1.4 Mb candidate gene region for X linked dyskeratosis congenita defined by combined haplotype and X chromosome inactivation analysis

S W Knight, T J Vulliamy, N S Heiss, G Matthijs, K Devriendt, J M Connor, M D'Urso, A Poustka, P J Mason, I Dokal

Abstract

Dyskeratosis congenita (DC) is a rare inherited disorder characterised by the early onset of reticulate skin pigmentation, nail dystrophy, and mucosal leucoplakia. In over 80% of cases bone marrow failure develops and this is the main cause of early mortality. The DC1 gene responsible for the X linked form (MIM 305000) of dyskeratosis congenita has been mapped to Xq28. In order to narrow the candidate gene region, genetic linkage analysis was performed in eight X linked pedigrees using a set of markers spanning Xq28. A maximum lod score of 5.31 with no recombinations was achieved with marker DXS1073. Two recombination events were identified; one of these uses X chromosome inactivation pattern analysis to determine carrier status and haplotype analysis to fine map the recombination breakpoint. The fine mapping of these recombination events has enabled the candidate gene region for X linked dyskeratosis congenita to be defined as the 1.4 Mb interval between Xq3274 and DXS1108.

(*J Med Genet* 1998;35:993–996)

Keywords: dyskeratosis congenita; X chromosome inactivation pattern; Xq28; DC1

Department of Haematology, Imperial College School of Medicine, Hammersmith Campus, London W12 0NN, UK
S W Knight
T J Vulliamy
P J Mason
I Dokal

Department of Molecular Genome Analysis, Deutsches Krebsforschungszentrum (DKFZ), Heidelberg, Germany
N S Heiss
A Poustka

Centre for Human Genetics, University Hospital Gasthuisberg, Leuven, Belgium
G Matthijs
K Devriendt

Institute of Medical Genetics, Yorkhill Hospitals Campus, Glasgow, UK
J M Connor

International Institute of Genetics and Biophysics, Naples, Italy
M D'Urso

Correspondence to:
Dr Knight.

Received 19 February 1998
Revised version accepted for publication 23 April 1998

Dyskeratosis congenita (DC) is a rare inherited disorder characterised by the early onset of reticulate skin pigmentation, nail dystrophy, and mucosal leucoplakia.[1 2] In over 80% of patients bone marrow failure develops and this is the main cause of early mortality (I Dokal, unpublished observation).[3] X linked inheritance (MIM 305000) is the most common form of the disease although both autosomal dominant (MIM 127550) and recessive (MIM 22430) forms are also recognised. DC appears to be distinct from the other more common inherited bone marrow failure syndrome, Fanconi's anaemia, with respect to the cellular sensitivity to clastogens.[4] The DC1 gene responsible for the X linked form of the disease was mapped to Xq28 and subsequently localised to a 3.5 Mb region between DXS1684 and DXS1108.[5 6] Obligate carriers of DC are largely asymptomatic, but they do occasionally manifest limited mucocutaneous signs.[3] In common with many other X linked diseases,[7] obligate carriers of X linked DC have been shown to have a completely skewed X chromosome inactivation pattern (XCIP).[8–10] This paper redefines the DC1 candidate gene region by the identification of two crucial recombination events. One of these uses XCIP analysis to determine carrier status and haplotype analysis to fine map the recombination breakpoint. The DC1 candidate gene region can now be defined as a 1.4 Mb region between Xq3274 and DXS1108 in the distal part of Xq28.

Materials and methods

FAMILIES

Eight X linked pedigrees of European origin were used in the genetic linkage analysis. Seven of these (DCR-001, DCR-002, DCR-005, DCR-006, DCR-007, DCR-025, and DCR-030) were recruited from the Dyskeratosis Congenita Registry and the other one (NIGMS 515) was from the NIGMS Coriell Cell Repository (URL: http://www.gdb.org/coriell.html). Ethical approval was obtained from the local medical ethical committee for the use of patient samples in this study. Pedigrees DCR-001, DCR-002, DCR-005, and DCR-006 have all been described previously,[6] together with DCR-025[8] and NGMS 515.[11] DCR-007 represents the updated and extended pedigree from which linkage analysis was originally obtained.[5] DNA was extracted from peripheral blood, or in the case of NIGMS 515, from lymphoblastoid cell lines in the standard way.[12] At risk subjects were

