Am. J. Hum. Genet. 46:492–501, 1990

A Long-Range Restriction Map of the Human Chromosome 19q13 Region: Close Physical Linkage between CKMM and the ERCC1 and ERCC2 Genes

Hubert Smeets,* Linda Bachinski,† Marga Coerwinkel,* Jan Schepens,* Jan Hoeijmakers,‡ Marcel van Duin,‡ Karl-Heinz Grzeschik,§ Christine A. Weber,‖ Pieter de Jong,‖ Michael J. Siciliano,† and Bé Wieringa*

*Department of Human Genetics, University of Nijmegen, Nijmegen, The Netherlands; †Department of Genetics, University of Texas M.D. Anderson Hospital and Tumor Institute, Houston; ‡Department of Cell Biology and Genetics, Erasmus University, Rotterdam; §Institut für Humangenetik, Universität Marburg, Marburg, Federal Republic of Germany; and ‖Biomedical Sciences Division, Lawrence Livermore National Laboratory, Livermore, CA

Summary

We report on the physical ordering of genes in a relatively small area of chromosome 19, segment q13, containing the locus for myotonic dystrophy (DM), the most frequent heritable muscular dystrophy of adulthood in man. DNAs from somatic cell hybrids with der 19q products that carry a breakpoint across the muscle-specific creatine kinase (CKMM) gene were analyzed by Southern blotting using probes for CKMM, APOC2, and the repair genes ERCC1 and ERCC2. Results were combined with data from CHEF and field inversion-gel-electrophoresis separation of large-sized DNA restriction fragments to establish a map localizing both DNA-repair genes and the CKMM gene within the same 250 kb of DNA, the order being cen–CKMM–ERCC2–ERCC1–ter, with APOC2 being at more than 260 kb proximal to CKMM. Transcriptional start sites of the CKMM and DNA-repair genes are all on the telomeric side of the genes. Our results provide a framework for the construction of a larger physical map of the area, which will facilitate the search for the DM gene.

Introduction

Presumably, more than 95% of the genome is now covered in the various genetic linkage maps available for every human chromosome. Adequate genetic maps with polymorphic markers located every 5–10 million bases can be used for localization and indirect diagnosis of a number of human disorders and will assist the construction of a complete physical map of the human genome. The latter goal has come into reach since the development of (1) novel cloning techniques to bridge large-sized DNA fragments (Collins and Weissman 1984; Poustka and Lehrach 1986; Burke et al. 1987), (2) methods for separation of large DNA fragments (Schwartz and Cantor 1984), and (3) new in situ hybridization techniques applicable to cytogenetic and somatic cell genetic studies (Lawrence et al. 1988; Pinkel et al. 1988). Once physical chromosome maps consisting of overlapping collections of cloned DNAs become available for every chromosome, they will greatly facilitate attempts to identify chromosomal regions involved in disease and development. Moreover, they are absolute prerequisites for obtaining a better understanding of the relationship between gene expression and subchromosomal location and for study of the evolutionary relationships between chromosome segments in mammals.

We—in collaboration with several other groups—have been involved in projects aiming at the complete physical and genetic characterization of the human chromosome 19q region around the locus involved in myotonic dystrophy (DM). DM is an autosomal dominant

Received August 15, 1989; revision received October 11, 1989.

Address for correspondence and reprints: Bé Wieringa, Ph.D., Department of Human Genetics, University of Nijmegen, P.O. Box 9101, 6500-HB Nijmegen, The Netherlands.

0002-9297/90/4603-0012$02.00

multisystemic disorder (Harper 1979) and is the most frequent muscular dystrophy in adults, occurring at a frequency of 1/7,500 individuals. As yet, no chromosomal aberrations in this region have been reported to be associated with DM, and no clues to the biochemical defect of the disease exist. We therefore decided to approach the defect by applying reversed genetics strategies. Our studies started with the cloning of several 19q-specific anonymous unique DNAs, dispersed and clustered repeat elements (Hulsebos et al. 1988; Smeets et al. 1989), CpG islands, and expressed genes (H. J. M. Smeets and E. C. M. Mariman, unpublished data). These DNAs were assigned to various intervals across 19q by use of somatic cell hybrids (Stallings et al. 1988; Schonk et al. 1989). Detailed genetic linkage analysis in DM families (Brunner et al. 1989*b*) showed that the markers most closely linked to DM reside in a chromosomal interval estimated to be, at most, 5,000 kb. This spans the area at 19q13.2-q13.3 just distal to the APOE–C1–C2 gene cluster and the muscle-specific creatine kinase (CKMM) gene, which is at the moment the nearest proximal polymorphic marker for DM diagnosis (Brunner et al. 1989*a*; Le Beau et al. 1989). Moreover, it has become evident that the DM gene is situated on a chromosome segment spanning one of the largest evolutionarily conserved mammalian linkage groups known (Lyon 1987; Thompson et al., in press).

As a first attempt to identify the DM gene, we decided to physically characterize the chromosomal area at 19q13.2-q13.3 by using the CKMM gene as a proximal demarcation point. Recent data from somatic cell studies (Stallings et al. 1988) indicated that APOC2, CKMM, and ERCC1 were in close proximity. We therefore concentrated on these genes and on the ERCC2 gene, which has recently also been mapped to this region of chromosome 19q (Mohrenweiser et al., in press; Thompson et al., in press).

Here we present evidence that the CKMM gene and both DNA-repair genes are physically linked within a segment of approximately 250 kb that does not include the APOC2 gene. Evidence for both the relative gene order and the chromosomal orientation of these genes are provided, and implications for further research into DM are discussed.

Material and Methods

DNA Clones

Preliminary data on the cloning of overlapping human CKMM-specific DNAs from a lambda EMBL3 phage library have been published (Mariman et al. 1987) or will be published elsewhere (E. C. M. Mariman, personal communication). Subclones pCKMM-BS2a (insert size 1.5 kb) and pCKMM 3′ (insert size 3 kb [Coerwinkel-Driessen et al. 1988]) contain, respectively, the 5′- and 3′-outermost gene regions and flanking genomic DNA elements. The cloning of the ERCC1-specific cosmid 43–34 and the generation of subclones pT56-4.2 (1.7-kb *Taq*I insert from the 5′ end of the gene) and pBa1.0 (1.0-kb *Bam*HI insert from the 3′ end) have been described in detail elsewhere (van Duin 1988). ERCC1 cDNA probes pE12-12 (900 bp) and pcd4a4, a 3′ end–specific probe of 640 bp, have been described by van Duin (1988; also see refs. therein). Clone pKER2 containing a 2.4-kb *Kpn*I insert from the 3′ end of ERCC2 in vector pKSV10 was derived from the p5T4-1-7 cosmid clone of the ERCC2 gene (Weber et al. 1988). All other DNAs used have been described elsewhere (Wieringa et al. 1988; Schonk et al. 1989).

