Am. J. Hum. Genet. 65:50–58, 1999

X-Linked Dyskeratosis Congenita Is Predominantly Caused by Missense Mutations in the DKC1 Gene

S. W. Knight,[1,*] N. S. Heiss,[2,*] T. J. Vulliamy,[1] S. Greschner,[2] G. Stavrides,[1] G. S. Pai,[3] G. Lestringant,[4] N. Varma,[5] P. J. Mason,[1] I. Dokal,[1] and A. Poustka[2]

[1]Department of Haematology, Imperial College School of Medicine, Hammersmith Hospital, London; [2]Department of Molecular Genome Analysis, Deutsches Krebsforschungszentrum, Heidelberg; [3]Department of Pediatrics, Division of Genetics and Child Development, Medical University of South Carolina, Charleston; [4]Al-Ain Medical District, Tawam Hospital, Abu Dhabi, United Arab Emirates; and [5]Department of Haemotology, Postgraduate Institute of Medical Education and Research, Chandigarh, India

Summary

Dyskeratosis congenita is a rare inherited bone marrow–failure syndrome characterized by abnormal skin pigmentation, nail dystrophy, and mucosal leukoplakia. More than 80% of patients develop bone-marrow failure, and this is the major cause of premature death. The X-linked form of the disease (MIM 305000) has been shown to be caused by mutations in the DKC1 gene. The gene encodes a 514-amino-acid protein, dyskerin, that is homologous to *Saccharomyces cerevisiae* Cbf5p and rat Nap57 proteins. By analogy to the homologues in other species, dyskerin is predicted to be a nucleolar protein with a role in both the biogenesis of ribosomes and, in particular, the pseudouridylation of rRNA precursors. We have determined the genomic structure of the DKC1 gene; it consists of 15 exons spanning a region of 15 kb. This has enabled us to screen for mutations in the genomic DNA, by using SSCP analysis. Mutations were detected in 21 of 37 additional families with dyskeratosis congenita that were analyzed. These mutations consisted of 11 different single-nucleotide substitutions, which resulted in 10 missense mutations and 1 putative splicing mutation within an intron. The missense change A353V was observed in 10 different families and was shown to be a recurring de novo event. Two polymorphisms were also detected, one of which resulted in the insertion of an additional lysine in the carboxy-terminal polylysine domain. It is apparent that X-linked dyskeratosis congenita is predominantly caused by missense mutations; the precise effect on the function of dyskerin remains to be determined.

Received November 16, 1998; accepted for publication April 22, 1999; electronically published May 27, 1999.

Address for correspondence and reprints: Dr. A. Poustka, Department of Molecular Genome Analysis, Deutsches Krebsforschungszentrum (DKFZ), Im Neuenheimer Feld 280, 69120 Heidelberg, Germany. E-mail: a.poustka@dkfz-heidelberg.de

* These authors contributed equally to this work

0002-9297/99/6501-0009$02.00

Introduction

X-linked dyskeratosis congenita (DC) is a rare bone marrow–failure syndrome characterized by the early onset of abnormal skin pigmentation, nail dystrophy, and mucosal leukoplakia (Drachtman and Alter 1992; Dokal 1996). The X-linked form of the disease (MIM 305000) is the most common, with 90% of patients being male, although both autosomal recessive (MIM 224230) and autosomal dominant (MIM 127550) forms are recognized. Progressive bone-marrow failure develops in >40% of patients by age 10 years, increases to >80% by age 30 years, and is the major cause of premature death (Knight et al. 1998*a*). There is also an increased incidence of a range of malignancies (Drachtman and Alter 1992). Chromosomal instability is observed in the fibroblasts and bone marrow of some patients (Pai et al. 1989; Kehrer et al. 1992; Dokal and Luzzatto 1994; Demiroglu et al. 1997), although lymphocytes are not sensitive to the clastogenic agents mitomycin C and diepoxybutane, with respect to increases in chromosome breakage (Coulthard et al. 1998). Carriers of the X-linked form of the disease show a skewed X-chromosome inactivation pattern in peripheral-blood lymphocytes (Vulliamy et al. 1997; Devriendt et al. 1997; Ferraris et al. 1997). The candidate-gene region for the X-linked form of the disease was originally mapped to Xq28 (Connor et al. 1986) and was further refined to a 1.4-Mb region in the distal end of Xq28 (Knight et al. 1998*b*). The DKC1 gene was subsequently positionally cloned as the result of the identification of a partial gene deletion in one male DC patient (Heiss et al. 1998). Five additional mutations were subsequently detected in the DKC1 cDNA of other patients (Heiss et al. 1998). The partial gene deletion was shown to result in the loss of the last exon of the gene, encoding the terminal 22 amino acids, but yields a viable transcript (Vulliamy et al., in press).

The DKC1 gene encodes a 514-amino-acid protein, dyskerin, that is highly conserved in evolution. Homologous proteins have been identified in other species, in-

