SHORT REPORT

Localisation of a gene for mucopolysaccharidosis IIIC to the pericentromeric region of chromosome 8

J Ausseil, J C Loredo-Osti, A Verner, C Darmond-Zwaig, I Maire, B Poorthuis, O P van Diggelen, T J Hudson, T M Fujiwara, K Morgan, A V Pshezhetsky

....

Mucopolysaccharidosis type IIIC (MPS IIIC, or Sanfilippo syndrome C) is a rare lysosomal storage disorder caused by a deficiency of acetyl-coenzyme A:α-glucosaminide-N-acetyltransferase. Patients develop progressive neuropsychiatric problems, mental retardation, hearing loss, and relatively minor visceral manifestations. The pattern of transmission is consistent with an autosomal recessive mode of inheritance. The aim of this study was to find a locus for MPS IIIC using a homozygosity mapping approach. A genomewide scan was performed on DNA from 27 affected individuals and 17 of their unaffected relatives. Additional patients were recruited, and DNA was obtained from a total of 44 affected individuals and 18 unaffected family members from 31 families from 10 countries. A working candidate interval was defined by looking for excess homozygosity in patients compared with their relatives. Additional markers were genotyped in regions of interest. Linkage analysis was performed to support the informal analysis. Inspection of the genomewide scan data showed apparent excess homozygosity in patients compared with their relatives for markers on chromosome 8. Additional genotyping identified 15 consecutive markers (from D8S1051 to D8S2332) in an 8.3 cM interval for which the genotypes of affected siblings were identical in state. A maximum multipoint lod score of 10.61 was found at marker D8S519. A locus for MPS IIIC maps to an 8.3 cM (16 Mbp) interval in the pericentromeric region of chromosome 8.

Mucopolysaccharidosis type IIIC (MPS IIIC, or Sanfilippo syndrome C; OMIM #252930; http://www.ncbi.nlm.nih.gov/omim) is caused by a deficiency of acetyl-coenzyme A (CoA):α-glucosaminide N-acetyltransferase (GNAT, EC 2.3.1.3.) which has properties of a lysosomal membrane transporter. The birth prevalence in Australia,[1] Portugal,[2] and the Netherlands[3] has been estimated to be 0.07, 0.12, and 0.21 per 100 000, respectively. Clinically the disease manifests similarly to other subtypes of Sanfilippo syndrome, and results in progressive neuropsychiatric problems, mental retardation, hearing loss, and relatively minor visceral manifestations, such as mild hepatomegaly, mild dwarfism with joint stiffness and biconvex dorsolumbar vertebral bodies, mild coarse facies, and hypertrichosis.[4] This subtype of mucopolysaccharidosis was first described by Kresse *et al*,[5] who found that three patients with the phenotype of Sanfilippo syndrome had a deficiency of an enzyme that transfers an acetyl group from cytoplasmically derived acetyl-CoA to terminal α-glucosamine residues of heparan sulphate within lysosomes. The enzyme catalyses the acetylation of heparan sulphate without transporting the intact molecule of acetyl-CoA into the lysosomal compartment, where it would be rapidly degraded.[6–8] Klein *et al*[9] [10] reported a similar deficiency in 11 patients diagnosed with Sanfilippo syndrome, therefore suggesting that the disease is a relatively frequent subtype. By studying two siblings diagnosed with MPS IIIC who had an apparently balanced Robertsonian translocation, Zaremba *et al*[11] suggested that the mutant gene might be located in the pericentric region of either chromosome 14 or chromosome 21, but no further confirmation of this finding was published. Multiple attempts to purify and clone GNAT as a candidate gene for MPS IIIC were unsuccessful because of the low tissue content, instability, and hydrophobic nature of the enzyme, suggesting that positional mapping may be a better approach.

Under the assumption that MPS IIIC is a rare autosomal recessive disease and that most of the patients are homozygous by descent, we attempted to map the locus for MPS IIIC by looking for chromosomal regions of shared homozygosity[12] [13] in a diverse collection of patients, and for regions at which genotypes of affected siblings were identical in state.