Cell Lines and Somatic Cell Hybrids

The 908K1 hybrid containing the cen-q13.2 (i.e., proximal) portion of the long arm of chromosome 19q, as well as the GM89A99C-7 hybrid containing the distal q13.2-ter portion of the long arm, have been described elsewhere (Hellkuhl and Grzeschik 1978; Mohandas et al. 1980; Hulsebos et al. 1986). These reciprocal human/hamster and human/mouse hybrids were derived from a patient with a 46,X t(X;19)(q22::q13) karyotype (NIGMS Human Genetic Mutant Cell Repository GM0089). The hybrid 20XP3542-1-4, containing a small segment of 19q, has been described by Stallings et al. (1988). All other hybrids used are described in detail elsewhere (Schonk et al. 1989).

Pulsed-Field Gel Electrophoresis

Field-inversion-gel-electrophoresis (FIGE) separation of large DNA fragments was carried out as described elsewhere (Cremers et al. 1989). CHEF electrophoresis was carried out at 14°C in 50 mM Tris-Borate, 1 mM EDTA in a homemade device using a hexagonal array of electrodes and constructed according to specifications given by Chu et al. (1986). Electric fields were alternated between orientations 120° apart; switch times were 120 s at 150 V and 170 mA. For detailed mapping around the ERCC1 gene, either conventional electrophoresis on 0.5% agarose gels or tranverse alternating-field electrophoresis (TAFE) (Gardiner et al. 1986) was used. DNA from lymphoblastoid cells and somatic cell hybrids was immobilized in agarose blocks (LGT-agarose; Biorad), and restriction-enzyme diges-

tions of agarose-immobilized DNAs were performed overnight with 10–20 U enzyme/agarose block, containing DNA from approximately 0.4 × 10^6 cells, at conditions given by the manufacturer.

Southern Blot Analysis

Southern blot analysis of DNA fragments resolved either on regular agarose gels or on CHEF or FIGE gels was carried out as described elsewhere (Hulsebos et al. 1986, 1988; Cremers et al. 1989). DNAs were immobilized on Gene Screen Plus or BioTrace nylon membranes. Probe insert DNAs were excised from plasmids by appropriate restriction enzymes, separated from vector DNA on 1% (w/v) low-gel-temperature agarose gels, and ^{32}P-labeled by random-primer synthesis (Feinberg and Vogelstein 1983). Prehybridization and hybridization were in 0.5 M NaH_2PO_4/Na_2HPO_4 buffer, 7% (w/v) SDS, and 0.5 mg sonicated herring sperm carrier DNA/ml at 65°C as described elsewhere (Schonk et al. 1989). Washing was at conditions equivalent to a stringency of 0.1 × SSC at a 65°C final temperature. All other procedures were according to standard protocols as given by Maniatis et al. (1982).

Results

Somatic Cell Hybrid Analysis

Hybridization analyses with single-copy probes for CKMM, ERCC1, and ERCC2 genes, both in somatic cell hybrids 908K1, 908K1B18, and GM89A99C-7, containing the reciprocal translocation products of the t(X;19) cell line GM0089, and in various other 19q-specific hybrids (Schonk et al. 1989) are shown in figure 1. With probe pCKMM 3′, originating from the 3′ end of the CKMM gene, a genomic *Eco*RI fragment with a rather unusual size of approximately 40 kb was observed in blood DNAs and in most of the somatic cell hybrids (fig. 1*A*). In the 908K1 and 908K1B18 hybrids (fig. 1*A*, lanes 6 and 10) we identified only a 9.4-kb *Eco*RI fragment. No signal was observed in hybrid GM89A99C-7 containing the distal segment of 19q (fig. 1*A*, lane 12). As shown in figure 1*B*, the reverse situation was seen with probe pCKMM-BS2a, derived from the 5′ end of the gene. With this probe we observed the same large-sized 40-kb *Eco*RI fragment as in figure 1*A* in blood DNA and in those cell lines that displayed the 40-kb fragment with the 3′ probe. A signal of almost similar size was seen in hybrid GM89A99C-7 (fig. 1*B*, lane 12). From hybrids 908K1 and 908K1B18 (fig. 1*B*, lanes 6 and 10) no signal was obtained.

These results indicate that both the chromosome 19 breakpoint in GM0089 and derived hybrids must be situated between the loci recognized by the two CKMM probes. The 5′ probe pCKMM-BS2a apparently detects the larger part of the broken *Eco*RI fragment. This translocation fragment in hybrid GM89A99C-7 cannot be differentiated from the normal 40-kb fragment in control DNAs, owing to the limited resolution of conventional agarose-gel electrophoresis. In addition to these findings, our results establish a 5′-to-3′ orientation of the CKMM transcription unit in the telomere-to-centromere direction on chromosome 19.

Further support for this arrangement was obtained from more detailed restriction-enzyme mapping of cosmids and phages containing CKMM gene inserts. We have determined that the only *Eco*RI site in the CKMM gene area is at approximately 3 kb 3′ downstream of the polyadenylation site. The entire human CKMM gene spans approximately 16 kb. Because hybridization with the pCKMM 3′ probe resulted in an altered signal of only 9.4 kb in all cellular DNAs with the translocation breakpoint, we can conclude that the breakpoint is within 9.4 kb upstream of the unique *Eco*RI site and that it therefore indeed must be situated within the CKMM-transcribed segment. Subsequent experiments (data not shown) involving Southern analyses of various restriction fragments from 908K1 and GM89A99C-7 DNAs by means of individual exon-specific probes allowed us to determine that sequences within CKMM intron 4, just upstream of exon 5, must have been involved in the original reciprocal X;19 translocation event. Probes for exons 2–8 of the CKMM gene were generated by PCR amplification (Saiki et al. 1988) using pairs of oligonucleotide primers which flank the intron-exon junctions. A schematic drawing summarizing our results and indicating the positions of the most important restriction sites in the CKMM gene relative to the X;19 translocation breakpoint is given at the bottom of figure 2.