Table 1

Primers Used in SSCP Analysis

Exon	Primer Designation, Sequence		Annealing Temperature (°C)	Size of PCR Fragment (bp)
	Forward	Reverse		
1	ex1F, TCGGCTGTGGACCGGGCG	ex1R, CGGGAACCAGAGGGAGGC	65	158
2	ex2F, AATCCATTTCCTACCTGCCC	ex2R, CAAAGCTGGCCCATTCCTTG	63	276
3	ex3F, AAAGGCATACATTTCCATGG	ex3R, CAAGGATGCCAGCAGTAAG	62	268
4	ex4F, GCCACATAGTGGTACTGACTC	ex4R, CCTGAATAGCTGATGTGAAAG	60	242
5	ex5F, GATTTGTTGTTTCACTGGAGC	ex5R, TTCACTCTAGCCAGTCCTTC	62	288
6	ex6F, GGAGTGACTGAGCATATAAG	ex6R, AACCCATCTCCAGATGTTTAG	60	219
7	ex7F, GCTGCAGCCAGCCTGGACC	ex7R, AGTCTTCAACTTCAAGGCATC	68	293
8	ex8F, GTGGCTCAGATGAAGGATAAC	ex8R, GAGTGCCATCAAAGGCTGGC	62	284
9	ex9F, GGTCTGATGGGCTGAGATAC	ex9R, GAGCAAGCGTCATCTTTGGAG	63	264
10	ex10F, CACTCCCTTGTTGTCCTCC	ex10R, TATATACACCTAGTATGTAAC	60	272
11	ex11F, TAAAGTGGCATACAACAGTAG	ex11R, ACCTGGCAGGGCACGCAAC	65	242
12	ex12F, ATTCTTTGTAGTCACCATGCC	ex12R, AGCAAGTGTGCCGTCTCTACC	64	227
13	ex13F, CTACATAACATCAGTACTGCC	ex13R, TAAGACGAATGCCAGTGCC	62	219
14	ex14F, TACCTTTTGACTCACTGAACC	ex14R, GGTACCACCTGGGTAATTC	62	288
15	ex15F, GAACTTTGTGTCACATGCAGC	ex15R, AACATGTTTTCTCAATAAGGC	58	278

cluding rat Nap57 (Meier and Blobel 1994), *Saccharomyces cerevisiae* Cbf5p (Jiang et al. 1993), *Drosophila melanogaster* Nop60B (Philips et al. 1998), and *Kluyveromyces lactis* Cbf5p (Winkler et al. 1998). Dyskerin also has regions of homology with the class of bacterial Trub proteins (Nurse et al. 1995) and the *S. cerevisiae* PUS4 protein (Becker et al. 1997). By analogy to the predicted functions of the orthologues in other species, dyskerin is likely to be a multifunctional nucleolar protein involved in centromere function (Jiang et al. 1993, 1995), nucleocytoplasmic trafficking (Meier and Blobel 1992, 1994), rRNA transcription (Cadwell et al. 1997), the stability of H + ACA snoRNA particles (Lafontaine et al. 1998), ribosome biosynthesis (Cadwell et al. 1997; Lafontaine et al. 1998; Luzzatto and Karadimitris 1998), and RNA pseudouridylation (Lafontaine et al. 1998). We have now determined the genomic structure of the DKC1 gene, which has allowed for the development of an SSCP mutation-screening strategy in male DC patients.

Patients and Methods

Patients

DNA samples from 37 families with DC were analyzed; 14 of these families were multiplex families, with two or more affected males, and the remaining 23 were sporadic families, with a single affected male patient. Of the 37 families, 36 were recruited from the Dyskeratosis Congenita Registry, and the other family, pedigree NIGMS 515, was from the Coriell Cell Repository. In family DCR-046, a sample from the proband was not available, so the mother of the proband was used in the screening. Ethical approval was obtained from the local medical ethical committee, for the use of patient samples in this study.

Determination of the Genomic Structure of the DKC1 Gene

The genomic structure of the DKC1 gene was determined by the sequencing of subcloned fragments from cosmid Qc11G6 and genomic PCR fragments. PCR was performed with an initial denaturation of 94°C for 2 min, then 35 cycles each of 96°C for 30 s, 60°C for 30 s, and 72°C for 3 min, followed by 72°C for 10 min. Intronic fragments were gel purified, TA-cloned (pGEM-TEasy; Promega), and sequenced. Sequences were edited and assembled by the GAP4 program (Staden). For accuracy, most sections of the gene were sequenced from at least two independent clones, and a minimum threefold sequence redundancy was ensured.

SSCP Analysis

Each of the exons was screened for sequence changes in every patient. Primers flanking each exon were designed; the primer sequences, the corresponding sizes of fragments, and the PCR annealing temperatures are given in table 1. A [^{32}P]-labeled fragment corresponding to each exon was generated by a 20-μl PCR reaction performed in the presence of 1 μM of each primer pair, together with 0.01 μl of [^{32}P]-dCTP. To 3.5 μl of the PCR reaction, 2 μl of formamide loading buffer (100% formamide, 1 mM EDTA, 1 mg of bromophenol blue/ml, and 1 mg of xylene cyanol/ml) was added. Each sample was heat-denatured, placed on ice, and then loaded onto a nondenaturing polyacrylamide gel by a vertical electrophoresis apparatus (S2; Gibco-BRL). Two sets of running conditions were used: (1) a 6% acrylam-

ide (99:1 acrylamide:bis-acrylamide), 1 × TBE, 10% glycerol gel run for 15–19 h at 15 mA at room temperature; and (2) a 4% acrylamide (37:1 acrylamide:bis-acrylamide), 1 × TBE, 10% glycerol gel run for 17-19 h at 15 mA at 4°C. The gels were then dried and exposed to x-ray film for 6–24 h. Any patient sample that showed a shift on the SSCP gel was then reamplified in the absence of [^{32}P]-dCTP, and the PCR fragment was purified by use of a Qiagen column. The samples were then sequenced on an ABI 373 automatic sequencer. The locations of the sequence changes were numbered with respect to the DKC1 cDNA sequence (AJ224481), with the A of the ATG numbered as "+1."

