its properties, it would be of great interest to know the chromosomal localization of isochore family H3, which has been proposed (2, 8) to correspond to thermal denaturation-resistant telomeric (T) bands (9) and chromomycin A3-positive 4′,6-diamidino-2-phenylindole-negative bands (ref. 10; these bands will be called chromomycin bands henceforth). This proposal was based on the fact that compositional mapping (2, 8) of the long arm of human chromosome 21 showed that single-copy sequences from the G+C-richest isochores are located in the telomeric reverse (R) band, which is a T-band (9) and a chromomycin band (10).

Partial evidence in favor of the location of isochores from the H3 family in T-bands has been obtained in recent investigations. Indeed, (*i*) a probe corresponding to the telomeric repeats present on all human chromosomes (11) only hybridized on isochore families H1, H2, and H3 (12); this finding indicates that the terminal 50–100 kilobases of telomeres (which correspond, in the vast majority of cases, to R- or T-bands) are formed by isochores belonging to families H1 and/or H2 and/or H3. (*ii*) Telomeric probes corresponding to R-bands or T-bands from eight telomeres hybridized on isochore families H1/H2 or H3, respectively (12). (*iii*) G+C levels of third codon positions of genes located in T-bands are higher than those of genes located in Giemsa (G) or R-bands (12, 13) and correspond to those of genes located in isochore family H3 (3). The data presented here provide, however, evidence that the isochores of the H3 family are located in the two coincident sets of T-bands and chromomycin bands that are either telomeric (which is the general case) or intercalary.

MATERIALS AND METHODS

DNA Preparation. Human DNA, extracted from placenta as described (14), was fractionated (15) by ultracentrifugation in a preparative Cs_2SO_4/3,6-bis(acetatomercurimethyl)-1,4-dioxane gradient at a ligand/nucleotide molar ratio (R_f) equal to 0.14. A DNA fraction, with a density of 1.7138 g/cm^3 in a CsCl gradient (i.e., a G+C level of 54.9%) that represents 3.5% of human DNA and corresponds to the isochore family H3 (with minor amounts of DNA from isochore family H2 and some rRNA-encoding DNA on its light and heavy sides, respectively), was isolated and used in our experiments. The H3 fraction was labeled by nick-translation with biotin-16-dUTP purified on Sephadex G-50, ethanol-precipitated with 50 times excess of salmon sperm DNA and yeast tRNA, and hybridized to metaphase spreads (see below) in the presence of a 1000 times excess of unlabeled sonicated total human DNA. In some cases, an excess (50 times) of a pBR322 plasmid containing the Blur 8 sequence (16) was added to the competitor DNA, but this only provided a supplementary 10 times excess of *Alu* sequences.

***In Situ* Hybridization and Detection.** The protocol was adapted from previous protocols (17–23). Metaphase spreads from human lymphocytes were treated with RNase A (100 μg/ml) for 1 h at 37°C and then incubated with proteinase K (0.1 μg/ml) for 7.5 min at 37°C. After fixation in 1% formaldehyde, chromosomes were denatured by immersion in 70% (vol/vol) formamide/2× standard saline citrate (SSC) at 70°C for 3 min and immediately dehydrated through a series of ice-cold ethanol washes. The hybridization solution contained biotinylated DNA (20 ng/μl), 50% formamide, 2× SSC, 10% (wt/vol) dextran sulfate, sonicated salmon sperm DNA (1 μg/μl), yeast tRNA (1 μg/μl), and unlabeled sonicated total human DNA (20 μg/μl). The probe mixture was

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: T-band, telomeric band; R-band, reverse band; G-band, Giemsa band.

‡To whom reprint requests should be addressed.

