HUMAN MUTATION Mutation in Brief 32: E2189-E2210 (2011) Online

MUTATION IN BRIEF

HUMAN MUTATION

IDUA Mutational Profiling of a Cohort of 102 European Patients with Mucopolysaccharidosis Type I: Identification and Characterization of 35 Novel α-L-iduronidase (*IDUA*) Alleles

Francesca Bertola[1*], Mirella Filocamo[2*], Giorgio Casati[1], Matthew Mort[3], Camillo Rosano[4], Anna Tylki-Szymanska[5], Beyhan Tüysüz[6], Orazio Gabrielli[7], Serena Grossi[2], Maurizio Scarpa[8], Giancarlo Parenti[9,10], Daniela Antuzzi[11], Jaime Dalmau[12], Maja Di Rocco[13], Carlo Dionisi Vici[14], Ilyas Okur[15], Jordi Rosell[16], Attilio Rovelli[17], Francesca Furlan[17], Miriam Rigoldi[17], Andrea Biondi[17], David N Cooper[3], and Rossella Parini[17]

[1]Consortium for Human Molecular Genetics, Milano Bicocca University, Monza, Italy; [2]S.S.D. Lab. Diagnosi Pre-Postnatale Malattie Metaboliche, IRCCS G. Gaslini, Genova, Italy;[3]Institute of Medical Genetics, School of Medicine, Cardiff University, Heath Park, Cardiff CF14 4XN, UK; [4]Nanobiotecnologie, Istituto Nazionale per la Ricerca sul Cancro, Genova, Italy; [5]Department of Metabolic Diseases, The Children's Memorial Health Institute, Warsaw, Poland; [6]Division of Genetics, Department of Pediatrics, Istanbul University, Cerrahpasa Faculty of Medicine, Istanbul, Turkey; [7]Division of Pediatrics, Polytechnic University of the Marche, Ospedali Riuniti, Presidio Salesi, Ancona, Italy; [8]Department of Pediatrics, University Children's Hospital, University of Padua, Italy; [9]Telethon Institute of Genetics and Medicine, Napoli, Italy; [10]Department of Pediatrics, Federico II University, Napoli, Italy; [11]Department of Pediatric Sciences, Catholic University, Roma, Italy; [12]Division of Metabolism, Hospital Infantil La Fe, Valencia, España;[13]S.S. Malattie Rare, U.O. Pediatria II, IRCCS G. Gaslini, Genova, Italy; [14]Division of Metabolism, Bambino Gesù Children's Hospital, Roma, Italy;[15]Department of Pediatric Nutrition and Metabolism, Gazi University Medical School, Ankara, Turkey; [16]Department of Genetics, Hospital Universitari Son Dureta, Palma de Mallorca, Spain; [17]Department of Pediatrics, University of Milano Bicocca, San Gerardo Hospital, Monza, Italy

*Correspondence to:
Dr Francesca Bertola, Consortium for Human Molecular Genetics, Milano Bicocca University, (20052) Monza, Italy.
Tel: +39 02 64488055; Fax: +39 02 64488361; E-mail: genemonza@unimib.it
Dr Mirella Filocamo, Lab Diagnosi Pre-Postnatale Malattie Metaboliche, Istituto G. Gaslini, Largo G. Gaslini 5 (16147) Genova, Italy. Tel: +39 010 5636792; Fax: +39 010 383983; E-mail: mirellafilocamo@ospedale-gaslini.ge.it

Communicated by Mark H. Paalman

ABSTRACT: Mutational analysis of the *IDUA* gene was performed in a cohort of 102 European patients with mucopolysaccharidosis type I. A total of 54 distinct mutant *IDUA* alleles were identified, 34 of which were novel including 12 missense mutations, 2 nonsense mutations, 12 splicing mutations, 5 micro-deletions, 1 micro-duplication 1 translational initiation site mutation, and 1 'no-stop' change (p.X654RextX62). Evidence for the pathological significance of all novel mutations identified was sought by means of a range of methodological approaches, including the assessment of evolutionary conservation, RT-PCR/in *vitro* splicing analysis, MutPred analysis and visual inspection of the 3D-model of the IDUA protein. Taken together, these data not only demonstrate the remarkable mutational heterogeneity characterizing type 1 mucopolysaccharidosis but also illustrate our increasing ability to make deductions pertaining to the genotype-phenotype relationship in disorders manifesting a high degree of allelic heterogeneity.

KEY WORDS: *IDUA*, mucoplysaccharidosis type I, Hurler disease, genotype-phenotype analysis, molecular modelling, MutPred analysis, in *vitro* splicing analysis

Received 18 November 2010; accepted revised manuscript 7 February 2011.

DOI: 10.1002/humu.21479

INTRODUCTION

Mucopolysaccharidosis type I (MPS I) is a rare autosomal recessive disorder resulting from deficiency of the lysosomal enzyme α-L-iduronidase (IDUA; E.C. 3.2.1.76) which is involved in the degradation of the glycosaminoglycans (GAGs), namely heparan sulfate and dermatan sulfate. IDUA deficiency leads to the intra-lysosomal accumulation of undegraded GAG substrates and this coincides with the onset of pathology [Neufeld and Muenzer, 2001; Clarke, 2007]. Although the MPS-I clinical phenotype represents a continuous spectrum from the severe to the attenuated forms, three distinct phenotypes have been classically distinguished: 1) severe (Hurler syndrome, MPS IH; MIM# 607014) when the onset of symptoms is before 12 months of age, with survival no more than 10 years and mental retardation manifesting before the age of three; 2) intermediate (Hurler/Scheie syndrome, MPS IH/S; MIM# 607015) when onset of symptoms is between 1 and 6 years, survival is variable and mental retardation is absent or mild, but never before 3 years of age; and 3) 'attenuated' (Scheie syndrome, MPS IS; MIM# 607016) when symptoms first become apparent after the age of 5, survival is normal and mental retardation is never present [Neufeld and Muenzer, 2001].

The gene encoding α-L-iduronidase (*IDUA*; MIM# 252800) maps to chromosome 4p16.3 and contains 14 exons; the cDNA open reading frame (ORF) is ~2 kb in length and encodes a polypeptide of 653 amino acids (Scott et al., 1990, 1991, 1992). So far, more than 100 different disease-causing *IDUA* mutations have been reported (Human Gene Mutation Database; http://www.hgmd.org; Stenson et al., 2009). Several different types of mutation have been documented: whereas missense mutations may allow for some residual enzyme activity and are associated with quite a variable clinical phenotype, those mutations that are likely to impact upon RNA processing (i.e. nonsense, frameshift and splice site mutations) almost invariably result in a more severe phenotype [Terlato and Cox, 2003]. Non-pathogenic functional polymorphisms may also have a role in modifying the expression of *IDUA* mutant alleles, thereby contributing to the phenotypic and clinical heterogeneity characteristic of MPS I [Scott et al., 1995; Beesley et al, 2001; Matte et al., 2003].

Overall, most mutations are 'private', with only four mutations (p.W402X, p.Q70X, p.P533R, p.G51D) being common in specific populations. The most common *IDUA* mutation is p.W402X which has a frequency of around 50% in Northern Europe, the United Kingdom, North America [Beesley et al, 2001; Clarke et al., 1994; Scott et al., 1995] and Spain [Gort et al., 1998], while in Russia, Italy and Brazil its frequency has been estimated to be 4% [Voskoboeva et al., 1998], 11% [Gatti et al., 1997; Venturi et al., 2002] and 20% [Matte et al., 2000], respectively . By contrast, the p.Q70X mutation is much more frequent in Scandinavia and Russia (around 50% of alleles) than in other countries [Bunge et al., 1994; Voskoboeva et al., 1998]. The p.P533R mutation, which probably had a North African origin, has spread to Mediterranean countries, representing 13% and 10% of *IDUA* mutant alleles in Italy and Spain respectively [Alif et al., 2000; Venturi et al., 2002; Voskoboeva et al., 1998]. Finally, the p.G51D mutation seems to be exclusively Italian with a relative frequency of 13% among *IDUA* mutant alleles [Venturi et al., 2002].

We have previously reported *IDUA* gene mutations from a series of 30 Italian patients [Venturi et al., 2002]. Here we have characterized the underlying *IDUA* mutations in a group of 102 newly studied European patients, including 37 Italians, whose condition has been clinically and biochemically diagnosed as MPS I.

MATERIALS AND METHODS

Patients

The present series comprises a total of 102 European patients (pts) affected by MPS I. They are from diverse ethnic backgrounds: 37 were of Italian origin, 23 Polish, 21 Turkish, 18 Spanish, and 3 patients each from Hungary, Serbia and Greece, (Table 1). The diagnosis of MPS I was confirmed biochemically in all patients, demonstrating a defect of IDUA activity either in leukocytes or a fibroblast cell line. As reported in Table 1, the clinical phenotype, defined as previously reported [Neufeld and Muenzer, 2003], was available for all but three patients: 59% of patients presented with the severe form of the disease (MPS IH), 22% with the intermediate form (MPS IH/S) and 15% with the attenuated form (MPS IS); the remaining two patients (2%) manifested a clinical phenotype ranging between MPS IH-H/S (pt #43) and MPS IH/S-S (pt #92).

Thirty seven Italian cases and one Serbian case were recruited with the help of their attending clinicians. The remaining 64 patients from different countries were recruited and collected through a collaboration with Genzyme Corporation (Cambridge, MA, USA) as part of a pan-European project with the aim of performing the molecular characterization of all MPS I patients in Europe and collecting these data in an MPS I Registry (https://www.lsdregistry.net/mpsiregistry) [Pastores et al., 2007].

