ORIGINAL ARTICLE

Identification of a locus for a form of spondyloepiphyseal dysplasia on chromosome 15q26.1: exclusion of aggrecan as a candidate gene

S Eyre, P Roby, K Wolstencroft, K Spreckley, R Aspinwall, R Bayoumi, L Al-Gazali, R Ramesar, P Beighton, G Wallis

See end of article for authors' affiliations

Correspondence to: Dr G A Wallis, School of Biological Sciences, University of Manchester, 2.205 Stopford Building, Oxford Road, Manchester M13 9PT, UK; g.wallis@man.ac.uk

J Med Genet 2002;**39**:634–638

Revised version received 11 March 2002
Accepted for publication 25 April 2002

We have investigated a family with an autosomal dominant form of spondyloepiphyseal dysplasia (SED) characterised by short stature and severe premature degenerative arthropathy. Previous studies have excluded linkage between this condition and the locus for the type II collagen gene. Here we report the identification of linkage between this disorder and a locus on the long arm of chromosome 15 between markers D15S979 and D15S1004. According to current linkage maps and sequence data, this locus includes that of the aggrecan gene (*AGC1*). Our linkage data from the SED family show, however, that *AGC1* maps to a locus that is proximal to D15S979. This proximal location for *AGC1* is further supported by linkage data from a second family with an autosomal recessive form of multiple epiphyseal dysplasia that also maps to the SED locus. In both families *AGC1* is therefore excluded as a candidate gene.

The spondyloepiphyseal dysplasias (SED) are a heterogeneous group of conditions with predominant involvement of the vertebral bodies and the proximal epiphyses of the long bones. SED is characterised by short stature, flattening of the vertebral bodies, a barrel shaped chest, kyphosis, and lumbar lordosis. Cleft palate, myopia, retinal detachment, hearing loss, club foot, and pectus carinatum are also variable manifestations. In SED congenita (MIM 183900) dwarfism is pronounced, whereas SED tarda (MIM 313400 and MIM 184100) is characterised by mild stunting of stature with truncal shortening. Autosomal dominant, autosomal recessive, and X linked forms of inheritance have been reported.[1] In most instances, the autosomal forms of SED are caused by mutations in the gene encoding type II collagen at the locus 12q13.11-q13.2 (*COL2A1*, MIM 120140), whereas, X linked SED tarda is caused by mutations in a gene termed "sedlin" (MIM 300202), which has a putative role in endoplasmic to Golgi vesicular transport[2] and maps to Xp22.2-p22.1.

We have previously reported the clinical and radiographic manifestations of a form of SED, which we named SED type Kimberley (SED-K), in a South African family of English stock.[3] Preliminary studies to locate the gene responsible for SED-K led to the exclusion of linkage to the *COL2A1* locus.[3] We therefore performed a genome wide scan to map the disease gene. We report here that SED-K maps to a locus on chromosome 15q26.1 that excludes the aggrecan gene. This locus is similar to that which has been previously described for an autosomal recessive (AR) syndrome of macrocephaly, multiple epiphyseal dysplasia (MED), and distinctive facies.[4]

PATIENTS AND METHODS

The clinical and radiographic manifestations of SED-K have been described in detail previously (fig 1).[3] Briefly, the phenotype of the affected subjects is of proportionate short stature (below the 5th centile for age), with a stocky habitus and progressive osteoarthropathy of the weight bearing joints. Radiographically there is prominent end plate irregularity and sclerosis of the vertebral bodies. Generalised epiphyseal changes are mild and variable. The phenotype of the family with the AR-MED syndrome has also been described previously.[4 5]