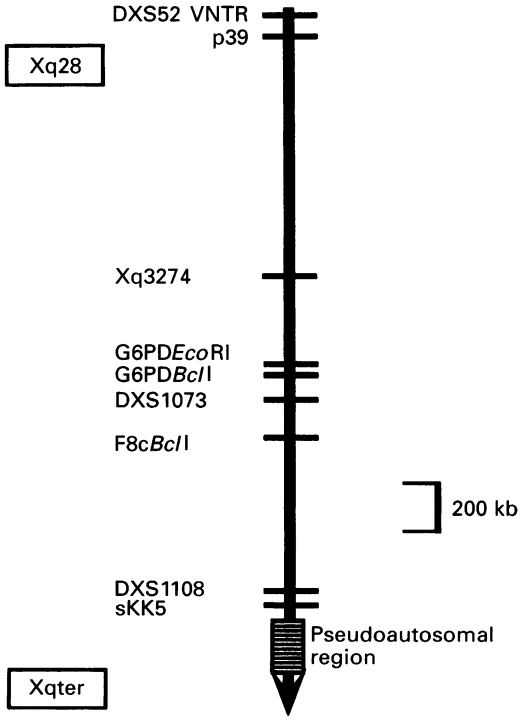

Figure 1 Schematic representation of distal Xq28 showing the relative locations of the polymorphic markers used in the genetic linkage analysis.

Table 1 Two point lod score between DC1 and five markers in distal Xq28

Locus	Lod score at recombination fraction of						
	0.00	0.01	0.05	0.10	0.20	0.30	0.40
DXS52VNTR	−∞	8.89	8.81	8.10	6.24	4.06	1.80
G6PD *BclI*	3.43	3.37	3.12	3.79	2.08	1.33	0.57
DXS1073	5.31	5.22	4.87	4.40	3.39	2.30	1.14
F8c *BclI*	4.13	4.05	3.74	3.32	2.41	1.41	0.44
DXS1108	−∞	4.17	4.45	4.21	3.36	2.28	1.09

considered to be unaffected if they were over 15 years of age and without nail dystrophy, leucoplakia, or skin abnormalities.

GENETIC LINKAGE ANALYSIS

The relative location of the polymorphic markers used are illustrated in fig 1. Details of primers and polymerase chain reaction (PCR) conditions for the polymorphic markers DXS52 VNTR, p39(9120/9121), DXS1073, F8c *BclI*, and DXS1108 are all from GDB (URL:http://gdbwww.gdb.org:80/gdb/gdbtop.html). Two single nucleotide polymorphisms (SNP) in the glucose-6-phophate dehydrogenase (G6PD) region were used: G6PD *BclI* was analysed with a *BclI* digestion following amplification with primers 1311 and F and G6PD *Eco*RI was analysed by an *Eco*RI digestion following amplification using the primers LPTVALU and RPTVALU, as described previously.[13 14] Primers were also designed to amplify an additional previously described CA_n repeat, sKK5, which was located distal to DXS1108 and proximal to the pseudoautosomal region (PAR) in Xqter.[15] The primers sKK5a 5'-TATGGCTGCATAGTATTCC-3' and sKK5 b5'-CAGATGATTAGATAAA GAAG-3' were used. PCR was performed using the standard buffer with 1 μmol/l of each primer and a small quantity of ^{32}P-dCTP. After an initial denaturation for four minutes at 94°C, 30 cycles of 94°C for 60 seconds, 52°C for 60 seconds, and 72°C for 60 seconds were performed, with a final extension cycle of 10 minutes at 72°C. The PCR products were analysed on a 6% urea/polyacrylamide/1 × TBE gel. A total of four alleles were identified of the size 96 ± 2n and these alleles were arbitrarily numbered. A G/A SNP Xq3274 close to the TKT2 gene at position 3631 in the sequence HS14B7 (Z49258) was also used (http://www.ibc.wustl.edu/SNP/) (P Y Kwok, personal communication). The PCR was performed as above with an annealing temperature of 55°C using the primers 8891 5'-GCCTGAGGAGGGATGATC-3' and 8892 5'-ACCATCAACCGCATACTAGC-3'. The primer 8891 was mutagenic and created a *BclI* site with the A allele; following *BclI* digestion the A allele produced bands of 122 bp and 18 bp, while the G allele gave an uncut band of 155 bp. The *BclI* digested PCR products were analysed by gel electrophoresis on a 2% agarose gel. Linkage analysis was performed using the MLINK programme of the LINKAGE package.[16] The DC1 locus was tested assuming that the disease was an X linked recessive trait with an allele frequency of the affected allele of 0.001 with complete penetrance. The family linkage data were managed using the IGD/XPED database system.[17]