Table 1. Clinical Phenotypes and Genotypes Encountered in 102 MPS Type 1 Patients with Distinct Ethnic Background

Pt	Ethnic origin	Phenotype	Allele 1[§]	Allele 2[§]
1	Italy	H/S	p.P496R	p.G265R
2	Italy	H	p.A327P	p.A327P
3	Italy	H	p.G51D	p.G51D
4	Italy	H	p.Q70X	p.Q70X
5	Italy	H/S	p.G51D	p.P533R
6	Italy	H	p.G51D	p.P496R
7	Italy	S	p.G51D	p.R89W
8	Italy	H/S	p.P533R	p.P533R
9	Italy	H/S	p.Q70X	?
10	Italy	H/S	p.W402X	**p.L535F**
11	Italy	H/S	p.W402X	p.E178K
12	Italy	S	**p.N348K**	**c.1727+5G>C**
13	Italy	H/S	p.R89W	p.P496R
14	Italy	S	p.G51D	c.1333_1335del3
15	Italy	S	p.Q70X	c.1333_1335del3
16	Italy	H/S	**p.[S423R;F188L]**	**p.[S423R;F188L]**
17	Italy	H	p.Q70X	p.P496R
18	Italy	H	p.W402X	p.W402X
19	Italy	H	**c.1727+1G>A**	p.P496R
20	Italy	H	p.G51D	p.G51D
21	Italy	S	**p.Y581X**	?
22	Italy	H	p.G51D	p.P496R
23	Italy	H	p.Q70X	p.P496R
24	Italy	H	c.46_57del12	**p.Y201X**
25	Italy	H	p.Q70X	p.Q70X
26	Italy	H	?	?
27	Italy	H/S	c.46_57del12	**c.1189+5G>A**
28	Italy	H	c.380_382del3	c.380_382del3
29	Italy	S	p.A327P	c.878_889dup
30	Italy	H	p.Q70X	p.P496R
31	Italy	H	**c.1166_1171dup**	**c.1166_1171dup**
32	Italy	H/S	p.W402X	p.G265R
33	Italy	H	p.Q70X	p.P496R
34	Italy	H	**p.T103P**	?
35	Italy	H	p.Q70X	c.398_403del6
36	Italy	H	p.W402X	p.W402X
37	Italy	H	p.P533R	?
38*	Poland	H	p.A327P	p.W402X
39	Poland	H	p.Q70X	p.W402X
40	Poland	H	p.Q70X	p.Q70X
41	Poland	H	p.Q70X	p.Q70X
42	Poland	H	**c.1045_1047del3**	**c.1045_1047del3**
43	Poland	H - H/S	p.Q70X	**c.1045_1047del3**
44	Poland	S	p.Q70X	p.G265R
45	Poland	S	p.Q70X	c.46_57del12
46	Poland	S	p.Q70X	**p.Y76C**
47	Poland	H	p.W402X	c.1650+5G>A
48	Poland	S	c.46_57del12	**p.X654RextX*62**
49	Poland	S	c.878_889dup	c.1181_1189dup
50	Poland	H	p.W402X	p.W402X
51	Poland	H	p.W402X	**c.1398delC**
52	Poland	H	p.Q70X	c.972+2T>C

Table 1 (continued)

53	Poland	H	p.Q70X	p.Q70X
54	Poland	H	p.Q70X	p.Q70X
55	Poland	H	p.Q70X	p.W402X
56	Poland	H	p.W402X	p.W402X
57	Poland	H	p.Q70X	p.Q70X
58	Poland	H	p.W402X	p.W402X
59	Poland	S	**p.G219E**	**p.W306L**
69	Poland	H/S	p.Q70X	**p.X654RextX*62**
61	Spain	S	p.R492P	?
62	Spain	H	p.Q70X	?
63	Spain	H/S	p.G51D	**c.1727+4C>T**
64	Spain	H/S	**c.1189+4A>G**	p.R621X
65	Spain	H	p.A327P	**p.P385R**
66	Spain	H	p.W402X	p.W402X
67	Spain	H	**p.M1?**	**p.M1?**
68	Spain	H	p.W402X	p.W402X
69	Spain	NA	p.W402X	**p.L396P**
70	Spain	NA	p.W402X	**c.1189+4A>G**
71	Spain	H	p.Q70X	**c.1190-1delG**
72	Spain	H/S	p.P533R	p.P533R
73	Spain	H	**c.574delT**	**c.574delT**
74	Spain	H	p.W402X	p.W402X
75	Spain	H	p.W402X	p.W402X
76	Spain	H	p.Q70X	p.W402X
77	Spain	H	p.W402X	p.W402X
78	Spain	H	p.W402X	**c.385+1G>A**
79	Turkey	NA	**c.1727+6T>A**	?
80	Turkey	H	p.P533R	p.P533R
81	Turkey	H/S	**c.494-1G>A**	**c.494-1G>A**
82	Turkey	S	p.L490P	p.L490P
83	Turkey	H/S	**c.493+1G>A**	**c.493+1G>A**
84	Turkey	H/S	c.46_57del12	c.46_57del12
85^	Turkey	H/S	?	?
86	Turkey	H	**c.494-1G>A**	**c.494-1G>A**
87^	Turkey	H	**c.956_972+9delinsTA**	**c.956_972+9delinsTA**
88^	Turkey	H	p.W402X	p.W402X
89^	Turkey	H	**c.826_828del3**	**c.826_828del3**
90°	Turkey	H	**c.494-1G>A**	**c.494-1G>A**
91	Turkey	H/S	**c.1727+6T>A**	p.Q70X
92	Turkey	H/S - S	**c.494-1G>A**	**p.E276K**
93	Turkey	S	**p.E276K**	**p.E276K**
94	Turkey	H/S	c.46_57del12	c.46_57del12
95	Turkey	H/S	**p.A436P**	**p.A436P**
96	Turkey	H	p.Q70X	p.Q70X
97	Turkey	H	p.W402X	p.W402X
98	Turkey	H	**c.494-1G>A**	**c.494-1G>A**
99	Turkey	H	**c.1650+5G>C**	**c.1893delC**
100	Hungary	H	p.Q70X	p.W402X
101	Serbian	H	p.Q70X	p.Q70X
102	Greece	H/S	**p.G84R**	**p.E276K**

H=Hurler; S=Scheie; NA=Not available; ?=Unknown; Pt=Patient; Novel mutations are marked in bold. [§]GenBank-EMBL accession no. NM_000203.3 (c.DNA considering the A of the ATG translation initiation start site as nucleotide +1) and GenBank-EMBL accession no. NP_000194 (protein); *Patient not analysed in our Lab; ^indicates parents' consanguinity; °Patient's parents belonging to the same village; Underlined patients belonging to the same geographical region.

Some of the samples from the Italian patients were supplied by the "Cell line and DNA Biobank from patients affected by Genetic Diseases" in Genova (http://dppm.gaslini.org/biobank/). Following ethical guidelines, all samples obtained for analysis and storage required prior written informed consent using a form approved by the Local Ethics Committee.

Cell culture

Fibroblast cells were cultured according to standard procedures. The cell lines were cultured and maintained in RPMI medium (EuroClone, Gibco, Paisley, UK) containing 15% FCS and penicillin/streptomycin, in a humidified atmosphere containing 5% CO_2 at 37°C.

Enzymatic assay

α-L-iduronidase activity was assayed in homogenates of leukocytes and/or fibroblast cell lines using 4-methylumbelliferyl-α-L-iduronide (Glycosynth, Cheshire, UK) as a substrate [Stirling et al., 1978]. Total protein was measured according to the Lowry method [Lowry et al., 1951].

Molecular analysis

Genomic DNA was extracted from circulating lymphocytes and/or cultured fibroblasts using either the Wizard Genomic DNA Purification kit (Promega, Madison, WI) or a standard phenol-chloroform-based method.

IDUA gene exons, exon–intron boundaries and part of the 5' untranslated region were PCR amplified using specific primers designed by reference to the genomic sequence (GenBank-EMBL Accession no. NG_008103.1). Intronic primers, designed for both PCR and sequence analyses, are listed in Supp. Table S1. PCR amplification reactions (except for exon 9) were carried out in 25μl vols containing 100ng genomic DNA, buffer 1X with 1.5mM $MgCl_2$, 200μM dNTPs mix, 15pmol each primer, GC-RICH solution 1X and 0.5U proofreading PWO SuperYield DNA polymerase (Roche, Monza, Italy). Exon 9 PCR-amplification was carried out in 25μl vols containing 100ng genomic DNA, buffer 1X with 1.5mM $MgCl_2$, 200μM dNTPs mix, 15pmol each primer and 0.5U GoTaq DNA polymerase (Promega, Madison, WI); owing to the use of a no-proofreading DNA polymerase, this reaction was performed in duplicate. Cycling conditions were: initial denaturation at 96°C for 5 min, 28 cycles denaturation at 96°C for 1 min, annealing at 63-70°C for 1 min, extension at 72°C for 1 min, followed by final extension at 72°C for 7 min. Control PCR reactions, in which no template was added, were included during each set of PCR reactions. PCR products were controlled for contamination by gel electrophoresis and were then purified using an enzymatic reaction containing 5U exonuclease I (Celbio, Italy) and 1U alkaline phosphatase (Promega, Madison, WI) under the following conditions: 15 min at 37°C followed by 15 min at 80°C.

Total RNA was prepared from peripheral blood lymphocytes isolated by the Fycoll method or from cultured fibroblasts using Trizol Reagent (Gibco, Paisley, UK) according to the manufacturer's instructions. First-strand cDNAs were synthesized using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA) and random hexamer primers. *IDUA* exonic fragments flanking splicing mutations were amplified and sequenced using primers and conditions listed in Supp. Table S1.

Sequence analysis of PCR and RT-PCR products was performed in the forward and reverse directions using BigDye v3.1 terminator technology (Applied Biosystems, Foster City, CA) and then purified with Wizard MagneSil Sequencing Reaction Clean-Up System (Promega, Madison, WI). Sequencing reactions were carried out, the products purified according to the manufacturer's instructions and analyzed on an ABI Prism 3130 Avant Automatic Sequencer (Applied Biosystems, Foster City, CA). Sequence alterations were confirmed by sequencing duplicate PCR products.

The question of the pathological authenticity of the novel *IDUA* sequence alterations detected was addressed by (i) searching dbSNP (http://www.ncbi.nlm.nih.gov/SNP) for their presence, (ii) screening 100 alleles from healthy control subjects for each alteration, (iii) modeling the amino acid changes into a homologous three-dimensional structure of the protein, (iv) employing the MutPred program [Li et al., 2009; Mort et al., 2010] for the missense mutations, (v) using ESRPred and a neural network for splice site prediction [Krawczak et al., 2007] for the coding region variants upon splicing, and (vi) using PhyloP method [Pollard et al., 2010] and ClustalW2 program [Larkin et al., 2007] to measure evolutionary conservation of nucleotides and residues, respectively.