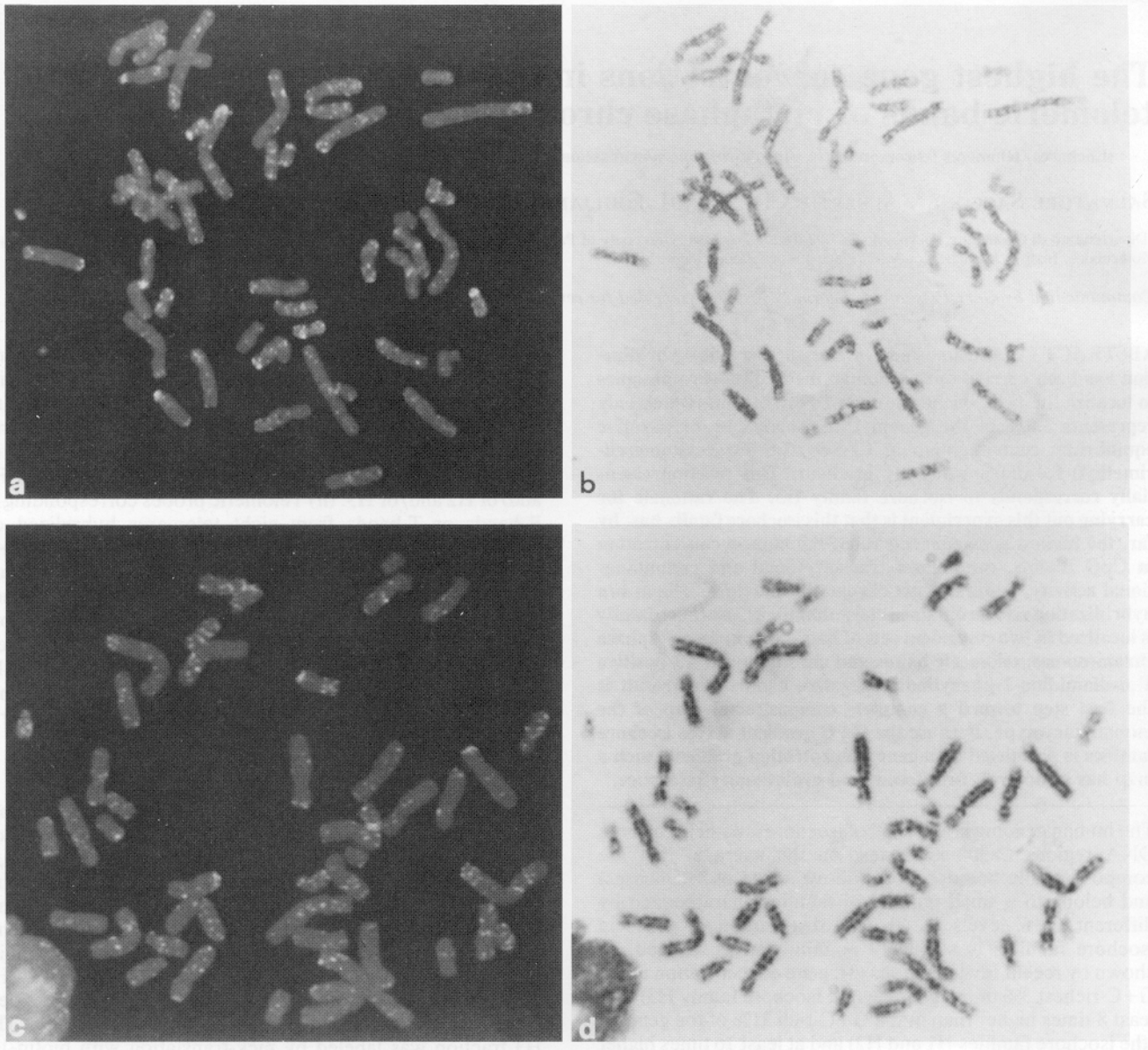

FIG. 1. *In situ* hybridization. The human DNA fraction corresponding to isochore family H3 was hybridized to human male metaphase chromosome spreads. (*a* and *c*) Detection of hybridized biotinylated H3 fraction DNA by fluorescent isothiocyanate-conjugated avidin; chromosomes are counterstained with propidium iodide. (*b* and *d*) G-banding of metaphases shown in *a* and *c*, respectively. In *a*, hybridization was carried out with the H3 DNA probe at a final concentration of 20 ng/μl in the presence of an excess (1000 times) of sonicated unlabeled total human DNA to suppress hybridization of repeated sequences present in the probe. In *c*, an excess (50 times) of a pBR322 plasmid containing the Blur 8 (16) sequence was added to the 1000 times excess of competitor DNA; the results were the same in both cases. Chromosomes were subsequently G-banded with Wright's stain.

