DOE/ER/60857--3

DOE/ER/60857-3 DE92 003412

HUMAN CHROMOSOME 21: MAPPING OF THE CHROMOSOMES AND CLONING OF cDNAs

Progress Report

for Period June 15, 1989 - August 31, 1991

Stylianos E. Antonarakis

Johns Hopkins University School of Medicine
Baltimore, Maryland 21205

September 1991

Prepared for

THE U.S. DEPARTMENT OF ENERGY
AGREEMENT NO. DE-FG02-89ER60857

ABSTRACT

The objective of the research funded by DOE grant DE-FG02-89ER60857 from 6/15/89 to 8/31/91 was to contribute to the physical mapping of human chromosome 21 (HC21) by cloning large fragments of DNA into Yeast Artificial Chromosomes (YACs) and identify YACs that map on HC21.

A total of 54 sequence tagged sites (STS) have been developed and mapped in our laboratory to HC21 and can be used as initial reference points for YAC identification and construction of overlapping clones. A small YAC library was constructed which is HC21 specific. DNA from somatic cell hybrid WAV17 or from flow-sorted HC21 was partially digested with EcoRI, ligated into vectors PUS97, PUS98, and YACs have been obtained with average size insert of more than 300 kb. This library has been deposited in D. Patteron's lab for the Joint YAC screening effort. Additional YAC libraries from ICI Pharmaceuticals or from Los Alamos National Laboratories have been screened with several STS and positive YACs have been identified.

Work in progress includes screening of YAC libraries in order to construct overlapping clones, characterization of the cloning ends of YACs, characterization of additional STS and cloning of HC21 specific cDNAs.

NOTICE

PROGRESS REPORT

The results described in this section have been obtained using funds from both the NIH and DOE. The DOE grant DE-FG02-89ER60857 mainly supported the efforts concerning YAC cloning and screening. However, it is difficult to separate the DOE from the NIH funded results since the various aspects of the "genomics" project of HC21 are interrelated. I will briefly describe the status of our data collection to date and our experience in generating and mapping STS, constructing YAC libraries and cloning methods for cDNA's.

1. **STS on HC21**
   A number of STS (sequenced tagged sites - see Olson et al. Science 1989) have been developed in our laboratory that map on HC21. These STS give a single amplification product and most of them define a polymorphic locus due to a short sequence repeat. Other STS are for well mapped loci on HC21 and for the 3' untranslated regions of cDNAs that map on HC21. Table 1 lists the STS developed and used in our laboratory (STS #9 was developed by R. Myers; #12 and 14 by M. Litt; #11 by C. vanBroekhoven)

2. **The HC21 YAC library from Johns Hopkins**
   A small YAC library specific for chromosome 21 has been constructed in our laboratories by Dr. M. C. McCormick during her graduate studies years. The starting DNA material was from hybrid WAV17 (which contains HC21 as the only human chromosome) and from flow-sorted chromosome 21 kindly supplied by Dr. Y.W. Kan. Description of the library, the methodology used, the vectors used, the advantages of the vectors can be found in publications in the appendix (McCormick et al. 1989, 1990, Shero et al. 1991, Hieter et al. 1990). The vectors used were constructed in Dr. Hieter's lab and facilitate cloning of the YAC ends into plasmids in E. Coli. All manipulations of DNA were performed in agarose. Polyamines have been used to increase the size of the inserts (Connelly et al. 1990). The vectors offer more selectable markers for YAC manipulations (see Pavan et al. 1990, Reeves et al. 1990).

The small library obtained after screening for human DNA insert sequences consists of 79 YACs with average insert size of 330 kb in the YACs from WAV17 and 190 kb in the YACs from flow-sorted HC21. A list of these YACs is provided below. Some YACs have been mapped to chromosome 21 using a series of somatic cell hybrids of HC21 kindly provided by D. Patterson. These hybrid cell lines divide HC21q into several intervals in which DNA probes can be localized. The somatic cell hybrids available to our laboratory are: WAV17, R2-10, t8;21, t21;8, t1;21, t21;1, t4;21, ACEM, 2FUR1(for description see Gardiner et al. 1990). A number of YAC ends have been cloned in E. Coli and a subset of those have been subjected to nucleotide sequencing for generation of YAC end STS. (See table 2).

Another subset of these YACs have been used as probes in FISH (fluorescent in situ hybridization) (Lichter et al. 1990) in collaboration with Dr. W. Kearns at Johns Hopkins and mapped to HC21 (see table 2, column FISH).

3. **Screening of YAC libraries with STS**

   3.1 After the construction and initial characterization of the Johns Hopkins HC21 YAC library, Dr. M.K. McCormick moved to the Los

Alamos National Laboratory. The decision was made to continue to work together towards the completion of a YAC physical map of HC21 ie. the production of a large contig from overlapping YAC clones. Dr. McCormick has taken the task of creating YAC libraries from flow-sorted HC21 since the facilities at Los Alamos were ideal for such experiments. Three complete digestion libraries were constructed (a) A ClaI complete digestion into vectors pJS97, pJS98. A total of 200 clones were obtained with average size of 200 kb. About 50% of the clones were positive for human sequences. (b) A Eag I complete digestion and (c) a NotI-NheI, complete digestion into the same vectors. From each of the last two libraries about 1500 clones have been obtained and about 50% of those had human DNA inserts. The average size was again about 300 kb. The construction of these libraries has been reported in the CSH 1991 meeting (McCormick et al. 1991). Pools from these libraries have been made and DNA from those pools have been screened in our laboratory with STS for the identification of YACs. The pools have been made from rows and columns of 96-microtiter well plates and from whole plates, so that in general no secondary screening is necessary for the identification of an individual YAC. The screening with STS was performed in my laboratory by Dr. M. Kalaitsidaki. The results of the ongoing screening of these libraries are given in the next table 3A. Unfortunately, only 25% of the STS used give positive results.

3.2 A collaborative agreement has been made with scientists from ICI Pharmaceutical to screen their total genomic YAC library for HC21 inserts. Dr. Anand and co-workers have constructed a 3.5X coverage total human genomic YAC library in vector pYAC4 using EcoRI partial digestion (Anand et al. 1990, Anand et al. 1991)). The co-cloning events in this library are not more than 10% and the average size of inserts is about 370 kb. For these reasons, we decided to screen this library as opposed to the CEPH YAC library for HC21. (Please remember that the St. Louis library is screened by the collaborative joint YAC screening effort in Dr. D. Patterson laboratory; please also note that all the YACs from the initial Johns Hopkins HC21 YAC library have been given to the Denver lab for screening). A total of 40 master pools of YACs, each one representing the DNA from 9 96-well microtiter plates (864 clones) have been given to our laboratory for STS screening. The agreement with ICI is that after the initial identification of positive pools, an investigator from our lab will isolate individual YACs in the ICI labs in England. The results of the ongoing screening of these YACs by Dr. M. Kalaitsidaki is also given in table 3B. It is encouraging that for about 65% of the STS used, a positive pool has been identified.