METHODS

Subjects

Cultured skin fibroblasts and blood samples of MPS IIIC patients, their relatives, and controls were obtained from cell depositories (Hôpital Debrousse, France; NIGMS Human Genetic Mutant Cell Repository, USA; Montreal Children's Hospital, Canada; and Department of Clinical Genetics, Erasmus University, the Netherlands) or collected after ethics approval from the institutional review board of Hôpital Sainte-Justine. Dutch patients from families F3, F4, F5, and F6 were followed by J J P van de Kamp, who also obtained the written consent and pedigree information.[14] The detailed clinical features of patients will be described elsewhere. For the genomewide scan, DNA was available for 27 patients and 17 unaffected parents or siblings from 19 families from Byelorussia, Canada, Finland, France, the Netherlands, North Africa, Portugal, and Turkey. As the study progressed, additional patients and family members were recruited to the study. Altogether we collected and analysed DNA samples from 44 affected individuals, and 18 of their unaffected relatives distributed among 31 families. Pedigrees are shown in fig 1 for 13 families; three pedigrees (F1, F3, and F13) had documented consanguinity. Leukocyte DNA samples were also obtained from 38 control individuals of the same self identified population origin as the MPS IIIC patients.

....

Abbreviations: ASP, affected sibling pair; CoA, coenzyme A; GNAT, α-glucosaminide N-acetyltransferase; MPS IIIC, mucopolysaccharidosis type IIIC
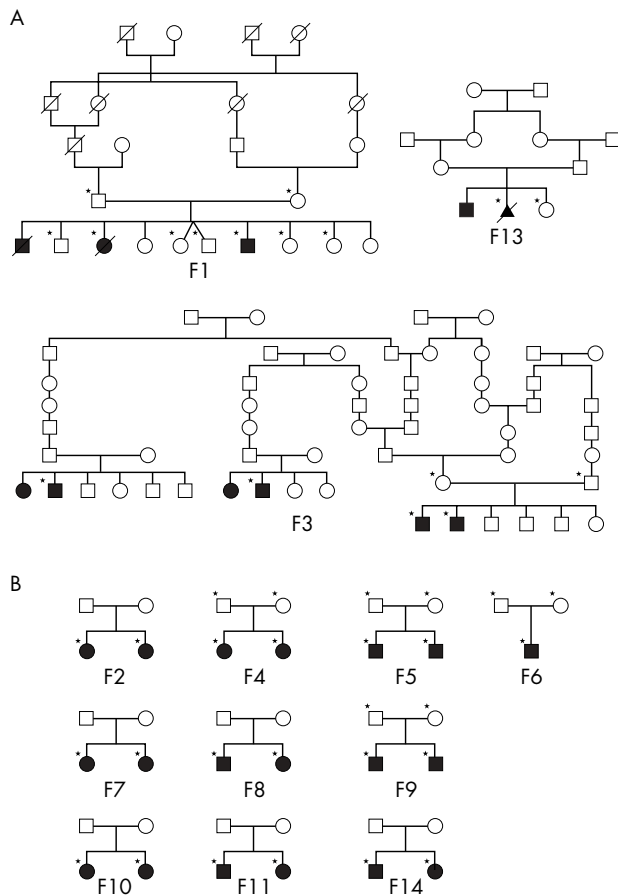
A

B

Figure 1 Pedigrees of MPS IIIC families. (a) Pedigrees of MPS IIIC families with known consanguinity from France (F1), the Netherlands (F3), and Spain (F13). (b) Partial pedigrees of other MPS IIIC families from France (F2), the Netherlands (F4, F5, F6), Finland (F7), Turkey (F8, F11), Byelorussia (F9), Portugal (F10), and Poland (F13). Solid symbol: individual with clinically and biochemically documented MPS IIIC; solid triangle: an affected fetus. *Individuals for whom DNA was available. In addition, DNA samples were obtained from 18 patients from Canada (I15, I8, I9), Finland (I13), France (I3–I6), the Netherlands (I12), North Africa (I7), Poland (I16–I20), Portugal (I10), Spain (I11), and Turkey (I2).

Genotyping and linkage analysis

Genotyping was performed at the McGill University and Genome Quebec Innovation Centre on an ABI 3700 DNA analyser platform essentially as described in Mira et al.[15] A panel of 392 highly informative, fluorescently labelled microsatellite markers, with an average interspacing of 10 cM, was derived from a modified version of the Cooperative Human Linkage Centre screening set (version 6.0), which also included Généthon markers. Alleles were assigned using Genotyper software (version 3.6; Applied Biosystems).