Segregation Analysis and Polymorphism Screening

Confirmation of any sequence changes detected was performed by the appropriate restriction-enzyme digestion, when available (table 2). For those pedigrees in which additional family members were available, it was possible to investigate the segregation of any mutation with the disease by the use of the appropriate restriction digestion. To investigate the possibility that any sequence change represented a polymorphism, 50 normal U.K. white women were screened; this was not done for those sequence changes shown to be de novo sequence changes. The recurrent A353V change was investigated by amplification using primers ex11F and ex11R, followed by a *Msp*A1I digestion; the substitution C to T at position 1058 results in the loss of a MspA1I restriction site. The polymorphism screening for the other sequence changes, in exons 1, 3, 4 , 11, and 12, was performed by SSCP as described above. In the case of the insertion polymorphism in exon 15, a set of individuals of either Greek or Turkish origin were also analyzed. The polymorphic substitution G→A in the 3′ UTR at position 1551 was investigated by amplification with primers ex15F and ex15R, followed by *Hae*III digestion. The presence of an additional AAG codon in exon 15 was investigated by [^{32}P]-labeled amplification using primers ex15F and ex15R, followed by analysis of the samples on a denaturing 6% polyacrylamide-urea gel. The Xq28 haplotype analysis was performed by use of polymorphic markers DXS52 and sKK5, as described elsewhere (Richards et al. 1991; Knight et al. 1998*b*).

Results

Genomic Structure of the DKC1 Gene

The DKC1 gene consists of 15 exons spanning almost to 15 kb (fig. 1). The size range of the internal exons is 65–185 bp. The first exon consists of a 92-bp 5′ UTR and 16 bp of coding region; the last exon is 908 bp long and comprises a 839-bp 3′ UTR (table 3 and fig. 1). Overall, the introns are small (size range 194–2251 bp) (table 3). A sequence gap of 600 bp remains in intron 11, and the size of this intron was estimated on the basis of the PCR product. The sequences of the splice junctions are given in table 3. The GenBank accession numbers for the entire genomic sequence are AJ0101395 (exons 1–11) and AJ0101396 (exons 12–15). By inspection, all of the 5′ splice-donor sites and 13 of the 14 splice-acceptor sites conform to the known consensus

Table 2

DKC1 Mutations in Families with DC

Family	Ethnic Origin	Family History	Exon	DNA Change	Protein Change	Restriction Enzyme	Inheritance
DCR-004	United Kingdom	Sporadic	11	1058C→T	A353V	*Msp*AI (−)	Inherited
DCR-005	United Kingdom	Multiplex	11	1058C→T	A353V	*Msp*AI (−)	Inherited
DCR-006	Italy	Multiplex	11	1058C→T	A353V	*Msp*AI (−)	Inherited
DCR-008	United Kingdom	Sporadic	3	115A→G	K39E	*Taq*I (+)	De novo in mother
DCR-009	Italy	Multiplex	10	961C→G	L321V	*Mbo*II (−)	De novo in grandparents
NIMGS 515	United States	Multiplex	4	196A→G	T66A	No sample available	Unknown
DCR-020	France	Sporadic	11	1058C→T	A353V	*Msp*AI (−)	De novo in proband
DCR-021	Austria	Sporadic	11	1050G→A	M350I	No sample available	Unknown
DCR-027	India	Multiplex	12	1204G→A	G402R	*Nla*III (−)	Unknown
DCR-029	United States	Sporadic	11	1058C→T	A353V	*Msp*AI (−)	De novo in proband
DCR-030	United Kingdom	Multiplex	11	1058C→T	A353V	*Msp*AI (−)	Inherited
DCR-031	United Kingdom	Sporadic	11	1049T→C	M350T	*Ase*I (−)	Inherited
DCR-033	Germany	Multiplex	4	194G→C	R65T	*Mae*III (+)	Inherited
DCR-038	Turkey	Sporadic	3	121G→A	E41K	*Hin*fI (−)	Inherited
DCR-041	United Arab Emirates	Multiplex	1	5C→T	A2V	*Aci*I (−)	Unknown
DCR-044	United States	Sporadic	11	1058C→T	A353V	*Msp*AI (−)	De novo in proband
DCR-046	United Kingdom	Sporadic	11	1058C→T	A353V	*Msp*AI (−)	De novo in grandparents
DCR-050	Spain	Sporadic	IVS2	−5C→G	…	*Taq*I (−)	Inherited
DCR-051	United States	Multiplex	11	1058C→T	A353V	*Msp*AI (−)	Inherited
DCR-055	Spain	Multiplex	11	1058C→T	A353V	*Msp*AI (−)	De novo in proband
DCR-056	United States	Sporadic	11	1058C→T	A353V	*Msp*AI (−)	De novo in proband

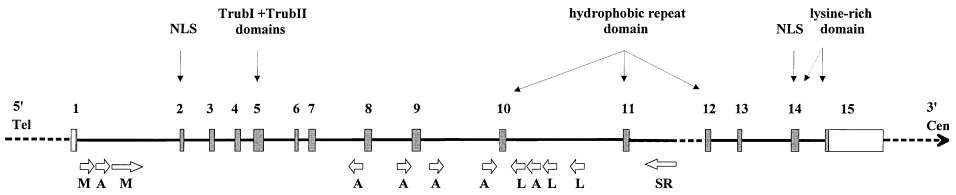

Figure 1 Schematic representation of gene structure of DKC1. The coding exons are represented by gray-shaded boxes, numbered "1"–"15," and the 5′ and 3′ UTRs are represented as unshaded boxes. The gap in intron 11 is indicated by a partially dashed line. The direction of the repetitive elements with respect to the gene are shown by the unblackened arrows; A = *Alu,* M = MER, L= LINE, and SR = SINE-R11. The locations of the NLSs, Trub domains, the hydrophobic repeat, and the lysine-rich domain are indicated. The gene is transcribed in a telomere-to-centromere direction.

sequence (table 3) (Mount 1982). The one exception was the 3′ splice-site acceptor for intron 2; in this case there is the rare occurrence of an adenine at the −3 position (table 3) (Mount 1982). The following repetitive elements were identified by RepeatMasker 2 (RepeatMasker documentation): two medium-reiteration-frequency repeats (MERs), six *Alu* elements, and the 3′ UTRs of three long interspersed repetitive elements (LINEs) (see fig. 1). The simple repeat sequence $(TA)_6AATA(TG)_4(TA)_3TG(TA)_2T(TA)_3(TGTA)_2T(TA)_2$-$(TGTA)_2TAT(TA)_6$ was found in intron 1.