3.3 The original Johns Hopkins HC21 YAC library has been screened with several STS. One YAC has been identified with D21S156 (by PCR) and several with alphoid sequences (by hybridization). The D21S156 positive YAC along with 2 other YACs from the St. Louis libraries which are positive for ETS2 and D21S156 comprise the first contig of 3 YACs which covers about 700 kbs of DNA. The

latter two YACs have been screened for the appropriate STS in Denver and supplied to our laboratory as part of the collaborative effort.

4. **DNA Polymorphisms due to GT dinucleotide and other short sequence repeats.**

The introduction of PCR amplification in the studies of variation of the human genome revealed a wealth of new polymorphic markers due to short sequence repeats (SSR)(Tautz 1989, Weber et al. 1989, Litt et al. 1989). These can be 2, 3, 4 or more nucleotide repeats or poly A tracts(Economou et al. 1990, Edwards et al. 1991). The most well studied SSR are those due to GT dinucleotide repeats. (Weber, et al. 1990)). We searched the GENBANK for all 2, 3 and 4 nt repeats and found that the most common SSR are the following; $(AC)_n$, $(AG)_n$, $(AGC)_n$, $(CGC)_n$ $(ACA)_n$, $(ACAA)_n$, $(AGAA)_n$, $(AGGA)_n$, $(ATTT)_n$. (See figure 1) Therefore, we have begun a systematic search for HC21 SSRs that will provide both excellent markers for the linkage map and STS for the physical map. The HC21 specific flow-sorted DNA library in plasmid PBS (Fuscoe et al. 1989) was used as the starting material. This library was constructed in Joe Gray laboratory and kindly provided to us. The library which has average size Hind III inserts of 2 kb was screened by Dr. A. Warren with oligonucleotide $(GT)_{10}$ and subsequently with $(ATTT)_5$, $(CGG)_7$. Positive clones were subjected to nucleotide sequencing surrounding the repetitive element, oligonucleotides were made that can be used to amplify the area of the SSR, amplification products have been mapped to chromosome 21 using somatic cell hybrids and selected SSRs have been mapped in the 40 CEPH pedigrees and placed in the linkage map. The computer linkage analysis has been performed in collaboration with Dr. Aravinda Chakravarti. Table 4 shows the results of the screening for $(GT)_n$ SSR. About 70% of the positive plasmid inserts map to HC21. Experiments using $(ATTT)_5$ and other oligonucleotide probes to detect different SSR are in preliminary stages. However, few polymorphisms for $(ATTT)_n$ and poly A tails of Alu sequences for HC21 have been described from our laboratory (Petersen et al. 1991, McInnis et al. 1991).

5. **DNA Polymorphisms from the 3 untranslated regions of cDNAs using SSGDE (Single-stranded DNA gel electrophoresis)**

Cloning of Chromosome specific cDNAs will be a major part of the Human Genome project. The mapping of cDNAs can be rapidly accomplished by PCR amplification of the 3 untranslated region (3UT) using DNAs from somatic cell hybrids. Since the majority of genes do not contain intron sequences in their 3UT region, oligonucleotides primers from the 3' UT regions of cDNAs will amplify the same size product from genomic DNA. In order to more precisely map the cDNAs and place them into the linkage maps of chromosomes, we began a systematic search for polymorphisms in the 3UT regions of HC21 genes. The rationale is that most of the 3UT regions are not conserved and accumulate neutral mutations and therefore polymorphisms will be easily identified. All HC21 cDNAs cloned to date have been examined using single stranded DNA gel electrophoresis (SSDGE) (Orita et al. 1989) of PCR amplified products. The products with variation after SSDGE were further analyzed for Mendelian inheritance and subjected to nucleotide sequencing for characterization of the polymorphisms. Mapping of these polymorphisms either by restriction digestion after PCR, or SSDGE after PCR using DNA from the 40 CEPH families is in progress. Table 5 shows the results of the analysis to date. DNA polymorphism have been found in 9 of 25 STS (Each STS is about 300 nt long and therefore there is one polymorphism in every 850 nucleotides. It is therefore expected that almost every 3UT region of a cDNA will be polymorphic since in most cases

the 3UT region is about 1 kilobase. Sequence analysis in 3 cases showed that the polymorphism was due to a single nucleotide substitution. In one cDNA out of 17 examined, an intron has been found in the 3UT region.

6. **The linkage map of human chromosome 21.**

Our linkage map is based on the CEPH pedigrees (Dausset J, et al. 1990) and is a product of collaboration between the laboratory that generates genotypes (Johns Hopkins) and the computer laboratory of Dr. A. Chakravarti (University of Pittsburgh) that performs the linkage analyses. We have recently published a map with 27 markers (Genomics 9:407, 1991). The average distance between adjacent loci was 6cM and the total map length was 120 cM for sex average. The female map was longer than the male map especially in the proximal part of HC21. Most of the recombination was observed in the distal part of the long arm. The published map contained only 2 loci that could be detected by PCR amplification. In addition, not all markers fulfilled the criteria of index markers i.e. did not have heterozygosity of at least 70%. A considerable effort is now under way to enrich the linkage map with index markers and to detect these markers with PCR amplification. The current map from our laboratory contains genotypes of 40 markers. All of the 12 recently added markers can be served as STS and the majority have heterozygosity greater than 60%. Great effort has been directed towards the correction of suspected errors that have been made in genotypes included in previous maps and could have affected both the order and the distances among loci. Figure 2 shows the current linkage map for HC21.

The average distance among loci is about 5cM and a total of 7 loci can be classified as index markers. Recently, two polymorphisms of note have been found. The first is due to a GT repeat at locus 21-GT14 and is the most centomeric placed marker. Locus 21-GT14 will therefore, be the proximal end of the map; in addition, this marker will be used in determining the meiotic origin of nondisjunction in Down syndrome and will also clarify the potential HC21 pericentomeric linkage of Alzheimer disease. The second marker is at the 3UT region of the S100B gene which by pulsed field gel analysis (Burmeister et al. 1991) maps about 200 kb from the telomeric sequences of HC21q. The mapping of this marker, which can only be done by SSDGE and is in progress, will provide the distal end of the map.

