J. Biochem. 122, 872-877 (1997)

Isolation of a Novel Human Gene from the Down Syndrome Critical Region of Chromosome 21q22.2[1]

Akiko Nakamura, Masahira Hattori, and Yoshiyuki Sakaki[2]

Human Genome Center, The Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108

Received for publication, June 23, 1997

Down syndrome is the most common birth defect, and is caused by trisomy 21. We identified a novel gene in the so-called Down syndrome critical region by means of computer-aided exon prediction and subsequent cDNA cloning. The gene, designated as *DCRA* (*D*own syndrome *C*ritical *R*egion gene *A*), consists of eight exons of 3,252 bp in total and encodes a large open reading frame of 297 amino acid residues. The open reading frame shows significant homology to Hβ58, a mouse gene essential for embryogenesis, PEP8, a yeast homologue of Hβ58, and an expressed sequence tag of *Arabidopsis thariana*, suggesting that DCRA has some important function that has been conserved during the course of evolution. *DCRA* is expressed in most tissues examined, including fetal and adult brain, heart, lung, liver, and kidney. The cDNA of the *DCRA* mouse homologue, *Dcra*, was also cloned. It is 2,157 bp long and has an open reading frame of 297 amino acid residues, which shows 92% identity to human DCRA. *Dcra* is expressed in all the embryo and adult tissues examined.

Key words: cDNA cloning, chromosome 21, computer-aided exon prediction, Down syndrome.

Down syndrome (DS) is the most common birth defect, and is caused by trisomy 21, and characterized by distinct facial and physical features, and mental retardation. Some patients have associated congenital heart disease or gut disease, immune deficiencies, or an increased rate of leukemia (*1, 2*). These complex features of DS imply the involvement of multiple genes in its pathogenesis.

Studies on partial trisomy 21 DS patients suggested that a region extending from DNA marker D21S55 to ERG of 21q22.2 is critical for DS (*2, 3*), which is now called the Down syndrome critical region (DCR). To understand the pathogenesis of DS, it is obviously important to identify and characterize the genes in DCR. We thus started the systematic analysis of the region to identify the genes in DCR, and previously reported the presence of a novel gene designated as *TPRD* (*4*). In this paper, we report another novel gene which might be involved in the pathogenesis of DS.

[1] This study was supported in part by Grants-in-Aid for Scientific Research on Priority Areas and for Creative Basic Research (Human Genome Project) from the Ministry of Education, Science, Sports and Culture of Japan, and a grant from the Science and Technology Agency (STA). The DDBJ, EMBL, and GenBank accession numbers for the complete DNA sequences of human *DCRA* and mouse *Dcra* are D87343 and AB001990, respectively.

[2] To whom correspondence should be addressed. Tel: +81-3-5449-5623, Fax: +81-3-5449-5445, E-mail: sakaki@hgc.ims.u-tokyo.ac.jp

Abbreviations: DCR, Down syndrome critical region; DS, Down syndrome; ESTs, expressed sequence tags; Mb, mega base pair; nt, nucleotide; ORF, open reading frame; RACE, rapid amplification of cDNA ends; RT-PCR, reverse transcription-polymerase chain reaction; SSC, sodium citrate-sodium chloride buffer.

MATERIALS AND METHODS

Genomic DNA Sequencing—The genomic DNA sequence (110,392 bp) of two overlapping P1 clones, S310 and D10 (*5*), in chromosome 21q22.2 was determined using a dideoxy method based on a novel nested deletion system, that was developed by our group (*6*).

Computer Analysis of Genomic Sequence Data—Exon prediction software, GRAIL (xgrail II), was used to identify putative exons in genomic DNA sequences (*7*). The Alu-masked sequence data were also compared with a non-redundant nucleotide database (constructed from GenBank, GenBank-upd, and EMBL) using a homology search program, BLASTN, to identify registered transcribed sequences (*8*).

RT-PCR Analysis—Reverse transcription was carried out using poly(A)$^+$ RNA from a 19-23-week male/female pool of human fetal brains, a 18-25-week male/female pool of human fetal hearts, a 22-26-week female pool of human livers, and a 37-year-old human adult male brain (Clontech), as templates. First strand cDNA was synthesized using 0.5 μg of poly(A)$^+$ selected RNA, SuperScript reverse transcriptase, random hexamers, and oligo(dT)$_{12-18}$, according to the manufacturer's recommendations (GIBCO BRL). The resultant cDNA (0.125 ng) was amplified by PCR in a reaction mixture (25 μl) comprising 0.5 U of ELONGASE Enzyme Mix, 0.5$\times$buffer A, 0.5$\times$buffer B (GIBCO BRL), 0.25 mM dNTP, and 0.2 μM primers (a: 5′-CTGAGAACACCAGCATCTG-3′, b: 5′-TGAGTGCTTGTCACACATG-3′, c: 5′-CATGTGTGACAAGCACTCA-3′, d: 5′-AGCTGCTCACCTCCTGCTG-3′, e: 5′-TCTGTA-