Genotyping

DNA samples from 14 subjects from the SED-K kindred (fig 2) including nine affected subjects, three unaffected related subjects, and two unrelated spouses were available for genotyping from the previously reported linkage study.[3] The DNA samples from the SED-K family were typed using a standard set of fluorescently labelled microsatellite markers that spanned the genome[6] and 10 additional markers spanning a 40 cM region of chromosome 15q26.1. DNA samples were available from the AR-MED family as described in Bayoumi *et al*[5] and genotyped for markers D15S979 and D15S202. Genotyping was performed using an ABI 373 sequencer and GENESCAN 1.2.2-1 and GENOTYPER 1.1.1 software. Two point lod score (Z) values were computed by the LINKAGE 5.1 MLINK program[7] for various recombination fraction (θ) values at a penetrance of 100% and a disease frequency of 0.0001. Multipoint linkage analysis was performed using the GENEHUNTER plus program.[8 9] The order of the markers was obtained from online genetic mapping data at the Centre for Medical Genetics, Marshfield Medical Research Foundation Website.

Genotyping of *AGC1* polymorphisms

The variable number of tandem repeats (VNTR) polymorphism within exon 12 of the *AGC1* gene was amplified using the primers described by Doege *et al*.[10] To type the VNTR, P^{32}dCTP was included in the PCR reaction, the PCR products

Abbreviations: SED, spondylometaphyseal dysplasia; SED-K, spondylometaphyseal dysplasia type Kimberley; AR, autosomal recessive; MED, multiple epiphyseal dysplasia; VNTR, variable number of tandem repeats; SSCP, single stranded conformational polymorphism analysis