7. **Cloning of cDNAs that map on HC21**

We began preliminary experiments towards the cloning of cDNAs that map to HC21. The cDNA libraries used for screening are (i) the two fetal brain libraries of R. Neve (Neve et al. 1986). The first, FB-1, comprises cDNAs from PolyA RNA from a 20-22 week fetal brain. The second, FB-2, was constructed from a 1:1 mixture of the FB-1 RNA and RNA from 18-19 week old fetus. Both libraries are in lambda gt11 and have a background of non-recombinant phage of not more than 10%. FB-1 contains 7.5×10^6 and FB-2 3×10^6 recombinants. Library FB-2 was methylated so that it contains full-length clones with internal EcoRI sites. The genes for APP, Tau and Map 2 have been cloned from these libraries (Neve et al. 1986) (ii) The retina cDNA library constructed by J. Nathans in lambda gt10 and has full length clones with internal methylated EcoRI sites. It has been used to clone several genes including the human visual pigment genes (Nathans et al. 1988) and GABA-rho channel genes (Cutting et al. 1991). (iii) The human liver cDNA library from the Genetics Institute Inc (Toole et al. 1984) that has been used to clone the full length human factor VIII gene.

The following methods have been tried to evaluate which one will give the best results for cloning chromosome specific cDNAs:

7.1 Alu-PCR of DNA from WAV17 and use the amplified product as a probe to screen cDNA libraries. The Alu-PCR method of Nelson et al. 1989 which uses the polymerase chain reaction to amplify DNA between Alu repetitive elements has been used in several laboratories for a variety of studies. Alu-PCR products from WAV17 were labeled with 32P and used as probes against 500,000 plaques from the J. Nathans retina library. The probe was prehybridized with total human genomic DNA to block the repetitive elements. A total of 20 positive clones were plaque-purified and the inserts of 5 of them have been used as probes against human genomic DNA, zoo blots and a mapping panel for HC21. All continued highly repetitive elements that prevent their mapping even after blocking with cold human DNA. The conclusion was that Alu-PCR of hybrid DNA is not the best method of preparing probes for cDNA library screening.

7.2 Alu-PCR of HC21 specific YACs.
DNA from YACs Y21.12, Y21.11, Y21.92 (see table 2) have been used for Alu-PCR and the amplification products have been pooled and used as probes against cDNAs from the J. Nathans retina library. In a preliminary experiment, a total of 10^4 pfu have been screened in duplicate and no positives have been identified. The conclusion from these experiments is that YAC PCR does not give positive results due to repetitive sequences. A larger number of pfus have to be screened before identifying positive cDNAs.

7.3 Use of CpG islands to screen cDNA libraries.
CpG islands are frequently found at the 5' end of genes in human DNA (Linday and Bird 1987). For restriction enzymes with 6-8 bp cutters and 2 CpGs in their recognition sequence, 75% of their target sites are in CpG islands. We have digested the HC21 phage library LL21NS02 with NotI and HindIII and subclone the NotI-Hind III fragments into plasmid vector bluescript. About 60 plasmid clones with NotI-Hind III inserts have been isolated. Similar experiments have been reported by Gao et al. 1991. These clones will be used as probes against cDNA libraries.

7.4 Chromosome 21 YAC DNA as probes to screen cDNA libraries.
Several investigators here at Hopkins have tried this method which was first used in F. Collins lab (Wallace et al. 1990) to identify cDNAs that map in the YAC DNA. Dr. M. Geraghty in D. Valle's lab used YACs from Xp to clone cDNAs close to OAT pseudogenes; W. Pavan in R. Reeves lab use YACs from mouse chromosome 16 to clone cDNAs from that mouse chromosome; Kinzler et al. 1991 and Joslyn et al. 1991 used YACs to clone colon specific cDNAs. A. Bergen in my laboratory used a human 650 YAC positive for DXS7 probe at Xp 11.3 to isolate cDNAs from a retina library which are candidates for the Norrie disease gene. The protocol for screening is included in the appendix. DNA from the yeast colony that contains the YAC is subjected to PFGE, the YAC is purified and used as a probe (after suppression

of repeated sequences with human genomic DNA) against the cDNA libraries. Dr. A. Warren had used 3 YACs that map on HC21, namely Y21.11, Y21.12, Y21.92 and screened 10^4 cDNAs. No duplicate positives have been identified; however, the number of pfus screened is very small and much more extensive screening is needed. Fortunately, no high background hybridization is observed due to repetitive elements.

8. **Collaboration with other investigators for the mapping of HC21.** The Human Genome project requires a high degree of collaboration among the laboratories in order to complete the short and long term goals. Our laboratory can only contribute to the project but the completion of HC21 map and sequencing will require a tremendous effort from both national and international laboratories. We have collaborated so far with all the major HC21 devoted labs. The following is a brief report on collaborative efforts.

   8.1 Dr. D. Patterson's lab. We deposited the JHU-HC21 YACs for screening. We also used somatic cell hybrids from that laboratory for our mapping.

   8.2 Dr. C. Smith's lab. We sent all mapped YACs or further characterization and mapping.

   8.3 Dr. J. Gray's lab. We sent all YACs for in situ hybridization. We are using their plasmid HC21 library for identifying short sequence repeat polymorphisms and develop STS.

   8.4 Dr. H. Lehrach's lab. We are participating in the effort to identify cosmids from the HC21 cosmid library. The cosmid filters, described in Nizetic et al. 1991, have been screened in our laboratory for D21S16, D21Z1, D21S8, PW245D, D21S82, FR8-77, MX1, D21S113, D21S110, D2153, S100B, PMCT 15 and cosmids positive for all these probes/loci have been identified. (See ICRF, reference library news update #1).

   8.5 Drs. Rouleau and Siddique. To map the Amyotrophic Lateral Sclerosis locus; Drs. Handy, van Broeckhoven, Schellenberg, Tanzi to map the Alzheimer locus. We have also collaborated with many clinical genetics centers as part of our effort to define Down syndrome minimal region (DSMR) (See McCormick et al. 1989 and McGinniss et al. 1991).