TGAGACGTATCATG-3′, f: 5′-CCAAGTTTCAAGTGACTCAG-3′, and g: 5′-ATTGGAGAGGCCTTCAATG-3′; Fig. 1). The DNA was incubated at 94°C for 3 min, followed by 30 cycles of denaturation at 94°C for 30 s, reannealing to primers at 53°C for 1 min, and incubation at 72°C for 5 min, and finally incubated at 72°C for 10 min in a thermal cycler (9600, Perkin-Elmer Cetus).

5′ and 3′ RACE—RACE was performed from 1 μg of poly(A)$^+$ RNA from the 19-23-week male/female pool of human fetal brains using a Marathon cDNA Amplification Kit (Clontech) according to the manufacturer's recommendations. Adopter ligated cDNA from 9-11-week male BALB/C mice (Marathon-Ready cDNA, Clontech) was used for cloning of the mouse homologue. These cDNAs (0.5 ng) were amplified by PCR in a reaction mixture (25 μl) comprising 0.5 U of Ex taq, 1×PCR buffer (Takara Shuzo), 0.25 mM dNTP, and 0.2 μM primers (human/5′ RACE gene-specific primer 1: 5′-CTGAATGTTGACAAACACGCCATG-3′ and 3′ RACE gene-specific primer 1: 5′-GGTTCGAAGGATATGGACTATTGC-3′, mouse/5′ RACE gene-specific primer 1: 5′-ACCGCCGCATGTCACAGCGCAGTG-3′, and 3′ RACE gene-specific primer 1: 5′-GACTGTAAACCTCCAGCTCAGTGCCA-3′, and the AP1 primer). DNA was incubated at 94°C for 1 min, followed by 30 cycles of denaturation at 94°C for 30 s, reannealing to primers, and incubation at 68°C for 4 min in a thermal cycler (9600, Perkin-Elmer Cetus). The second PCR reaction was performed for the first PCR products using nested primers (human/5′ RACE gene-specific primer 2: 5′-GCCATGATACGTCTCATACAGAAC-3′ and 3′ RACE gene-specific primer 2: 5′-TTCTTCCTTCAAATCCTGCCACTG-3′, mouse/5′ RACE gene-specific primer 2: 5′-TCACAGCGCAGTGTGTACTG-3′ and 3′ RACE gene-specific primer 2: 5′-ATCCAGATTATCAACAGCACC-3′, and the AP2 primer).

Sequencing of cDNAs—To sequence cDNAs, we employed the cycle sequencing reaction using an ABI PRISM Dye Primer Cycle Sequencing Ready Reaction Kit or an ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction Kit (Applied Biosystems). The sequences were determined with an automated DNA sequencer, ABI Model 373S DNA Sequencing System. Sequence data were analyzed using GENETYX-MAC (Japan Software, Tokyo). The complete cDNA sequence data will appear in the DDBJ (accession numbers, D87343 for *DCRA* and AB001990 for *Dcra*).

Northern Blot Analysis—Northern blots (Clontech) containing poly(A)$^+$ mRNAs from human fetal and adult tissues or mouse embryo and adult tissues were hybridized at 65°C with a human partial cDNA probe (nucleotide positions 551-1014) labeled with [α-^{32}P]dCTP by random oligonucleotide priming. The blots were washed in 2×SSC/1% SDS at 65°C.