X CHROMOSOME INACTIVATION ANALYSIS

The X chromosome inactivation pattern (XCIP) was analysed, using DNA extracted from whole blood, by investigating the methylation status of the HUMARA locus as described previously.[18] X chromosome inactivation results in the hypermethylation of the genomic DNA. The proportion of paternally and maternally inherited X chromosomes which are active can be assessed by digestion with *Hpa*II, a methylation sensitive enzyme, followed by PCR using primers which span the polymorphic HUMARA locus. The HUMARA locus contains an *Hpa*II site, such that after *Hpa*II digestion the allele from the active X chromsome is digested and cannot be amplified, and only the allele from the inactive X chromosome can be amplified. Subjects were scored as completely skewed when one allele was present at more than 98% of the total signal as measured by densitometry.[9]

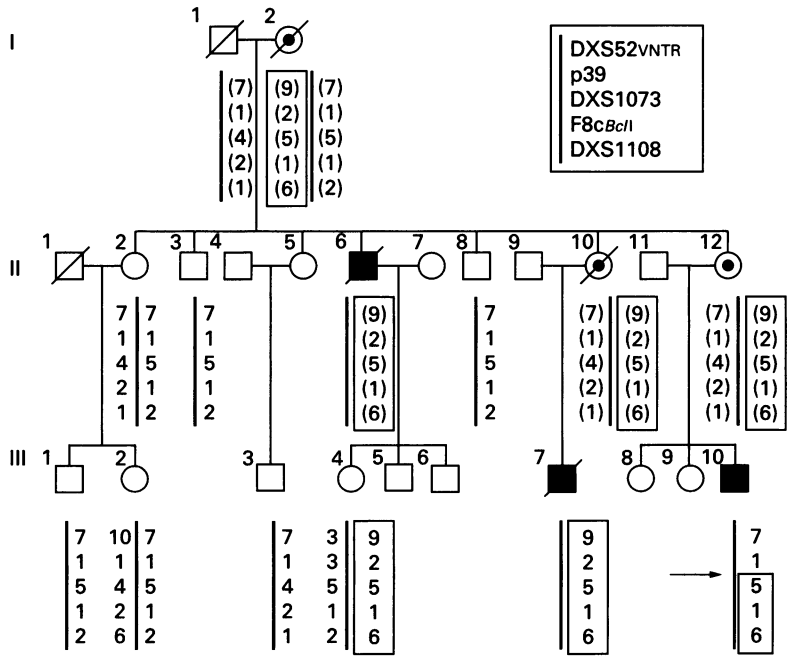

Figure 2 Haplotype analysis in pedigree DCR-025. The haplotypes in brackets were inferred minimising recombinations. The disease associated haplotype is boxed and the recombination event is marked with an arrow.