Molecular modeling of missense mutations

For modeling of the *IDUA* missense mutations, an homology model of the IDUA protein was used. This model was built by Rempel et al. (2005) using the atomic coordinates of β-xylosidase (E.C.3.2.1.37) from *Thermoanaerobacterium saccharolyticum* as a template [PDB code 1Y24]. Visual inspection and graphical representations were then performed using the programs *Coot* and *Chimera*, respectively [Emsley and Cowtan, 2004; Pettersen et al., 2004].

Bioinformatic analysis of IDUA variants

Missense variants in the *IDUA* gene were analysed with a computational model, MutPred [Li et al., 2009, Mort et al., 2010]. MutPred is designed to assess the likely phenotypic consequences of missense substitutions occurring within sites of structural and/or functional importance, on the wild-type protein sequence. MutPred can be used to generate hypotheses as to the underlying molecular mechanism(s) responsible for disease pathogenesis.

The effect of the coding region variants upon splicing [splice site disruption, cryptic splice site activation and exon skipping via loss of exonic splicing enhancers (ESE) and/or gain of exonic splicing silencers (ESS)] was ascertained using ESRPred [M. Mort, unpuplished] and a neural network for splice site prediction [Krawczak et al., 2007].

Evolutionary sequence conservation across an alignment of IDUA orthologues from 44 vertebrate species was measured using the phyloP method [Pollard et al., 2010]. PhyloP can measure accelerated evolution (more rapid evolution than expected under neutral drift) as well as evolutionary conservation (slower than expected evolution). A positive phyloP score represents an evolutionarily conserved nucleotide and a negative phyloP score indicates that the nucleotide has experienced more rapid evolution than would be expected under neutral drift.

Mutation nomenclature

Mutations are described according to current mutation nomenclature guidelines (http://www.hgvs.org/mutnomen), ascribing the A of the first ATG translational initiation codon as nucleotide +1 [den Dunnen and Antonarakis, 2000; den Dunnen and Paalman, 2003].

RESULTS AND DISCUSSION

A total of 102 European patients affected by MPS I were recruited by our centre. All patients underwent molecular characterization and the *IDUA* genotype was completely ascertained in 93 of the 102 patients.

Mutation detection

All 14 exons, splice junctions and proximal portions of the 5' and 3' untranslated regions of the *IDUA* genes of the 102 unrelated MPS I patients were investigated by DNA sequence analysis. Putatively pathological *IDUA* mutations were identified in 193 of 204 alleles. The genotypes, clinical phenotypes and geographic origins of the 102 patients studied here are listed in Table 1. The characteristics of the 55 distinct *IDUA* mutations identified are reported in Table 2. The mutational spectrum comprised 22 missense mutations, 14 splice site alterations, 9 micro-deletions, 5 nonsense mutations, 3 micro-duplications, 1 translational initiation site mutation and 1 no-stop mutation. More than 60% (35/55) of these mutant alleles have not been previously reported, attesting to the extensive allelic heterogeneity of MPS I. The novel lesions resulted from (i) 11 missense mutations (p.Y76C, p.G84R, p.T103P, p.G219E, p.E276K, p.W306L, p.N384K, p.P385R, p.L396P, p.A436P, p.L535F) as well as two missense mutations located in *cis* to each other (p.[F188L;S423R]); (ii) 12 splicing mutations (c.385+1G>A, c.493+1G>A, c.494-1G>A, c.956_972+9delinsTA, c.1189+4A>G, c.1189+5G>A, c.1190-1delG, c.1650+5G>C, c.1727+1G>A, c.1727+4C>T, c.1727+5G>C, c.1727+6T>A); (iii) 5 micro-deletions (c.574delT, c.826_828del3, c.1045_1047del3, c.1398delC and c.1893delC), 1 micro-duplication (c.1166_1171dup), 2 nonsense mutations (p.201X and p.Y581X), 1 translational initiation site mutation (p.M1?) and 1 'no-stop' mutation (p.X654RextX62).

In addition to these putative pathological mutations, a total of 37 putative intragenic *IDUA* polymorphisms were identified within either the exons or introns; eight novel SNPs were noted within the coding region (4) and in the introns (4) (Supp. Table S2). The allele frequencies of the novel variants, derived from 100 normal control alleles, are given in Supp. Table S2. Figure 1 depicts the location of the various mutations and polymorphisms detected in relation to the structure of the *IDUA* gene.

Table 2. Characteristics of the *IDUA* gene mutations identified in the 102 MPS type I patients and MutPred analysis of the missense mutations

Loc.	Site of nucleotide substitution*	Predicted effect on protein structure **	Type of mutation	No. of alleles (%)	Evolutionary Conservation (phyloP)	MutPred analysis of missense mutations #		References
						Deleterious mutation probability	Summary of confident *in silico* hypotheses	
Ex. 1	**c.1A>C**	**p.M1?**	**Translational initiation site**	2(0.01)	Acc (-0.3)	0.99		Present study[§]
	c.152G>A	p.G51D	Missense	10(4.9)	Cons (2.7)	0.97	Protein structure disruption	Bunge et al. (1994)
	c.46_57del12	p.S16_A19del	In-frame del	8(3.9)	.	.		Bunge et al. (1994)
Ex. 2	c.208C>T	p.Q70X	Nonsense	38(18)	Cons (2.1)	.		Scott et al. (1992)
	c.227A>G	**p.Y76C**	**Missense**	1(0.05)	Cons (2.7)	0.73		Present study
	c.250G>C	**p.G84R**	**Missense**	1(0.05)	Cons (5.5)	0.93	Gain of binding site and disruption to protein structure	Present study
	c.265C>T	p.R89W	Missense	2(0.01)	Cons (0.3)	0.96		Bunge et al. (1995)
Ex. 3	**c.307A>C**	**p.T103P**	**Missense**	1(0.05)	Cons (0.4)	0.36		Present study
	c.380_382del3	p.L127del	In-frame del	2(0.01)	.	.		Venturi et al. (2002)
Intr. 3	**c.385+1G>A**	**r.spl?**	**Splicing**	1(0.05)	Cons (5.1)	.		Present study
Ex. 4	c.398_403del6	p.M133_G134del	In-frame del	1(0.05)	.	.		Venturi et al. (2002)
Intr. 4	**c.493+1G>A**	**r.spl?**	**Splicing**	2(0.01)	Cons (6.2)	.		Present study
	c.494-1G>A	**p.R166TfsX27**	**Splicing**	9(4.4)	Cons (5.7)	.		Present study
Ex. 5	c.532G>A	p.E178K	Missense	1(0.05)	Cons (4.5)	0.97	Gain of functional sites including PTMs	Venturi et al. (2002)
	c.[562T>C;1269C>A]⁺	**p.[F188L; S423R]⁺**	**Missense**	2(0.01)	Cons(3.0;1.2)	0.95;0.92	Gain of methylation at S423	Present study
	c.574delT	**p.S192PfsX2**	**Frameshift**	2(0.01)	.	.		Present study
Ex. 6	**c.603C>G**	**p.Y201X**	**Nonsense**	1(0.05)	Cons (0.7)	.		Present study
	c.656G>A	**p.G219E**	**Missense**	1(0.05)	Cons (5.9)	0.89	Protein structure disruption	Present study
Ex. 7	c.793G>C	p.G265R	Missense	3(1.4)	Cons (6.2)	0.95	Secondary structure disruption, gain of methylation at G265	Yogalingam et al.(2004)
	c.826G>A	**p.E276K**	**Missense**	4(1.9)	Cons (6.2)	0.76		Present study
	c.826_828del3	**p.E276del**	**In-frame del**	2(0.01)	.	.		Present study
	c.878_889dup	p.T293_Y296dup	In-frame dup	2(0.01)	.	.		Bunge et al. (1995)
	c.956_972+9delinsTA	**r.spl?**	**Splicing**	2(0.01)	.	.		Present study
Intr. 7	c.972+2T>C	r.spl?	Splicing	1(0.05)	Cons (1.6)	.		Scott et al. (1993)
Ex. 8	**c.917G>T**	**p.W306L**	**Missense**	1(0.05)	Cons (6.0)	0.82	Loss of catalytic residue at P309	Present study
	979G>C	p.A327P	Missense	5(2.4)	Cons (1.2)	0.84	Protein structure disruption and gain of methylation at K324	Bunge et al. (1995)
	c.1044C>G	**p.N348K**	**Missense**	1(0.05)	Acc (-0.4)	0.85	Decrease in protein stability and gain of PTMs	Present study
	c.1045_1047del3	**p.D349del**	**In-frame del**	3(1.4)	-	.		Present study
	c.1154C>G	**p.P385R**	**Missense**	1(0.05)	Cons (5.4)	0.82	Protein structure disruption	Present study
	c.1166_1171dup	**p.A389_M390dup**	**In-frame dup**	2(0.01)		.		Present study
	c.1187T>C	**p.L396P**	**Missense**	1(0.05)	Cons (3.1)	0.83	Decrease in protein stability and secondary structure disruption	Present study
	c.1181_1189dup	p.A394_L396dup	In-frame dup	1(0.05)	.	.		Bunge et al. (1995)
Intr. 8	**c.1189+4A>G**	**p.V371MfsX43**	**Splicing**	2(0.01)	Cons (4.3)	.		Present study
	c.1189+5G>A	**p.V371MfsX43**	**Splicing**	1(0.05)	Cons (2.9)	.		Present study
	c.1190-1delG	**r.spl?**	**Splicing**	1(0.05)	-	.		Present study
Ex. 9	c.1205G>A	p.W402X	Nonsense	37(18)	Cons (1.8)	.		Scott et al. (1992)
	c.1306G>C	**p.A436P**	**Missense**	2(0.01)	Cons (2.8)	0.45		Present study
	c.1333_1335del3	p.D445del	In-frame del	2(0.01)	.	.		Bunge et al. (1995)
	c.1398delC	**p.P467RfsX58**	**Frameshift**	1(0.05)	.	.		Present study
Ex. 10	c.1469T>C	p.L490P	Missense	2(0.01)	Cons (0.4)	0.86	Decrease in protein stability and protein structure disruption	Tieu et al. (1995)
	c.1475G>C	p.R492P	Missense	1(0.05)	Cons (0.7)	0.83	Loss of binding site	Tieu et al. (1995)
	c.1487C>G	p.P496R	Missense	9(4.4)	Cons (4.4)	0.95	Gain of binding site	Beesley et al. (2001)
Ex. 11	c.1598C>G	p.P533R	Missense	8(3.9)	Cons (4.1)	0.94	Gain of binding site	Scott et al. (1992)
	c.1603C>T	**p.L535F**	**Missense**	1(0.05)	Cons (1.7)	0.59		Present study
Intr. 11	c.1650+5G>A	r.spl?	Splicing	1(0.05)	Cons (4.7)	.		Venturi et al. 2002
	c.1650+5G>C	**r.spl?**	**Splicing**	1(0.05)	Cons (4.7)	.		Present study
Intr. 12	**c.1727+1G>A**	**r.spl?**	**Splicing**	1(0.05)	Cons (3.8)	.		Present study
	c.1727+4C>T	**p.L578VfsX14**	**Splicing**	1(0.05)	Cons (0.4)	.		Present study
	c.1727+5G>C	**Exon 12 skipping**	**Splicing**	1(0.05)	Cons (5.4)	.		Present study
	c.1727+6T>A	**Exon 12 skipping**	**Splicing**	2(0.01)	Cons (1.1)	.		Present study
Ex. 13	**c.1743C>G**	**p.Y581X**	**Nonsense**	1(0.05)	Acc (-1.1)	.		Present study
Ex. 14	c.1861C>T	p.R621X	Nonsense	1(0.05)	Cons (1.5)	.		Bunge et al. (1994)
	c.1893delC	**p.F632SfsX105**	**Frameshift**	1(0.05)		.		Present study
	c.1960T>C	**p.X654RextX62**	**No-stop**	2(0.01)	Acc (-0.4)	.		Present study