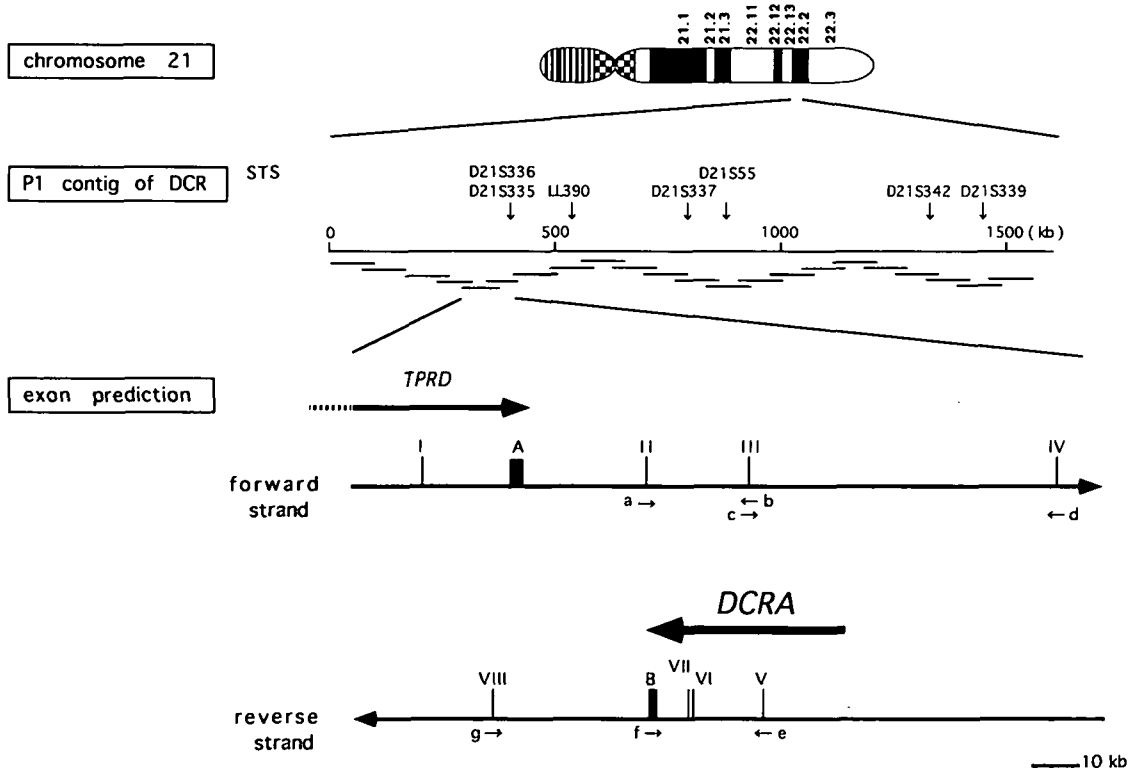
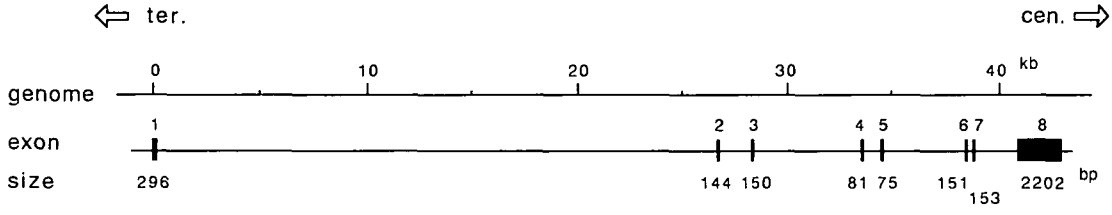


Fig. 1. **Exon prediction using GRAIL and BLASTN of the 110 kb DNA sequence in DCR.** The DNA sequence was directly determined from two P1 clones, S310 and D10. GRAIL predicted four "excellent" exons (vertical bars) in each strand (I-VIII). BLASTN was used for the homology search. Fifteen ESTs were mapped in region A (black box). Their accession numbers in GenBank are HSC1YA021, H29370, R60437, R92635, R57164, R91381, T65893, R60204, HSC1YA022, HUMNK715, R05530, T65777, T07926, H29282, and HUM21ES84 (*9*). Seven ESTs were mapped in region B (black box). Their accession numbers in GenBank are HUM424C06B, HUM-413A01B, H05324, R94225, HSC2RG062, R43971, and R41105. Predicted exon I and region A are parts of the *TPRD* gene (*4*). PCR primers (a-g) for RT-PCR were prepared from other predicted exons (II-VIII) and region B. Their positions and directions are shown by small arrows. A novel gene designated as *DCRA* was identified using primers e and f.

RESULTS AND DISCUSSION

Exon Prediction from a Large Genomic Sequence—We previously constructed a 1.6 Mb P1 contig of the DCR (*5*), and identified a novel gene using the exon trapping method (*4*). To identify the gene more systematically, we started sequencing of the region, followed by computer-aided sequence analysis. As the first step, we analyzed a 110 kb region (corresponding to P1 clones S310 and D10; DDBJ accession number, D87676) with exon prediction software, GRAIL (xgrail II, Ref. 7), and a homology search program, BLASTN (*8*). GRAIL predicted four "excellent" exons in the forward strand (5′ to 3′ = centromere to telomere), and four "excellent" exons in the reverse strand (5′ to 3′ = telomere to centromere). On the other hand, BLASTN identified sequences homologous to fifteen overlapping expressed sequence tags (ESTs) in the forward strand and ones homologous to seven overlapping ESTs in the reverse strand (Fig. 1). The predicted exon I and all the identified ESTs in region A were found to be parts of the *TPRD* gene (*4*). The other predicted exons (II–VIII) and all the identified ESTs in region B showed no homology to known sequences.