Results

A maximum lod score of 5.31 at zero recombination was obtained for the marker DXS1073 (table 1). A recombination between the marker DXS52 and the disease locus was observed in III.10 from pedigree DCR-025 (fig 2). By comparison of the haplotypes of the two affected

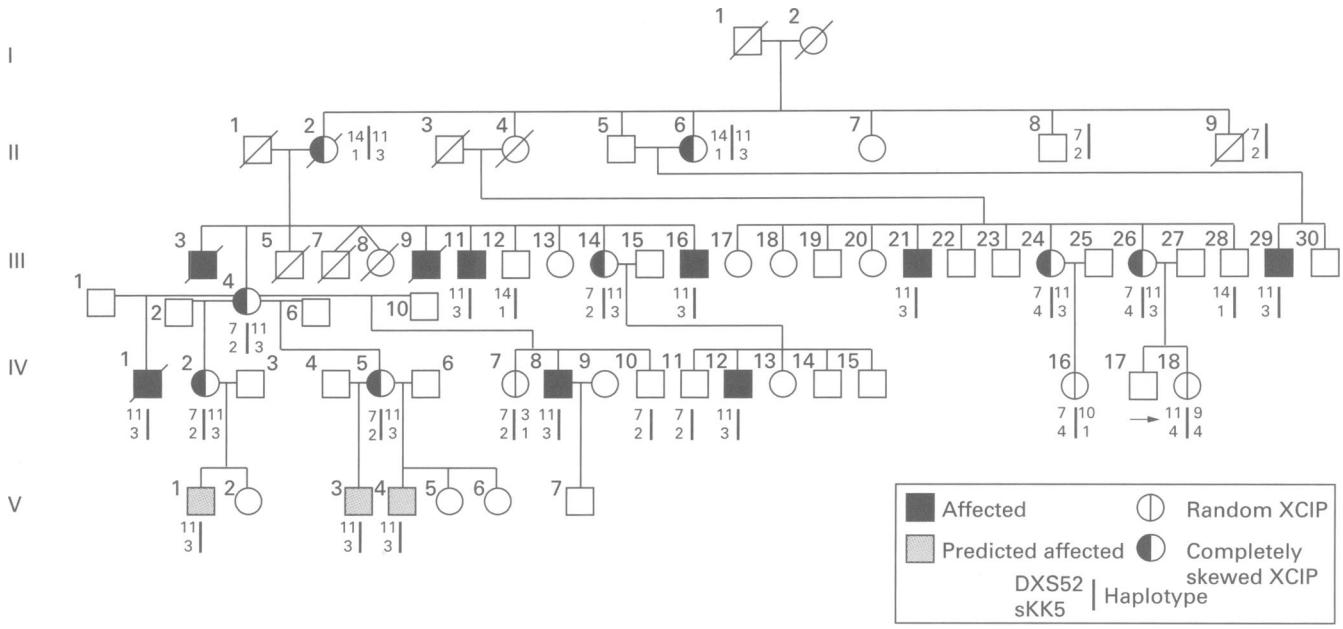

Figure 3 Segregation of haplotype and XCIP in pedigree DCR-007. Affected subjects are represented as black squares and shaded squares represent subjects predicted to be affected. Circles with a vertical line represent women who have a random XCIP and half blackened circles denote women who have a skewed XCIP. The haplotype for the markers DXS52VNTR and sKK5 is given below those subjects for whom samples were available. The recombination event is marked with an arrow.

subjects III.7 and III.10, it was apparent that the recombination breakpoint was located between p39 and DXS1073; additional markers between these loci were uninformative. From this family the proximal boundary of the DC1 candidate gene region was defined at p39.