As shown in figures 1*C* and 1*D*, analysis of our set of cell hybrids with probes specific for the ERCC2 and ERCC1 gene regions yielded a picture similar to that obtained with the 5′ CKMM probe, indicating that both genes are located distal to the translocation breakpoint on 19q. With all enzymes tested, including *Eco*RI, the ERCC1-specific pT56-4.2 probe (fig. 1*D*) and pBa1.0 probes (not shown) gave signals which were clearly different from the signals obtained with the 5′ CKMM probe. The size of the *Eco*RI fragment detected with the ERCC2 pKER2 probe (fig. 1*C*) was similar to the unusually large 40-kb fragment seen with both CKMM

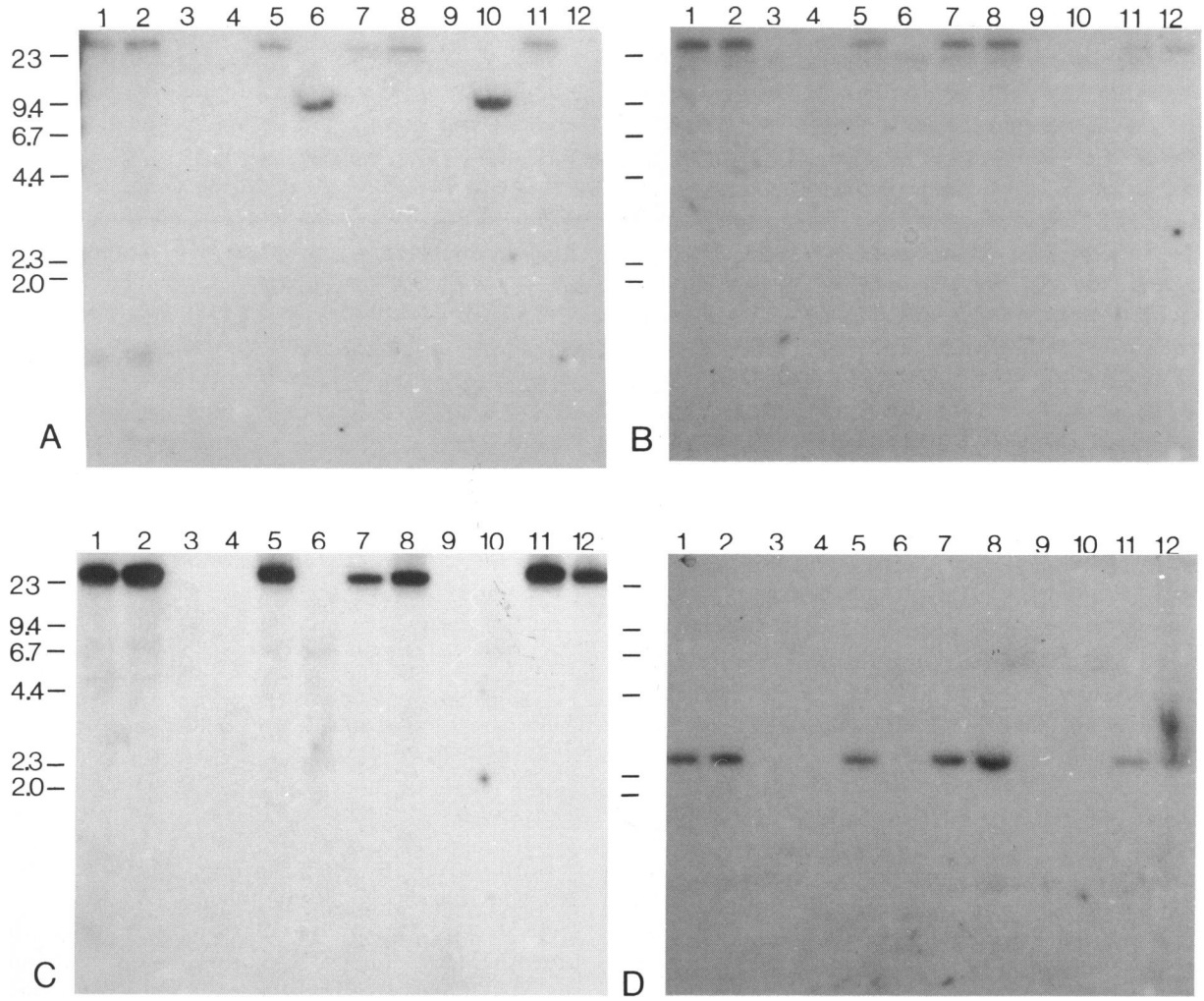

Figure 1 Regional assignment of CKMM, ERCC1, and ERCC2 on 19q by using a panel of somatic cell hybrids. Southern blot hybridization patterns obtained from (*A*) CKMM 3′ (probe pCKMM 3′), (*B*) CKMM 5′ (probe pBS2a), (*C*) ERCC2 (probe pKER2), and (*D*) ERCC1 (probe T56-4.2) DNA probes. Genomic DNAs from blood (lanes 1 and 2), cell line 1023 (hamster only; lanes 3) and hybrids 578 (X, hamster; lanes 4), 1040A6 (19q; lanes 5), 908K1 (rearranged 19p + 19cen-q13.2; lanes 6), ORIM 7.1 (19q13.1-qter; lanes 7), 1219G2 (19q1.1-qter; lanes 8), 908K1A1 (rearranged 19p; lanes 9), 908K1B18 (19cen-q13.2; lanes 10), 20XP3542-1-4 (19q13.2-q13.3; lanes 11), and GM89A99C-7 (19q13.2-qter; lanes 12) were digested with *Eco*RI, resolved on 0.8% agarose gels, blotted, and hybridized. Cell lines not further specified in Material and Methods are as described by Schonk et al. (1989).

probes, suggesting that CKMM and ERCC2 are on the same fragment.

Previously we had found that the APOC2 gene region, which is genetically closely linked to the DM and CKMM loci (Brunner et al. 1989*b*), is physically located just proximal to the breakpoint in cell hybrid 908K1 (Stallings et al. 1988; Schonk et al. 1989). From conventional Southern analyses of the various somatic cell hybrids no further proof for possible overlap between restriction fragments specific for APOC2 and CKMM or ERCC1/2 was obtained (data not shown). We therefore decided to test this issue further by employing long-range restriction-site mapping.