DNA samples from 55 individuals (37 patients and 18 unaffected relatives) were genotyped for 36 markers on chromosome 8 (19 genomewide scan markers and 17 additional markers). Twelve families were potentially informative for linkage all except F6; three families (F1, F3, and F13) had extended pedigrees. The positions of the 36 marker loci were based on deCODE genetic map locations where available (Kong et al[16]; http://www.nature.com/ng/journal/v31/n3/suppinfo/ng917_S1.html); otherwise genetic map positions were interpolated using deCODE and Marshfield genetic maps (http://www.marshfieldclinic.org/research/genetics), along with NCBI MapViewer (http://www.ncbi.nlm.nih.gov/mapview) (build 31) and the UCSC Human

Genome Browser (November 2002 assembly; http://genome.ucsc.edu). Two point linkage analysis for each of the 36 markers on chromosome 8 was performed using FASTLINK/MLINK (version 4.1P).[17 18] The genetic model was a single gene completely penetrant autosomal recessive trait with a disease allele frequency of 0.0045 (the square root of an incidence of 1/50 000). Marker allele frequencies were estimated by counting the number of instances of an allele in the sample of 55 individuals for whom genotype data was available. Multipoint lod scores were calculated by SUPERLINK/GH (version 1.4),[19] using nuclear family pedigrees because computation over the entire candidate interval was not feasible using extended pedigrees.

For 20 markers spanning the candidate interval, genotype data were available on 100 individuals (44 patients, 18 unaffected relatives of patients, and 38 controls). Twelve families were potentially informative for linkage (all except F6). For each marker, the maximum two point lod score and the corresponding maximum likelihood estimate of the recombination fraction (θ) were calculated using ILINK and MLINK. The genetic model and map locations were as described above. Marker allele frequencies were estimated from the number of alleles in the controls plus an additional count of 1 for any allele that was present in patients or their relatives but absent from the controls. We performed a multipoint analysis employing a sliding window of six consecutive markers across the candidate interval and SUPERLINK/GH to calculate multipoint lod scores with five equally spaced positions between adjacent markers. The three extended pedigrees were retained in both the two point and the multipoint analysis.

RESULTS AND DISCUSSION

Inspection of the results of the genomewide scan on 27 patients and 17 unaffected family members showed apparent excess homozygosity in patients compared with their unaffected relatives for eight genomic regions. Based on these observations, affected sibling pairs (ASPs) were genotyped for additional markers on chromosomes 1 (3 markers), 2 (2), 3 (2), 7 (2), 8 (3), 10 (2), 14 (1), and 16 (2). Apparent excess homozygosity was greatest for chromosome 8, and 17 additional markers on chromosome 8 were genotyped (in addition to the 19 markers in the genomewide scan) for the samples in the genomewide scan and for 11 samples obtained subsequently. We undertook exploratory data analysis of homozygosity in patients and identity in state of genotypes among affected relatives to delineate a candidate interval in the pericentromeric region of chromosome 8. A working candidate interval of 8.9 cM between D8S532 and D8S1816 was based on inspection for loss of identity in state within ASPs. The largest two point lod score was 5.5 at genome scan marker D8S1110, a genomewide scan marker in the candidate interval. A maximum multipoint lod score of 6.74 was found at a position between D8S509 and D8S1816 (fig 2).

Thirteen additional markers were identified in public databases to comprise a set of 20 microsatellites spanning the candidate interval (table 1; marker 20 was included to confirm that D8S1816 defined the distal boundary). The largest gap (1.7 cM) was between markers 15 and 16. We compared the average homozygosity of markers in the candidate interval with genomewide or ''background'' homozygosity. All individuals were genotyped for the 20 candidate interval markers and 57 genome scan markers that were about 50 cM or greater apart and not on chromosome 8. Five recently recruited patients from Poland were not genotyped for markers outside the candidate interval. In order to obtain less biased estimates, the average homozygosity of a subset of 33 of the 44 patients (one sibling from
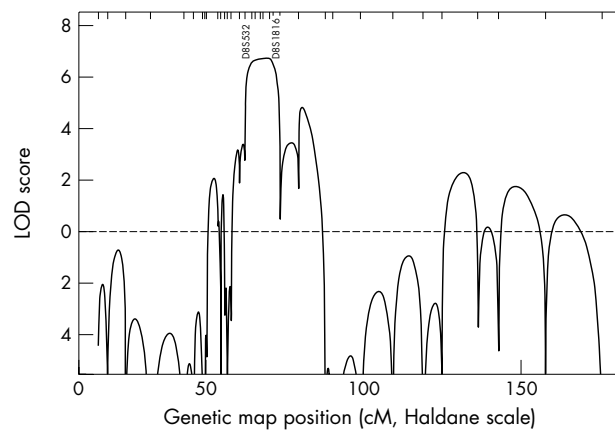
Figure 2 Multipoint lod scores of the disease locus across chromosome 8.

