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


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 (GNAT, EC 2.3.1.3.) which has properties of a lysosomal membrane transporter. The birth prevalence in Australia, Portugal, and the Netherlands 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 mucopolysaccharidosis IIIC by looking for chromosomal regions of shared 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 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.

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. 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
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 Genotypy 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; http://www.nature.com/ng/journal/v31/n3/supplinfo/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 using 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 (theta) 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
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 for the patients and controls for each marker is given in table 1.

We calculated two point lod scores between the disease locus and the 20 markers in the candidate interval in 12 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 (D8S319). 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 pericentric 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.

![Figure 2](https://example.com/figure2.png)

**Figure 2** Multipoint lod scores of the disease locus across chromosome 8.

<table>
<thead>
<tr>
<th>Order</th>
<th>Locus</th>
<th>Marker</th>
<th>Genetic location (cM)</th>
<th>Homozygosity†</th>
<th>Maximum lod score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>D8S532</td>
<td>AFM081yd11</td>
<td>59.76</td>
<td>0.48</td>
<td>4.74</td>
</tr>
<tr>
<td>2</td>
<td>D8S268</td>
<td>AFM156ox3</td>
<td>60.01</td>
<td>0.67</td>
<td>3.93</td>
</tr>
<tr>
<td>3</td>
<td>D8S1051</td>
<td>U7847</td>
<td>60.92</td>
<td>0.58</td>
<td>4.16</td>
</tr>
<tr>
<td>4</td>
<td>D8S1460</td>
<td>ATA38C03</td>
<td>61.85</td>
<td>0.64</td>
<td>5.56</td>
</tr>
<tr>
<td>5</td>
<td>D8S519</td>
<td>AFM234hx8</td>
<td>62.08</td>
<td>0.48</td>
<td>7.25</td>
</tr>
<tr>
<td>6</td>
<td>D8S531</td>
<td>AFM081xe9</td>
<td>62.26</td>
<td>0.79</td>
<td>4.19</td>
</tr>
<tr>
<td>7</td>
<td>D8S1745</td>
<td>AFM200x12</td>
<td>62.89</td>
<td>0.48</td>
<td>5.77</td>
</tr>
<tr>
<td>8</td>
<td>D8S589</td>
<td>GATA12Dh01</td>
<td>62.89</td>
<td>0.52</td>
<td>4.87</td>
</tr>
<tr>
<td>9</td>
<td>D8S1831</td>
<td>AFM007yi9</td>
<td>63.65</td>
<td>0.48</td>
<td>6.45</td>
</tr>
<tr>
<td>10</td>
<td>D8S538</td>
<td>AFM274ve5</td>
<td>63.89</td>
<td>0.43</td>
<td>5.68</td>
</tr>
<tr>
<td>11</td>
<td>D8S517</td>
<td>AFM224x6</td>
<td>64.48</td>
<td>0.91</td>
<td>1.49</td>
</tr>
<tr>
<td>12</td>
<td>D8S524</td>
<td>AFM242x6e</td>
<td>64.48</td>
<td>0.58</td>
<td>3.83</td>
</tr>
<tr>
<td>13</td>
<td>D8S1815</td>
<td>AFM004w9</td>
<td>64.73</td>
<td>0.73</td>
<td>5.48</td>
</tr>
<tr>
<td>14</td>
<td>D8S1110</td>
<td>GATA8G10</td>
<td>65.47</td>
<td>0.56</td>
<td>7.02</td>
</tr>
<tr>
<td>15</td>
<td>D8S601</td>
<td>AFM275xe9</td>
<td>65.75</td>
<td>0.44</td>
<td>7.05</td>
</tr>
<tr>
<td>16</td>
<td>D8S509</td>
<td>AFM203ec1</td>
<td>67.47</td>
<td>0.72</td>
<td>4.51</td>
</tr>
<tr>
<td>17</td>
<td>D8S2332</td>
<td>GATA15Hd01</td>
<td>68.19</td>
<td>0.58</td>
<td>6.46</td>
</tr>
<tr>
<td>18</td>
<td>D8S285</td>
<td>AFM255yb9</td>
<td>68.31</td>
<td>0.48</td>
<td>2.31</td>
</tr>
<tr>
<td>19</td>
<td>D8S1816</td>
<td>AFM021e1</td>
<td>68.63</td>
<td>0.70</td>
<td>2.57</td>
</tr>
<tr>
<td>20</td>
<td>D8S166</td>
<td>M6d159A</td>
<td>68.92</td>
<td>0.30</td>
<td>4.60</td>
</tr>
</tbody>
</table>

*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

The authors thank the patients and their families for participating in our study. We also acknowledge N Gusina, J Clarke, and T Rupar for providing cell lines from MPS IIIC patients; D Frapper for help in interpreting genotyping experiments; D Labuda for helpful advice; N Roslin for assistance in evaluating the genomewide scan data; M Fischelson and D Geiger for help with SUPERLINK and for providing the updated executable code for version 1.4; and L Gallant for help in preparing the manuscript. This work was supported by operating grants from the Sanfilippo Children’s Research Foundation, Canadian Institutes of Health Research (CIHR) MT-38107 to A V Pshezhetsky, and the Canadian Networks of Centres of Excellence Program (The Canadian Genetic Diseases Network; CGDN) and the Mathematics of Information Technology and Complex System to K Morgan, and by an equipment grant from the Canadian Foundation for Innovation to A V Pshezhetsky. J Ausseil acknowledges postdoctoral fellowships from Vaincre les Maladies Lysosomales and Fondation de l’Hôpital Sainte-Justine, and J C Leredo-Osti acknowledges CGDN and a CIHR Strategic Training Program Grant for postdoctoral support. T J Hudson is a CIHR Investigator, and A V Pshezhetsky is a National Investigator of the Fonds de la Recherche en Santé du Québec.

Authors’ affiliations

J Ausseil, A V Pshezhetsky, Hôpital Sainte-Justine and Département de Pédiatrie, Université de Montréal, Montréal, Québec, Canada
J C Leredo-Osti, T J Hudson, T M Fujisawa, K Morgan, Department of Human Genetics, McGill University, Montréal, Québec, Canada
A Verner, C Damond-Zwaag, T J Hudson, McGill University and Genome Quebec Innovation Centre, Montréal, Québec, Canada
I Maire, Hôpital Debrusac, 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 Fujisawa, 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.pshejetski@umontreal.ca

Received 6 April 2004
Accepted for publication 18 June 2004

REFERENCES