FIGE and CHEF Analysis

Inspection of the restriction-enzyme maps and of the partial nucleotide sequences generated from the ERCC1, ERCC2, and CKMM gene areas published earlier (Mariman et al. 1987, and unpublished data; van Duin 1988 [and refs. therein]; Weber et al. 1988) revealed that cleavage sites for restriction enzymes *Sal*I, *Xho*I, and *Cla*I were rarely present. The presence of

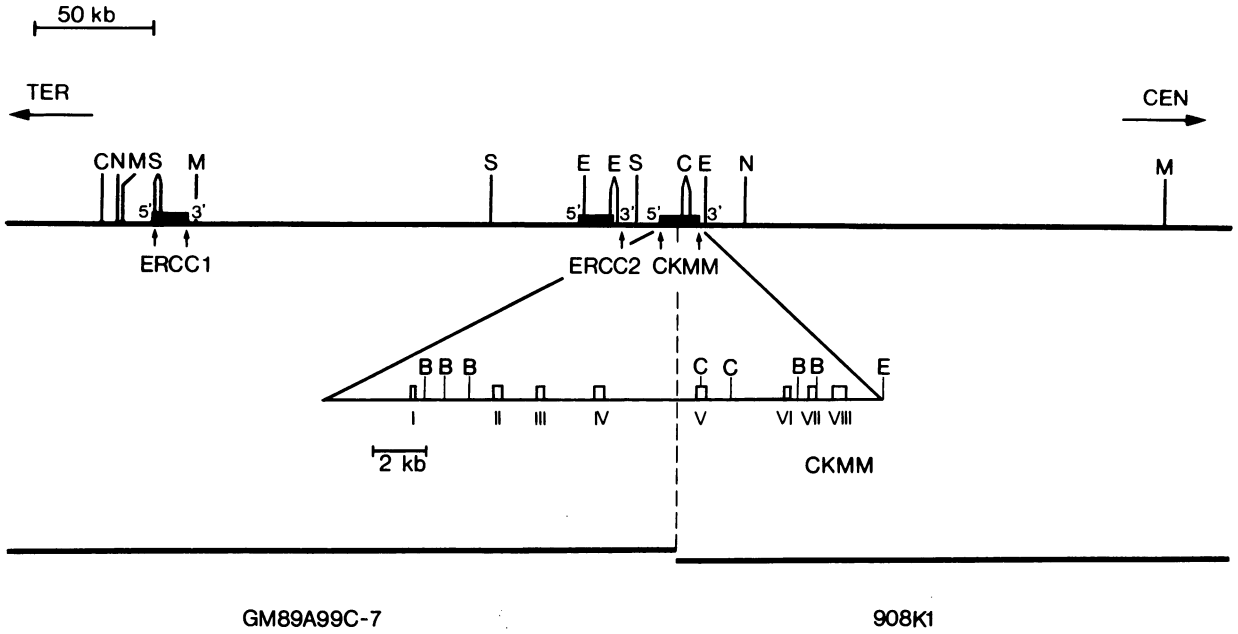

Figure 2 CHEF and FIGE map of the CKMM, ERCC2, and ERCC1 gene area. The genes are represented by filled boxes, and the chromosomal orientation of the cluster is indicated. The location of the probes used is shown by arrows. A detailed map of the CKMM region showing both the restriction sites relevant for Southern analysis and the location of the GM0089 breakpoint in the CKMM gene is given at the bottom. Exons are indicated by Roman numerals. C = *Cla*I; N = *Not*I; M = *Mlu*I; S = *Sfi*I (ERCC1 contains three *Sfi*I sites; see detailed map in fig. 5*b*); E = *Eco*RI (not all sites shown); B = *Bam*HI.

sites for rare-cutting enzymes *Sfi*I, *Not*I, and *Mlu*I was tested directly by appropriate digestions of cosmid (ERCC1 clone 43-34 [Westerveld et al. 1984]; ERCC1 cosmid clones F8697 and F15123 [de Jong et al. 1989]), phage (CKMM phage lambda 3-1 and overlapping phages; ERCC1 and -2 phages [H. J. M. Smeets, unpublished data]), or plasmid (pKER2; ERCC2 [Weber et al. 1988]) clones. Digestion with *Not*I revealed that this enzyme does not cleave any of the DNAs. Two *Sfi*I sites were detected near and within the first exon of the ERCC1 gene, and another site was detected within the third exon. An *Mlu*I site was found in the region immediately flanking the 3′ end of this gene.

Figure 3 shows the results from resolving, on CHEF and FIGE gels, *Not*I digests of genomic DNAs isolated from Epstein-Barr virus–immortalized lymphoblastoid cells of three unrelated individuals. On a CHEF blot, the ERCC1-specific probe pT56-4.2 (fig. 3*b*), as well as the ERCC2 (fig. 3*c*) probe and pCKMM 3′ probe (fig. 3*a*) identify an identical 250-kb band. Though, on FIGE gels, the *Not*I fragment had migrated to a position corresponding to a somewhat different molecular weight of 300 kb, again complete identity was found with all three probes (only two are shown [figs. 3*d* and 3*e*]). We consider this to be strong evidence that all three genes are physically closely linked.

As shown in the upper half of figure 4, with *Mlu*I on both total genomic DNAs as well as on DNA from the 20XP3542-1-4 cell hybrid containing the pertinent 19q area, an identical 400-kb fragment was identified with the 5′ and 3′ CKMM probes and with ERCC2 probe pKER2. The fragment identified with ERCC1 probe pT56-4.2 clearly had a much smaller size of approximately 30–40 kb. With *Sfi*I (fig. 4, lower half) no similarity was observed between CKMM-, ERCC1-, or ERCC2-specific fragments with sizes ranging between 60 and 200 kb. Partial *Cla*I digestion generated identical fragments of 200 kb and 260 kb at rather different intensity, for both ERCC probe pT56-4.2 and ERCC2 probe pKER2. When the same blot was hybridized with the pCKMM 3′ probe, only a band of 260 kb was observed (fig. 4, lower half). With the 5′ CKMM probe pBS2a the same 200-kb and 260-kb fragments were detected as with the ERCC probes, confirming that all genes are closely adjacent. The results seen for the partial digest can be explained by incomplete cleavage at the *Cla*I site, which separates the 5′ and 3′ ends of the CKMM gene and is positioned in exon V (see fig. 2).