A 643-bp segment in intron 11 of the DKC1 gene showed homology to the short interspersed repeat (SINE)–type family of nonviral retroposons, called "SINE-R" (fig. 1) (Ono et al. 1987; Zhu et al. 1992). The DKC1 retroposon shows highest identity to SINE-R11 (95%). SINE-R retroposons have a 5′→3′ direction, and the DKC1 SINE-R element lies in an orientation the reverse of that of the DKC1 gene. It consists of a GC-rich stretch of 40-bp tandem repeats followed by a polypurine tract, a glucocorticoid-responsive element, a poly(A) signal, and a poly(A) tail. This is characteristic of SINE-R elements, which are derived largely from the human endogenous retrovirus HERV-K10 long terminal repeats (Zhu et al. 1992). In intron 11 of DKC1 a 21-bp direct repeat sequence (CCAGCATTGACACCA-TTCTAA) flanks the GC-rich stretch and the boundaries of the sequence gap. This may explain why it was not possible to sequence this section of the gene.

SSCP Screening of the DKC1 Gene in Patients with DC

Primers were designed from the genomic sequence of the DKC1 gene such that each exon could be amplified and screened, by SSCP, for sequence changes (table 1). Every sample from patients was screened, in each exon, with two different sets of SSCP conditions. Mutations were detected in 21 of the 37 families analyzed. All of these were single-nucleotide substitutions and comprised 10 different missense coding changes and 1 change within an intron. For those mutations that were not de novo, 50 normal white U.K. women were screened to exclude polymorphisms; in each case, none of the normal women carried these sequence changes (data not shown).

The nucleotide substitution C→T at position 1058 in exon 11, resulting in the missense mutation A353V, was detected in 11 different families (DCR-004, DCR-005, DCR-006, DCR-021, DCR-030, DCR-031, DCR-044, DCR-046, DCR-051, DCR-055, and DCR-056) (table 3). These pedigrees were from disparate geographical locations. This change was shown to be a de novo event in the probands in families DCR-020, DCR-029, DCR-044, DCR-055, and DCR-056 (fig. 2). In addition, this change was also shown to be de novo in the mother of the proband in family DCR-046. Also in exon 11, two different nucleotide substitutions affecting the same codon were detected: nucleotide 1049T→C (M350T) in pedigree DCR-031 and nucleotide 1050G→A (M350I) in pedigree DCR-021. A maternal sample was available only from pedigree DCR-031, and in this case the mutation was shown to have been inherited from the mother. Three different single-nucleotide substitutions were detected by primers spanning exon 3, which resulted in the missense mutations K39E in family DCR-008, E41K in family DCR-038, and a C→G substitution at position −5 of intron 2 in family DCR-050 (table 2). The mutations in families DCR-008, DCR-038, and DCR-050 were each shown to have been maternally inherited. Two different missense mutations were detected in exon 4; these were single-nucleotide substitutions resulting in amino changes R65T (in family DCR-033) and T66A (in family NIGMS-515). Two further single-nucleotide substitutions leading to missense mutations were observed: L321V in exon 10 (family DCR-009) and G402A in exon 12 (family DCR-027). In family DCR-009 (fig. 3), the mutation was detected in the proband's mother (II-8) and in one of his aunts (II-1) but was not observed in either of the grandparents (I-1 and I-2), which suggested that it was a de novo event. Haplotype analysis using the markers DXS52 and sKK5 indicated that the proband's X chromosome had been inherited from the maternal grandfather, and it also

confirmed that a cousin of the proband (III-3) shared the haplotype with the proband (III-5) but did not have the L321V mutation (fig. 3). Considered together, these observations suggest that the maternal grandfather was a germline mosaic for the mutation.

Polymorphisms in the DKC1 Gene

A nucleotide substitution, G→A at position 1551 in exon 15, located 6 nucleotides downstream from the TAG stop codon in the 3′ UTR, was observed in three families with DC. This sequence change was shown to be polymorphic, with a frequency of 11/100 for the 1551A allele when a normal white U.K. population was screened. In family DCR-008, the 1551A allele, resulting in the loss of a *Hae*III restriction site, was observed, in *cis,* with the missense coding change K39E, in the proband. The combination of the DXS52 haplotype and the DKC1 mutation analysis in this family confirmed that an unaffected cousin of the proband had inherited the same DXS52 allele and the 1551A allele without the K39E. These data confirm that the 1551A was not disease causing and that the K39E mutation was a de novo event.

A second polymorphism was due to an insertion of an additional AAG triplet in a stretch of seven AAGs within exon 15 (nucleotides 1494–1514). This polymorphism was observed, in *cis,* with the coding change E41K, in family DCR-038 from Turkey. The $(AAG)_8$ allele was not observed in 100 normal X chromosomes from a white U.K. population. However, the $(AAG)_8$ allele was observed in 3 of 53 X chromosomes from individuals of Greek origin and in 3 of 29 X chromosomes from individuals of Turkish origin, including a hemizygote normal male (data not shown).