denatured at 80°C for 8 min, preannealed at 37°C for 30 min, and applied to the samples. The washings were in 50% formamide/2× SSC (for three 5-min periods) and in 2× SSC (for three 5-min periods) at 42°C. Hybridization was detected using avidin conjugated with fluorescein isothiocyanate. The signal was amplified by three series of incubations with avidin-conjugated fluorescein isothiocyanate and biotinylated goat anti-avidin antibodies. Chromosomes were counterstained with propidium iodide and photographed. The same metaphase spreads were G-banded with Wright's stain and rephotographed.

A total of 3421 hybridization signals scored in 16 well-spread metaphase plates from two males were localized on G-banded chromosomes (≈400 bands for haploid chromosome complement). The estimates of location of the fluorescent signals along chromosomes were mainly limited by two factors. (*i*) The chromosome extension and band resolution are somewhat variable from one metaphase to another, as is the usual case. (*ii*) There are differences in size and intensity of the fluorescent spots. In spite of this, we observed that the vast majority of the signals were clustered on chromosomal regions no wider than 0.5% of the total haploid chromosome complement. For this reason, the histogram showing the distribution of the signals (see Fig. 2) was constructed using bars corresponding to 0.5% intervals of the haploid chromosome complement. Each bar was centered on the peak of the signal distribution. Only clusters with >0.1% of the total signal scored are shown in Fig. 2.

RESULTS

Human DNA was fractionated by preparative centrifugation in a Cs_2SO_4/3,6-bis(acetatomercurimethyl)-1,4-dioxane density gradient (examples of such fractionations can be found in

refs. 5, 12, and 24). A fraction corresponding to isochore family H3 (and containing minor amounts of DNA from the H2 family and some rRNA-encoding DNA) was labeled with biotin and hybridized to human male metaphase chromosomes. Fig. 1 shows two metaphase plates with fluorescent hybridization signals and G-banded chromosomes, respectively. Label is concentrated in several telomeric regions and in a few intercalary bands; rRNA-encoding DNA, present in the H3 fraction used (see above and refs. 5 and 24), was also labeled.

An assessment of the distribution of hybridization signals is presented in Fig. 2 (see above for the criteria used to construct the histogram). Apart from the expected positive bands located on the short arms of the acrocentric chromosomes, due to rRNA-encoding DNA, labeling was essentially found at T-bands. The strongest signals were located at telomeric T-bands on 1p, 2q, 4p, 5q, 7p, 8q, 9q, 10q, 11p, 12q, 16q, 17q, 19p, 20q, 21q, and 22q and at intercalary T-bands located on chromosomes 1, 3, 11, 15, 19, and 22. A low labeling level was found on a number of R-bands, in all likelihood because of the presence of DNA derived from isochore family H2 in the fraction used as a probe (see above and ref. 12).

When a comparison with previous results concerning T-bands (9) and chromomycin bands (10) was made, an excellent agreement was found among the three sets of data (Table 1). At first sight, this may seem an obvious coincidence because (*i*) T-bands are the most heat-denaturation-resistant R-bands (9); (*ii*) a positive reaction with chromomycin A3 is due to the interaction of this compound with G·C base pairs (10); and (*iii*) the H3 isochore family is the G+C-richest component in the human genome. In fact, however, the coincidence is far from obvious, because T-bands and chromomycin bands might be due to tandem or interspersed repeated sequences (or, conceivably, even to protein–DNA complexes in the case of T-bands), whereas the *in situ* hybridization results are due to single-copy DNA.

