Electronic-Database Information

Accession numbers and URLs for data used in this study are as follows:

Center for Medical Genetics, Marshfield Medical Research Foundation, http://www.marshmed.org/genetics/ (for marker order)
Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nim.nih.gov/Omim (for BHD [MIM 142669], pseudoachondroplasia [MIM 177170], spondyloepiphyseal dysplasia [MIM 183900 and MIM 184100], and multiple epiphyseal dysplasia [MIM 226900])
Whitehead Institute/MIT Genome Sequencing Project, http://carbon.wi.mit.edu (for mapping of ESTs)

References

Al-Ali M, Beighton P, Read A, Warman M, Donn R, Boot-Handford R, Wallis GA (1994) Exclusion of candidate genes in two families with inherited forms of osteoarthritis. Int J Exp Pathol 75:A71
Beighton P, Cilliers HJ, Ramesar R (1994) Autosomal dominant (Beukes) premature degenerative osteoarthropathy of the hip joint unlinked to COL2A1. Am J Med Genet 53: 348–351
Briggs MD, Mortier GR, Cole WG, King LM, Golik SS, Bonaventure J, Nuytinck L, et al (1998) Diverse mutations in the gene for cartilage oligomeric matrix protein in the pseudoachondroplasia–multiple epiphyseal dysplasia disease spectrum. Am J Hum Genet 62:311–319
Cilliers HJ, Beighton P (1990) Beukes familial hip dysplasia: an autosomal dominant entity. Am J Med Genet 36:386–390
Davies JL, Kawaguchi Y, Bennett ST, Copeman JB, Cordell HJ, Pritchard LE, Reed PW, et al (1994) Genome-wide search for human type 1 diabetes susceptibility genes. Nature 371: 130–136
Kuivaniemi H, Tromp G, Prockop DJ (1997) Mutations in fibrillar collagens (types I, II, III and XI), fibril-associated collagen (type IX) and network-forming collagen (type X) cause a spectrum of diseases of bone, cartilage and blood vessels. Hum Mutat 9:300–315
Lathrop GM, Lalouel JM (1984) Easy calculations of LOD scores and genetic risks on small computers. Am J Hum Genet 36:460–465

Address for correspondence and reprints: Dr. Gillian A. Wallis, School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, United Kingdom. E-mail: gwallis@fs1.scg.man.ac.uk

0002-9297/99/6403-0031$02.00

Am. J. Hum. Genet. 64:908–910, 1999

Common Fragile Sites: G-Band Characteristics within an R-Band

To the Editor:

Common fragile sites are chromosomal loci prone to breakage and rearrangement and are considered to be part of the normal chromosome structure. They are visualized as constrictions, gaps, or breaks on metaphase chromosomes from cells exposed to specific tissue-culture conditions (Sutherland and Richards 1995). Three common fragile sites—FRA3B, FRA7H, and FRA7G—were recently cloned and identified at the molecular level (Boldog et al. 1997; Inoue et al. 1997; Huang et al. 1998; Mishmar et al. 1998). Sequence analysis of these three common fragile sites revealed no CGG or other expanded repeated sequences, such as have been found in rare fragile sites (Sutherland and Richards 1995). DNA sequence analysis of FRA3B, FRA7H, and FRA7G did not reveal any obvious feature that could account for the fragility of these sites. To shed light on the mechanism of fragility, we undertook a new approach and analyzed the available sequences of FRA3B, FRA7H, and FRA7G, for DNA structural characteristics that might be associated with their fragility (Mishmar et al. 1998). The analysis revealed several regions with a potential to form unusual DNA structures, including high flexibility, low stability, and non–B DNA–forming sequences. Thus, these unusual DNA characteristics are possibly intrinsic properties of common fragile sites, which may affect their replication, condensation, and organization and may lead to fragility.

While analyzing the sequences of FRA3B, FRA7H, and FRA7G, we noticed several features that are characteristic of G-bands (Gardiner 1995). The three cloned fragile sites have high (>57%) A/T content, and are all gene poor. FRA3B and FRA7H are rich in LINE sequences. FRA3B and markers proximal to FRA7G were shown to replicate late during S-phase (Selig et al. 1992; Huang et al. 1998; Le Beau et al. 1998). G-bands and R-bands correspond to functional subregions, represented as stained bands, that apparently reveal the basic structural organization of chromosomes. G-bands are characterized as regions with high A/T content that replicate late during S-phase, are insensitive to DNase-I, and are gene poor, Alu poor, and LINE rich. In contrast, the complementary R-bands are regions with high G/C content that replicate early during S-phase, are DNase sensitive, and are gene rich, Alu rich, and LINE poor (Gardiner 1995). Most (76/89 [>85%]) of the common fragile sites, including the cloned sites, map to R-bands (according to our analysis of the Genome Database data). These characteristics suggest that fragile sites

Table 1

Clusters of Regions with High (>13.7%) Flexibility in Common Fragile Sites and in R- and G-Band Sequences