Identification of a Novel Gene from Predicted Exons—PCR primers were prepared based on the predicted exons and identified ESTs in region B, and inter exon RT-PCR analysis was performed using mRNAs from human fetal brain, heart, and liver, and adult brain. For the forward strand, a PCR product was not detected with any primer combination (a/b, c/d, and a/d) in any of the examined tissues. For the reverse strand, an approximately 2 kb PCR product was obtained with primers e and f in all cases examined (data not shown). A PCR product was not detected using primers e and g. Sequencing of the 2 kb fragment showed that it contained predicted exons V, VI, and VII, and region B. As described below, the corresponding transcript had a length of 3.3 kb. By means of further 5′



B

```
GGGGTGAACCGTGACAGCGCTTCGGCCCGCACTAAGGCCGGCTCTTGTGCCGGAAGGAGG    60
AAGGCGTGGGGCATTCGCCCCTCGGAGCTAGGGAGTGTGTGCGACGCCGCTGCGAGGTCA   120
CGTGAGCCACTGCCGGCAGAGAGGGAAAGGGGCGGGGCCCAGAACGAAGCGGGGAGGCGC   180
CCCTTGTTTCCCTGGGGTCACGCGCAGCCGGAAGTGGCGGCTGCTGCGGAGAATTGGAGA   240
                                                          M     1

TGGGGACCGCCCTGGACATCAAGATTAAAAGAGCGAATAAAGTTTATCACGCCGGGGAAG   300
  G  T  A  L  D  I  K  I  K  R  A  N  K  V  Y  H  A  G  E  V    21

TGCTCTCTGGCGTGGTGGTCATATCGAGTAAGGATTCAGTCCAACACCAGGGAGTGTCTT   360
  L  S  G  V  V  V  I  S  S  K  D  S  V  Q  H  Q  G  V  S  L    41

TGACCATGGAAGGAACTGTAAACCTCCAGCTCAGTGCCAAAAGTGTGGGTGTGTTTGAAG   420
  T  M  E  G  T  V  N  L  Q  L  S  A  K  S  V  G  V  F  E  A    61

CTTTTTATAATTCTGTTAAGCCTATCCAGATTATCAACAGCACCATAGAAATGGTGAAGC   480
  F  Y  N  S  V  K  P  I  Q  I  I  N  S  T  I  E  M  V  K  P    81

CGGGGAAATTTCCCAGCGGCAAAACAGAAATCCCTTTTGAATTTCCTCTGCACTTGAAGG   540
  G  K  F  P  S  G  K  T  E  I  P  F  E  F  P  L  H  L  K  G   101

GTAACAAAGTTCTGTATGAGACGTATCATGGCGTGTTTGTCAACATTCAGTATACACTGC   600
  N  K  V  L  Y  E  T  Y  H  G  V  F  V  N  I  Q  Y  T  L  R   121

GCTGTGACATGAAGCGGTCTCTGTTGGCCAAGGACTTGACAAAGACCTGTGAATTTATCG   660
  C  D  M  K  R  S  L  L  A  K  D  L  T  K  T  C  E  F  I  V   141

TTCACTCCGCTCCTCAGAAGGGGAAGTTTACTCCCAGTCCCGTGGACTTCACGATTACAC   720
  H  S  A  P  Q  K  G  K  F  T  P  S  P  V  D  F  T  I  T  P   161

CTGAAACCTTACAGAACGTCAAAGAGAGAGCTTTGCTTCCCAAATTTCTCCTTCGAGGAC   780
  E  T  L  Q  N  V  K  E  R  A  L  L  P  K  F  L  L  R  G  H   181

ATCTCAACTCAACAAACTGTGTCATCACGCAGCCACTAACGGGAGAGCTGGTGGTGGAGA   840
  L  N  S  T  N  C  V  I  T  Q  P  L  T  G  E  L  V  V  E  S   201

GCTCGGAAGCCGCCATCAGAAGCGTGGAGCTGCAGCTGGTGCGCGTGGAGACGTGCGGGT   900
  S  E  A  A  I  R  S  V  E  L  Q  L  V  R  V  E  T  C  G  C   221

GTGCAGAAGGCTATGCCCGCGACGCCACGGAGATTCAGAACATTCAGATCGCCGACGGGG   960
  A  E  G  Y  A  R  D  A  T  E  I  Q  N  I  Q  I  A  D  G  D   241