Legend: Loc= Location; Ex=exon; Intr=intron; del=deletion; dup=duplication; Acc= Accelerated; Cons= Conserved; PTMs= Post-translational modifications **IDUA* gene GenBank-EMBL accession no. NM_000203.3; ***IDUA* gene GenBank-EMBL accession no. NP_000194; the novel mutations are given in bold; #Li et al. (2009) and Mort et al. (2010). ⁺Allele carrying 2 novel mutations in *cis;* note that the second change occurs in exon 9 and it was previously reported as single allele as due to c.1269C>G by Yogalingam et al. (2004); [§]mutation previously reported as due to a c.3G>A by Lee-Chen and Wang (1997)

Missense mutations

A total of 14 novel *IDUA* missense mutations were detected during the course of this study; two being located in *cis* on the same *IDUA* allele (p.[F188L;S423R]). Consistent with their putative pathological authenticity, none of the underlying nucleotide changes were found in a screen of 100 control alleles. Nor were any of these substitutions listed as polymorphisms in dbSNP (http://www.ncbi.nlm.nih.gov/projects/SNP).

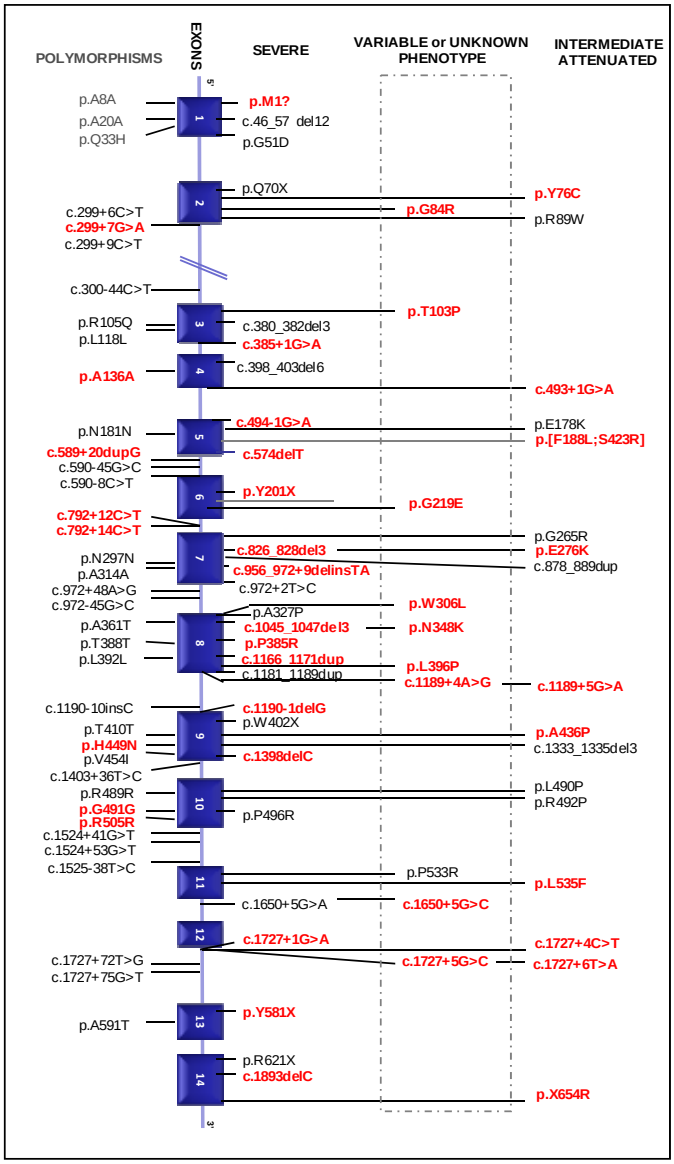


Figure 1. Distribution of the 54 detected mutant alleles and 37 putatively neutral polymorphisms in relation to the structure of the *IDUA* gene. The schematic map of the *IDUA* gene depicts the positions of the 14 numbered exons (blue boxes) and associated unnumbered introns (blue lines). An indication of the clinical phenotype associated with each of the 54 different mutant alleles identified in the 102 patients is provided. Thus, on the right of the diagram are listed those mutations associated with severe and attenuated (intermediate+mild) phenotypes respectively. Mutations associated with variable or unknown clinical phenotypes are given in the gray dotted box. On the left of the diagram, the polymorphic variants detected during the course of this study are listed. Novel alleles (34) and novel polymorphic variants (8) are marked in red.

Bioinformatic analysis of missense variants

MutPred predicted that most of the missense mutations listed in Table 2 are deleterious (91%; 21 out of 23; MutPred general score > 50). Further, confident *in silico* hypotheses for the underlying mechanism of pathogenesis were generated for 14 of the missense mutations (summarised in Table 2 and reported in full in Supp. Table S3).

The two missense mutations (p.T103P and p.A436P) predicted not to be deleterious by MutPred, were analysed further (Supp. Table S4). PhyloP scores were used to measure evolutionary conservation at the nucleotide sequence level across IDUA orthologues derived from 44 vertebrate species. Both mutations (p.T103P and p.A436P) exhibited positive phyloP scores indicating that these mutations are located at conserved nucleotides. It should be noted that p.T103P displays only marginal conservation, with a phyloP score of 0.40. Analysis with SIFT [Ng and Henikoff, 2001] predicted that both mutations would be tolerated in the protein. PolyPhen2 [Adzhubei et al., 2010] however predicted that whilst p.T103P was benign, p.A436P mutation was likely to be potentially damaging. Neither mutation was predicted to disrupt splicing using a neural network [Krawczak et al., 2007]. MutPred analysis was also performed on the six polymorphic missense variants identified (listed in Supp. Table S2) but none of these were predicted to be deleterious (MutPred general scores invariably < 50).

Further evidence for the pathological/functional significance of the missense mutations was then sought from the analysis of the extent of evolutionary conservation of the mutated residues in 6 orthologous (vertebrate) IDUA proteins from human to *Tetraodon nigroviridis*. The computational analysis, carried out using ClustalW2 [http://www.ebi.ac.uk/clustalw; Larkin et al., 2007], revealed that among the 5 other vertebrate IDUA protein sequences examined, Y76, G84, F188, G219, R276, W306, N348, P385, L396, S423, L535 were invariant at these positions; such conservation over 500 Myrs of evolutionary time is supportive of the functional significance of these residues and hence the direct involvement of the respective mutations in disease causation. In agreement with the PolyPhen score results, residue A436 was found to be partially conserved (chimpanzee, dog and chicken) but T103 was not evolutionarily conserved. (Supp. Figure S1). It may be that the p.A436P mutation is mildly deleterious, whereas p.T103P may be tolerated and could therefore represent a rare polymorphism.

3D modelling of IDUA missense mutations

To shed further light on the deleterious consequences of the novel missense *IDUA* mutations at the protein level, we employed the homology model of IDUA constructed by Rempel et al. [2005]. Since the IDUA model starts at residue 36 and terminates at residue 522, the p.L535F substitution could not be included in this analysis. As shown in Figure 2, and detailed in Supp. Table S5, several substitutions (p.G219E, p.P385R and p.S423R) appear to affect a hydrophobic buried region in the protein, one substitution occurs in the predicted active site residue (p.F188L), and one substitution in the active site cleft (p.W306L) [Rempel et al., 2005]. The remaining amino acid replacements (p.G84R, p.Y76C, p.T103P, p.E276K, p.N348K, p.L396P and p.A436P) affected solvent accessible residues. Consistent with previous findings [Rempel et al., 2005; Vazna et al., 2009; Sugawara et al., 2008], those mutations that occur in the protein core (including the predicted active site), were associated with the severe (H) form of MPS I, whereas the missense mutations which affected surface-located residues were associated with the more attenuated clinical phenotypes (H/S, S) (Figure 2 and Supp. Table S5). Despite obtaining data from both structural and bioinformatic analyses, the pathological significance of the p.T103P, which occurs in a long loop of unknown function that protrudes into the solvent, remains unclear.