This conclusion is supported by the following points. (*i*) Hybridization was carried out in the presence of a 1000 times excess of unlabeled human DNA, which is a very strong competitor for *Alu* sequences (and for any repeated sequence that is widely distributed in the human genome) but is a very weak one for single-copy and rRNA-encoding DNA sequences, as present in H3 DNA, which only represents 3.5% of the total human DNA. Indeed, since the H3 fraction contains (see figure 5 of ref. 24) *Alu* sequences at a level 1.5–2 times the average level in total human DNA (namely, 10%), the competitor DNA contains an *Alu* sequence excess of 500–700 times. In contrast, if the H3 fraction contains 50–70% single-copy DNA (like total human DNA; see ref. 25),

FIG. 2. Histogram (solid bars) showing the distribution of sequences hybridizing with the H3 DNA fraction (see legend of Fig. 1) on human G-banded chromosomes. Open bars correspond to rRNA-encoding DNA. Scales are percentage of total number of signals. Neglecting rRNA-encoding DNA, 42% of the signals was found on telomeric T-bands, 31% was on intercalary T-bands, and 27% was scattered on a large number of R-bands.

Table 1. Comparison of the *in situ* hybridization intensities of a human DNA fraction corresponding to the H3 isochore family with T-bands and chromomycin bands

Chromosome region	T-bands	Chromomycin bands	*In situ* hybridization	Chromosome region	T-bands	Chromomycin bands	*In situ* hybridization
1p	+++	+++	+++	13p	var	+	R
1q	–	–	–	13q	+	–	+
1i(p)	++	+++	+++	13i(q)	–	–	–
1i(q)	+	+	+	14p	var	+	R
2p	–	–	–	14q	++	++	+
2q	+	++	++	14i(q)	–	–	–
2i(p)	–	–	+	15p	var	+	R
3p	–	–	–	15q	–	–	–
3q	–	–	–	15i(q)	+	+	++
3i(p)	++	++	+++	16p	++	+	+
3i(q)	+	+	++	16q	++	+	++
4p	+++	+++	+++	16i	–	–	–
4q	–	–	–	17p	–	++	+
4i(q)	–	–	+	17q	+++	+++	++
5p	+	–	+	17i(p)	–	–	+
5q	+	+	++	17i(q)	+	++	+
5i(q)	–	–	+	18p	–	–	–
6p	–	–	–	18q	–	–	–
6q	–	–	–	18i	–	–	–
6i(p)	+	+++	++	19p	+++	+++	+++
7p	+++	+++	++	19q	–	++	–
7q	+	++	+	19i(p)	++	++	++
7i(p)	–	–	+	19i(q)	++	++	+++
7i(q)	++	++	+	20p	–	–	–
8p	–	–	–	20q	++	++	++
8q	+++	+++	++	20i(q)	–	+	–
8i(p)	–	–	+	21p	–	+	R
9p	–	–	–	21q	++	++	++
9q	+++	+++	+++	21i(q)	–	–	–
9i	–	–	–	22p	–	+	R
10p	–	–	–	22q	++	+++	+++
10q	++	+	++	22i(q)	++	++	+++
10i(q)	–	+	+	Xp	–	–	–
11p	+++	++	++	Xq	–	–	–
11q	–	–	+	Xi	–	–	–
11i(q)	+++	+++	+++	Yp	–	–	–
12p	–	+	–	Yq	–	–	–
12q	+	++	++	Yi	–	–	–
12i(q)	+	+	+				

T-bands were evaluated by Dutrillaux (ref. 9; var, variability); intercalary T-bands and chromomycin bands were evaluated by us using the data of Dutrillaux (9) and Ambros and Sumner (10), respectively. For *in situ* hybridization data, –, values <0.6% of total signals; +, ++, and +++, values between 0.6% and 1.2%, between 1.2% and 1.8%, and >1.8%, respectively; p and q, localization on the telomeres of short or long chromosome arms, respectively; i, localization on intercalary bands; R, rRNA-encoding DNA.

the competitor DNA contains a 30–40 times excess of those single-copy sequences. Likewise, since the H3 fraction contains 10–15% rRNA-encoding DNA (as judged from figure 4 of ref. 24), the competitor DNA contains a 30–50 times excess of rRNA-encoding DNA sequences. In other words, the excess of competitor DNA is 10–20 times higher for *Alu* sequences than for single-copy and rRNA-encoding DNA sequences.