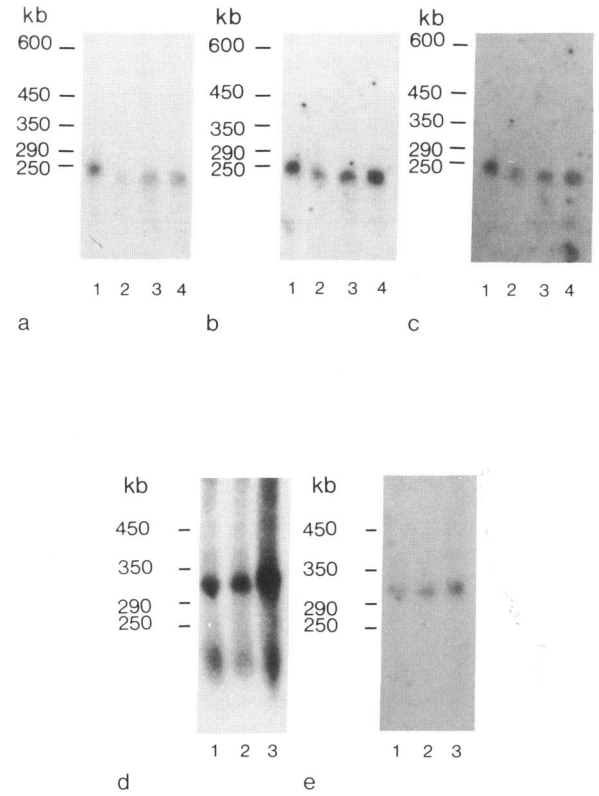

Figure 3 CHEF and FIGE analysis of *Not*I digestions of genomic DNAs. The probes CKMM 3′ (panels *a* and *d*), ERCC1 pT56-4.2 (panels *b* and *e*), and ERCC2 pKER2 (panel *c*) were hybridized to blots of genomic DNAs digested with *Not*I and resolved by CHEF electrophoresis (panels *a–c*) or FIGE (panels *d–e*). Lanes 1 and 2 contain DNA of the same individual, but at different concentrations corresponding to the use of 0.4 × 10^6 and 0.2 × 10^6 EBV-immortalized lymphoblastoid cells, respectively. Lanes 3 and 4 contain DNA from 0.4 × 10^6 cells of two other unrelated individuals.

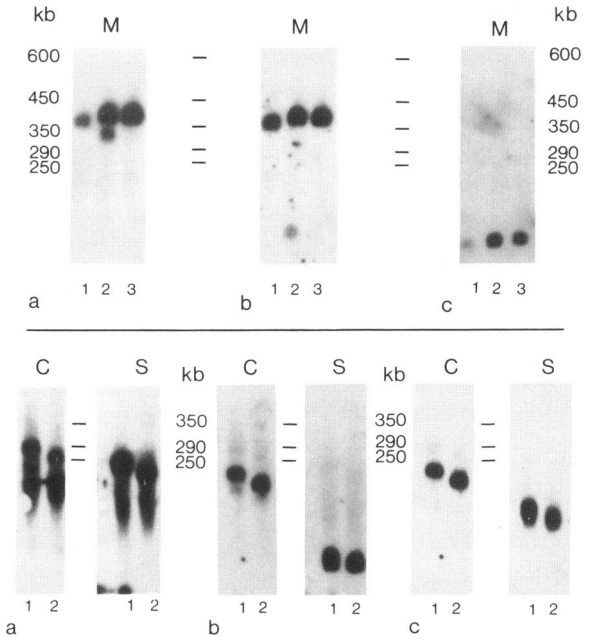

Figure 4 CHEF and FIGE analysis of *Mlu*I (M), *Cla*I (C) and *Sfi*I (S) digestions of cell hybrid and genomic DNAs. Hybridization patterns of CKMM 3′ (panel *a*), pKER2 (ERRC2; panel *b*), and T56-4.2 (ERCC1; panel *c*), on a CHEF blot of *Mlu*I(M)-digested DNAs (upper half of fig.), derived from the somatic cell hybrid 20XP3542-1-4 (lane 1) and from lymphoblastoid cell lines of two unrelated individuals (lanes 2 and 3), and on a FIGE blot of *Cla*I(C)- and *Sfi*I (S)-digested DNA (lower half of fig.) from the two individuals.

An additional—presumably polymorphic—*Cla*I site is present at 1,800 bp downstream of exon V (E. P. M. van Kerkhoff, personal communication). Whether this latter *Cla*I site is present in the DNAs tested is not known at the moment. Furthermore, our results are compatible with the fact that no other *Cla*I sites were identified either in any of the CKMM-specific phages with inserts extending 5′ and 3′ of the transcribed region or in the regions cloned from the ERCC1 and -2 loci.

The different findings for the 5′ and 3′ CKMM probes corroborate the results obtained from the t(X;19) hybrid analysis in that the 5′ end of the CKMM gene is oriented toward the DNA-repair genes. The similarities between fragments detected with the 5′ probe of CKMM and fragments detected with the ERCC2 probe pKER2 are in agreement with the gene order CKMM–ERCC2–ERCC1. A map summarizing the results for the entire area is shown in figure 2. It should, however, be noted that sizes of various restriction fragments could not be determined very precisely because of the conspicuous difference in mobility between fragments resolved on the CHEF system versus the FIGE system. The overall arrangement, however, is clearly correct, as also with enzymes such as *Sal*I, *Xho*I, and *Pvu*I (data not shown), evidence for (partial) identity of fragments bearing CKMM or ERCC sequences was obtained. For most enzymes the fragments migrated at the lower zone of the resolution range, indicating the presence, across this area of proximal 19q13.3, of multiple cleavage sites for rare-cutting enzymes.

In order to determine more precisely both the position of these sites around the ERCC1 gene and the orientation of the transcribed region with respect to the ERCC2 and CKMM genes, we performed double digestions with either *Not*I or *Mlu*I and several restriction enzymes (i.e., *Kpn*I, *Sal*I, *Sfi*I, and *Xho*I [van Duin 1988]) that cleave only rarely within the ERCC1 gene and analyzed the digests on conventional agarose gels.

In this way we could position the *Not*I site flanking the common 250-kb fragment close to an *Mlu*I site at 17 kb upstream of the transcriptional start site, establishing a ter–5′ 3′–cen orientation for ERCC1. Another *Mlu*I site was mapped just downstream of ERCC1.

To further investigate the possibility of a CpG island in the 5′ region of ERCC1, we took advantage of a large, approximately 20-kbp *Hin*dIII fragment that overlaps the upstream area and almost the entire gene (van Duin 1988).

Double digests were done with *Hin*dIII and a battery of rare-cutting enzymes such as *Sfi*I, *Sac*II, *Mlu*I, *Nru*I, and *Not*I. The band shifts seen when these digests were compared with the *Hin*dIII digest revealed presence of *Sac*II and *Sfi*I sites (fig. *5a*). From sequencing data and cosmid mapping (M. van Duin, unpublished data) we know that, for *Sac*II, two sites closely adjacent to each other and to the transcriptional start site are present. The resultant map for the region around ERCC1, showing a clustering of rare sites at the 5′ region, is shown in figure *5b*.