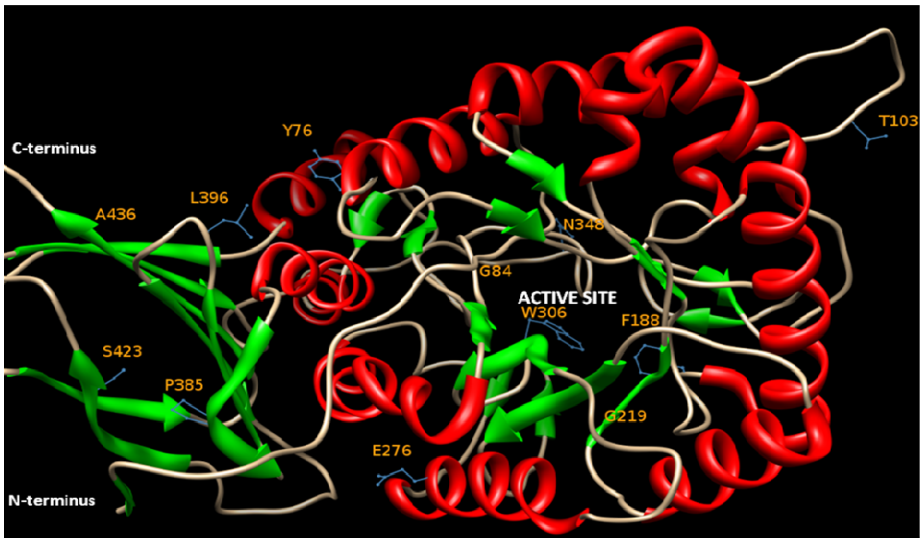


Figure 2. Ribbon representation of the three dimensional structure of (the homology model of) wild-type IDUA [Rempel et al., 2005]. The α helices are colored red and the β sheets green; the wild-type residues analysed are depicted as blue sticks.

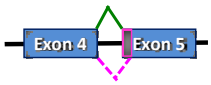
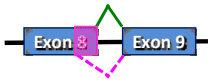
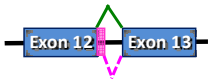
Splicing mutations

A substantial proportion of the novel *IDUA* mutations detected (34.3%) occurred in either the donor (c.385+1G>A, c.493+1G>A, c.1727+1G>A) or acceptor (c.494-1G>A, c.1190-1delG) splice sites or in the surrounding splicing motifs (c.1189+4A>G, c.1189+5G>A, c.1650+5G>C, c.1727+4C>T, c.1727+5G>C, c.1727+6T>A) (Table 2). All the single base-pair substitutions predicted to disrupt splicing were located at nucleotides which were evolutionarily conserved across an alignment of *IDUA* gene sequences from 44 vertebrate species. We also identified an indel mutation (c.956_972+9delinsTA) as the cause of aberrant splicing in one case. At the genomic level, the mutation comprised the deletion of 26 nucleotides and the insertion of the dinucleotide TA in a region overlapping the exon-intron 7 donor splice site.

Supp. Figure S2 graphically represents the intronic location of the 12 novel splicing mutations identified in MPS I patients in the present study. To assess the potential effect of these mutations on mRNA splicing, the relative strengths of the 5' and 3' splice site signals were evaluated using the NetGene2 server [http://www.cbs.dtu.dk/services/NetGene2]. According to this program, all 12 mutations would be predicted to impact upon RNA processing. In particular, donor splice sites would have been abrogated in the presence of c.385+1G>A, c.493+1G>A, c.1650+5G>C, c.1727+1G>A, c.1727+5G>C and c.1727+6T>A as would acceptor sites in the presence of c.494-1G>A and c.1190-1delG. The same analysis predicted a considerable decrease in the donor splice site confidence score in the cases of c.1189+4A>G, c.1189+5G>A and c.1727+4C>T.

These predictions were then followed up by performing *in vitro* splicing analysis for those mutations identified in patients from whom appropriate samples were available. The functional relevance of six intronic mutations (c.494-1G>A, c.1189+4A>G, c.1189+5G>A, c.1727+4C>T, c.1727+5G>C and c.1727+6T>A) was assessed by carrying out reverse transcriptase-polymerase chain reaction (RT-PCR) analysis on mRNA extracted from either lymphocytes (3 pts) or fibroblast cell lines (3 pts). In the case of c.1727+5G>C and c.1727+6T>A, the RT-PCR results clearly confirmed the abrogation of the expected donor splice site leading to the skipping of exon 12 or the activation of alternative splice sites (exonic or intronic) leading to frameshifts with consequent premature truncation of the protein (p.Y167TfsX, p.V371MfsX, p.L578VfsX). Full details are given in Table 3.

Table 3. Summary of the results of the reverse transcriptase-PCR analysis on mRNA samples

MUTATION	INTRON	EFFECT ON SPLICING	PREDICTED EFFECT ON PROTEIN	PREDICTED TRANSCRIPTS
c.494-1G>A	4	Abrogation of intron 4 acceptor 3'ss Activation of alternative 3'ss in exon 5 (4 nucleotides downstream of the canonical 3'ss)	Alteration of ORF and introduction of premature stop codon ↓ **p.R166TfsX27**	
c.1189+4A>G **c.1189+5G>A**	8	Abrogation of intron 8 donor 5'ss Activation of alternative 5'ss in exon 8 (79 nucleotides upstream of the canonical 5'ss)	Alteration of ORF and introduction of premature stop codon ↓ **p.V371MfsX43**	
c.1727+4C>T	12	Abrogation of intron 12 donor 5'ss Activation of alternative 5'ss in intron 12 and retention of the first 4 nucleotides of intron 12	Alteration of ORF and introduction of premature stop codon ↓ **p.L578VfsX14**	
c.1727+5G>C **c.1727+6T>A**		Abrogation of intron 12 donor 5'ss	**Skipping of the exon 12**	

ss=splice site; ORF=open reading frame. On the right: normal splicing is depicted as unbroken green lines and abnormal splicing as dotted fucsia lines; blue boxes indicate normally spliced exons and fucsia boxes genomic (exonic and intronic) regions involved in the altered splicing process

Genotype-phenotype relationship

Genotypes, clinical phenotypes and geographic origins of the 102 patients studied here are listed in Table 1. Some 45 (44%) of the patients (pts) were found to be either homozygous for the pan-ethnic *IDUA* mutations, p.W402X (12 pts), p.Q70X (9 pts) and p.P533R (3 pts), other known mutations (7 pts) or novel lesions (14 pts).

The remaining 54% of patients were compound heterozygotes (including 7 in whom the second mutant *IDUA* allele could not be identified). Finally, the mutant *IDUA* alleles in two patients (2%), in whom a defect in IDUA enzyme activity had been documented, remained unidentified.

Overall, we found 68 distinct genotypes including those (7) that were partially characterized. Our results confirm the high degree of mutational heterogeneity characteristic of MPS I. Extensive allelic heterogeneity often precludes the recognition of correlations between mutant genotypes and variant clinical phenotypes. This notwithstanding, in an attempt to obtain new insights into the genotype-phenotype relationship in MPS I, we have compared the genotypes present in our patient series with those of previously reported patients. Supp. Table S6, reports the comparable genotypes (13/68) and the respective number of patients per genotype reported in previous studies on this topic.

Among the analyzed group, 25 patients were either homozygous or compound heterozygous for the two common deleterious nonsense mutations (p.Q70X and p.W402X); their severe clinical phenotype concurred with those of 91 previously reported patients with comparable genotypes [Bunge et al., 1994; Gort et al., 1998; Hein et al., 2003; Li et al., 2002; Matte et al., 2003; Scott et al., 1992; Vazna et al., 2009; Venturi et al., 2002; Voskoboeva et al. 1998]. In agreement with previously reported genotype-phenotype correlations, a severe phenotype was not only associated with the p.W402X mutation in compound heterozygosity with both c.1650+5G>A and p.A327P [Bunge et al., 1994; Venturi et al., 2002; Vazna et al., 2009] but also with homozygosity for p.A327P and p.G51D [Gatti et al., 1997]. Also consistent with previous data [Beesley et al, 2001; Tieu et al., 1995], homozygosity for p.L490P and compound heterozygosity for p.Q70X and c.1333_1335del3 yielded mild phenotypes in two other patients (Supp. Table S6). Although for these 38 patients (37%), the predicted genotype-phenotype relationship may be considered informative, for the remaining group with previously described genotypes, the correlations only partially concurred (Supp. Table S6). This latter group comprised 6 patients, 3 of whom were homozygous either for the p.P533R mutation or for the microdeletion c.46_57del12 (2 pts), and another patient who was compound heterozygous for p.G51D and p.P533R. Comparison between the genotype-phenotype correlation data from our 6 patients and 23 patients from other studies [Alif et al., 1999; Hein et al., 2003; Gatti et al., 1997; Hein et al., 2003, Laradi et al., 2005; Matte et al., 2003; Scott et al., 1992; Venturi et al., 2002] confirmed the phenotypic variability of these mutations (p.P533R and c.46_57del12). Similarly, in our patient series (Table 1), these two latter mutations (p.P533R and c.46_57del12), although present in both the homozygous or compound heterozygous state, and in a significant number of individuals (viz. pts #5, #8, #37, #72, #80 and pts #24, #27, #45, #48, #84, #94, respectively), were found to be associated with a wide range of clinical severity.

Additional conclusions were drawn by individually comparing some mutations in our series (Table 1) with those from previous studies. According to Hein et al. [2003], mildly deleterious consequences could be confirmed for the mutation p.R89W, contributing to the mild form of MPS I noted in pts #7 and #13, in whom this allele was in association with the already known severe mutations, p.G51D and p.P496R, respectively. In our own patients, this latter allele (p.P496R) was invariably found to be associated with the severe form of MPS I (pts #6, #17, #19, #22, #23, #30, #33); this finding confirms the assessment of its detrimental nature, made in *vitro,* by Beesley et al. [2001]. The only patient harbouring p.P496R to exhibit an intermediate clinical form of MPS I (pt #1) was a compound heterozygote in whom the clinical effect of this detrimental allele was probably attenuated by the relative benign influence of p.G265R. In support of this latter supposition was the compound heterozygosity (p.G265R+p.Q70X) observed in the Scheie patient #44 (Table 1). Conversely, conclusions as to the genotype-phenotype relationship manifested by mutation c.494-1G>A, found in either the homozygous (pts #81, #86, #90, #98) or compound heterozygous (pt #92) state, were equivocal since this lesion was associated with very different degrees of clinical severity, even in patients who appeared to share the same genetic background (Table 1). It may be that other genetic or non-genetic factors can modulate the clinical phenotype in MPS I. Among the possible genetic factors involved, as previously reported in the same MPS I [Yogalingam et al., 2004] or in other MPSs such as MPS II [Lualdi et al., 2006], are differentially spliced-RNA transcripts that might result from the splicing

mutation c.494-1G>A, and could potentially be considered as modifiers of the clinical phenotype [Nissim-Rafinia and Kerem, 2002]. Unfortunately, RNA samples were not available to permit experimental verification.