ATGTGTGCAGGGGCCTCTCTGTCCCCATCTACATGGTCTTCCCTAGGCTGTTCACCTGCC  1020
  V  C  R  G  L  S  V  P  I  Y  M  V  F  P  R  L  F  T  C  P   261

CTACACTGGAGACCACCAACTTCAAAGTGGAATTTGAGGTTAACATCGTGGTGCTGCTTC  1080
  T  L  E  T  T  N  F  K  V  E  F  E  V  N  I  V  V  L  L  H   281

ACCCTGACCACCTCATCACGGAGAACTTCCCGCTGAAGCTCTGCAGGATATAGCCCGGAG  1140
  P  D  H  L  I  T  E  N  F  P  L  K  L  C  R  I  *            297
```

```
GAGGGAAGCATAGAGAACGGGAGTGGCCATCTGGAAATCCAGCTGGTTATCCAAATCCTA  1200
AGGGGAGCTACAGCCAGCGGCATATACTTGTTTTTGTGATTATTCTGTATCAGAAATGAA  1260
ACAGACCCTCAAATTAACTTTCCTTCCTCATTTCTTGAGGCTTCTGCTTCCAACAGGCAC  1320
CTCTAATCAGACCTTTTCTTTGAAATTCAACAAGATTTCTTAATGCTATTTGCCAAGACC  1380
ATTTCACAGAAAACATTGACTGTGGCTCTTGCCTTATCTGTTCCTTTTTAGGTACAGTAA  1440
AACAATTGTGACAGCAGTTTGAGCTTGCTGGAGAGTGGCATCATGGGGACAAAAGGAAAC  1500
CTCTGACTTGCTAATGGATGTAGCCAGGGACTCCCCATAGCAAAGGGTCTGTGGCCAGTT  1560
GACATCCAGGATGGCTGCAAGCGCACTTGATGGTCAGGAAGTTTGCAGATACTCGCCAAG  1620
GCAGACGCCAAAGTGCTAGCCACTGGAAATGCATGACTTCCCTCCACCCCTACTCTATTC  1680
TGTAGTTTTTTGGTTTTGTTTCTGAGACGGAGTCTCAGTCTGTCACCCAGGCTGGAGTGA  1740
TCTCAGCTCACTGCAACCTCCACCTCCCAGGTTCAAGCGACTCTCCTGCCTCAGCCTCCC  1800
GAGTAGTTGGGATTACAGGTGACTGCCACCGTGCCCGGCTAATGTTTGTATTTTTAGTAG  1860
AGACGGGGCTTCACCATCTTGGCCAGGCTGGTCTTGAACTCCTGACCTCGTGACCCACCC  1920
GCCTTGGCCTCCCAAAGTGCTGGGATTACAGGTGTGAGCCACCACACCCAGCCTCTGTAG  1980
TTCTTTTTACAACATTTTTCATTATAACTTTAAATTTTTTAAGCAACTGGAAAAGTGTTC  2040
CTTGCTCTCTTGGGGGGATTTGGCTGGTGCCGAAGTGTTTCTGAAGTCTCAAGAACTGCC  2100
ATAAAATCTCACGCTGCCATTTCCCTGAACAGATACATACATAGAGAGAGACAGTTTTCC  2160
AAACTGTGTCACGCAGGCTGAGTGCACTGGCAGGATCACAGCTCACGGCAGCCTCAACCT  2220
CCCTGGCTCAAGCGATCCCTCCCCTCAGCCTCCTGAGTAGCTGAGACTACAGGTGAGTGC  2280
CACCACACTCAGCTAATTTTCAAATTTTTTGTAGACAGGGTCTCCCTATGTTGCCCAGGC  2340
TGGTCTTGAACTCCTAGACTCAAGTGATCCTCCTGTCTTGGCCTCCCAAAGTGCTGAGAT  2400
TACAGGTGTGAGCCACTGTGCCCAGCAGTTTCCCAGAATATATTTAAATGCAAAGTTACA  2460
TGAGGGGAAAACATGTATGTTTGCTCCTGTTGTTACTGGGTAGGTTCTGAACAGCAGAAA  2520
CCCATGTGCAGGGTGGGCTGGTGAAGGCCCCTCTCCGCAAGGTGGTAGCAGGAAAAGGTC  2580
CTTGACTTGATGAATTTGGTCTGCCTCTGAGCCACTGGAGGAAGCTGTTTTGAGCCAGGG  2640
TTTTTTGGCCTAAAGCCAGCATTTCCTCAGTCTCCCTTTGTGGTTCGAAGGATATGGACT  2700
ATTGCAATACATTTCTTCCTTCAAATCCTGCCACTGTTTTGTTGGCCCACAACTAATAGG  2760
ACCTCAAAATAAGCCATGCTGCTTTGCACACACACTAGCCTTCTTTTGTACTTTTCATTC  2820
TGGATGGGCTTGGCCAAAACAGGCTCAGGCCAAAGACCTCCCAAGCTGTATGTACTTCCA  2880
GTATCCTGAAACAGTGTTTGGTGACATAATGCCAAGGGTAAACAAGCCTGATTTAGGCAC  2940
TGCTTTATCCAGGGGCTTCACCCATGAAATTAATAAAACTTATCTGAGTCACTTGAAACT  3000
TGGTTCCCAGAAAACACATTTCTGGTTTATAATCTCCTTTTATGCTCACCTGACATTAAT  3060
TATCTATCCTTGATGATGTGTTTAAACTGAGTAGCAGAAAACAGAGGCCACACTTTCTGG  3120
GAAATTTTAAAGGAAGAAACCATTTTTAATGAGATGAAAATATTTAACGAATTTAAAAAG  3180
CTAATGACAATTTTGAGAAAAGGTTTGGGATGTATATTGCTATGTAATTTAATAAACTGA  3240
TTTTATGGATAT
```