With none of the enzymes tested was any overlap seen between the APOC2-hybridizing and CKMM-, ERCC1, or ERCC2-hybridizing fragments. These results demonstrate that the APOC2 gene region is at least 260 kb (i.e., the size of the CKMM 3′ *Cla*I fragment) proximal to the 3′ end of the CKMM gene. Since probes specific for the APOC2 gene recognize a 1,000-kb *Not*I fragment (not shown), we estimate that a substantially larger chromosome segment may separate these loci.

Discussion

By Southern analysis of DNAs from somatic cell hy-

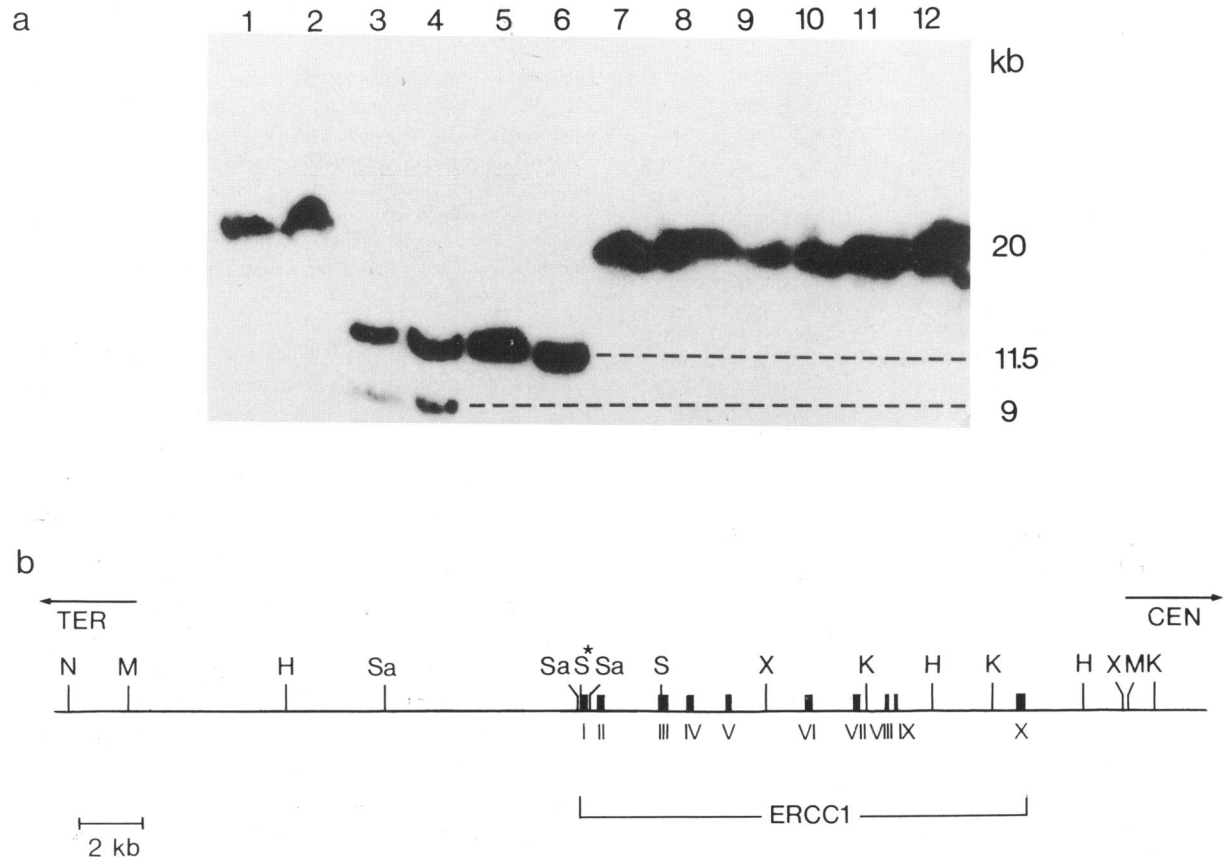

Figure 5 Localization of a CpG island in the vicinity of the ERCC1 gene. *a*, Southern blot probed with the cDNA probe pE12-12. Odd-numbered lanes contain DNA from the somatic cell hybrid 20XP3542-1-4. Even-numbered lanes contain Hela DNA. All lanes were digested with *Hin*dIII. Lanes 3–13 were redigested with *Sfi*I (lanes 3 and 4), *Sac*II (lanes 5 and 6), *Mlu*I (lanes 7 and 8), *Nru*I (lanes 9 and 10), or *Not*I (lanes 11 and 12). *b*, Restriction map in the region of the ERCC1 gene. Exons are represented by filled boxes (Roman numerals), and the chromosomal orientation is indicated. N = *Not*I; M = *Mlu*I; H = *Hin*dIII; Sa = *Sac*II; X = *Xho*I; K = *Kpn*I; S = *Sfi*I; S* = two *Sfi*I sites only 32 nucleotides apart.

brids as well as by FIGE and CHEF analysis of genomic human DNAs, we have shown that the CKMM gene and the ERCC1 and ERCC2 repair genes are all located within a relatively small (250–300 kb) segment of 19q13. Thus far—perhaps because of the low mutation frequency and the high evolutionary conservation of the repair-gene region (van Duin 1988; Thompson et al., in press)—RFLPs have not been found for either ERCC gene or in the immediate vicinity of the gene loci (R. Hermens, personal communication). This precluded a reliable estimation of the genetic distance between these loci and CKMM or DM. It was both (*a*) the serendipitous finding of similarity between unusually large 40-kb *Eco*RI fragments recognized by CKMM- and ERCC2-specific probes and (*b*) the assignment of the CKMM, ERCC1, and APOC2 genes to a relatively small area of 19q in somatic cell hybrid 20XP3542-1-4 (Stallings et al. 1988) that initiated the studies presented here.

By identification of a translocation breakpoint across the fourth intron of the CKMM gene, we have established a fixed demarcation point that may serve to position all other loci in the DM-relevant area. Combining all our results yields the following absolute gene order: cen–APOC2–CKMM–ERCC2–ERCC1–ter. The APOC2 gene maps proximal to the t(X;19) breakpoint across the CKMM gene, and both repair genes map distal to that breakpoint. Similarities between fragments determined by the CKMM and ERCC2 probes (i.e., *Eco*RI-, *Mlu*I-, and *Sal*I-generated fragments), which were not detected by ERCC1, establishes the location of ERCC2 between CKMM and ERCC1. The cen–3′ 5′–ter orientation of the CKMM gene was determined from analysis of derivatives of the GM0089 cell line. The orientation of the ERCC2 gene was inferred from the positions of *Eco*RI restriction sites in the CKMM and ERCC2 genes, as indicated in figure 2. Restriction-site mapping of the CKMM gene has revealed the presence of an *Eco*RI cleavage site at about 1.5 kb downstream from the CKMM polyadenylation site. In the ERCC2 gene no less than three *Eco*RI sites, all 5′ upstream of the region recognized by probe pKER2, have been demonstrated (see Weber et al. 1988, fig. 2). This determines a tandem arrangement of cen–3′ CKMM 5′–ERCC2 5′–ter.