Type of Sequence and Variable Locus	Length (bp)	Region	Flexibility Cluster
Common fragile sites, >57% A/T:			
FRA3B (U66722 and AF020503)	276,822	3p14.2	1
FRA7G (AC002066)	151,770	7q31.2	1
FRA7H (AF017104)	161,115	7q32.3	1
G-band, >57% A/T:			
HS79C4	158,548	1q24	1
HS106H8	175,825	1q24	None
HS206D15	101,574	1q24	1
HS117P20	141,589	1q24	None
AC004615	235,141	5p15.2	3
HS451B15	186,510	6p24	None
AC003091	137,817	7p21	1
AC003075	123,336	7p21	1
AC004741	112,242	7p21	None
AC004492	165,608	7q31.1	1
U66059	267,156	7q35	1
U66060	215,422	7q35	1
U66061	232,650	7q35	1[a]
AC002526	134,580	Xq23	1
HSAC002086	112,686	Xq23	None
AC005191	185,893	Xq23	None
AC002476	147,102	Xq23	2
HS75N13	141,851	Xq21.1	1
G-band, <57% A/T:			
HS934G17	107,603	1p36.21	None
HS232K4	198,161	6p22.3	None
HS257A7	127,917	6p24	None
HS445C9	131,398	22q12.1	None
R-band, >57%:			
AC003099	95,129	4q25	None
AC003100	90,430	4q25	1
HS111M5	107,526	6p21.3	None
HS172K2	131,234	6p21.3	None
HS265J14	90,547	6p21.3	None
HS167A14	132,790	6q27	None
HS155D22	148,851	6q27	2
L11910	180,388	13q14.1	None
HSU91325	135,046	16p13.11	None
HS393P12	104,597	Xp11.21	None
HSU82696	158,898	Xq28	None
HSU40455	106,000	Xq28	None
AF003627	81,007	Xq28	1
HS884M20	114,173	Xq28	None
R-band, <57% A/T:			
HSU91326	150,296	16p11.2	None
HSAF001548	145,831	16p13.11	None
HSU95738	171,368	16p13.11	None
HSU95740	149,490	16p13.1	None
HSU95737	93,481	16p13.1	None
HS339A18	132,805	Xp11.2	None
HSU07000	152,141	22q11	None
HSU52111	153,460	Xq28	None
HSU52112	174,424	Xq28	None
HSU82672	156,854	Xq28	None
AF002992	104,037	Xq28	1
AF003628	86769	Xq28	None

[a] Same cluster as in U66060 (U66060 and U66061 are in contig).

might share, with the chromatin of G-bands, structural features (as well as other features) that distinguish the fragile sites from their flanking R-band sequences. This different chromatin organization might affect the replication and condensation of the fragile sequences and thus may contribute to the fragility.

Here we focus on one of the characteristics identified in common fragile sites: high DNA flexibility. We assessed the flexibility by measuring potential local variations in the DNA structure at the twist angle (for details, see Mishmar et al. 1998). A region with potential high flexibility was defined as a region deviating significantly from the average value of the entire analyzed sequence (Mishmar et al. 1998). Because DNA flexibility appears to play an important role in protein-DNA interactions, it could well affect chromatin condensation and organization (Sarai et al. 1989). Our previous flexibility analysis of FRA7H, FRA3B, and FRA7G revealed impressive clusters of regions with high flexibility, at all three sites. A cluster of regions with high flexibility is defined as at least three high-flexibility peaks in a region of <40 kb, on the basis of the flexibility pattern identified in the three common fragile sites (Mishmar et al. 1998). In control sequences comprising 14 genomic sequences that map to chromosomal bands in which fragile sites were not described (1.1 Mb in all), regions with high flexibility appeared approximately every 100 kb. No clusters of regions with high flexibility were identified in these control sequences (Mishmar et al. 1998).

Because fragile sites appeared to share several features with G-bands, we decided to extend our analysis of flexibility patterns (Mishmar et al. 1998) to sequences of known band localization and base composition. For this purpose we have developed a user-friendly computer program (FlexStab) that enables flexibility analysis of sequences as much as 350 kb in length. The program is available from our Website at Hebrew University of Jerusalem. The analysis was performed on available GenBank sequences that map to chromosomal bands in which fragile sites have not been described. The mapping information was drawn from Genome Database 6.0, on the basis of the resolution of ~500 chromosomal bands. Our search identified ~6.9 Mb of DNA sequences of >80 kb each, 3.53 Mb from G-bands and 3.35 Mb from R-bands. Most (2.97 Mb [89%]) of the G-band sequences had, as expected, high (>57%) A/T content (table 1). Fourteen clusters of regions with high flexibility were identified in this group. No clusters of regions with high flexibility were identified in the available G-band sequences (0.56 Mb) with low (<57%) A/T content table 1). In R-band sequences with high A/T content (1.68 Mb), four clusters were identified (table 1). In R-band sequences with low A/T content (1.67 Mb), only one cluster was identified (table 1). There were significantly more high-flexibility clusters in A/T-rich G-band se-

quences than in A/T-rich R-band sequences ($P = .009$). Thus, the flexibility pattern is one of the features that differentiate R- and G-bands.