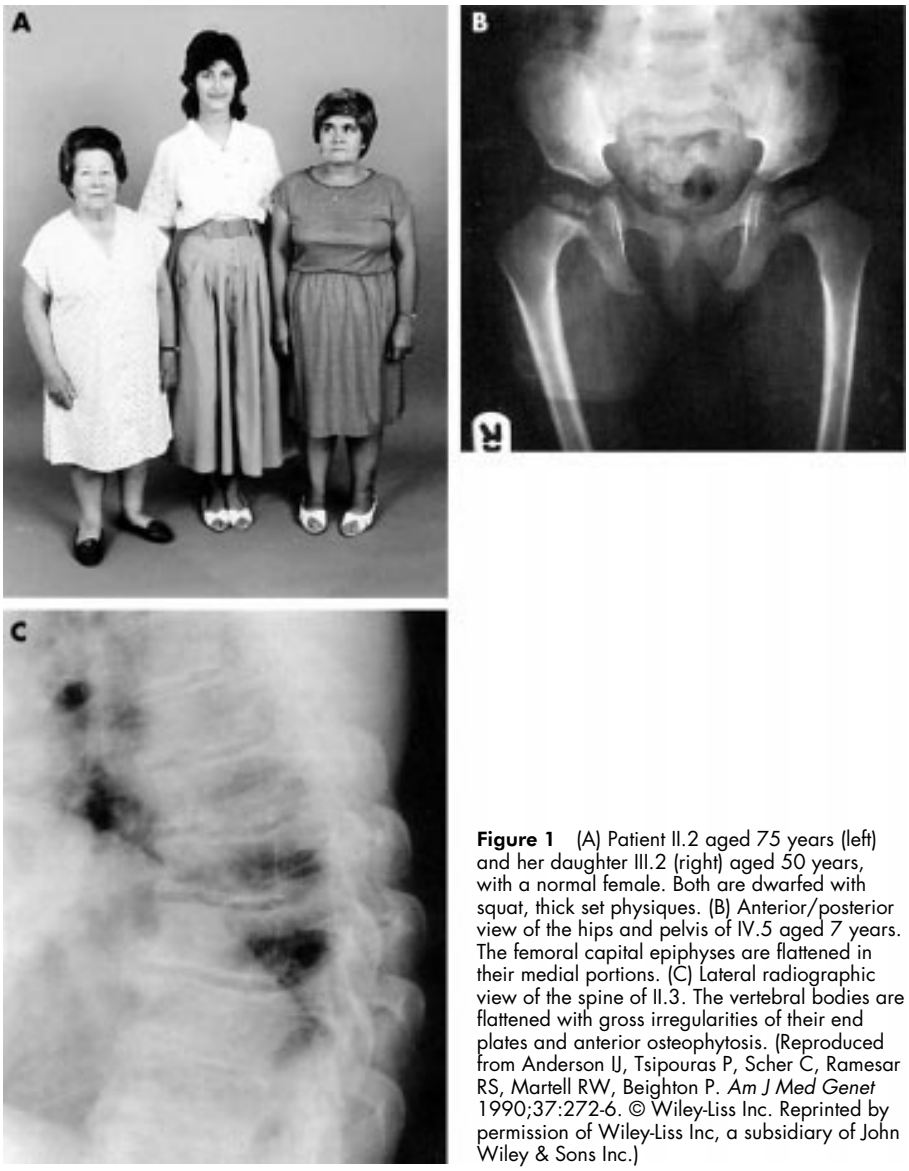

Figure 1 (A) Patient II.2 aged 75 years (left) and her daughter III.2 (right) aged 50 years, with a normal female. Both are dwarfed with squat, thick set physiques. (B) Anterior/posterior view of the hips and pelvis of IV.5 aged 7 years. The femoral capital epiphyses are flattened in their medial portions. (C) Lateral radiographic view of the spine of II.3. The vertebral bodies are flattened with gross irregularities of their end plates and anterior osteophytosis. (Reproduced from Anderson IJ, Tsipouras P, Scher C, Ramesar RS, Martell RW, Beighton P. *Am J Med Genet* 1990;37:272-6. © Wiley-Liss Inc. Reprinted by permission of Wiley-Liss Inc, a subsidiary of John Wiley & Sons Inc.)

separated by PAGE (6% w/v), and visualised by autoradiography. Four alleles for the VNTR were identified containing from 27 to 30 repeats of the 57 bp sequence. The restriction fragment length polymorphism (RFLP) identified in exon 18 (see Results) was amplified by PCR using the primers: 5′ GTCATCCCAGGAGACCCTATG and 5′ TAACCCTGTGCTCAGCGAGAT. The resultant 270 bp fragment was restriction enzyme digested with *Dde*I and the products separated on a 2% (w/v) agarose gel. In the absence of the *Dde*I RFLP, three bands of 212 bp, 48 bp, and 14 bp were generated, whereas in the presence of the *Dde*I RFLP, four bands of 167 bp, 48 bp, 45 bp, and 14 bp were generated. The accuracy of the genotyping of the *Dde*I RFLP was confirmed by sequence analysis of the PCR products using the BigDye Terminator Sequencing Ready Reaction Kit (Perkin Elmer Co, Foster City, CA, USA) on an ABI 377 sequencer.

Screening of the *AGC1* gene

Primers were designed to available intronic sequence[11] in order to generate PCR fragments that spanned each of the 19 exons of the *AGC1* gene and included the intron/exon boundaries. The PCR fragments generated from affected and unaffected subjects from the SED-K family were screened for mutations by single stranded conformational polymorphism analysis (SSCP) using previously published methods[12] and by sequence analysis.

Bioinformatics

Human genome sequence data were analysed by BLAST (version 2.0.6 and the NCBI advanced version 2.1),[13] GENESCAN,[14] and "Electronic PCR".[15 16]

RESULTS

Analysis of the genotype data generated from the SED-K family identified one marker on chromosome 15, D15S127, with a maximum Z (Zmax) value of 2.93 at θ=0.00. Further analysis with markers in this region of chromosome 15 identified a two point Zmax value of 3.01 for marker D15S116 (table 1). Multipoint analysis was performed using the marker order: cen - D15S131 - 4.57 cM - D15S211 - 2.45 cM - D15S206 - 0.62 cM - D15S205 - 4.48 cM - D15S979 -2.24 cM - D15S202 - 0.01 cM - D15S116 - 1.17 cM - D15S127 - 0.01 cM - D15S158 - 11.63 cM - D15S1004 - 2.15 cM - D15S1038 - 0.01 cM - D15S130 - 11.99 cM - D15S120 - tel (fig 3). This analysis mapped the SED-K locus to between markers D15S979 and D15S1004 with a maximum multipoint lod score of 3.3. No other markers in the screen showed suggestive evidence of linkage. However, the presence of a cosegregating locus could not be excluded formally because of the relatively sparse marker density and small family size.