families (see table 1). Recombination was detected proximally for markers 1 (D8S532) and 2 (D8S268), and distally for marker 18 (D8S285). Recombination with marker 20 (D8S166) was inferred because the maximum lod score was at a recombination fraction of 0.018. For multipoint analysis of the candidate interval, we used a sliding window analysis of six consecutive markers. A maximum multipoint lod score of 10.61 was found in the first window at marker 5 (D8S519). The next largest multipoint lod score was 9.94 located at marker 8 (D8S589) at the end of the third sliding window. Thus, the disease locus is linked to the entire candidate interval with the exception of recombination with markers 1, 2, 18, and 20.

For 15 consecutive markers from D8S1051 to D8S2332, the genotypes of affected siblings with MPS IIIC were identical in state. This interval spans 8.3 cM (16 Mbp) in the pericentromeric region of chromosome 8, and contains 72 identified genes and open reading frames. Some of these genes encode proteins with predicted multiple transmembrane domains, consistent with the previous finding that GNAT is an integral membrane protein. Others contain conserved amino acid motifs common for the enzymes from the acetyltransferase family, PF00583. These data will be further used for identification of the *GNAT* gene by an integrative biology approach, which combines information about the gene position with biochemical data and clues about its intracellular localisation and biological function. The mapping of the MPS IIIC locus is an important step toward understanding this severe disorder. Further progress in linkage mapping may be made by obtaining additional pedigree information and by sampling relatives of the single patients, of whom 12 of the 18 had low homozygosity in the candidate interval, the average ranging from 0.53 down to 0.20. In addition, sampling the parents of all the patients would facilitate determination of haplotypes for subsequent fine mapping. Narrowing the candidate interval might be accomplished by searching the 16 Mbp region for additional polymorphisms to reduce the distal gap and detect crossovers.

each of 11 ASPs, the two distantly related cousins of an ASP, and 20 other patients without an affected sibling) was compared with those of a set of 38 controls. The average homozygosity of the 20 candidate interval markers was 0.75 or greater for one third (12/33) of the patients compared with none of the controls (this was anticipated because of ascertainment of an increased proportion of inbred patients with this rare recessive disease). The average homozygosity for the patients and controls for each marker is given in table 1. Fig 3 shows that there was no apparent correlation between average homozygosity of the candidate interval and background homozygosity for 57 genome scan markers. This was not unexpected as the mean heterozygosity of offspring of, for example, first cousins would be decreased by only 1/16, an amount hardly detectable by averaging over a small sample of unlinked markers. However, inbreeding would be detected as long homozygous chromosomal segments.[20][21]

We calculated two point lod scores between the disease locus and the 20 markers in the candidate interval in 12

Table 1 Linkage analysis of MPS IIIC and markers in the chromosome 8 candidate interval

Order	Locus	Marker	Genetic location (cM)*	Homozygosity†		Maximum lod score	θ
				Patients	Controls		
1	D8S532	AFM081yd11	59.76	0.48	0.16	4.74	0.032
2	D8S268	AFM156xa3	60.01	0.67	0.43	3.93	0.028
3	D8S1051	UT7847	60.92	0.58	0.42	4.16	0.000
4	D8S1460	ATA38C03	61.85	0.64	0.32	5.56	0.000
5	D8S519	AFM234vh8	62.08	0.48	0.21	7.25	0.000
6	D8S531	AFM081xe9	62.26	0.79	0.45	4.19	0.000
7	D8S1745	AFM200xh12	62.89	0.48	0.18	5.77	0.000
8	D8S589	GATA12H01	62.89	0.52	0.37	4.87	0.000
9	D8S1831	AFMa071yf9	63.65	0.45	0.45	6.45	0.000
10	D8S538	AFM274ve5	63.89	0.43	0.32	5.68	0.000
11	D8S517	AFM224xh6	64.48	0.91	0.81	1.49	0.000
12	D8S524	AFM242zf6	64.48	0.58	0.47	3.83	0.000
13	D8S1815	AFMa046wg9	64.73	0.73	0.26	5.48	0.000
14	D8S1110	GATA8G10	65.47	0.56	0.21	7.02	0.000
15	D8S601	AFM276xe9	65.75	0.44	0.27	7.05	0.000
16	D8S509	AFM203zc1	67.47	0.72	0.45	4.51	0.000
17	D8S2332	GATA156H01	68.19	0.58	0.47	6.46	0.000
18	D8S285	AFM255yb9	68.31	0.48	0.11	2.31	0.090
19	D8S1816	AFM021te1	68.63	0.70	0.68	2.57	0.000
20	D8S166	Mfd159A	68.92	0.30	0.22	4.60	0.018