Fig. 2. **Organization, nucleotide sequence, and predicted amino acid sequence of the *DCRA* gene.** A: Genomic organization of the *DCRA* gene. Exons are shown as black boxes. The sizes of the exons are shown. The eight exons extended for 44,109 bp of the genomic DNA. B: Nucleotide sequence and predicted amino acid sequence of the *DCRA* cDNA. The termination codon is indicated by an asterisk. A possible polyadenylation signal is underlined. The numbers on the right refer to the nucleotide and amino acid sequences.

RACE and 3′ RACE experiments, we finally obtained a full- or nearly full-length cDNA clone. The PCR products were directly sequenced and the data were compared with the genomic DNA sequence data. The cDNA sequence data completely matched those of the genomic DNA except for an additional G in the cDNA at position 68. This insertion is located in the 5′ UTR, and it might be a polymorphism, although further studies are required. The 3,252 bp cDNA consists of eight exons and extends for 44,109 bp of the genomic DNA (Fig. 2A). An open reading frame (ORF) of 297 amino acid residues was predicted from the cDNA sequence starting at nucleotide position 240 and terminating at position 1131. Three in-frame stop codons are located at 117 nt, 207 nt, and 228 nt upstream of the initiation codon. The poly A signal (AATAAA) is located at 25 nt upstream of the poly (A) sequence (Fig. 2B). The gene was designated as *DCRA* (*D*own syndrome *C*ritical *R*egion gene *A*).

Isolation of the Mouse Homologue of DCRA—The cDNA sequence of *DCRA* showed significant homology to that of a mouse EST (GenBank accession No. AA051569) on a homology search of non-redundant nucleotide database using the BLASTN program. We isolated the mouse homologue of *DCRA* (designated as *Dcra*) from mouse brain by the RACE method using primers prepared from the mouse EST sequence. The *Dcra* cDNA has a length of 2,157 bp, which is significantly shorter than that of the human *DCRA*. The mouse *Dcra* mRNA was found to have

Fig. 3. **Structures and amino acid comparison of *DCRA* and *Dcra*.** For amino acid comparison of *DCRA* and *Dcra*, identical amino acids are shown by black boxes.