TABLE 1 STS on HC21 from JHU

	LOCUS / STS	5' OLIGO 5' TO 3'	3' OLIGO 5' TO 3'	POLY
1	D21S156	ATCCAGCCTGTAACACATTC	GTCAACATAGTGAGACCCCA	Yes GT
2	HMG14 sts1	TTACAGTAGATCTCAAGATG	TTACAGTAGATCTCAAGATG	Yes GT
3	HMG14 sts2	GAGGTGGAGGATGCAGTGAG	GAGGTGGAGGATGCAGTGAG	Yes AluVpA
4	HMG14 sts3	GGCACATTACTTGTCTGACA	GGCACATTACTTGTCTGACA	Yes GT
5	D21S171	CTTTATCTTCACACAGCTTC	CTTTATCTTCACACAGCTTC	Yes GT
6	IFNAR sts1	CACACTATGTAATACTATGC	TGCTTACTTAACCCAGTGTG	Yes AluVpA
7	IFNAR sts2	GACCAGAAATGAACTGTGTC	GAATGTAGGAAACGTTTGTC	No
8	IFNAR sts3	GACCAGAAATGAACTGTGTC	CTCGATCTCTTGACCTCATG	No
9	D21S120	GTGTGTCTGCCATTTCTGGGTGTAG	GATCCTGGGACAAAGTAGTCTCTAA	Yes GT
10	D21S13 sts1	ATCTGATGCCTTTGAATCT	TATGCCATTCACTGTGATA	Yes GT
11	D21S13 sts2	ATCCATTCATCCATTCTCCC	CAACATCAGGTCAACCAGAG	Yes EcoRI
12	D21S172	TACAAGGCAACCACATAATT	CTTTTCCTAATTGTGTGTGT	Yes GT
13	D21S3	GATCCAAGCAGTCTGTGTCC	ACACATTGGAAGCTCCACGG	No
14	D21S55	GGTTTGAGGGAACACAAAGC	CCACAGAGCTACAGCCTCTG	No
15	GT09	TTCTACCCAGTATTTCCCTG	AGGAAAGGATAAACATGGAG	Yes GT
16	GT16	CTGGCCAACATGGTGAAACCT	CATTTTAATGAACACCGCTC	Yes GT
17	GT12	CTGGAGAGAACTCTGATATG	AGATATTGGATGCTCAACTC	Yes GT
18	GT23	CCCAAAAGTAATTTGAGATG	ATGCTATACATTGCATTTAC	NT (GT)n
19	GT05	GTGTTTCCATTAGACACACACCC	TAGAGGCTTGAATTGGCTGG	Yes GT
20	GT07	ATAGTATCTTTAAGAACTTT	GATGGATATACAAGAAGGCT	NT (GT)n
21	GT15	AATGAACATACAATGAGCAC	CGTCAGCCTAGCTAATCATG	NT (GT)n
22	GT20	CAAGAATTAGAGAAAGTTGG	CAGCAGTGTGAAAATGGACT	NT (GT)n
23	GT11	AATGCTATCTTTGGAGTCAT	CCAAAGCAAGCAGCTTACAC	Yes GT
24	GT19	AGTAACTGTCCTGAAATAGC	GCACAGAGATCTGTATCATT	Yes GT
25	GT02	CTTCGTATAACGGGTACTAC	GTTAAACCAAGTACTACAAC	Yes GT
26	GT14	TCTAAAACAGTGTGTCTAGC	GCTGACGTGACAGTTGTGAG	Yes GT
27	GT10	GGCCTCCTGGAATAATTCTC	GATTTTAATGAACACCGCTC	Yes GT
28	GT25a	AACCATGTGGTAATTCTCTG	CAGATATTTCAGCCAATTAG	Yes GT
29	GT31	GACTAATAGGATTTTTGG	GGTCAGAGATACTAGTGCTC	Yes GT
30	GT34	CTCTCTCTGTCTATTGTTAC	TCAGAGAGGACACCACATTC	NT (GT)n
31	SOD1 sts1	TCCCTTGGATGTAGTCTGAG	CAGGGTTTTTATTCACAGGC	No
32	SOD1 sts2	GCCTGTGAATAAAAACCCTG	GGACAGCCTATTTGTCTAAG	No
33	PFKL	GCTTCTGAGGCCAGCCATGC	CCTGGGCAGGAAGGTGGATG	Yes SSDGE
34	GART	CTGCAGCCCTTCTGCTGGTG	AAAAATAGATGAAGTAAGGG	No
35	CD18	GAGCACTTGGTGAAGACAAG	GGATGTCATTTTATACCCTG	Yes SSDGE
36	COL6A1	ACCAAACCCTGTCCTCCCAC	AGGAGGGCAGGGCAGGAGAG	Yes SSDGE
37	COL6A2	TCCGCTGGATCTGCTAGCGC	TTTATTGGGTCAGGGCTTGC	No
38	MX1 sts1	GTTAACCACACTCTGTCCAG	GATTCTGAGGGCTGAAAATC	No
39	MX1 sts2	GGATGCTGTCTTCGTACTGG	GGCTAGAAATGAGTTTATTAC	No
40	ETS2	GCACAGCTAATTCTACTCAC	TGTTAAGGGATTCTGAGAAC	Yes SSDGE, Msp
41	ERG sts1	AAAGGGAAGTAGTAGAATTC	AACTCGTAGTGTATAAATGC	No
42	ERG sts2	GGAGGCTTTTCCCATCAGCG	CCTTAAGACTTCATGCTTC	No
43	ERG sts3	GAAGCATGAAGTCTTAAGG	CACAGCTGTCCATCAAACGG	No
44	APP sts1 ex16	ACAAACAGTAGTAGTGGAAAGAG	AGAAAGAAGTTTTGGGTAGG	No
45	APP sts2 ex17	CCTCATCCAAATGTCCCCCGTCATTG	CCTAATTCTCTCATAGTCTTAATTCC	No
46	APP sts3 3ut	CAGATGCAGAACTAGACCCC	GGAGAGAATCTATTCATGCA	No
47	CBS	TGGGTGGCTCAGCATCCTCC	TCCATTTATATGAAATGTCC	Yes SSDGE
48	CBR	GAGAAGAGAGTTGAACAGTG	CTCCTGCATCAGAGGAAATC	Yes SSDGE, Rsa
49	CRYA1	GATGCCCTTTCTTGAATTGC	CTCTAAAATCCTTGGAGCCC	No
50	S100B sts1	GCCTGCATCATTCTTTCTGC	GCATTTAAAGAACAGCAGGTG	No
51	S100B sts2	CACTGCTGTTCTTTAAATGC	GTGCTGGAGGCACGTTGGAG	No
52	S100B sts3	GGCCTGCTGTCAATAAAAAG	GCGCTGGATCGCTGTTGCTG	Yes SSDGE
53	BCEI	GCTTCTATCCTAATACCATC	TAGGCCAATTTTGAGTAGTC	No
54	21-SSR01	TGCTGCTGTGGGACTATTTG	CAGATGCACACAGCTGAGCC	No