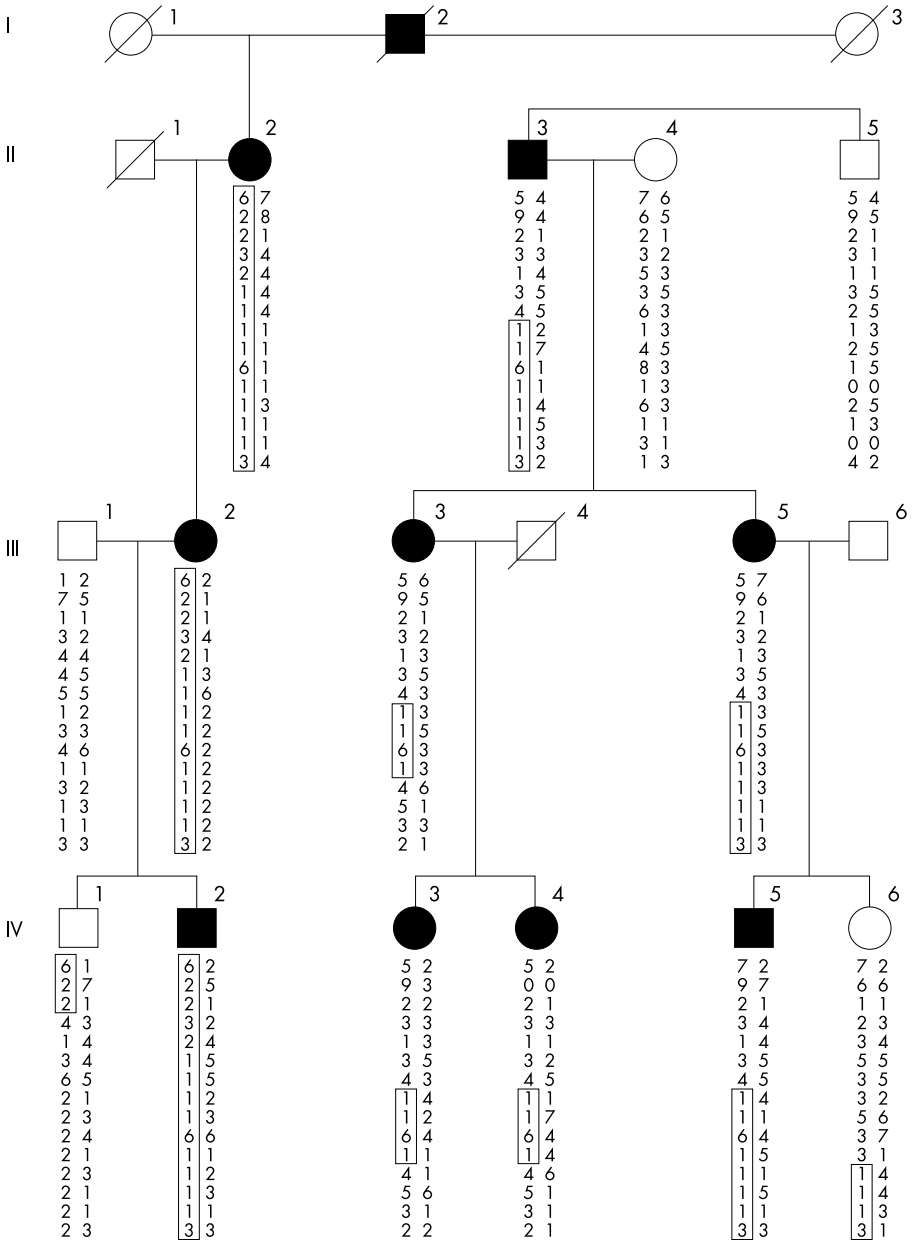

Figure 2 SED-K pedigree, showing disease linked haplotype. Blackened circles and squares represent affected females and males, respectively. The haplotypes for subjects who were genotyped are given under the symbols. The disease linked haplotype is on the left and is boxed. The marker order, from top to bottom, is D15S131, D15S211, *AGC1.2*, *AGC1.1*, D15S206, D15S205, D15S979, D15S202, D15S116, D15S127, D15S158, D15S1004, D15S1038, D15S130, and D15S120.