As mentioned, common fragile sites were found to be A/T rich. The pattern of high-flexibility clusters found in the identified common fragile sites (see table 1) (Mishmar et al. 1998) was significantly different ($P = .02$) from that of A/T-rich control sequences mapped to R-bands. This pattern was not different from that of A/T-rich control sequences mapped to G-bands ($P = .85$). These results might indicate that common fragile sites mapped to R-bands have the flexibility patterns characteristic of G-bands with the same A/T content.

Our previous analysis of potential unusual DNA structures in FRA7H revealed a cluster of regions with potential to form triple helixes (Mishmar et al. 1998). Previous studies, using monoclonal antibodies to triple-helix DNA, showed that G-bands are rich in triple-helix DNA (Burkholder et al. 1991). Thus, clusters of regions with potential to form triple-helix DNA might be added to the G-band characteristics found in common fragile sites.

Together, all the known molecular features of common fragile sites indicate that they might consist of DNA sequences with characteristics of G-bands embedded within R-bands. Of what significance could this feature be to the mechanism of fragility? We think that delayed replication and aberrant condensation of fragile sites might be involved. Chromosomal bands apparently represent regions with several origins of replication that are coordinately controlled to initiate the replication process. The presence of a relatively small region consisting of a common fragile site with G-band characteristics might lead to disturbances in the regional control of replication. This might involve inappropriate initiation of replication in the fragile region. The addition of aphidicolin, which inhibits DNA elongation, might further add to the interference in replication at fragile sites, leading to unreplicated sequences that might adopt abnormal chromatin organization, resulting in fragility.

DAN MISHMAR,[1] YAEL MANDEL-GUTFREUND,[2] HANAH MARGALIT,[2] AYELET RAHAT,[1] AND BATSHEVA KEREM[1]

[1]*Department of Genetics, The Hebrew University, and* [2]*Department of Genetics and Biotechnology, The Hebrew University, Hadassah, Jerusalem*

Electronic-Database Information

URLs for data in this article are as follows:

Genome Database, http://www.gdb.org

Hebrew University of Jerusalem, http://leonardo.ls.huji.ac.il/departments/genesite/faculty/bkerem.htm

References

Boldog F, Gemmill RM, West J, Robinson M, Robinson L, Li E, Roche J, et al (1997) Chromosome 3p14 homozygous deletions and sequence analysis of FRA3B. Hum Mol Genet 6:193–203

Burkholder GD, Latimer LJ, Lee JS (1991) Immunofluorescent localization of triplex DNA in polytene chromosomes of Chironomus and Drosophila. Chromosoma 101:11–18

Gardiner K (1995) Human genome organization. Curr Opin Genet Dev 5:315–322

Huang H, Qian J, Proffit J, Wilber K, Jenkins R, Smith DI (1998) FRA7G extends over a broad region: coincidence of human endogenous retroviral sequences (HERV-H) and small polydispersed circular DNAs (spcDNA) and fragile sites. Oncogene 16:2311–2319

Inoue H, Ishii H, Alder H, Synder E, Druck T, Huebner K, Croce CM (1997) Sequence of the FRA3B common fragile region: implications for the mechanism of FHIT deletion. Proc Natl Acad Sci USA 94:14584–14589

Le Beau MM, Rassool FV, Neilly ME, Espinosa III R, Glover TW, Smith DI, McKeithan TW (1998) Replication of a common fragile site, FRA3B, occurs late in S phase and is delayed further upon induction: implications for the mechanism of fragile site induction. Hum Mol Genet 7:755–761

Mishmar D, Rahat A, Scherer SW, Nyakatura G, Hinzmann B, Kohwi Y, Mandel-Gutfreund Y, et al (1998) Molecular characterization of a common fragile site (FRA7H) on human chromosome 7 by the cloning of an SV40 integration site. Proc Natl Acad Sci USA 95:8141–8146

Sarai A, Mazur J, Nussinov R, Jernigan RL (1989) Sequence dependence of DNA conformational flexibility. Biochemistry 28:7842–7849

Selig S, Okumura K, Ward DC, Cedar H (1992) Delineation of DNA replication time zones by fluorescence in situ hybridization. EMBO J 11:1217–1225

Sutherland GR, Richards RI (1995) The molecular basis of fragile sites in human chromosomes. Curr Opin Genet Dev 5:323–327

Address for correspondence and reprints: Dr. Batsheva Kerem, Department of Genetics, The Hebrew University, Jerusalem, Israel, 91904. E-mail: kerem@leonardo.ls.huji.ac.il

0002-9297/99/6403-0032$02.00

Am. J. Hum. Genet. 64:910–915, 1999

Finite-Sample Properties of Family-Based Association Tests

To the Editor:

During the past few years, there has been much interest in the use of family-based association tests to detect linkage between marker and disease loci, since these methods avoid the problems of ascertaining the appropriate pop-