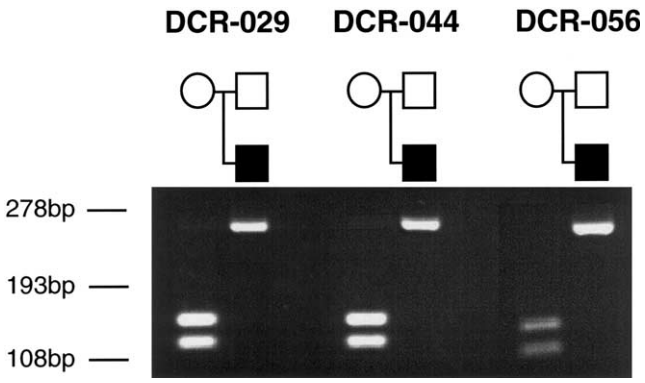

Figure 2 Illustration of recurrent de novo A353V mutation. DNA from the proband and the mother was amplified in sporadic families DCR-029, DCR-044, and DCR-056 and then was digested with *Msp*A1I. The presence of the mutation results in the loss of the restriction site. The size of molecular-weight makers is given.

Discussion

The DKC1 gene consists of 15 exons spanning a 15-kb region (fig. 1) and is located 1 kb from the MPP1 gene, in a tail-to-tail configuration (Vulliamy et al., in press). The identification of the genomic structure of the DKC1 gene has facilitated the screening of 37 male patients with DC for mutations. SSCP analysis identified DKC1 mutations in 21 of these 37 patients. A total of 11 different single-nucleotide substitutions were detected, which represented 10 missense coding mutations and 1 intron mutation (table 2). The failure to detect mutations in the other patients may be due to either inefficiency of the SSCP analysis or the fact that some of these other patients have DKC1 mutations outside the coding and splice-site regions that were the subject

Table 3

Sequence of Splice-Acceptor and Splice-Donor Sites of the DKC1 Gene

Splice-Acceptor Sequence	Exon	Exon Size (bp)	Splice-Donor Sequence	Intron Size (bp)
	1	108	ATGGCGGATGCGGAAG/gtaagggctgcaggct	1,915
gtgtttgtttttttag/TAATTATTTTGCCAAA	2	68	AGAAGAAGATGTAGCC/gtgagtagtaatgtgt	477
ttctgtcctgtcgaag/GAAATACAACACGCTG	3	87	GCCCCTTTTGCTAAAG/gtatgtggttaaaatc	376
tgtcgttgctttgtag/AATTTTGATAAGCTGA	4	92	TTGGGGACTATATCAG/gtaagtgttgggagga	217
ttttttttccattccag/GACAGGTTTCATTAAT	5	185	TCACAACAGAGTGCAG/gtatgtgggagaggga	596
ttttgtcatttttcag/GCAAAGAGTATGTGGG	6	65	GACCCAGCTTTCTAGG/gtaagtctgcaattgt	194
ttgtgtttgttcttag/GCCCTAGAAACTCTGA	7	127	CCTGAAAGAAGATTAG/gtgagtcattaggcct	917
tgtgtgtgtatcttag/GAATCTTTTGGGTGAG	8	131	AGTCATGAGTGAAAAG/gtatgtgttacggggc	733
catggctgcttttcag/GACCACATGGTGACAA	9	144	TATGAAAGACAGTGCA/gtaagttccgggttag	1,446
ggtgccatctctgcag/GTAAATGCCATCTGCT	10	121	GAAGCAATCTGCATGG/gtaagaggggatgttt	2,251
tattcaatgcctgtag/CTATTGCATTAATGAC	11	119	GGGTTTAGGCCCAAAG/gtaagtggtactgggt	1,350
cttgatgttccaccag/GCAAGTCAGAAGAAGC	12	104	AGTATGTTGACTACAG/gtgagggcaggatgtt	489
gctacctctttttcag/TGAGTCTGCCAAAAAA	13	79	AGCAAAAACTGCGAAG/gtgagtggcatgctgt	913
attgctttcatctcag/CGGAAGCGAGAGAGTG	14	138	CGAGCCTGGAGATGGG/gtgtgtggaaacacgt	474
ttctctttctttctag/GACAGTGATACCACCA	15	908		

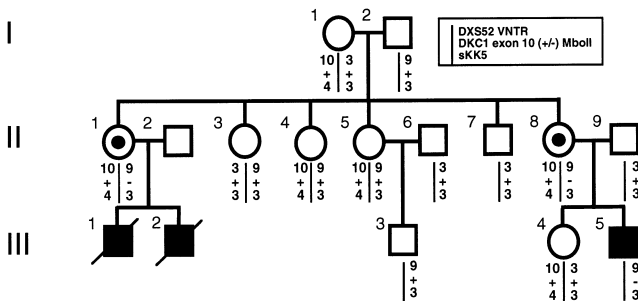

Figure 3 Haplotype and mutation segregation in family DCR-009, showing segregation of the L321V mutation, which mutation results in the loss (−) of a *Mbo*II site in exon 10. A haplotype of the markers DXS52 and sKK5 is also indicated.

of this study. Of the 16 patients in whom DKC1 mutations were not identified by SSCP, 13 were male sporadic cases. It is possible that some of the male sporadic cases have the autosomal form of the disease. Evidence for the putative autosomal DKC2 gene comes both from reports of women who have the identical phenotype (see review by Drachtman and Alter [1992]) and case reports of male-to-male transmission of DC (Gasparini et al. 1985).

One particular mutation, A353V (1058C→T), was detected in 11 families (table 3). This was the only mutation identified in those families in which it was found; and, as well, a screen of 100 normal X chromosomes failed to detect A353V as a polymorphism. In terms of its geographical distribution, there was no evidence for a founder effect, and, in fact, it was shown to be a de novo event in six of the families; however, it is possible that some of the other five families with DC and the A353V mutation share a common ancestor. This mutation represents an example of a CpG→TpG change that is known to occur at a high frequency as the result of the deamination of the methylated cytosine. The A353V mutation was detected in 30% (11/37) of the male patients with DC who were analyzed in this study; this represents 26% (11/43) of all male patients with DC who have been analyzed so far (Heiss et al. 1998). The occurrence of a common A353V mutation has a significant diagnostic consequence for the future screening for DKC1 mutations in other patients with DC.