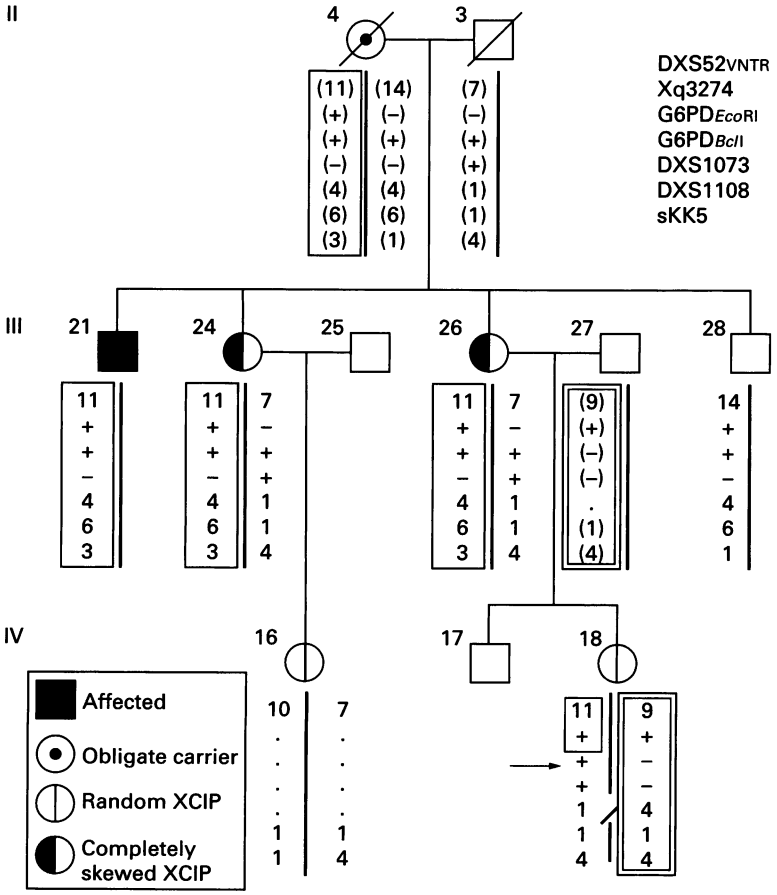

Figure 4 Haplotype analysis in part of pedigree DCR-007. Those haplotypes which were inferred are denoted by brackets. The disease associated haplotype is boxed and the inferred haplotype from III.27 is double boxed. The genotype of the marker DXS1073 for III.27 is unknown. The phase of the marker DXS1073 in IV.18 is unknown, which is denoted by the diagonal line. The recombination breakpoint is indicated with an arrow.

The markers DXS52 and sKK5 were shown to be completely informative in pedigree DCR-007 (fig 3) and it was apparent that three boys, under the age of 7 years (V.1, V.3, and V.4), had all inherited the affected haplotype. However, each was asymptomatic and consequently the disease status was defined as unknown (fig 3). An additional recombination event was observed in IV.18 from pedigree DCR-007 (fig 3). It was possible to infer the phase of the haplotypes of her mother (III.26) from the haplotypes of the mother's brothers (III.21 and III.28) (fig 4). It can be seen that the recombination breakpoint was between Xq3274 and DXS1108 (fig 4). A sample from the father, III.27, was not available and therefore it was not possible unambiguously to assign the phase of the markers G6PD *BclI*, DXS1073, and F8c in IV.18. However, from the known linkage disequilibrium of SNPs in the G6PD region it was predicted that the father, III.27, had the genotype G6PD *BclI*(-) (fig 4). This assumption was based on the fact that the haplotype G6PD *Eco*RI(-)/G6PD *BclI*(+) had not been observed in a population screen of white subjects from southern Italy.[14] By deduction, the maternally inherited haplotype in IV.18 was G6PD *Eco*RI(+)/G6PD *BclI*(+) which implied that the recombination breakpoint was between Xq3274 and G6PD *BclI* (fig 4).

The usefulness of this recombination event in defining the DC1 critical region was dependent on determining the carrier status of IV.18. As a method of determining the carrier status, X chromosome inactivation analysis was performed in all available women in pedigree DCR-007. All obligate carriers of the disease (II.2, II.6, III.4, and III.14) and those predicted to be carriers on the basis of inheritance of the affected DXS52/sKK5 haplotype (III.24, III.26, IV.2, and IV.5) had a completely skewed XCIP, as illustrated by III.26 in fig 5. In addition, those

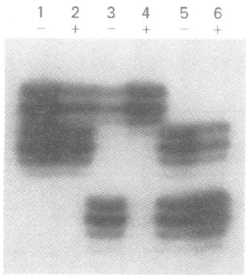

Figure 5 X chromosome inactivation pattern analysis in pedigree DCR-007. Amplification of the HUMARA polymorphism without (-) or with (+) previous digestion with HpaII restriction enzyme, followed by analysis of the ^{32}P labelled products on a 6% urea-polyacrylamide gel. Lanes 1 and 2: IV.16, lanes 3 and 4: III.26, lanes 5 and 6: IV.18.