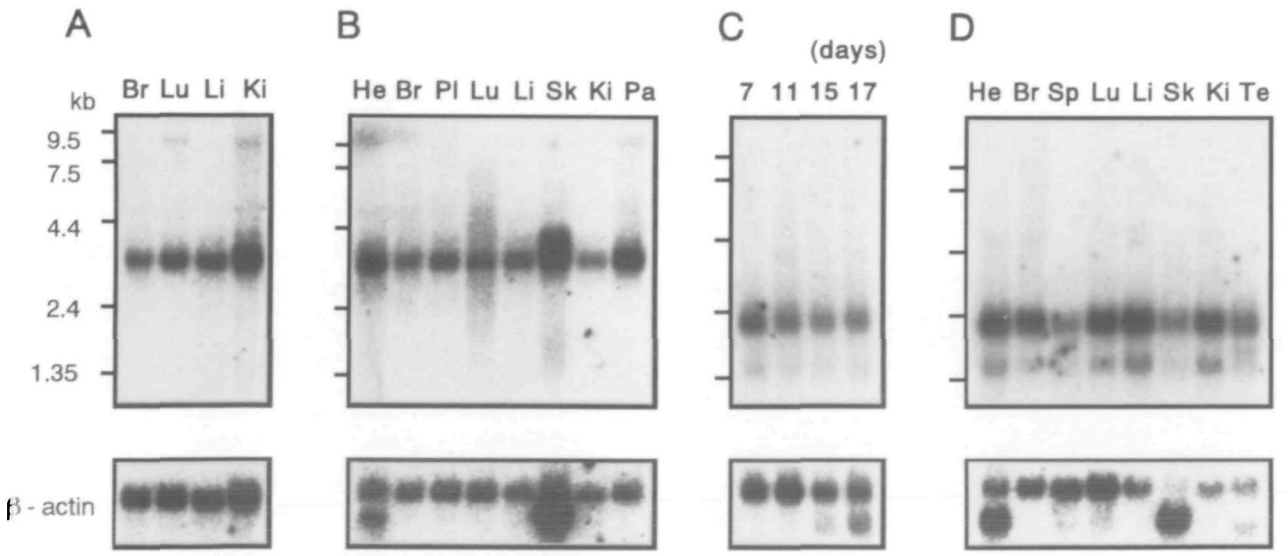


Fig. 4. **Northern blot analysis of the *DCRA* gene.** A ^{32}P-labeled PCR product of *DCRA* (nucleotide positions 551-1014) was hybridized to multiple tissue Northern blots (Clontech) containing 2 μg of poly(A)$^+$ RNA in each lane. A, human fetus (Br, brain; Lu, lung; Li, liver; Ki, kidney); B, human adult (He, heart; Br, brain; Pl, placenta; Lu, lung; Li, liver; Sk, skeletal muscle; Ki, kidney; Pa, pancreas); C, mouse embryo; D, mouse adult (He, heart; Br, brain; Sp, spleen; Lu, lung; Li, liver; Sk, skeletal muscle; Ki, kidney; Te, testis).

Fig. 5. **Comparison of the amino acid sequences of DCRA, mouse Hβ58, and *Arabidopsis thariana* EST (110D21T).** Identical amino acids are shown by black boxes.

shorter 5′ and 3′ UTRs. The poly A signal (AATAAA) is located 23 nt upstream of the poly(A) sequence. Its ORF was predicted from nucleotide positions 51 to 941 and has 297 amino acid residues, as in the case of the human *DCRA*, that shows 92% homology to the human DCRA in amino acid residues (Fig. 3).

Expression Patterns of DCRA and Dcra—Northern blot analysis revealed that a *DCRA* transcript of approximately 3.3 kb in length was expressed in a variety of tissues, including fetal brain, heart, lung, liver, and kidney, and adult brain, heart, placenta, lung, liver, skeletal muscle, kidney, and pancreas. In addition, an approximately 4 kb transcript was detected in adult skeletal muscle (Fig. 4, A and B). In mouse, an approximately 2.2 kb transcript was detected in embryo (at least from day 7 to 17) and adult heart, brain, spleen, lung, liver, skeletal muscle, kidney, and testis using the human probe (Fig. 4, C and D). Another shorter transcript (1.4 kb) was detected in all mouse tissues examined. It might be alternatively spliced, or transcribed from a different start site.

Homology Search of DCRA—The amino acid sequence of DCRA was subjected to a homology search in a non-redundant protein database (constructed from SWISS-PROT, PIR, PRF, GenPept, and GenPept-upd) and a non-redundant nucleotide database using the BLASTP and TBLASTN programs, respectively. The results showed that DCRA exhibits significant homology to mouse Hβ58 (68/297=23% identity, Ref. *10*), and *Saccharomyces cerevisiae* PEP8 (50/297=17% identity, Ref. *11*). *Arabidopsis thariana* cDNA clone 110D21T7 (*12*) also shows significant homology to the carboxy-terminal region of DCRA (34/117=29% identity, Fig. 5). The probability of identity to the carboxy-terminal 91 amino acids of the DCRA was 3.8e-05 (Hβ58), 0.0077 (PEP8), and 2.5e-12 (the partial sequence of the *A. thariana* cDNA clone 110D21T7). Mouse Hβ58 has been shown to be essential for embryogenesis in the mouse (*10, 13*). The significant homology of DCRA to Hβ58 implies that DCRA and Hβ58 belong to the same gene family. PEP8 is the yeast homologue of Hβ58, and its product plays a role in the protein sorting pathway from the cytoplasm to vacuoles. These observations suggested some important role of DCRA in human development.