Another factor limiting our ability to define effective genotype-phenotype correlations was probably our difficulty in attributing one or other clinical phenotype in a given case, bearing in mind that the clinical spectrum of disease in MPS I is in reality a continuum of phenotypes with gradually changing severity. This is particularly true for the so called 'intermediate form', where it is difficult to be truly objective, especially when the patients are young at the time of diagnosis and there is no standardized scoring index of severity.

Frequency and geographical distribution of common and rare mutations in the European polulation

As reported in Table 2, in the whole series comprising patients with diverse genetic backgrounds, the frequency of the most common mutations was: p.Q70X (18.6%), p.W402X (18%), p.G51D (4.9%), p.P496R (4.4%) p.P533R (3.9%), p.A327P (2.4%) and c.46_57del12 (3.9%). Altogether, they represented 56% of *IDUA* alleles. Rare or private mutations comprised 49.5% of *IDUA* lesions, and 28.4.% of these were novel. However, when we analysed the geographical distribution and relative frequencies of the five most frequent mutations in different areas of Europe, we observed significant differences in the distribution of these mutations (Figure 3).

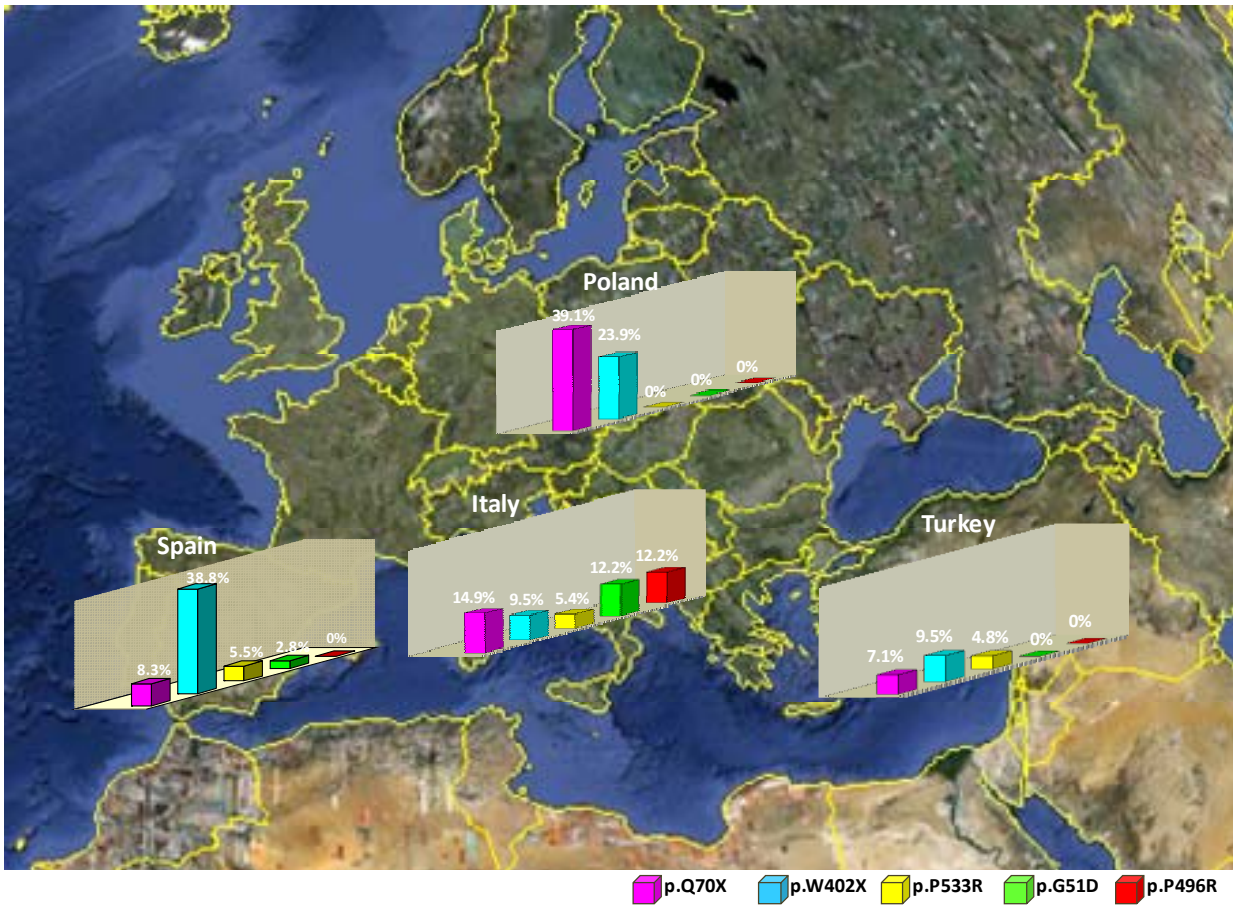


Figure 3. Geographical-distribution and relative proportion of the common mutations found in the present series.

In Western Europe, the most frequent mutation was p.W402X, representing 38.8% of alleles in Spain. Although its relative frequency in the other Mediterranean countries decreased sharply to 9.5% in both Italy and Turkey, p.W402X maintained a relatively high frequency in Northern Europe, accounting for 24% of *IDUA* alleles in our series of patients from Poland. Conversely, the p.Q70X mutation accounted for 39% of *IDUA* alleles in Poland but

progressively decreased in terms of its frequency in a Southerly direction, constituting 14.9% in Italy, 8.3% in Spain and 7.1% in Turkey. On the basis of these results, and supported by previous data [Terlato and Cox, 2003], it could be speculated that the European geographical diffusion of p.W402X occurred from the South-West in a Northerly direction, while that of p.Q70X occurred from the North-East in a Southerly direction. Mediterranean countries appear to have balanced the relatively low frequency of these two common mutations with other mutations specific to this geographical region viz. p.P533R, p.P496R and p.G51D, the latter two being specific to (and recurrent in) the Italian population, and c.494-1G>A, a Turkish-specific mutation (not shown in the figure). The relatively low frequency of the two usually common mutations (p.Q70X and p.W402X) in Italy and Turkey, results in an increase in *IDUA* allelic heterogeneity in these two countries as compared to Northern European countries. Thus, rare mutations accounted for 50% and 71.4% of alleles in Italy and Turkey, respectively, as compared to Poland and Spain where they were 37% and 41.6% respectively. The finding that European countries vary widely in terms of both the distribution and frequency of *IDUA* gene mutations has important implications for the selection of screening strategies in individual countries. The existence of dramatic differences in mutational heterogeneity and mutation prevalence highlights the importance of multi-national screening studies in helping to elucidate the genotype-phenotype relationship in disorders such as MPS I that are characterized by extensive allelic heterogeneity.

ACKNOWLEDGMENTS

European patient samples were obtained through a collaboration with Genzyme Corporation, Cambridge, MA, USA. Some samples were also obtained from the "Cell Line and DNA Biobank from Patients Affected by Genetic Diseases" (G. Gaslini Institute) – Telethon Genetic Biobank Network (Project No. GTB07001A). We also thank Prof. Generoso Andria (Napoli), Dr. Rita Barone (Catania), Prof. Cesare Danesino (Pavia), Prof. Paola Di Natale (Napoli), Dr. Graziella Uziel (Milano), for their contribution of clinical information and patient samples.

REFERENCES

Adzhubei IA, Schmidt S, Peshkin L, Ramensky VE, Gerasimova A, Bork P, Kondrashov AS, Sunyaev SR. 2010. A method and server for predicting damaging missense mutations. Nat Methods 7:248-249.

Alif N, Hess K, Straczek J, Sebbar S, Belahsen Y, Mouane N, Abkari A, Nabet P, Gelot MA. 2000. Mucopolysaccharidosis type I in Morocco: clinical features and genetic profile. Arch Pediatr 7:597-604.

Beesley CE, Meaney CA, Greenland GA, Adams V, Vellodi A, Young EP, Winchester BG. 2001. Mutational analysis of 85 mucopolysaccharidosis type I families: frequency of known mutations, identification of 17 novel mutations and in vitro expression of missense mutations. Hum Genet 109:503-511.

Bunge S, Kleijer WJ, Steglich C, Beck M, Schwinger E, Gal A. 1995. Mucopolysaccharidosis type I: identification of 13 novel mutations of the alpha-L-iduronidase gene. Hum Mutat 6:91-94.

Bunge S, Kleijer WJ, Steglich C, Beck M, Zuther C, Morris CP, Schwinger E, Hopwood JJ, Scott HS, Gal A. 1994. Mucopolysaccharidosis type I: identification of 8 novel mutations and determination of the frequency of the two common alpha-L-iduronidase mutations (W402X and Q70X) among European patients. Hum Mol Genet 3:861-866.

Cartegni L, Chew SL, Krainer AR. 2002. Listening to silence and understanding nonsense: exonic mutations that affect splicing. Nat Rev Genet 3:285-298.

Clarke LA, Nelson PV, Warrington CL, Morris CP, Hopwood JJ, Scott HS. 1994. Mutation analysis of 19 North American mucopolysaccharidosis type I patients: identification of two additional frequent mutations. Hum Mutat 3:275–282.

Clarke LA (Updated [September 21, 2007]). Mucopolysaccharidosis Type I. In: GeneReviews at GeneTests: Medical Genetics Information Resource (database online). Copyright, University of Washington, Seattle. 1993-2007. Available at http://www.genetests.org

den Dunnen JT, Antonarakis SE. 2000. Mutation nomenclature extensions and suggestions to describe complex mutations: a discussion. Hum Mutat 15:7-12.

den Dunnen JT, Paalman MH. 2003. Standardizing mutation nomenclature: why bother? Hum Mutat 2:181-182.