AGC1 genotyping and screening

As the *AGC1* gene was purported to reside within the SED-K locus, we genotyped the members of the SED-K family with the *AGC1* VNTR. This VNTR was not fully informative in this family but no recombination events were detected (see *AGC1.1* in fig 2 and table 1). We therefore began the screening of *AGC1* for mutations in genomic DNA from affected members of the SED-K family. The repeat region of exon 12 proved intractable to analysis but all other exonic sequence of the *AGC1* gene was screened for mutations. This analysis did not identify any potential mutations but detected a previously identified SNP (NCBI dbSNP 2280467) in exon 18 (see sequence accession J05062; nucleotides 1184 to 1366) that replaced a guanine with an adenine (at nucleotide number 1349) that would lead to the substitution of a glutamine with an arginine residue and created a *Dde*I restriction site. The genotypes of the members of the SED-K family for this RFLP are indicated in fig 2 (*AGC1.2*, where alleles 1 and 2 represent the presence and absence of the *Dde*I restriction site, respectively). The genotypes were also confirmed by sequence analysis. Examination of the haplotypes of *AGC1.1* and *ACG1.2* indicated that an intragenic recombination event had occurred between these two markers in IV.1 (fig 2), which was reflected by a comparison of the two point lod scores for the *AGC1* markers (table 1). This recombination event placed *AGC1* between markers D15S211 and D15S206 (fig 2).

Genotype analysis of the AR-MED family

As the position of *AGC1* that we had identified differed from that on the GeneMap'99 available at that time, we genotyped DNA from the AR-MED family where the *AGC1* locus had been shown to define the proximal limit of linkage to the disorder. All family members were genotyped with the markers D15S979 and D15S202 purported to flank the *AGC1* locus.

Table 1 Two point Z values between SED-K and chromosome 15q26 markers

Marker	Zmax	(θmax)	Z at θ = 0.00	0.01	0.05	0.10	0.20	0.30	0.40
D15S131	0.09	0.37	~	−3.47	−1.49	−0.73	−0.13	0.06	0.09
D15S211	0.63	0.20	~	−1.32	−0.05	0.39	0.63	0.58	0.36
AGC1.2	−0.02	0.40	~	−1.33	−0.67	−0.42	−0.20	−0.08	−0.02
AGC1.1	1.77	0.00	1.77	1.74	1.59	1.41	1.03	0.65	0.30
D15S206	0.49	0.19	~	−0.45	0.19	0.40	0.48	0.40	0.22
D15S205	0.42	0.13	~	−0.29	0.27	0.40	0.38	0.28	0.15
D15S979	1.55	0.10	~	0.97	1.49	1.55	1.35	0.7	0.51
D15S202	1.76	0.00	1.76	1.73	1.62	1.48	1.18	0.83	0.45
D15S116	3.01	0.00	3.01	2.96	2.77	2.51	1.95	1.34	0.68
D15S127	2.93	0.00	2.93	2.88	2.66	2.38	1.78	1.14	0.53
D15S158	1.73	0.00	1.73	1.70	1.58	1.43	1.11	0.77	0.41
D15S1004	0.75	0.19	~	−1.03	0.21	0.6	0.75	0.6	0.3
D15S1038	0.37	0.22	~	−0.64	0.01	0.24	0.37	0.34	0.21
D15S130	0.24	0.20	~	−0.73	−0.1	0.12	0.24	0.23	0.14
D15S120	0.42	0.17	~	−0.45	0.17	0.36	0.41	0.32	0.18

All values are calculated under the assumption of 100% penetrance and a disease allele frequency of 0.0001.