TABLE 2 "HC21" YAC Library from JHU

YAC name	DNA Source	HUMAN	Size (kb)	Cloned ends	Sequenced ends	FISH	MAP In CELL HYBRIDS
Y21.1	**WAV-17**	**5Alu +**	**135**	**pw1**		**21**	**21pter-q22.2**
Y21.2	**WAV-17**	**5Alu +**	**760**	**pw2a pw2c**	**YES**	**21q11**	**21pter-q21**
Y21.3	**WAV-17**		**420**	**wr3a**	**YES**	**21**	
Y21.4	**WAV-17**	**5Alu +**	**100**	**pw4c**	**YES**	**21**	**21pter-q22.2**
Y21.5	**WAV-17**	**5Alu +**	**1000**	**wr5a**	**YES**	**21**	
Y21.6	**WAV-17**	**5Alu +**	**280, 180**	**pw6a pw6c**		**21**	**21q22.3-qter**
Y21.7	**WAV-17**	**5Alu -**				**21**	
Y21.8	WAV-17	5Alu +	360	wr8a wr8c	YES		
Y21.9	WAV-17	5Alu -		pw9a	YES		
Y21.10	WAV-17	5Alu +					
Y21.11	**WAV-17**	**5Alu +**	**200**	**pw11a pw11c**	**YES**	**21**	**21q22.3-qter**
Y21.12	**WAV-17**	**5Alu +**	**230**	**pw12a**	**YES**	**21**	**21q21-q22.2**
Y21.13	**WAV-17**	**5Alu +**	**320**			**21**	
Y21.14	WAV-17	5Alu +	800				
Y21.15	WAV-17	5Alu +	120	pw15c			
Y21.16	**WAV-17**	**5Alu +**	**340**			**21**	
Y21.17	WAV-17	5Alu -	290				
Y21.18	WAV-17		160				
Y21.19	**WAV-17**	**5Alu +**	**280**	**pw19a**	**YES**		**21q21-q22.2**
Y21.20	WAV-17	5Alu +	750	pw20a	YES		
Y21.21	WAV-17	5Alu +	340				
Y21.22	WAV-17	5Alu +	420				
Y21.23	WAV-17	5Alu +	440	pw23			
Y21.24	WAV-17	5Alu +		pw24			
Y21.25	WAV-17	5Alu -		pw25			
Y21.26	WAV-17	5Alu +		pw26			
Y21.27	WAV-17			pw27			
Y21.28	WAV-17			pw28			
Y21.29	WAV-17			pw29			
Y21.30	WAV-17			pw30			
Y21.31	**WAV-17**	**5Alu +**	**600**	**pw31**	**YES**		**21q21-q22.2**
Y21.32	WAV-17	5Alu +					
Y21.33	WAV-17			pw33			
Y21.34	WAV-17			pw34	YES		
Y21.35	WAV-17			pw35			
Y21.36	WAV-17	5Alu +	<370	pw36			
Y21.37	WAV-17			pw37			
Y21.38	WAV-17	5Alu +		pw38			
Y21.39	WAV-17			pw39			
Y21.40	WAV-17			pw40			
Y21.41	WAV-17			pw41			
Y21.42	WAV-17			pw42			
Y21.43	WAV-17	5Alu +	360				
Y21.44	WAV-17	5Alu +	360				
Y21.45	WAV-17	5Alu +	800				
Y21.46	WAV-17	5Alu +	300				
Y21.47	WAV-17	5Alu +	800				
Y21.48	WAV-17	5Alu +	750				
Y21.49	WAV-17	5Alu +					
Y21.50	WAV-17	5Alu +					

TABLE 2 "HC21" YAC Library from JHU

Y21.51	WAV-17	5Alu +					
Y21.52	WAV-17	5Alu +	700				
Y21.53	WAV-17		220				
Y21.54	WAV-17	5Alu +					
Y21.62	**WAV-17**		**100**	**p62**	YES		**21pter-q22.2**
Y21.92	**WAV-17**	**5Alu +**	**80**	**p92**	YES		**21q22.3-qter**
Y21.2Nss	sorted HC21			p2Not			
Y21.6Nss	sorted HC21			p6Not			
Y21.17 fs	sorted HC21			fs17end			
Y21.17N	sorted HC21			p17Not			
-Y21.22fs	sorted HC21			p22fs			
Y21.28fs	sorted HC21			p28fs			
Y21.32fs	sorted HC21			p32fs			
Y21.34fs	sorted HC21			p34fs			
Y21.36 fs	sorted HC21			p36fs			
Y21.40 fs	sorted HC21						
Y21.42 fs	sorted HC21	5Alu + weak					
Y21.43 fs	sorted HC21	5Alu +					
Y21.44 fs	sorted HC21	5Alu + weak					
Y21.45 fs	sorted HC21	5Alu + weak					
Y21.47 fs	sorted HC21	5Alu + weak					
Y21.50 fs	**sorted HC21**	**5Alu +**	**140**	**p50fsa p50fsc**	YES		**21q22.3-qter**
Y21.59 fs	**sorted HC21**	**5Alu +**	**110**	**p59fsa p59fsc**	YES		**21**
Y21.60 fs	sorted HC21	5Alu +		p60fsa p60fsc	YES		
Y21.67 fs	sorted HC21	5Alu + weak		p67fsa			
Y21.69 fs	**sorted HC21**	**5Alu +**	**500**	**p69fsa p69fsc**	YES		**21q21-q22.2**
Y21.70 fs	sorted HC21			p70fsa			
Y21.71 fs	sorted HC21	5Alu + weak					
Y21.72 fs	sorted HC21						

TABLE 3A Screening for HC21 YACs in the McCormick et al Los Alamos Libraries

STS	POS / NEG	POSITIVE POOLS
GT12	POS	3 ; M ; C1-4
HMG14	POS	18 ; B ; C1-4
SOD1	POS	5 ; N ; C1-4
D21S13	POS	2 ; J ; C1-4
D21S120	POS	5 ; 6 ; 19 ; F ; J ; C1-4 ; E7-12 ; N7-12
D21S55	NEG	
D21S156	NEG	
ETS2	NEG	
APP	NEG	
GT05	NEG	
CBR	NEG	
ERG	NEG	
CRYA1	NEG	
S100B	NEG	
GT20	NEG	
GT11	NEG	
GT19	NEG	
GT02	NEG	
GT14	NEG	
IFNAR	NEG	
CD18	NEG	
MX1	NEG	

TABLE 3B Screening for HC21 YACs in the ICI Library

STS	ALL	STRONGLY	WEAKLY
	POOLS	POSITIVE	POSITIVE
		POOLS	POOLS
ETS2	POS	M15	M19 M23
D21S55	POS	M3 M9	M10 M11 M17 M18 M25 M27
SOD1	POS	M3 M10	M4 M15 M35
GT20	POS	in progress	
GT19	POS	in progress	
GT02	NEG		
S100B	NEG		
ERG2	NEG		
CD18	POS	in progress	
D21S13	POS	M5	M1
MX1	NEG		
APP	POS	M14 M32	M12 M15 M24 M28 M30 M31 M37 M39
IFNAR	NEG		
GT05	POS	in progress	
GART	NEG		
CRYA1	POS	M21 M26 M33 M36	M32 M35 M38 M39
CBR	POS	in progress	