*Kosambi scale, see text for details. According to the UCSC Genome Browser (July 2003 assembly), the centromere is located between markers 2 and 3. †Homozygosity was estimated as the proportion of homozygous genotypes observed. Because of a few missing data, the sample sizes ranged from 30 to 33 patients and 37 to 38 controls. See text for description of the subset of 33 patients.
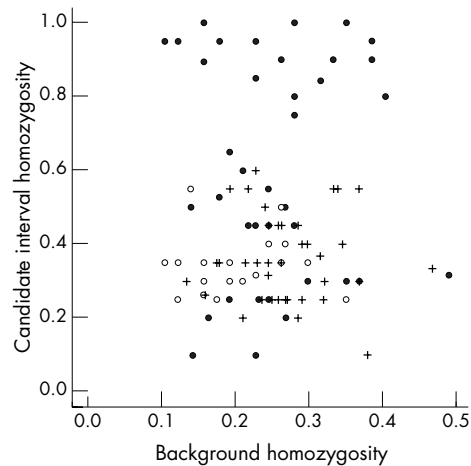
Figure 3 Homozygosity of patients, unaffected relatives and controls. The proportion of homozygous genotypes for 20 markers spanning the candidate interval is compared with the proportion of homozygous genotypes for 57 unlinked markers distributed across the genome (an estimate of background homozygosity but with chromosome 8 excluded). Solid circles: 39 patients; open circles: 18 unaffected relatives; +: 38 control individuals. The patients cluster into two groups: those with candidate interval homozygosity of 0.75 or greater and those with smaller values. The distribution of candidate interval homozygosity of the second group of patients overlaps the distributions of the relatives and controls. In addition, there appears to be no correlation between candidate interval and background homozygosity. There were some missing data, and some data points overlap and are therefore not distinguishable. Five additional patients from Poland with candidate interval homozygosity values of 0.30, 0.35, 0.39, 0.40, and 0.50 were not included because they were not genotyped for the background markers.

ACKNOWLEDGEMENTS

. . . .

Authors' affiliations

J Ausseil, A V Pshezhetsky, Hôpital Sainte-Justine and Département de Pédiatrie, Université de Montréal, Montréal, Québec, Canada
J C Loredo-Osti, T J Hudson, T M Fujiwara, K Morgan, Department of Human Genetics, McGill University, Montréal, Québec, Canada
A Verner, C Darmond-Zwaig, T J Hudson, McGill University and Genome Quebec Innovation Centre, Montréal, Québec, Canada
I Maire, Hôpital Debrousse, Lyon Cedex 05, Lyon, France
B Poorthuis, Department of Pediatrics and Clinical Genetics, Leiden University Medical Center, Leiden, the Netherlands

O van Diggelen, Department of Clinical Genetics, Erasmus University, Rotterdam, the Netherlands
T J Hudson, T M Fujiwara, K Morgan, Department of Medicine, McGill University, Montréal, Québec, Canada

Conflict of interest: none declared

The first two authors contributed equally to this work, and the last two authors share senior authorship.

Correspondence to: Dr A V Pshezhetsky, Service de Génétique Médicale, Hôpital Sainte-Justine, 3175 Côte Ste-Catherine, Montréal, Québec H3T 1C5, Canada; alexei.pchejetski@umontreal.ca