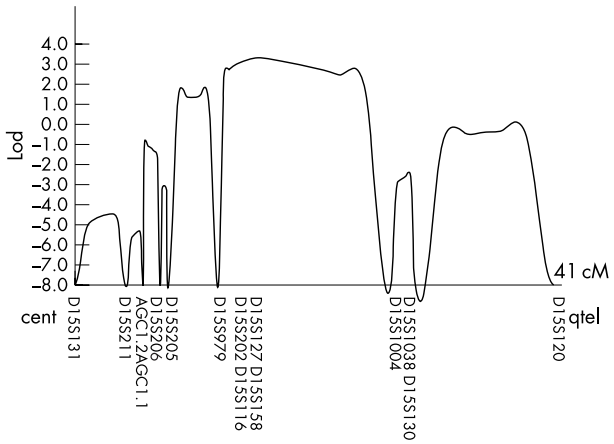

Figure 3 Graphical representation of the multipoint linkage analysis of chromosome 15 markers spanning the SED-K locus.

Both D15S979 and D15S202 were fully informative in this family and no recombination events were detected between these markers and the disorder (genotypes available on request), again placing *AGC1* proximal to D15S979.

Analysis of human genome sequence data

As data from both the SED-K and AR-MED families supported a locus for *AGC1* that was proximal to D15S979, we examined, in detail, the available raw sequence data for the linked region from the first draft of the human genome.[17] To examine the sequence data, BLAST searches were performed for each of the marker sequences used in the linkage analysis and for *AGC1* against both the unfinished High Throughput Genomic Sequence (HTGS) database[18] and the working draft sequences from the human genome to identify the contigs from which they were derived. As the sequencing of chromosome 15 is not yet complete, almost all of the markers occurred on unordered contigs. We found that the contigs containing *AGC1*, AC068969 and AC067805, did not contain any of the marker sequences. However, in the working draft sequence they had been incorporated into a much larger contig, NT_010356, which contained the markers D15S202 and D15S116 but not D15S979. This places *AGC1* in a position distal to D15S979. Since this was contrary to the linkage data, we sought to establish the way in which the large NT_010356 contig had been assembled. We therefore looked for contigs that overlapped the *AGC1* contigs and the marker contigs. Since many of the contigs identified were unordered and contained many repeated sequences, we could not find overlaps simply by alignment with other sequences in the databases. Instead, the gene prediction program GENSCAN[14] was run on each contig to identify any genes present. Further BLAST analyses determined which other clones or contigs contained each of the predicted genes. When the same group of genes was predicted on two different contigs, a potential region of overlap was indicated. After these investigations, a rough gene map of the whole area was constructed (data available on request). The order of this map was further verified using the program "Electronic PCR"[15 16] to identify all other marker sequences on each of the potentially overlapping contigs. The sequence map constructed in this way was consistent with the working draft map assembly but identified gaps in the linked region and multiple repeat sequences. One of the largest gaps occurred in a location directly proximal to *AGC1*.

DISCUSSION

We have identified linkage between the SED-K phenotype and a locus on the long arm of chromosome 15 between markers D15S979 and D15S1004. According to linkage maps available at the time of this finding and now more current linkage maps, this locus spans that of *AGC1*, which is located between markers D15S1046 (a marker distal to D15S979) and D15S202.

AGC1 (MIM 155760) was considered a likely candidate gene for SED-K as it encodes the major proteoglycan of the extracellular matrix of hyaline cartilage[19] and plays an important role in cartilage biology and limb development. Mutations within *AGC1* have been identified to cause forms of chondrodysplasia in both the mouse and chick but not as yet in humans. In the chick, *nanomelia* is a lethal disorder characterised by shortened and malformed limbs that is caused by homozygosity for a premature stop codon within *AGC1* leading to the truncation of the core protein which is neither processed nor secreted from the chondrocyte.[20] Similarly, mouse *cartilage matrix deficiency* (*cmd*), which is characterised by cleft palate and short limbs, tail, and snout, is caused by homozygosity for a 7 bp deletion in exon 5 of *AGC1* and premature chain termination.[21] Examination of the heterozygote *cmd* mice has shown that they appear normal at birth but dwarfism and spinal degeneration are age related changes,[22] a phenotype which is comparable with that of the SED-K family.