In this study, the one sequence change detected that was not a missense coding mutation was a C→G substitution at position −5 in intron 2 in family DCR-050. It is interesting to note that the splice-acceptor site in intron 2 has an adenine at position −3 (table 3); this is unusual, since purine nucleotides at this position have been observed in only 5% of splice acceptors (Mount 1982). This sequence change was not found in 50 white U.K. women, although the possibility remains that it represents a rare polymorphism in Spain, whence this family originated. An RNA sample was not available from the proband in pedigree DCR-050; therefore, the precise affect that this sequence change has on splicing could not be determined. The C→G change at position −5 does change a pyrimidine to a purine in the consensual polypyrimidine track. Such polypyrimidine-track mutations have been described in the β-globin gene (Beldjord et al. 1998; Murru et al. 1991), the ATR-X gene (Villard et al. 1996), and the steroidogenic acute-regulatory-protein gene (Tee et al. 1995).

Two polymorphisms were observed in the DKC1 gene; one was an A→G at position 1551 within the 3′ UTR, and the other was the presence of an additional AAG triplet within exon 15. The AAG insertion was not observed in a white U.K. populations but was shown to be polymorphic in people of Greek origin and in people of Turkish origin. The AAG insertion results in the presence of an additional lysine in a stretch of eight lysines encoded by $(AAG)_7$ AAA (amino acids 498–505), which implies that the precise number of lysines in this region is not essential for the normal function of dyskerin. It is possible that this insertion is the result of a triplet-expansion mechanism similar to that which results in the expansions of the identical triplet within intron 1 of the FRDA gene (Campuzano et al. 1996).

The mutations detected in this study, together with the four missense mutations, one single-amino-acid deletion, and one carboxy-terminal deletion identified in previous studies (Heiss et al. 1998) imply that X-linked DC is predominantly caused by missense dyskerin mutations: of the 17 known mutations of dyskerin, 14 are missense changes (fig. 4). It remains a possibility that additional mutations affecting splicing and level of expression will be detected in other patients with DC. To date, no premature terminations, frameshifts, or complete deletions have been identified in the DKC1 gene, and it can be speculated that such large-scale loss of dyskerin activity would be lethal to the cells. All the missense mutations so far identified are nonconservative changes (fig. 4). It is interesting to note that two pairs of mutations affect the same codons: (1) M350T and M350I and (2) G402E and G402R. The changes cluster in two regions: 7 of 14 missense mutations are in exons 3 and 4 (amino acids 36–72), and 3 of 14 mutations, including the recurrent mutation, are in exon 11 (amino acids 350–353) (fig. 4). The precise function(s) of dyskerin remains to be determined, and therefore the effect of each of the mutations cannot be accurately assessed.

The very high level of amino acid–sequence identity between the *S. cerevisiae* Cbf5p protein and dyskerin (Heiss et al. 1998) suggests that these proteins are functional orthologues and that, therefore, analysis of the yeast protein is extremely pertinent to an understanding of the function of dyskerin. Except for A2V and A353V,

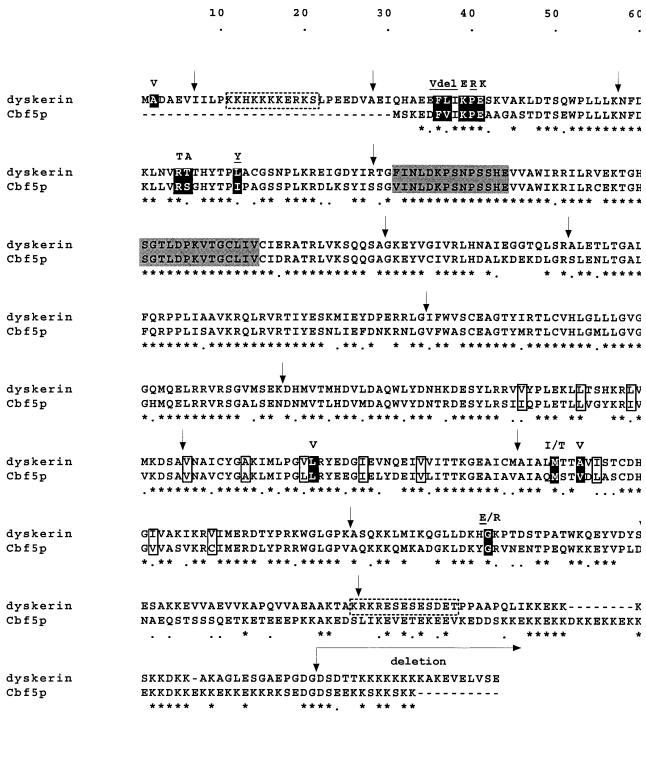

Figure 4 CLUSTAL alignment of dyskerin and Cbf5p, showing locations of all coding mutations. Mutations affecting single amino acids are represented as white letters in black boxes, and the corresponding amino acid changes are shown above the alignment. Previously identified single amino acid mutations are underlined (Heiss et al. 1998). The location of the carboxy-terminal deletion is indicated by a horizontal line (Vulliamy et al., in press). The exon boundaries are demarcated by vertical arrows. NLSs are indicated by dashed boxes, the two Trub domains are gray shaded, and amino acids forming the hydrophobic repeat domain are marked by open boxes. Stars indicate amino acids that are identical between dyskerin and Cbf5p; dots indicate different but conserved amino acids.

all the amino acid changes observed in patients with DC occur at positions that are conserved with respect to the Cbf5p protein (fig. 4). In the case of the recurrent A353V mutation, Cbf5p has a valine at the homologous position (fig. 4). In view of the fact that dyskerin and Cbf5p are so similar, it is surprising that a mutation that makes dyskerin more similar to Cbf5p would have a deleterious effect on its activity and cause the disease. None of the missense mutations alter the sequence of any of the putative phosphorylation sites or nuclear-localization signals (NLSs) (fig. 4). In addition, all of the missense mutations are located outside the bipartite Trub motif (amino acids 91–104 and 121–134), which has been shown to be conserved in other pseudouridine synthases and is thought to represent the active site of the enzyme (Koonin 1996) (fig. 4). In the case of the A2V change, the reason that it is pathogenic is not obvious.