The physical distance between the *Eco*RI sites in the CKMM and ERCC2 genes–i.e., the length of the common hybridizing fragment—was roughly estimated to be 40 kb. This means that the cap site of the CKMM gene and the polyadenylation site of the ERCC2 gene are no more than 25 kb apart. Attempts to confirm this result by isolation of cosmid clones that overlap both genes are in progress.

The orientation of the ERCC1 gene can be inferred from the location of the *Not*I site in its 5′ region, which flanks the 250-kb fragment. This means that the ERCC2 and CKMM genes are in the 3′ region of ERCC1, establishing a ter–5′ ERCC1 3′–cen orientation. The distance separating the ERCC1 and CKMM genes can be calculated from the fragment sizes given in figure 2 and is currently estimated to be a maximum of 250 kb (i.e., the size of the common *Not*I fragments) and a minimum of about 200 kb (i.e., the size of the *Sfi*I fragment specific for ERCC2 plus the *Sfi*I fragment detected by the ERCC1 3′ probe Ba1.0).

It appears likely that there is a CpG island at the 5′ end of ERCC1. The presence of recognition sites for enzymes such as *Not*I, *Mlu*I, *Sac*II, and *Sfi*I in such close proximity are consistent with the definition of such islands. In addition, doing the same analysis with cDNA from ERCC2 and CKMM, which also reveal large *Hin*dIII fragments, shows no deviation from the normal *Hin*dIII pattern, suggesting the absence of such an island for these two genes (data not shown). Whether (*a*) the presence of CpG islands 5′ of these closely spaced genes and (*b*) their common head-to-tail arrangement are relevant for their regulation is an interesting question that remains to be resolved. Transcription interference phenomena have been observed for gene pairs with opposite transcription directions (Adhya and Gottesman 1982; Cullen et al. 1984; Kadesch and Berg 1986) and may have played a role in the selection of the most appropriate orientation during evolution of this gene cluster.

The physical distance between CKMM and APOC2 could not be determined but is at least 260 kb. On the basis of genetic studies, these genes are considered to be extremely tightly linked at a recombination fraction of .01–.02 and a lod score (*Z*) of more than 140 (Le Beau et al. 1989). This indicates either that any straightforward comparison between genetic and physical distances in the q13 area is impossible or, a more trivial possibility, that the distance separating both genes simply cannot be bridged because too many nonoverlapping small-sized restriction fragments are generated from this area. A map of the entire CKMM–ERCC2–ERCC1 region, summarizing our results, is depicted in figure 2.

From linkage studies in DM families it has become evident that the mutation(s) underlying DM are located distal to APOC2 and CKMM and proximal to the loci recognized by probe pEFD4.2 (D19S22; Nakamura et al. 1988; Brunner et al. 1989*b*) and probe pEWRB1

(MacKenzie et al. 1989). These loci are at 13 cM (Z_{max} > 10) and at 9 cM (Z_{max} = 15.45) from DM, respectively. Between CKMM and DM, as yet only three confirmed crossovers have been found in affected family members, in about 300 meioses tested (Brunner et al. 1989*a*; Le Beau et al. 1989). We consider this strong evidence for any DM candidate gene to be situated within a relatively small segment of 19q13, telomeric from CKMM. Candidate genes from this area—such as the antisense transcribed sequence, which overlaps the ERCC1 3′ region and is located between ERCC1 and ERCC2 (van Duin 1988), or various DNA elements associated with CpG-islands—have to be tested by performing detailed comparisons between DNAs from DM patients and normal individuals.

Somatic cell hybrids—including hybrid 20XP3542-1-4 (Stallings et al. 1988), which contains the greater part of the segment of interest, and hybrids GM8A89A99C-7 (Mohandas et al. 1980) and 908K1 (Hulsebos et al. 1986), which contain in the reciprocal 19;X translocation products with the breakpoint in the CKMM gene—will prove particularly valuable as starting points or reference material for attempts to approach the DM gene by molecular cloning. The CKMM-ERCC area characterized here should serve not only as an ideal starting point to further complete the physical map of 19q13 but also as an easily distinguishable landmark in genetic studies in DM families once new markers are indentified.

Acknowledgments

This work was supported by the Dutch Praeventiefonds grant 28–1165, by Deutsche Forschungsgemeinschaft grant Ro 389/15-2 to H. H. Ropers, by Prinses Beatrixfonds grant 87-2694, by NIH grant CA 34936, by a personal gift from Kenneth D. Muller, by the American Muscular Dystrophy Association, and by the Piton Foundation. Work by C.A.W. and P.D.J. was performed at Lawrence Livermore National Laboratory under the auspices of the U.S. Department of Energy under contract W-7405-ENG-48.

References

Adhya S, Gottesman M (1982) Promoter occlusion: transcription through a promoter may inhibit its activity. Cell 29: 939–944

Brunner HG, Korneluk RG, Coerwinkel-Driessen M, MacKenzie A, Smeets H, Lambermon HMM, van Oost BA, et al (1989*a*) Myotonic dystrophy is closely linked to the gene for muscle-type creatine kinase (CKMM). Hum Genet 81:308–310

Brunner HG, Smeets H, Lambermon HHM, Coerwinkel-Driessen M, van Oost BA, Wieringa B, Ropers HH (1989*b*) A multipoint linkage map around the locus for myotonic dystrophy on chromosome 19. Genomics 5:589–595

Burke DT, Carle GF, Olson MV (1987) Cloning of large segments of exogeneous DNA into yeast by means of artificial chromosome vectors. Science 236:806–812

Chu G, Vollrath D, Davis RW (1986) Separation of large DNA molecules by contour clamped homogeneous electric fields. Science 234:1582–1585

Coerwinkel-Driessen M, Schepens J, van Zandvoort P, van Oost B, Mariman E, Wieringa B (1988) NcoI RFLP at the creatine kinase-muscle type gene locus (CKMM, chromosome 19). Nucleic Acids Res 16:8743