(*ii*) The hybridization intensities of *Alu* probes (26) and of the H3 fraction did not coincide; indeed, the latter is strong in T-bands that show low *Alu* intensity (4p, 9q, 10q, 16q, and 21q) and in telomeric T-bands (1p, 17q, 19p, and 22q) and intercalary T-bands (like 3p21, 11q13, 19p13.1, 19q13.3, and 22q11.2) that show high *Alu* intensity. The data of Table 1 indicate, therefore, that T-bands and chromomycin bands do contain the G+C-richest single-copy sequences of the human genome.

Some striking features of the data of Table 1 are the very high G+C levels (*i*) of telomeric and subtelomeric R-bands on the short arm of chromosome 1, (*ii*) of all R-bands of chromosomes 19 and 22, and (*iii*) of the intercalary bands in chromosomes 3 and 11. In this connection, it is worth mentioning that band 1p36 is highly recombinogenic (27), that chromosomes 19 and 22 are globally G+C-rich (28), and that the intercalary T-band 11q13 is the result of a pericentric inversion juxtaposing the telomeric region to the centromere (29). As far as chromosome 19 is concerned, biotin labeling covers the telomere of the long arm; a small G-band is, however, present at the telomere; labeling is, therefore, attributed to the R-band immediately above (19q13.3), in accordance with the interpretation given for the corresponding T-band (9).

DISCUSSION

The major interest of this work is the identification of a very specific chromosomal location for the G+C-richest isochore family H3. This family, which was originally defined on a purely compositional ground (1, 2, 24), is in fact characterized by the following outstanding properties. (*i*) It is, by far, the richest in genes (1–3), CpG doublets, and CpG islands (5, 6); about a third of human genes are currently estimated to be

located in the H3 family; this may, however, be a serious underestimation for reasons given elsewhere (3, 6). (*ii*) It is the most actively transcribed (4–6). (*iii*) It is the most recombination-prone (2, 4). (*iv*) It is the most susceptible to integration by the majority of retroviral sequences (ref. 4; S. Zoubak, A. Rynditch, G.B., unpublished data). Moreover, this family is largely endowed with an open chromatin structure, characterized by DNase sensitivity (2), nucleosome-free regions, scarcity of histone H1, and acetylation of histones H3 and H4 (ref. 7; see also ref. 6). Finally, isochore family H3 is characterized, like the other isochore families, by a remarkable conservation in mammalian evolution (ref. 30 and unpublished data).

In conclusion, the present results demonstrate that two overlapping features of human chromosomes, whose significance was totally obscure so far, T-bands and chromomycin bands, correspond to the strongest concentration of genes and to the highest transcriptional and recombinational activities. Because of these properties, they deserve a special attention in Human Genome Projects.

The present results also represent a step toward a complete compositional mapping of human chromosomes. This endeavor (attempted before, see ref. 26), which involves the individual hybridization of all other compositional DNA fractions on metaphase chromosomes, is expected to lead to a compositional map that will have structural, functional, and evolutionary relevance. Indeed, the compositional (G+C) gradient across isochore families is paralleled by a gene concentration gradient (3). In other words the compositional map of the human karyotype is at the same time a map of gene concentration and expression.

We thank J. Couturier, B. Dutrillaux, and A. T. Sumner for useful discussions and for their help in evaluating the intensities of bands of Table 1, and Mario Negri for skillful technical assistance. This work was supported by grants from the Progetto Finalizzato "Ingegneria Genetica", Consiglio Nazionale delle Ricerche, Rome, the Italian Association of Cancer Research (AIRC), and the Italian "Ministero dell' Università e della Ricerca Scientifica e Tecnologica" to G.D.V.; from the French Ministry for Research and Technology (MRT), the Association Française contre les Myopathies (AFM), and the Association pour la Recherche sur le Cancer (ARC) to G.B. S.S. was supported by the Italian Ph.D. Program.