Emsley, P. & Cowtan, K. 2004. Coot: model-building tools for molecular graphics. Acta Crystallogr. D 60:2126-2132

Gatti R, DiNatale P, Villani GR, Filocamo M, Muller V, Guo XH, Nelson PV, Scott HS, Hopwood JJ. 1997. Mutations among Italian mucopolysaccharidosis type I patients. J Inherit Metab Dis 20:803-806.

Gort L, Chabas A, Coll MJ. 1998. Analysis of five mutations in 20 mucopolysaccharidosis type 1 patients: high prevalence of the W402X mutation. Hum Mutat 11:332–333.

Hein LK, Hopwood JJ, Clements PR, Brooks DA. 2003. The alpha-L-iduronidase mutations R89Q and R89W result in an attenuated mucopolysaccharidosis type I clinical presentation. Biochim Biophys Acta 1639:95–103.

Krawczak M, Thomas NS, Hundrieser B, Mort M, Wittig M, Hampe J, Cooper DN. 2007. Single base-pair substitutions in exon-intron junctions of human genes: nature, distribution, and consequences for mRNA splicing. Hum Mutat 28:150-158.

Laradi S, Tukel T, Erazo M, Shabbeer J, Chkioua L, Khedhiri S, Ferchichi S, Chaabouni M, Miled A, Desnick RJ. 2005. Mucopolysaccharidosis I: alpha-L-Iduronidase mutations in three Tunisian families. J Inherit Metab Dis 28:1019-1026.

Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J. and Higgins D.G. 2007. ClustalW and ClustalX version 2. Bioinformatics 23(21): 2947-2948.

Li B, Krishnan VG, Mort ME, Xin F, Kamati KK, Cooper DN, Mooney SD, Radivojac P. 2009. Automated inference of molecular mechanisms of disease from amino acid substitutions. Bioinformatics 25:2744-2750.

Li P, Wood T, Thompson JN. 2002. Diversity of mutation and distribution of single nucleotide polymorphic alleles in the human alpha-L-iduronidase (IDUA) gene. Genet Med 4:420-426

Lowry OH, Rosebrough NJ, Farrand AL, Randall RJ. 1951. Protein measurement with the Folin-Phenol reagents. J Biol Chem 193:265-275.

Lualdi S, Pittis MG, Regis S, Parini R, Allegri A, Furlan F, Bembi B, Filocamo M. 2006. Multiple cryptic splice sites can be activated by *IDS* point mutations generating misspliced transcripts. J Mol Med 84:692-700.

Matte U, Yogalingam G, Brooks D, Leistner S, Schwartz I, Lima L, Norato DY, Brum JM, Beesley C, Winchester B, Giugliani R, Hopwood JJ. 2003. Identification and characterization of 13 new mutations in mucopolysaccharidosis type I patients. Mol Genet Metab 78: 37-43.

Mort M, Evani US, Krishnan VG, Kamati KK, Baenziger PH, Bagchi A, Peters BJ, Sathyesh R, Li B, Sun Y, Xue B, Shah NH, Kann MG, Cooper DN, Radivojac P, Mooney SD. 2010. In silico functional profiling of human disease-associated and polymorphic amino acid substitutions. Hum Mutat 31:335-346.

Nalla VJ and Rogan PK. 2005. Automated splice site analysis by information theory. Hum Mut 25:334-342.

Neufeld EF, Muenzer J. 2001 The mucopolysaccharidoses. in Scriver CR, Beaudet AL, Sly WS, Valle D (eds): The metabolic and molecular bases of inherited disease, 8th Ed McGraw-Hill, , New York pp 3421-3452.

Ng PC, Henikoff S. 2001. Predicting deleterious amino acid substitutions. Genome Res. 11:863-874.

Nissim-Rafinia M, Kerem B. 2002. Splicing regulation as a potential genetic modifier. Trends Genet 18:23–27

Pastores GM, Arn P, Beck M, Clarke JT, Guffon N, Kaplan P, Muenzer J, Norato DY, Shapiro E, Thomas J, Viskochil D, Wraith JE. 2007. The MPS I registry: design, methodology, and early findings of a global disease registry for monitoring patients with mucopolysaccharidosis type I. Mol Genet Metab 91:37-47.

Pettersen EF, Goddard, TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. 2004. UCSF Chimera - a visualization system for exploratory research and analysis. J Comput. Chem. 25:1605-1612.

Pollard KS, Hubisz MJ, Rosenbloom KR, Siepel A. 2010. Detection of nonneutral substitution rates on mammalian phylogenies. Genome Res. 20:110-121.

Rempel BP, Clarke LA, Withers SG. 2005. A homology model for human α-l-iduronidase: Insights into human disease. Mol Genet Metab 85:28-37

Scott HS, Anson DS, Orsborn AM, Nelson PV, Clements PR, Morris CP, Hopwood JJ. 1991. Human alpha-L-iduronidase: cDNA isolation and expression. Proc Natl Acad Sci USA 88:9695-9699.

Scott HS, Ashton LJ, Eyre HJ, Baker E, Brooks DA, Callen DF, Sutherland GR, Morris CP, Hopwood JJ. 1990. Chromosomal localization of the human alpha-L-iduronidase gene (IDUA) to 4p16.3. Am J Hum Genet 47:802-807.

Scott HS, Bunge S, Gal A, Clarke LA, Morris CP. Hopwood JJ. 1995. Molecular genetics of mucopolysaccharidosis type I: diagnostic, clinical, and biological implications. Hum Mutat 6:288-302.

Scott HS, Guo XH, Hopwood JJ, Morris CP. 1992. Structure and sequence of the human alpha-L-iduronidase gene. Genomics 13: 1311-1313.

Scott HS, Litjens T, Nelson PV, Brooks DA, Hopwood JJ, Morris CP. 1992. alpha-L-iduronidase mutations (Q70X, and P533R) associate with a severe Hurler phenotype. Hum Mutat 1:333–339.

Scott HS, Litjens T, Nelson PV, Thompson PR, Brooks DA, Hopwood JJ, Morris CP. 1993. Identification of mutations in the alpha -L-lduronidase gene (IDUA) that cause Hurler and Scheie syndromes. Am J Hum Genet 53:973-986.

Stenson PD, Mort M, Ball E, Howells K, Phillips A, Thomas NST, Cooper DN. 2009. The Human Gene Mutation Database: 2008 update. Genome Med 1:13.

Stirling JL, Robinson D, Fensom AH, Benson PF, Baker JE. 1978. Fluorimetric assay for prenatal detection of Hurler and Scheie homozygotes or heterozygotes. Lancet. 1:147.

Sugawara K, Saito S, Ohno K, Okuyama T, Sakuraba H. 2008. Structural study on mutant alpha-L-iduronidases: insight into mucopolysaccharidosis type I. J Hum Genet 53:467-474.

Terlato NJ, Cox GF. 2003. Can mucopolysaccharidosis type I disease severity be predicted based on a patient's genotype? A comprehensive review of the literature. Genet Med 5:286-294.

Tieu PT, Bach G, Matynia A, Hwang M, Neufeld EF. 1995. Four novel mutations underlying mild or intermediate forms of alpha-L-iduronidase deficiency (MPS IS and MPS IH/S). Hum Mutat 6:55-59.

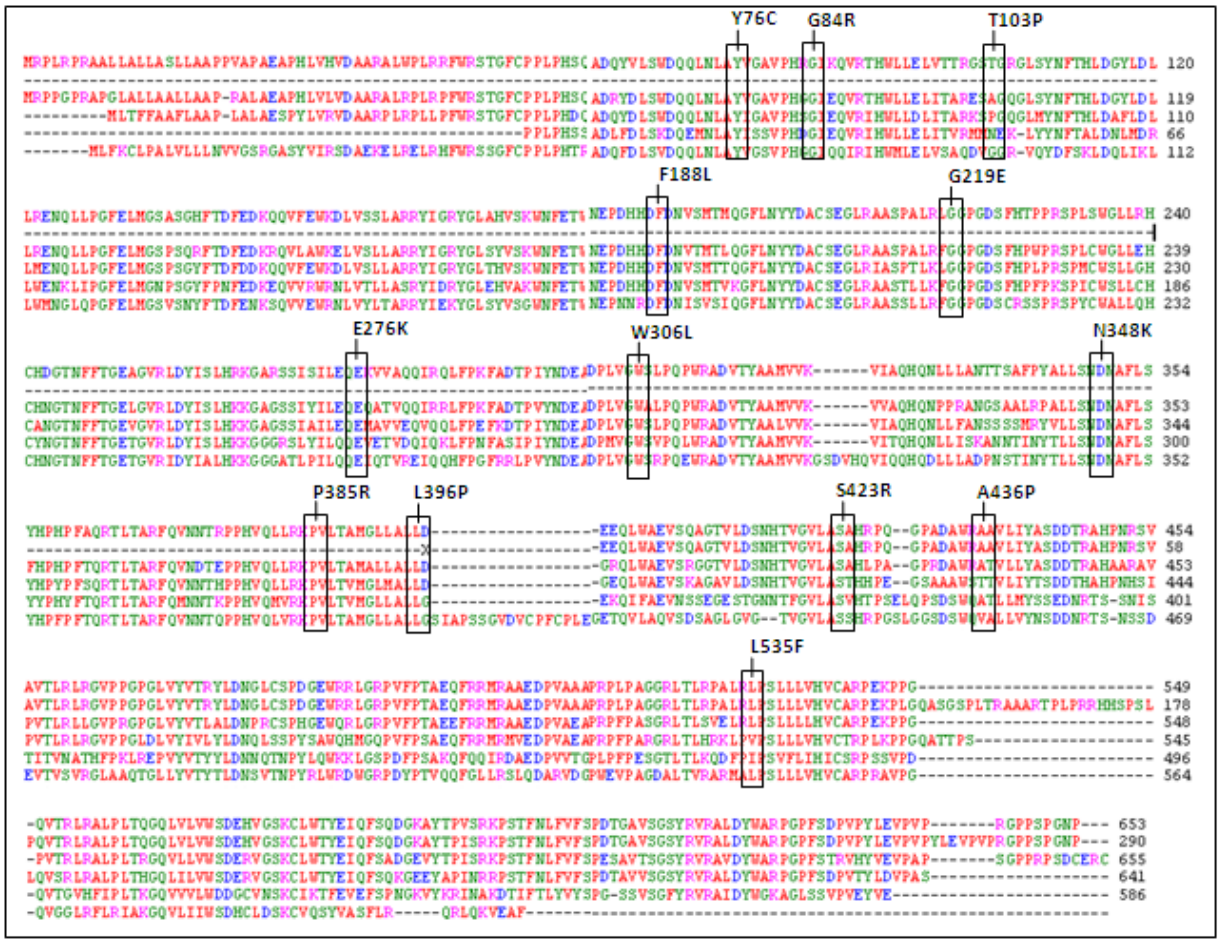
Vazna A, Beesley C, Berna L, Stolnaja L, Myskova H, Bouckova M, Vlaskova H, Poupetova H, Zeman J, Magner, Hlavata A, Winchester B, Hrebicek M, Dvorakova L. 2009. Mucopolysaccharidosis type I in 21 Czech and Slovak patients: mutation analysis suggests a functional importance of C-terminus of the IDUA protein. Am J Med Genet Part A 149A:965–974.