REFERENCES

1 **Meikle PJ**, Hopwood JJ, Clague AE, Carey WF. Prevalence of lysosomal storage disorders. *JAMA* 1999;**281**:249–54.
2 **Pinto R**, Caseiro C, Lemos M, Lopes L, Fontes A, Ribeiro H, Pinto E, Silva E, Rocha S, Marcao A, Ribeiro I, Lacerda L, Ribeiro G, Amaral O, Sa Miranda MC. Prevalence of lysosomal storage diseases in Portugal. *Eur J Hum Genet* 2004;**12**:87–92.
3 **Poorthuis BJ**, Wevers RA, Kleijer WJ, Groener JE, de Jong JG, van Weely S, Niezen-Koning KE, van Diggelen OP. The frequency of lysosomal storage diseases in the Netherlands. *Hum Genet* 1999;**105**:151–6.
4 **Bartsocas C**, Grobe H, van de Kamp JJP, von Figura K, Kresse H, Klein U, Giesberts MAH. Sanfilippo type C disease: clinical findings in four patients with a new variant of mucopolysaccharidosis III. *Europ J Pediat* 1979;**130**:251–8.
5 **Kresse H**, Von Figura K, Bartsocas C. Clinical and biochemical findings in a family with Sanfilippo disease, type C. *Clin Genet* 1976;**10**:364.
6 **Bame KJ**, Rome LH. Genetic evidence for transmembrane acetylation by lysosomes. *Science* 1986;**233**:1087–9.
7 **Bame KJ**, Rome LH. Acetyl-coenzyme A: alpha-glucosaminide N-acetyltransferase; evidence for a transmembrane acetylation mechanism. *J Biol Chem* 1985;**260**:11293–9.
8 **Bame KJ**, Rome LH. Acetyl-coenzyme A: alpha-glucosaminideN-acetyltransferase: evidence for an active site histidine residue. *J Biol Chem* 1986;**261**:10127–32.
9 **Klein U**, Kresse H, von Figura K. Sanfilippo syndrome type C: deficiency of acetyl-CoA: alpha-glucosaminide N-acetyltransferasein skin fibroblasts. *Proc Natl Acad Sci* 1978;**75**:5185–9.
10 **Klein U**, van de Kamp JJP, von Figura K, Pohlmann R. Sanfilippo syndrome type C: assay for acetyl-CoA: alpha-glucosaminide N-acetyltransferase in leukocytes for detection of homozygous and heterozygous individuals. *Clin Genet* 1981;**20**:55–9.
11 **Zaremba J**, Kleijer WJ, Huijmans JGM, Poorthuis B, Fidzianska E, Glogowska I. Chromosomes 14 and 21 as possible candidates for mapping the gene for Sanfilippo disease type IIIC. *J Med Genet* 1992;**29**:514.
12 **Smith CAB**. The detection of linkage in human genetics. *J R Stat Soc B* 1953;**15**:153–192.
13 **Lander ES**, Botstein D. Homozygosity mapping: a way to map human recessive traits with the DNA of inbred children. *Science* 1987;**236**:1567–70.
14 **van de Kamp JJP**. *The Sanfilippo syndrome: a clinical and genetical study of 75 patients in the Netherlands*, PhD Thesis. Leiden: Pasmans, 's-Gravenhage, 1979.
15 **Mira MT**, Alcais A, Nguyen VT, Moraes MO, Di Flumeri C, Vu HT, Mai CP, Nguyen TH, Nguyen NB, Pham XK, Sarno EN, Alter A, Montpetit A, Moraes ME, Moraes JR, Dore C, Gallant CJ, Lepage P, Verner A, Van De Vosse E, Hudson TJ, Abel L, Schurr E. Susceptibility to leprosy is associated with PARK2 and PACRG. *Nature* 2004;**427**:636–40.
16 **Kong A**, Gudbjartsson DF, Sainz J, Jonsdottir GM, Gudjonsson SA, Richardsson B, Sigurdardottir S, Barnard J, Hallbeck B, Masson G, Shlien A, Palsson ST, Frigge ML, Thorgeirsson TE, Gulcher JR, Stefansson K. A high-resolution recombination map of the human genome. *Nat Genet* 2002;**31**:241–7.
17 **Cottingham RW Jr**, Idury RM, Schäffer AA. Faster sequential genetic linkage computations. *Am J Hum Genet* 1993;**53**:252–63.
18 **Schäffer AA**, Gupta SK, Shriram K, Cottingham RW Jr. Avoiding recomputation in linkage analysis. *Hum Hered* 1994;**44**:225–37.
19 **Fishelson M**, Geiger D. Exact genetic linkage computations for general pedigrees. *Bioinformatics* 2002;**18**(Suppl 1):S189–98.
20 **Broman KW**, Weber JL. Long homozygous chromosomal segments in reference families from the centre d'étude du polymorphisme humain. *Am J Hum Genet* 1999;**65**:1493–500.
21 **Clark AG**. The size distribution of homozygous segments in the human genome. *Am J Hum Genet* 1999;**65**:1489–92.