1. Bernardi, G., Olofsson, B., Filipski, J., Zerial, M., Salinas, J., Cuny, G., Meunier-Rotival, M. & Rodier, F. (1985) *Science* **228**, 953–958.
2. Bernardi, G. (1989) *Annu. Rev. Genet.* **23**, 637–661.
3. Mouchiroud, D., D'Onofrio, G., Aïssani, B., Macaya, G., Gautier, C. & Bernardi, G. (1991) *Gene* **100**, 181–187.
4. Rynditch, A., Kadi, F., Geryk, J., Zoubak, S., Svoboda, J. & Bernardi, G. (1991) *Gene* **106**, 165–172.
5. Aïssani, B. & Bernardi, G. (1991) *Gene* **106**, 173–183.
6. Aïssani, B. & Bernardi, G. (1991) *Gene* **106**, 185–195.
7. Tazi, J. & Bird, A. P. (1990) *Cell* **60**, 909–920.
8. Gardiner, K., Aïssani, B. & Bernardi, G. (1990) *EMBO J.* **9**, 1853–1858.
9. Dutrillaux, B. (1973) *Chromosoma* **41**, 395–402.
10. Ambros, P. F. & Sumner, A. T. (1987) *Cytogenet. Cell Genet.* **44**, 223–228.
11. Moyzis, R. K., Buckingham, J. M., Scott Cram, L., Dani, M., Deaven, L. L., Jones, M. D., Meyne, J., Ratliff, R. L. & Wu, J.-R. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6622–6626.
12. De Sario, A., Aïssani, B. & Bernardi, G. (1991) *FEBS Lett.* **295**, 22–26.
13. Ikemura, T. & Wada, K. (1991) *Nucleic Acids Res.* **19**, 4333–4339.
14. Cuny, G., Soriano, P., Macaya, G. & Bernardi, G. (1981) *Eur. J. Biochem.* **115**, 227–233.
15. Cortadas, J., Macaya, G. & Bernardi, G. (1977) *Eur. J. Biochem.* **76**, 13–19.
16. Schmid, C. V. & Jelinek, W. R. (1982) *Science* **216**, 1065–1070.
17. Pinkel, D., Straume, T. & Gray, J. (1986) *Proc. Natl. Acad. Sci. USA* **83**, 2934–2938.
18. Pinkel, D., Landegent, J., Collins, C., Fuscoe, J., Segraves, R., Lucas, J. & Gray, J. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 9138–9142.
19. Landegent, J. E., Jansen in de Wal, N., Dirks, R. W., Baas, F. & van der Ploeg, M. (1987) *Hum. Genet.* **77**, 366–370.
20. Lichter, P., Cremer, T., Borden, J., Manuelidis, L. & Ward, D. C. (1988) *Hum. Genet.* **80**, 224–234.
21. Wada, M., Little, D. R., Abidi, F., Porta, G., Labella, T., Cooper, T., Della Valle, G., D'Urso, M. & Schlessinger, D. (1990) *Am. J. Hum. Genet.* **46**, 95–106.
22. Montanaro, V., Casamassimi, A., D'Urso, M., Yoon, J.-Y., Freije, W., Schlessinger, D., Muenke, M., Nussbaum, R. L., Saccone, S., Maugeri, S., Santoro, A. M., Motta, S. & Della Valle, S. (1991) *Am. J. Hum. Genet.* **48**, 183–194.
23. Lichter, P., Chang-Tang, C. J., Call, K., Hermanson, G., Evans, G. A., Housman, D. & Ward, D. C. (1990) *Science* **247**, 64–69.
24. Zerial, M., Salinas, J., Filipski, J. & Bernardi, G. (1986) *Eur. J. Biochem.* **160**, 479–485, and erratum (1991) *Eur. J. Biochem.* **197**, 819.
25. Soriano, P., Macaya, G. & Bernardi, G. (1981) *Eur. J. Biochem.* **115**, 235–239.
26. Korenberg, J. R. & Rykowski, M. C. (1988) *Cell* **53**, 391–400.
27. Romani, M., De Ambrosis, A., Alhadeff, B., Purrello, M., Gluzman, Y. & Siniscalco, M. (1990) *Gene* **95**, 231–241.
28. Korenberg, J. R. & Engels, W. R. (1978) *Proc. Natl. Acad. Sci. USA* **75**, 3382–3386.
29. Dutrillaux, B. (1979) *Hum. Genet.* **48**, 251–314.
30. Bernardi, G., Mouchiroud, D., Gautier, C. & Bernardi, G. (1988) *J. Mol. Evol.* **28**, 7–18.