The fact that DC is caused predominantly by missense mutations within a restricted number of regions of the protein raises the possibility that mutations affecting different regions of dyskerin may produce different disease phenotypes. For this reason, the DKC1 gene may be seen as a potential candidate for other syndromes that map to Xq28 and that have some phenotypic overlap with DC—for example, incontinentia pigmenti (IP2 [MIM 308310]) (Smahi et al. 1994; Jouet et al. 1997). In addition, DKC1 mutations may account for some of the male patients who develop idiopathic aplastic anemia, particularly those presenting at a very young age. It is hoped that a better molecular understanding of the underlying cause of DC will enable the development of therapeutic strategies and also provide insights into the mechanism of aplastic anemia in general.

Acknowledgments

We wish to thank the following physicians for providing us with samples from the families: Drs. C. Bunch, E. Burchby, H. Demiroglu, G. Forni, E. Gluckman, S. Lampert, and G. Lowe; Prof. Janka-Schaub; and Drs. J. Macrco Buades, A. Miguel-Sosa, A. Pazsy, F. Peteiro-Cartelle, I. Rodrigues, I. Roberts, Y. Ross, C. Steward, D. Schiller and A. Toutain. This work was funded by grants from the Wellcome Trust, Action Research, and Deutsche Forschungsgemeinschaft. We also want to thank Maria Mukherjee, for help with the Dyskeratosis Congenita Registry, and Prof. John Goldman, for his support of this project.

Electronic-Database Information

Coriell Cell Repository, http://locus.umdnj.edu/nigms (for family NIGMS 515)

Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nlm.nih.gov/Omim (for autosomal dominant [MIM 127550], autosomal recessive [MIM 224230], and X-linked [MIM 305000] DC and IP2 [MIM 308310])

RepeatMasker documentation, http://ftp.genome.washington.edu/RM/RepeatMasker.html (for RepeatMasker 2)

References

Becker HF, Motorin Y, Planta RJ, Grosjean H (1997) The yeast gene YNL292w encodes a pseudouridine synthase (Pus4) catalyzing the formation of Psi(55) in both mitochondrial and cytoplasmic tRNAs. Nucleic Acids Res 25:4493–4499

Beldjord C, Lapoumeroulie C, Pagnier J, Benabadji M, Krishnamoorthy R, Labie D, Bank A (1988) A novel beta-thalassemia gene with a single base mutation in the conserved polypyrimidine sequence at the 3′ end of IVS-2. Nucleic Acids Res 16:4927–4935

Cadwell C, Yoon HJ, Zebarjadian Y, Carbon J (1997) The yeast nucleolar protein Cbf5p is involved in rRNA biosynthesis and interacts genetically with the RNA polymerase I transcription factor RRN3. Mol Cell Biol 17:6175–6183

Campuzano V, Montermini L, Molto MD, Pianese L, Cossee

M, Cavalcanti F, Monros E, et al (1996) Friedreich's ataxia autosomal recessive disease caused by an intronic GAA triplet repeat expansion. Science 271:1423–1427

Connor JM, Gatherer D, Gray FC, Pirrit LA, Affara NA (1986) Assignment of the gene for dyskeratosis congenita to Xq28. Hum Genet 72:348–351

Coulthard S, Chase A, Pickard J, Goldman J, Dokal I (1998) Chromosomal breakage analysis in dyskeratosis congenita peripheral blood lymphocytes. Br J Haematol 102: 1162–1164

Demiroglu H, Alikasifoglu M, Dundar S (1997) Dyskeratosis congenita with an unusual chromosomal abnormality. Br J Haematol 97:243–244

Devriendt K, Matthijs G, Legius E, Schollen E, Blockmans D, van Geet C, Degreef H, et al (1997) Skewed X-chromosome inactivation in female carriers of dyskeratosis congenita. Am J Hum Genet 60:581–587

Dokal I (1996) Dyskeratosis congenita—an inherited bone-marrow failure syndrome. Br J Haematol 92:775–779

Dokal I, Luzzatto L (1994) Dyskeratosis congenita is a chromosomal instability disorder. Leuk Lymphoma 15:1–7

Drachtman RA, Alter BP (1992) Dyskeratosis congenita: clinical and genetic heterogeneity: report of a new case and review of the literature. Am J Pediatr Hematol Oncol 14: 297–304

Ferraris AM, Forni GL, Mangerini R, Gaetani GF (1997) Nonrandom X-chromosome inactivation in hemopoietic cells from carriers of dyskeratosis congenita. Am J Hum Genet 61:458–461

Gasparini G, Sambvani N, Cuidarelli C, Sarchi G, Di-Pietro A, Raimondi E, Romagnoni MM, et al (1985) Zinsser-Cole-Engman syndrome: two cases transmitted by a dominant autosomic gene. G Ital Dermatol Venereol 120:429–433

Heiss NS, Knight SW, Vulliamy TJ, Klauck SM, Wiemann S, Mason PJ, Poustka A, et al (1998) X-linked dyskeratosis congenita is caused by mutations in a highly conserved gene with putative nucleolar functions. Nat Genet 19:32–38