TABLE 4 Short Sequence Repeat Polymorphisms on HC21

LOCI	Rec/nt	MAP	STS	Size	Heteroz %	Alleles	CEPH Closest Marker (O ; z)	REPEAT
GT01	phage 20	NON 21	YES	210				GT19
21-GT02	phage 12	HC21	YES	220	84	9	D21S8 (5% ; z=16.5)	GT18
GT03	phage 18							GT17
GT04	AW 3A9		YES	205				GT22
21-GT05	AW 1A7	HC21	YES	280	55	7	D21S82 (1% ; z=49.5)	CA15GA17
GT06	AW4A3	NON 21	YES	220				GT13GA12
21-GT07	AW 4A1	HC21	YES	195			in progress	GT15
GT08 A	AW 1B6							AC16
GT08 B	AW 2B6		YES					GT18
21-GT09	AW 2B3	HC21	YES	120			in progress	GT18
21-GT10	AW 3A5	HC21	YES	240	86	11	CBS (5% ; z=19.75)	GT15
21-GT11	AW 4A8	HC21	YES	290	72	7	in progress	GT20
21-GT12	AW 2A1	HC21	YES	160	86	14	APP (0% ; z=64.0)	TA11GT17
GT13	AW Z4	NON 21	YES	230				CA12GA14
21-GT14	AW Z1	HC21	YES	170	65	6	D21S120 (1% ; z=14.2)	CA16
GT15	TA 5B		YES	190				GT14
GT15A	TA 5A		YES	110				TA7CA9
21-GT16	AW X10	HC21	YES	280			in progress	CA7NNTA5CA8
GT17	AW 5A6							CA18
GT18	AW 5A1?							
21-GT19	AW 8AD	HC21	YES	130	54	3	IFNAR (0.03 ; 52.5)	GT18
21-GT20	AW 5A23	HC21	YES	115			in progress	CA11
GT21	AW 2A2							
GT22	AW 8A7	NON 21	YES	100				CA14
GT23	AW 1A5		YES	110				TA5CA11
GT24	AW 5A1-5A11		YES	210				CA24
GT25								
GT25A	AW 5A7		YES	160	58	5	in progress	GT10
GT26								CA15
GT27	AW X5							GT14
GT28	AW 7A5		YES					

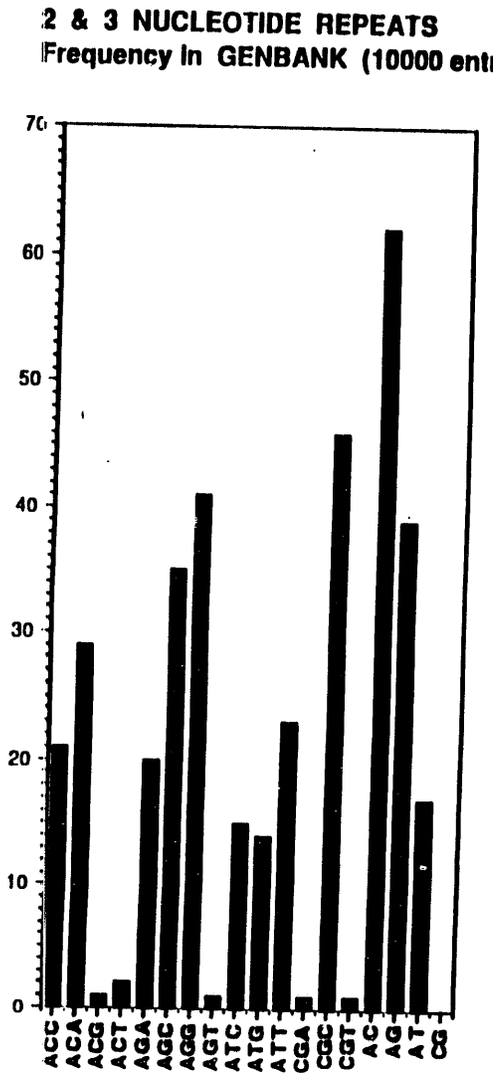
TABLE 4 Short Sequence Repeat Polymorphisms on HC21

GT29	AW 5A2							GT13GA11
GT30	AW 1A2							GT16
21-GT31	AW X3	HC21	YES	160	40	3	in progress	GT14
GT32	AW 5A21							
GT33	AW Z5							GT8TA2GT6TA
GT34	AW X7							TA16GT10
ATTT 01	D36	in progress						tetra
ATTT 02	AW1a							
ATTT 03	U21							
ATTT 04	J1-1							
ATTT 05	MK8-1							
ATTT 06	ROCI 1-4							
ATTT 07	SEA 10-1							
ATTT 08	SEA 12-2							
ATTT 09	SEA 8-1							
ATTT 10	SEA 9-1							