However, our genotype analysis of both the SED-K and AR-MED families using microsatellite markers spanning the purported *AGC1* locus and polymorphic markers within the *AGC1* gene support a locus for *AGC1* that is proximal to

D15S979. This locus for *AGC1* is inconsistent with available sequence data. Our detailed analysis of the available sequence data did not resolve this inconsistency. However, as yet the human genome sequence is a dynamic structure, still containing gaps and ambiguities, which are being resolved and updated continuously. Indeed the repeat sequences and gaps we identified in the chromosome 15 linked region indicate at least the potential for errors in contig assembly. Equally, the repeat sequences contained within this region may also lead to an increased frequency of recombination or double recombination events. It appears though that *AGC1* is an unlikely candidate gene for the SED-K and AR-MED phenotypes. However, the inconsistencies between the linkage and sequence data clearly need to be resolved before the targeted analysis of other genes within the linked region can be systematically performed.

ACKNOWLEDGEMENTS

We thank Dr Mike Briggs for useful discussions. This work was supported by grants from the Royal Society (UK), the Arthritis Research Campaign (UK), the University of Cape Town Staff Research Fund (SA), the Mauerberger Foundation (SA), and the Orthopaedic Association (SA).

Electronic database information. Accession numbers and URLs for data used in this study are as follows: Biocomputing Service Group, Heidelberg, http://genome.dkfz-heidelberg.de/cgi-bin/GENSCAN/genscan.welcome.pl (for GENSCAN program). Centre for Medical Genetics, Marshfield Medical Research Foundation, http://www.marshmed.org/genetics/ (for marker order). National Centre for Biotechnology Information (NCBI), http://www.ncbi.nih.gov (for GenBank, GeneMap'99, advanced BLAST version 2.1, Electronic PCR and HTGS database). Online Mendelian Inheritance in Man (OMIM), http://www.ncbi.nim.nih.gov/Omim (SED congenita (MIM 183900), SED tarda (MIM 313400 and MIM 184100), sedlin (MIM300202), and aggrecan (MIM 155760)). The Sanger Centre/European Bioinformatics Institute, http://www.ensembl.org (for ENSEMBL V1.1.0). University of California Santa Cruz (UCSC), http://genome.ucsc.edu/index.html (For human genome project working draft, April 2001 assembly).

Authors' affiliations

S Eyre, P Roby, K Spreckley, R Aspinwall, G Wallis, The Wellcome Trust Centre for Cell-Matrix Research, University of Manchester, Manchester, UK
S Eyre, P Roby, Arthritis Research Campaign Epidemiology Research Unit, University of Manchester, Manchester, UK
K Wolstencroft, Bioinformatics Unit, University of Manchester, Manchester, UK
G Wallis, Department of Medicine, University of Manchester, Manchester, UK
R Bayoumi, Department of Biochemistry, Sultan Qaboos University, Muscat, Sultanate of Oman
L Al-Gazali, Department of Paediatrics, United Arab Emirates University, Al Ain, United Arab Emirates
R Ramesar, P Beighton, Department of Human Genetics, University of Cape Town, Cape Town, South Africa