Jiang WD, Lim MY, Yoon HJ, Thorner J, Martin GS, Carbon J (1995) Overexpression of the yeast mck1 protein-kinase suppresses conditional mutations in centromere-binding protein genes cbf2 and cbf5. Mol Gen Genet 246:360–366

Jiang WD, Middleton K, Yoon HJ, Fouquet C, Carbon J (1993) An essential yeast protein, cbf5p, binds in-vitro to centromeres and microtubules. Mol Cell Biol 13:4884–4893

Jouet M, Stewart H, Landy S, Yates J, Yong SL, Harris A, Garret C, et al (1997) Linkage analysis in 16 families with incontinentia pigmenti. Eur J Hum Genet 5:168–170

Kehrer H, Krone W, Schindler D, Kaufmann R, Schrezenmeier H (1992) Cytogenetic studies of skin fibroblast cultures from a karyotypically normal female with dyskeratosis congenita. Clin Genet 41:129–134

Knight SW, Vulliamy T, Copplestone A, Gluckman E, Mason PJ, Dokal I (1998*a*) Dyskeratosis Congenita (DC) Registry: identification of new features of DC. Br J Haematol 103: 990–996

Knight SW, Vulliamy TJ, Heiss NS, Matthijs G, Devriendt K, Connor JM, D'Urso M, et al (1998*b*) 1.4 Mb candidate gene region for X-linked dyskeratosis congenita defined by combined haplotype and X chromosome inactivation analysis. J Med Genet 35:993–996

Koonin EV (1996) Pseudouridine synthases: four families of enzymes containing a putative uridine-binding motif also conserved in dUTPases and dCTP deaminases. Nucleic Acids Res 24:2411–2415

Lafontaine DLJ, Bousquet-Antonelli C, Henry Y, Caizergues-Ferrer M, Tollervey D (1998) The box H + ACA snoRNAs carry Cbf5p, the putative rRNA pseudouridine synthase. Genes Dev 12:527–537

Luzzatto L, Karadimitris A (1998) Dyskeratosis and ribosomal rebellion. Nat Genet 19:6–7

Meier UT, Blobel G (1992) Nopp140 shuttles on tracks between nucleolus and cytoplasm. Cell 70:127–138

Meier UT, Blobel G (1994) Nap57, a mammalian nucleolar protein with a putative homolog in yeast and bacteria. J Cell Biol 127:1505–1514

Mount SM (1982) A catalog of splice junction sequences. Nucleic Acids Res 10:459–472

Murru S, Loudianos G, Deiana M, Camaschella C, Sciarratta GV, Agosti S, Parodi MI, et al (1991) Molecular characterization of beta-thalassemia intermedia in patients of Italian descent and identification of 3 novel beta-thalassemia mutations. Blood 77:1342–1347

Nurse K, Wrzesinski J, Bakin A, Lane BG, Ofengand J (1995) Purification, cloning, and properties of the tRNA Ψ *55* synthase from *Escherichia coli*. RNA 1:102–112

Ono M, Kawakami M, Takezawa T (1987) A novel human nonviral retroposon derived from an endogenous retrovirus. Nucleic Acids Res 15:8725–8737

Pai GS, Yan Y, DeBauche DM, Stanley WS, Paul SR (1989) Bleomycin hypersensitivity in dyskeratosis congenita fibroblasts, lymphocytes, and transformed lymphoblasts. Cytogenet Cell Genet 52:186–189

Philips B, Billin AN, Cadwell V, Buchholz R, Erickson C, Merriam JR, Carbon J, et al (1998) The *Nop60B* gene of *Drosophila* encodes an essential nucleolar protein that functions in yeast. Mol Gen Genet 260:20–29

Richards B, Heilig R, Oberle I, Storjoham L, Horn GT (1991) Rapid PCR analysis of the st14 (DXS52) VNTR. Nucleic Acids Res 19:1944

Smahi A, Hydengranskog C, Peterlin B, Vabres P, Heuertz S, Fulchignonilataud MC, Dahl N, et al (1994) The gene for the familial form of incontinentia pigmenti (IP2) maps to the distal part of Xq28. Hum Mol Genet 3:273–278

Tee MK, Lin D, Sugawara T, Holt JA, Guiguen Y, Buckingham B, Strauss JF, et al (1995) T-A transversion 11 bp from a splice acceptor site in the human gene for steroidogenic acute regulatory protein causes congenital lipoid adrenal-hyperplasia. Hum Mol Genet 4:2299–2305

Villard L, Toutain A, Lossi AM, Gecz J, Houdayer C, Moraine C, Fontes M (1996) Splicing mutation in the ATR-X gene can lead to a dysmorphic mental retardation phenotype without α-thalassemia. Am J Hum Genet 58:499–505

Vulliamy TJ, Knight SW, Dokal I, Mason PJ (1997) Skewed X-inactivation in carriers of X-linked dyskeratosis congenita. Blood 90:2213–2216

Vulliamy TJ, Knight SW, Heiss NS, Smith OP, Poustka A, Dokal I, Mason PJ. Dyskeratosis congenita caused by a 3′ deletion: germ line and somatic mosaicism in a female carrier. Blood (in press)

Winkler AA, Bobok A, Zonneveld B, Steensma HY, Hooykaas

P (1998) The lysine-rich C-terminal repeats of the centromere-binding factor 5 (Cbf5) of *Kluyveromyces lactis* are not essential for function. Yeast 14:37–48

Zhu ZB, Hsieh S-L, Bentley DR, Campbell RD, Volanakis JE (1992) A variable number of tandem repeat locus within the human complement C2 gene is associated with a retroposon derived from a human endogenous retrovirus. J Exp Med 175:1783–1787