TABLE 5 HC21 cDNA 3'UT Polymorphisms

GENE	5' OLIGO 5' TO 3'	3' OLIGO 5' TO 3'	PCR bp	Poly-morphism	Frequency	Sequenced	Mapped in CEPH
CBR	GAGAAGAGAGTTGAACAGTG	CTCCTGCATCAGAGGAAATC	235	YES	20% (16/80)	TGT to TAT RsaI	IFNAR 5% z=15.98
SOD1	TCCCTTGGATGTAGTCTGAG	CAGGGTTTTTATTCACAGGC	300	No			
	GCCTGTGAATAAAAACCCTG	GGACAGCCTATTTGTCTAAG	240	No			
ERG2	AAAGGGAAGTAGTAGAATTC	AACTCGTAGTGTATAAATGC	327	No			
	GGAGGCTTTTCCCATCAGCG	CCTTAAGACTTCATGCTTC	270	No			
	GAAGCATGAAGTCTTAAGG	CACAGCTGTCCATCAAACGG	315	No			
	AGCTTTCTCAAACTGTGAAG	CCATTACGCTGTGTCCTTTC	305	No			
	CATTTATACACTACGAGTTG	GTGCCAAACATCCTATTTCC	326	ND			
ETS2	GCACAGCTAATTCTACTCAC	TGTTAAGGGATTCTGAGAAC	309	YES	50% (15/30)	CG to TG MspI	
S100B	GCCTGCATCATTCTTTCTGC	GCATTTAAAGAACAGCAGGTG	255	No			
	CACTGCTGTTCTTTAAATGC	GTGCTGGAGGCACGTTGGAG	280	No			
	GGCCTGCTGTCAATAAAAAG	GCGCTGGATCGCTGTTGCTG	305	YES	7% (6/80)		D21S171 0% z=10.54
CRYA1	GATGCCCTTTCTTGAATTGC	CTCTAAAATCCTTGGAGCCC	249	No			
HMG14	GGAGAGAAAGAAGCCAAGTC	GACACCCGAGACAGTCAGAG	290	YES	25%		
PFKL	GCTTCTGAGGCCAGCCATGC	CCTGGGCAGGAAGGTGGATG	261	YES ?	20%		
GART	CTGCAGCCCTTCTGCTGGTG	AAAAATAGATGAAGTAAGGG	223	No			
APP	CAGATGCAGAACTAGACCCC	GGAGAGAATCTATTCATGCA	308	No			
	TGCATGAATAGATTCTCTCC	GCAGAAGCAGCAATCTGTAC	280	No			
MX1	GTTAACCACACTCTGTCCAG	GATTCTGAGGGCTGAAAATC	271	No			
	GGATGCTGTCTTCGTACTGG	GGCTAGAAATGAGTTTATTAC	210	No			
CD18	GAGCACTTGGTGAAGACAAG	GGATGTCATTTTATACCCTG	341	YES	35% (7/20)	CG to TG	
IFNAR	GACCAGAAATGAACTGTGTC	CTCGATCTCTTGACCTCATG	268	No			
	GACCAGAAATGAACTGTGTC	GAATGTAGGAAACGTTTGTC	262	No			
	GACAAACGTTTCCTACATTC	ACAGAATGTTAGGAATGGCG	242	ND			
CBS	TGGGTGGCTCAGCATCCTCC	TCCATTTATATGAAATGTCC	323	YES	50% (40/80)		D21S113 1% z=36.29
COL6A1	ACCAAACCCTGTCCTCCCAC	AGGAGGGCAGGGCAGGAGAG	295	YES			
COL6A2	TCCGCTGGATCTGCTAGCGC	TTTATTGGGTCAGGGCTTGC	320	YES			
BCEI	GCTTCTATCCTAATACCATC	TAGGCCAATTTTGAGTAGTC	Intron				
							18-Aug-91

FIGURE 1


2 & 3 NUCLEOTIDE REPEATS
Frequency in GENBANK (10000 entries)
70
60
50
40
30
20
10
0
ACC
ACA
ACG
ACT
AGA
AGC
AGG
AGT
ATC
ATG
ATT
CGA
CGC
CGT
AC
AG
AT
CG

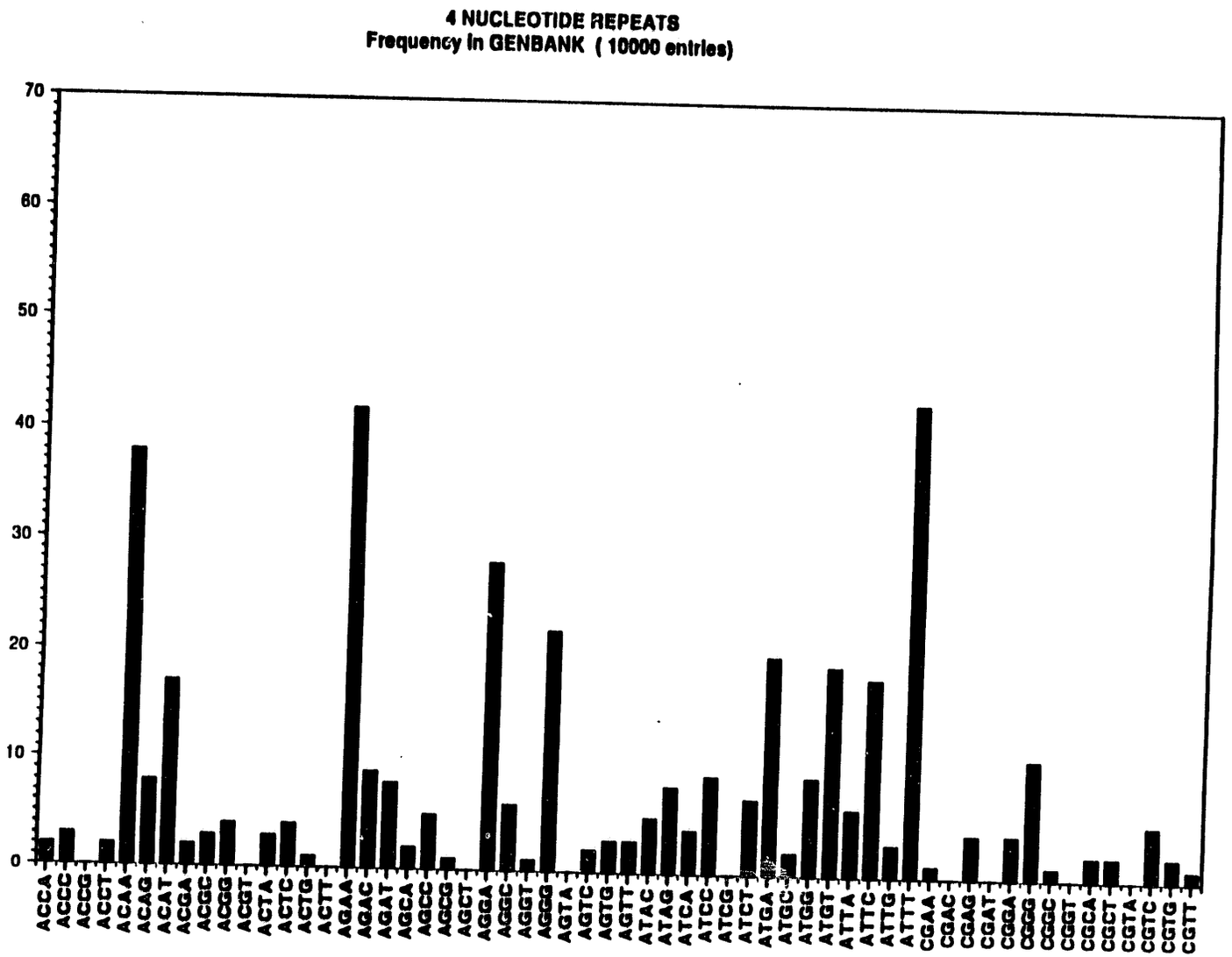
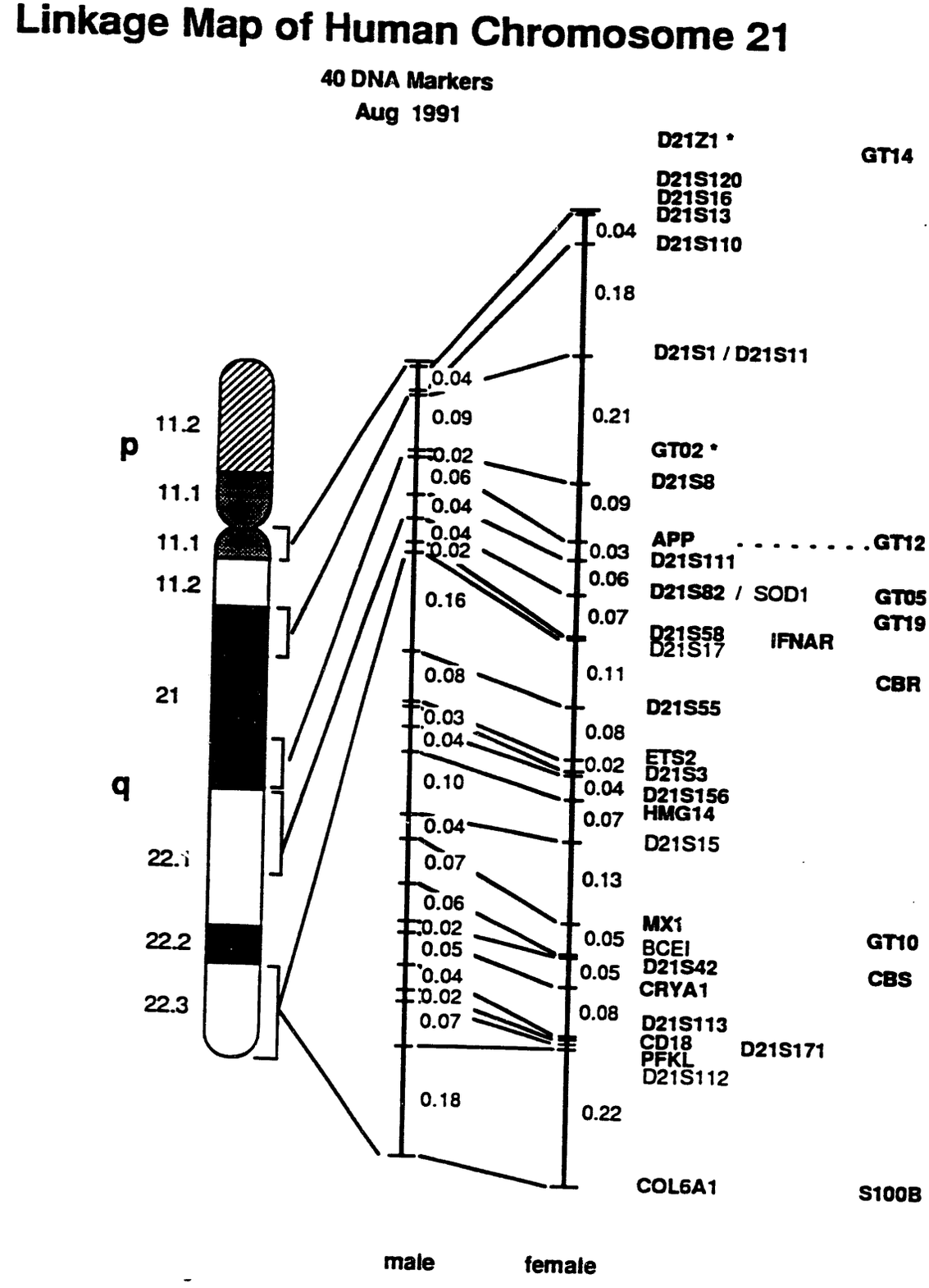