women predicted not to be carriers did not have a completely skewed XCIP (IV.7 and IV.16) (fig 5). There was, therefore, a cosegregation of affected haplotype and skewed XCIP in pedigree DCR-007. IV.18 had a random XCIP as illustrated in fig 5 and was therefore predicted not to be a carrier of the disease. This must have been as a result of the loss of the mutated gene owing to the recombination event as described above. The location of the breakpoint and the carrier status of IV.18 mapped the DC1 gene distal to Xq3274.

Discussion

The recombination event in pedigree DCR-025 unambiguously placed the gene responsible for X linked dyskeratosis congenita distal to p39. This is the first report of a recominbination with DXS52. The marker DXS52 had previously had the highest lod score for linkage with DC1 and was used as the method of determining carrier status by haplotype analysis.[6] The DXS52 VNTR was shown to be informative in pedigree DCR-025 unlike the DXS52 RFLP which was uninformative.[8] The marker DXS1073 now has a maximum lod score of 5.31 for linkage with DC1. The marker DXS1073 was uninformative in the largest pedigree (DCR-007) and therefore the maximum lod score represents an underestimation of the linkage between DXS1073 and the disease (table 1). The recombination event with DXS52 in pedigree DCR-025, together with that previously reported with DXS1108 in pedigree DCR-006, narrows the DC1 candidate gene region to a 2.8 Mb interval between p39 and DXS1108.[6]

Haplotype analysis in pedigree DCR-007 detected an additional recombination event in IV.18 and also identified three young boys (V.1, V.3, and V.4) who were currently asymptomatic but predicted to develop the disease (fig 3). The DC1 candidate gene region was refined further when the recombination breakpoint in IV.18 was fine mapped and XCIP data used to analyse carrier status (fig 4). In pedigree DCR-007 there was a cosegregation of skewed XCIP with the affected haplotype (fig 3). This observation was in agreement with that found in other multiplex X linked DC families where there was cosegregation of carrier status and XCIP data.[8-10] The recombination breakpoint in IV.18 was fine mapped to distal to Xq3274 and when the known linkage disequilibrium between the G6PD *Eco*RI and G6PD *Bcl*I SNPs was used, the proximal boundary can be defined as G6PD *Bcl*I (fig 4).[14] The absence of a sample from the father, IV.27, made the mapping of the recombination breakpoint slightly ambiguous because it relied on linkage disequilibrium to define the proximal boundary.

The DC1 candidate gene region has now been defined as the 1.4 Mb interval between Xq3274 and DXS1108 and may be further delimited as additional polymorphic markers become available. An early study reported an X linked DC family with non-segregation of G6PD deficiency which can be used to define the DC1 candidate gene region as either the interval between Xq3274 and G6PD or between G6PD and DXS1108.[19] The recombination events with DXS52 have now excluded a large number of genes as potential candidates[20] and focuses attention on the 28 known positional candidate genes in the distal portion of Xq28.[21-23]

We wish to thank the following physicians for providing samples: Drs G L Forni, D Oscier, A Fryer, A Copplestone, J Ross, J Raeburn, S Davies, and C Bunch. We acknowledge Faith Bellis and Maria Mukherjee for helping with the Dyskeratosis Congenita Registry and also the support of Professors Lucio Luzzatto and John Goldman. This work was funded in part by Wellcome Trust project grant 051181. We also wish to acknowledge the resources and facilities of the MRC Human Genome Mapping Project.