Collins FS, Weissman SM (1984) Directional cloning of DNA fragments at a large distance from an initial probe: a circularization method. Proc Natl Acad Sci USA 81:6812–6816

Cremers FPM, van de Pol DJR, Diergaarde PJ, Wieringa B, Nussbaum RL, Schwartz M, Ropers HH (1989) Physical fine mapping of the choroideremia locus using Xq21 deletions associated with complex syndromes. Genomics 4: 41–46

Cullen BR, Lomedico PT, Ju G (1984) Transcriptional interference in avian retroviruses—implications for the promoter insertion model of leukaemogenesis. Nature 307:241–245

Duin M van (1988) Cloning and characterization of the human DNA-excision repair gene ERCC-1. Thesis, University of Rotterdam

Feinberg AP, Vogelstein B (1983) A technique for radiolabeling DNA restriction endonuclease fragments to high specific activity. Anal Biochem 132:6–13

Gardiner K, Laas W, Patterson D (1986) Fractionation of large mammalian DNA restriction fragments using vertical pulsed field gradient gel electrophoresis. Somatic Cell Mol Genet 12:185–195

Harper PS (1979) Myotonic dystrophy. WB Saunders, Philadelphia

Hellkuhl B, Grzeschik KH (1978) Partial reactivation of a human inactive X chromosome in human-mouse somatic cell hybrids. Cytogenet Cell Genet 22:527–530

Hulsebos T, Schonk D, van Dalen I, Coerwinkel-Driessen M, Schepens J, Ropers HH, Wieringa B (1988) Isolation and characterization of alphoid DNA sequences specific for the pericentric regions of chromosomes 4, 5, 9, and 19. Cytogenet Cell Genet 47:144–148

Hulsebos T, Wieringa B, Hochstenbach R, Smeets D, Schepens J, Oerlemans F, Zimmer J, et al (1986) Toward early diagnosis of myotonic dystrophy: construction and characterization of a somatic cell hybrid with a single human der(19) chromosome. Cytogenet Cell Genet 43:47–56

Jong PJ de, Yokobata K, Chen C, Lohman F, Pederson L, McNinch J, Van Dilla M (1989) Human chromosome-specific partial digest libraries in lambda and cosmid vectors. Cytogenet Cell Genet 51:985

Kadesch T, Berg P (1986) Effects of the position of the simian virus 40 enhancer on expression of multiple transcription units in a single plasmid. Mol Cell Biol 6:2593–2601

Lawrence JB, Villnave CA, Singer RH (1988) Sensitive, high-resolution chromatin and chromosome mapping in situ: presence and orientation of two closely integrated copies of EBV in a lymphoma line. Cell 52:51–61

Le Beau MM, Ryan D Jr, Pericak-Vance MA (1989) Report of the Committee on the Genetic Constitution of Chromosomes 18 and 19. Cytogenet Cell Genet 51:338–358

Lyon MF (1987) Mouse chromosome atlas. Mouse News Lett 787:12–38

MacKenzie AE, Nakamura Y, Dubé ID, Korneluk (1989) A seven locus chromosome 19 linkage group encompassing the myotonic dystrophy locus. Cytogenet Cell Genet 51:1036

Maniatis T, Fritsch EF, Sambrook J (eds) (1982) Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, NY

Mariman ECM, Broers CAM, Claesen CAA, Tesser GI, Wieringa B (1987) Structure and expression of the human creatine kinase B gene. Genomics 1:126–137

Mohandas T, Sparkes RS, Hellkuhl B, Grzeschik KH, Shapiro LJ (1980) Expression of an X-linked gene from an inactive human X chromosome in mouse-human hybrid cells: further evidence for the noninactivation of the steroid sulfatase locus in man. Proc Natl Acad Sci USA 77:6759–6763

Mohrenweiser HW, Carrano AV, Fertitta A, Perry B, Thompson LH, Tucker JD, Weber CA. Refined mapping of the three DNA repair genes, ERCC1, ERCC2 and XRCC1, on human chromosome 19. Cytogenet Cell Genet (in press)

Nakamura Y, Lathrop M, O'Connell P, Leppert M, Lalouel JM, White R (1988) A primary map of ten DNA markers and two serological markers for human chromosome 19. Genomics 3:67–71

Pinkel D, Landegent J, Collins C, Fuscoe J, Segraves R, Lucas J, Gray J (1988) Fluorescence in situ hybridization with human chromosome-specific libraries: detection of trisomy 21 and translocations of chromosome 4. Proc Natl Acad Sci USA 85:9138–9142

Poustka A, Lehrach H (1986) Jumping libraries and linking libraries: the next generation of molecular tools in mammalian genetics. Trends Genet 2:174–179

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, et al (1988) Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239:487–491

Schonk D, Coerwinkel-Driessen M, van Dalen I, Oerlemans F, Smeets B, Schepens J, Hulsebos T, et al (1989) Definition of subchromosomal intervals around the myotonic dystrophy gene region at 19q. Genomics 4:384–396

Schwartz DC, Cantor CL (1984) Separation of yeast chromosome sized DNAs by pulsed field gel electrophoresis. Cell 37:67–75

Smeets HJM, Brunner HG, Ropers HH, Wieringa B (1989) Use of variable simple sequence motifs as genetic markers: application to study of myotonic dystrophy. Hum Genet 83:245–251

Stallings RL, Olson E, Strauss AW, Thompson LH, Bachinski LL, Siciliano MJ (1988) Human creatine kinase genes on chromosome 15 and 19, and proximity of the gene for the muscle form to the genes for apolipoprotein C2 and excision repair. Am J Hum Genet 43:144–151

Thompson LH, Bachinski LL, Stallings RL, Doif G, Weber C, Westerveld A, Siciliano MJ. Complementation of repair gene mutations on the hemizygous chromosome 9 in CHO: a third repair gene on human chromosome 19. Genomics (in press)

Weber CA, Salazar EP, Stewart SA, Thompson LH (1988) Molecular cloning and biological characterization of a human gene, ERCC2, that corrects the nucleotide excision repair defect in CHO UV5 cells. Mol Cell Biol 8:1137–1146

Westerveld A, Hoeijmakers JHJ, van Duin M, de Wit J, Odijk H, Pastink A, Wood RD, et al (1984) Molecular cloning of a human DNA repair gene. Nature 310:425–429

Wieringa B, Brunner H, Hulsebos T, Schonk D, Ropers HH (1988) Genetic and physical demarcation of the locus for dystrophia myotonica. Adv Neurol 48:47–49