4 NUCLEOTIDE REPEATS
Frequency in GENBANK (10000 entries)
70
60
50
40
30
20
10
0
ACCA
ACCC
ACCG
ACCT
ACAA
ACAG
ACAT
ACGA
ACGC
ACGG
ACGT
ACTA
ACTC
ACTG
ACTT
AGAA
AGAC
AGAT
AGCA
AGCC
AGCG
AGCT
AGGA
AGGC
AGGT
AGGG
AGTA
AGTC
AGTG
AGTT
ATAC
ATAG
ATCA
ATCC
ATCG
ATCT
ATGA
ATGC
ATGG
ATGT
ATTA
ATTC
ATTG
ATTT
CGAA
CGAC
CGAG
CGAT
CGGA
CGGG
CGGC
CGGT
CGCA
CGCT
CGTA
CGTC
CGTG
CGTT

FIGURE 2


Linkage Map of Human Chromosome 21
40 DNA Markers
Aug 1991
p
q
11.2
11.1
11.1
11.2
21
22.1
22.2
22.3
D21Z1 *
GT14
D21S120
D21S16
D21S13
D21S110
D21S1 / D21S11
GT02 *
D21S8
APP
GT12
D21S111
D21S82 / SOD1
GT05
GT19
D21S58
D21S17
IFNAR
CBR
D21S55
ETS2
D21S3
D21S156
HMG14
D21S15
MX1
BCEI
GT10
D21S42
CBS
CRYA1
D21S113
CD18
D21S171
PFKL
D21S112
COL6A1
S100B
0.04
0.18
0.04
0.09
0.21
0.02
0.06
0.09
0.04
0.04
0.02
0.03
0.06
0.16
0.07
0.08
0.11
0.03
0.04
0.08
0.02
0.10
0.04
0.07
0.04
0.07
0.13
0.06
0.02
0.05
0.05
0.04
0.02
0.08
0.07
0.18
0.22
male
female

APPENDIX

A. Papers *

McCormick et al. PNAS 86:9991-9995, 1989
McCormick et al. Technique 2:65-71, 1990
Hieter et al. Genome Analysis 1:83-120, 1991
McCormick et al. Gata 7:114-118, 1990
Shero et al. Genomics 10:505-508, 1991
Connelly et al. Genomics 10:10-16, 1991
Economou et al. PNAS 87:2951-2954, 1990
Petersen et al. Genomics 9:407-419, 1991
Petersen et al. Genomics 7:136-138, 1990
Lewis et al. Genomics 8:400-402, 1990
Petersen et al. Hum. Genet. in press 1991
Jabs et al. Genomics 9:141-146, 1991
Francomano et al. Hum. Genet. 87:162-166, 1991
McInnis et al. Genomics in press 1991
Paran et al. PNAS 87:1300-1304, 1990

* (cycled separately)

END

DATE FILMED

12/31/91