1 Drachtman RA, Alter BP. Dyskeratosis congenita: clinical and genetic heterogeneity. Report of a new case and review of the literature. *Am J Pediatr Hematol Oncol* 1992;**14**:297-304.
2 Dokal I. Dyskeratosis congenita: an inherited bone marrow failure syndrome. *Br J Haematol* 1996;**92**:775-9.
3 Knight SW, Vulliamy T, Mason PJ, *et al.* Initial analysis from the dyskeratosis congenita registry (DCR). *Blood* 1997;**90**: S1-441.
4 Coulthard S, Chase A, Pickard J, *et al.* Chromosomal breakage analysis in dyskeratosis congenita peripheral blood lymphocytes. *Br J Haematol* 1997;**97**:S1-12.
5 Connor JM, Gatherer D, Gray FC, *et al.* Assignment of the gene for dyskeratosis congenita to Xq28. *Hum Genet* 1986; 72:348-51.
6 Knight SW, Vulliamy T, Mason PJ, *et al.* Fine mapping of the dyskeratosis congenita locus in Xq28. *J Med Genet* 1996;**33**:993-5.
7 Belmont JW. Genetic-control of X-inactivation and processes leading to X-inactivation skewing. *Am J Hum Genet* 1996;**58**:1101-8.
8 Devriendt K, Matthijs G, Legius E, *et al.* Skewed X-chromosome inactivation in female carriers of dyskeratosis congenita. *Am J Hum Genet* 1997;**33**:581-7.
9 Vulliamy TJ, Knight SW, Dokal I, *et al.* Skewed X-inactivation in carriers of X-linked dyskeratosis congenita. *Blood* 1997;**90**:2213-16
10 Ferraris AM, Forni GL, Mangerini R, *et al.* Nonrandom X-chromosome inactivation in hemopoietic cells from carriers of dyskeratosis congenita. *Am J Hum Genet* 1997;**61**: 458-61.
11 Sirinavin C, Trowbridge AA. Dyskeratosis congenita: clinical features and genetic aspects. Report of a family and review of the literature. *J Med Genet* 1975;**12**:339-54.
12 Sykes BC. DNA in heritable disease. *Lancet* 1983;**12**:787-9.
13 Kurdi-Haidar B, Mason PJ, Berrebi A, *et al.* Origin and spread of the glucose-6-phosphate dehydrogenase variant (G6PD-Mediterranean) in the Middle East. *Am J Hum Genet* 1990;**47**:1013-19.
14 Filosa S, Calabro V, Lania G, *et al.* G6PD haplotypes spanning Xq28 from F8C to red/green color vision. *Genomics* 1993;**17**:6-14.
15 Kvaloy K, Galvagni F, Brown WR. The sequence organization of the long arm pseudoautosomal region of the human sex chromosomes. *Hum Mol Genet* 1994;**3**:771-8.
16 Lathrop GM, Lalouel JM, Julier C, *et al.* Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 1984;**81**:3443-6.
17 Bryant SP, Spiridou A, Spurr NK. The Integrated Genomic Database (IGD): enhancing the productivity of gene mapping projects. In: Suhai S, ed. *Proceedings of the 1996 International Symposium on Theoretical and Computational Genome Research.* New York: Plenum Press, 1997:117-34.
18 Allen RC, Zoghbi HY, Moseley AB, *et al.* Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. *Am J Hum Genet* 1992;**51**:1229-39.
19 Gutman A, Frumkin A, Adam A, *et al.* X-linked dyskeratosis congenita with pancytopenia. *Arch Dermatol* 1978;**114**: 1667-71.
20 Heiss NS, Rogner UC, Kioschis P, *et al.* Transcriptional mapping in a 700-kb region around the DXS52 locus in Xq28 - isolation of 6 novel transcripts and a novel ATPase isoform (hPMCA5). *Genome Res* 1996;**6**:478-91.
21 Heiss NS, Korn B, Rogner UC, *et al.* Generation of specific exon trap probes from YACs using Alu long-range PCR products. *Meth Mol Cell Biol* 1995;**5**:337-44.
22 Rogner UC, Heiss NS, Kioschis P, *et al.* Transcriptional analysis of the candidate region for incontinentia pigmenti (IP2) in Xq28. *Genome Res* 1996;**6**:922-34.
23 Chen EY, Zollo M, Mazzarella R, *et al.* Long-range sequence-analysis in Xq28 - 13 known and 6 candidate genes in 219.4 kb of high GC DNA between the RCP/GCP and G6PD loci. *Hum Mol Genet* 1996;**5**:659-68.