REFERENCES

1 **International Working Group on Constitutional Diseases of Bone**. International nomenclature and classification of the osteochondrodysplasias (1997). *Am J Med Genet* 1998;**79**:376-82.
2 **Gedeon AK**, Colley A, Jamieson R, Thompsom EM, Rogers J, Sillence D, Tiller GE, Mulley JC, Gecz J. Identification of the gene (SEDL) causing X-linked spondyloepiphyseal dysplasia tarda. *Nat Genet* 1999;**22**:400-4.
3 **Anderson IJ**, Tsipouras P, Scher C, Ramesar RS, Martell RW, Beighton P. Spondyloepiphyseal dysplasia, mild autosomal dominant type is not due to primary defects of type II collagen. *Am J Med Genet* 1990;**37**:272-6.
4 **Bayoumi R**, Saar K, Lee YA, Nurnberg G, Reis A, Nur-E-Kamal M, Al-Gazali LI. Localisation of a gene for an autosomal recessive syndrome of macrocephaly, multiple epiphyseal dysplasia, and a distinctive facies to chromosome 15q26. *J Med Genet* 2001;**38**:369-73.
5 **Al-Gazali LI**, Bakalinova D. Autosomal recessive syndrome of macrocephaly, multiple epiphyseal dysplasia and distinctive facial appearance. *Clin Dysmorphol* 1998;**7**:177-84.
6 **Roby P**, Eyre S, Worthington J, Ramesar R, Cilliers H, Beighton P, Grant M, Wallis G. Autosomal dominant (Beukes) premature degenerative osteoarthropathy of the hip joint maps to an 11 cM region on chromosome 4q35. *Am J Hum Genet* 1999;**64**:904-8.
7 **Lathrop G**, Lalouel J, Julier C, Ott J. Strategies for multilocus linkage analysis in humans. *Proc Natl Acad Sci USA* 1984;**81**:3443-6.
8 **Kruglyak L**, Daly MJ, Reeve-Daly MP, Lander ES. Parametric and nonparametric linkage analysis; a unified multipoint approach. *Am J Hum Genet* 1996;**58**:1347-63.
9 **Kong A**, Cox NJ. Allele-sharing models, LOD scores and accurate linkage tests. *Am J Hum Genet* 1997;**61**:1179-88.
10 **Doege KJ**, Coulter SN, Meek LM, Maslem K, Wood JG. A human-specific polymorphism in the coding region of the aggrecan gene. Variable number of tandem repeats produce a range of core protein sizes in the general population. *J Biol Chem* 1997;**272**:13974-9.
11 **Valhmu WB**, Palmer GD, Rivers PA, Ebara S, Cheng JF, Fischer S, Ratcliffe A. Structure of the human aggrecan gene: exon-intron organization and association with the protein domains. *Biochem J* 1995;**309**:535-42.
12 **Sweetman W**, Sykes B, Rash B, Beighton P, Zabel B, Thomas T, Boot-Handford R, Grant M, Wallis G. SSCP and segregation analysis of the human type X collagen gene (COL10A1) in heritable forms of chondrodysplasia. *Am J Hum Genet* 1992;**51**:841-9.
13 **Altschul SF**, Madden TL, Schaffer AA, Zhang J, Zang Z, Miller W, Lipman DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 1997;**25**:3389-402.
14 **Burge C**, Karlin S. Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 1997;**268**:78-94.
15 **Schuler GD**. Sequence mapping by electronic PCR. *Genome Res* 1997;**7**:541-50.
16 **Schuler GD**. Electronic PCR: bridging the gap between genome mapping and genome sequencing. *Trends Biotechnol* 1998;**16**:456-9.
17 **International Human Genome Sequencing Consortium**. Initial sequencing and analysis of the human genome. *Nature* 2001;**409**:860-921.
18 **Ouellette BF**, Boguski MS. Database divisions and homology search files: a guide for the perplexed. *Genome Res* 1997;**7**:952-5.
19 **Hardingham TE**, Fosang AJ. Proteoglycans: many forms and functions. *FASEB J* 1992;**6**:861-70.
20 **Li H**, Schwartz NB, Vertel BM. cDNA cloning of chick cartilage chondroitin sulfate (aggrecan) core protein and identification of a stop codon in the aggrecan gene associated with the chondrodystrophy, nanomelia. *J Biol Chem* 1993;**268**:23504-11.
21 **Watanabe H**, Kimata K, Line S, Strong D, Gao L, Kozak CA, Yamada Y. Mouse cartilage matrix deficiency (cmd) caused by a 7 bp deletion in the aggrecan gene. *Nat Genet* 1994;**7**:154-7.
22 **Watanabe H**, Nakata K, Kimata K, Nakanishi I, Yamada Y. Dwarfism and age-associated spinal degeneration of heterozygote cmd mice defective in aggrecan. *Proc Natl Acad Sci USA* 1997;**94**:6943-7.