These results imply that over expression of *DCRA* causes certain physiological changes which might be involved in the pathogenesis of the Down syndrome, although the real function of *DCRA* remains to be elucidated.

We wish to thank Dr. S. Kuhara for calculation of the amino acid sequence homology. We also thank K. Murakami and Dr. H. Tei for their help in the computer analysis.

REFERENCES

1. Hassold, T.J. and Jacobs, P.A. (1984) Trisomy in man. *Annu. Rev. Genet.* **18**, 69-97
2. Korenberg, J.R., Chen, X.-N., Schipper, R., Sun, Z., Gonsky, R., Gerwehr, S., Carpenter, N., Daumer, C., Dignan, P., Disteche, C., Graham, Jr., J.M., Hugdins, L., McGillivray, B., Miyazaki, K., Ogasawara, N., Park, J.P., Pagon, R., Pueschel, S., Sack, G., Say, B., Schuffenhauer, S., Soukup, S., and Yamanaka, T. (1994) Down syndrome phenotypes: The consequences of chromosomal imbalance. *Proc. Natl. Acad. Sci. USA* **91**, 4997-5001
3. Korenberg, J.R., Kawashima, H., Pulst, S.-M., Ikeuchi, T., Ogasawara, N., Yamamoto, K., Schonberg, S.A., West, R., Allen, L., Magenis, E., Ikawa, K., Taniguchi, N., and Epstein, C.J. (1990) Molecular definition of a region of chromosome 21 that causes features of the Down syndrome phenotype. *Am. J. Hum. Genet.* **47**, 236-246
4. Tsukahara, F., Hattori, M., Muraki, T., and Sakaki, Y. (1996) Identification and cloning of a novel cDNA belonging to tetratricopeptide repeat gene family from Down syndrome-critical region 21q22.2. *J. Biochem.* **120**, 820-827
5. Ohira, M., Ichikawa, H., Suzuki, E., Iwaki, M., Suzuki, K., Saito-Ohara, F., Ikeuchi, T., Chumakov, I., Tanahashi, H., Tashiro, K., Sakaki, Y., and Ohki, M. (1996) A 1.6-Mb P1-based physical map of the Down syndrome region on chromosome 21.

Genomics **33**, 65-74
6. Hattori, M., Tsukahara, F., Furuhata, Y., Tanahashi, H., Hirose, M., Saito, M., Tsukuni, S., and Sakaki, Y. (1997) A novel method for making nested deletions and its application for sequencing of 300 kb region of human APP locus. *Nucleic Acids Res.* **25**, 1802-1808
7. Uberbacher, E.C. and Mural, R.J. (1991) Locating protein-coding regions in human DNA sequences by a multiple sensor-neural network approach. *Proc. Natl. Acad. Sci. USA* **88**, 11261-11265
8. Altschul, S.F., Gish, W., Miller, W., Myers, E.W., and Lipman, D.J. (1990) Basic local alignment search tool. *J. Mol. Biol.* **215**, 403-410
9. Cheng, J.-F., Boyartchuk, V., and Zhu, Y. (1994) Isolation and mapping of human chromosome 21 cDNA: progress in constructing a chromosome 21 expression map. *Genomics* **23**, 75-84
10. Radice, G., Lee, J.J., and Costantini, F. (1991) Hβ58, an insertional mutation affecting early postimplantation development of the mouse embryo. *Development* **111**, 801-811
11. Bachhawat, A.K., Suhan, J., and Jones, E.W. (1994) The yeast homolog of H$\langle\beta\rangle$58, a mouse gene essential for embryogenesis, performs a role in the delivery of proteins to the vacuole. *Genes Dev.* **8**, 1379-1387
12. Newman, T., de Bruijn, F.J., Green, P., Keegstra, K., Kende, H., McIntosh, L., Ohlrogge, J., Raikhel, N., Somerville, S., Thomashow, M., Retzel, E., and Somerville, C. (1994) Genes galore: a summary of methods for accessing results from large-scale partial sequencing of anonymous Arabidopsis cDNA clone. *Plant Physiol.* **106**, 1241-1255
13. Lee, J.J., Radice, G., Perkins, C.P., and Costantini, F. (1992) Identification and characterization of a novel, evolutionarily conserved gene disrupted by the murine *Hβ58* embryonic lethal transgene insertion. *Development* **115**, 277-288