Venturi N, Rovelli A, Parini R, Menni F, Branbillasca F, Bertagnolio F, Uziel G, Gatti R, Filocamo M, Donati MA, Biondi A, Goldwurm S. 2002. Molecular analysis of 30 mucopolysaccharidosis type I patients: evaluation of the mutational spectrum in Italian population and identification of 13 novel mutations. Hum Mutat 20:231.

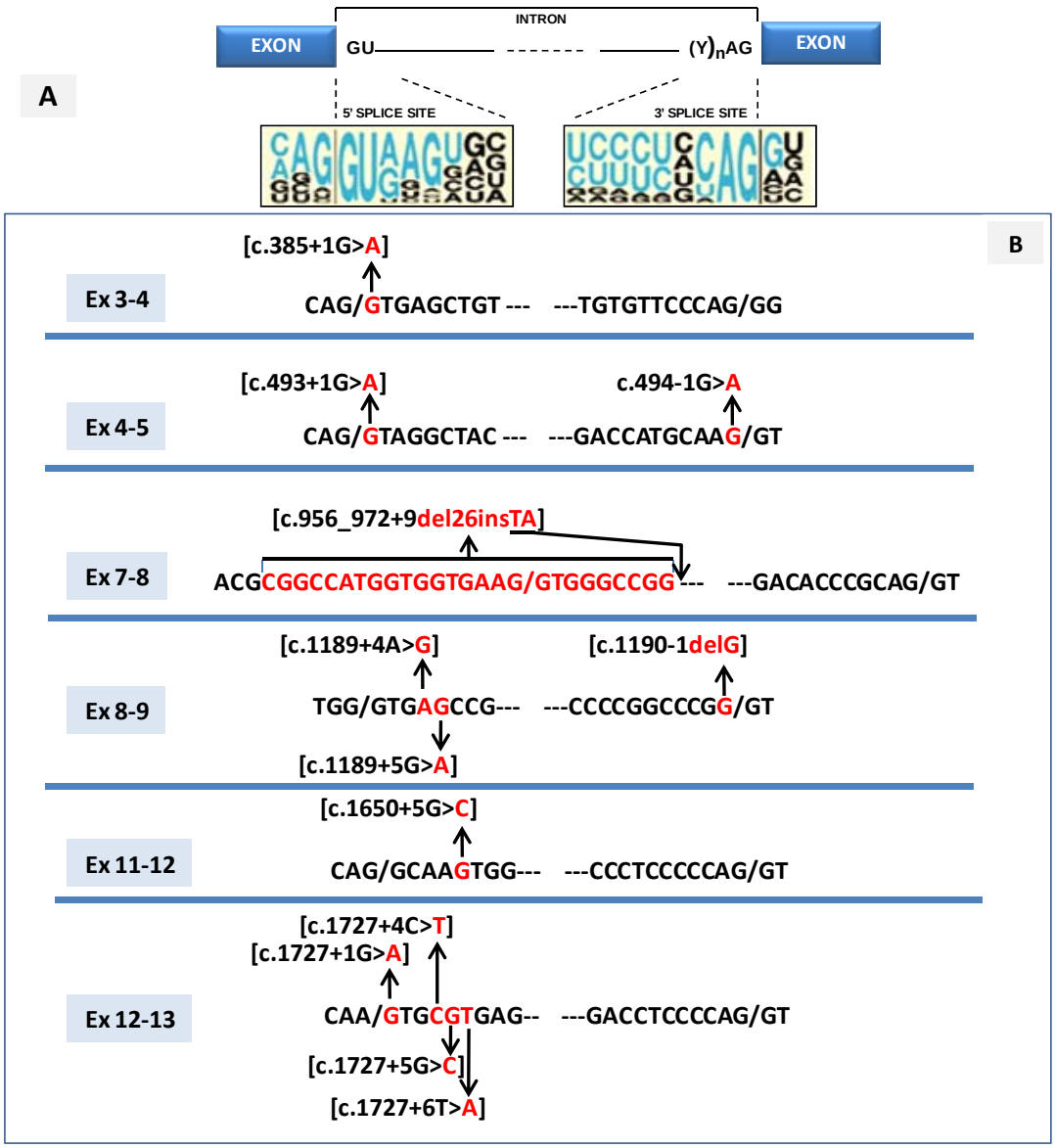
Voskoboeva E, Krasnopolskaya X, Mirenburg T, Weber B, Hopwood J. Molecular genetics of mucopolysaccharidosis type I: mutation analysis among the patients of the former Soviet Union. Mol Genet Metab 1998;65:174–180

Yogalingam G, Guo X, Muller V, Brooks DA, Clements PR, Kakkis ED, Hopwood JJ. 2004. Identification and molecular characterization of alpha-L-iduronidase mutations present in mucopolysaccharidosis type I patients undergoing enzyme replacement therapy. Hum Mutat 24: 199-207.

SUPPORTING INFORMATION



Supp. Figure S1. Evolutionary comparison of the protein sequences flanking the missense mutations identified in the human IDUA protein with their orthologous counterparts in 5 vertebrates (in *silico* data from http://www.ebi.ac.uk/clustalw/). From the top: From the top: *Homo sapiens*, *Pan troglodytes*, *Canis familiaris*, *Mus musculus*, *Gallus gallus*, *Tetraodon nigrovisidis*. Positions of mutations are marked by vertical boxes. Note that the translational initiation site substitution, predicted to produce either no protein or to move the initiation site up/downstream (p.M1?), was excluded from the analysis.



Supp. Figure S2. Graphical representation of the 12 *IDUA* splicing mutations identified in MPS I patients in the present study. A) Canonical consensus sequence of acceptor and donor splice sites (adapted from Cartegni et al, 2002); B) Positions of the *IDUA* splicing mutations in relation to the acceptor and donor splice sites. Note that only the involved sequences are reported. Affected nucleotides are given in red. The mutations are reported in square brackets.

Supp. Table S1. List of primers used for polymerase chain reaction (PCR) amplification of *IDUA* gene exons and exon-intron boundaries

Exons amplified	Forward (5'>3')*	Nucleotide position	Reverse (5'>3')*	Nucleotide position[a]	PCR product size (bp)	Annealing Temperature (°C)
Genomic DNA Amplification[a]						
1	AGGCGGAACCGGCAGTGCAG	g.5035-5054	TCCGGTCTCTGAAGCTCTCA	g.5338-5319	304	67
2	CGCTGCCAGCCATGCTGAGGCTCG	g.5760-5783	ACAAGGGGTCTTCCGAGC	g.6045-6028	286	67
3-4**	ATCGGAGTCCTGTGTGGCACCT	g.18497-18518	TGCCCACCAATGTATCTCCTGA	g.19101-19080	605	65
5	CATCACCTTGCACCCTCC	g.19391-19408	CGTCTACACCTGCCCTGG	g.19663-19646	273	65
6	GAGGAAGGCAGGAGCAGAG	g.19611-19629	CTCAGCACCACCAGGGTC	g.19969-19952	359	65
7-8**	TGCGGCTGGACTACATCTC	g.19854-19872	CTGGAGGAAGTGCGCTCC	g.20584-20567	731	63
***	CCACCTTCCTCCCGAGAC (8F)	g.20199-20216	AGTAGCAGGTTCTGATGCTGC (7R)	g.20301-20281	-	-
9-10**	TCCTTCACCAAGGGGAGG	g.20631-20648	TCCTCAGGGTTCTCCAGG	g.21331-21314	701	63
***	GGTGACCCTGCGGCTG (10F)	g.20911-20926	CTGACACTCAGGCCTCGG (9R)	g.21030-21013	-	-
11-12**	GTGTGGGTGGGAGGTGGA	g.21244-21261	GAGCCCGAATGTCCATCTCATGTC	g.21853-21830	610	67
13-14**	GGGCTTGAGGGAATGAGG	g.21911-21928	GGTGATGGGAGGGCAGCA	g.22483-22466	573	65
cDNA Amplification[b]						
3-4-5	CGGCATCTTGCTGGTCCGGACCC	c.337-359	CGCTGTCCCTCCGAGCAGGC	c.717-698	381	65
8-9	GGGCTGGTCCCTGCCACAGC	c.1000-1019	CGTTGTCCAGGTAGCGCGTGAC	c.1523-1502	524	67
12-13-14	TGCTGGTGCACGTGTGTGCGC	c.1695-1715	ATGGGAGGGCAGCATGGGCAC	c.2132-2112	474	67

IDUA gene: [a]GenBank-EMBL accession no. NG_008103.1; [b]GenBank-EMBL accession no. NM_000203.3; * the primers are located 5' (upstream) and 3' (downstream) to the intronic sequences flanking each exon; **in this case, the size also includes intronic regions; ***internal primers used for sequencing

Supp. Table S2. Characteristics of putatively neutral SNPs identified in the *IDUA* gene

Location	dbSNP no.	Site of nucleotide substitution* (amino acid change)**	Minor allele frequency§ ( )&	MutPred analysis ♦		Polymorphic background in the Patients# or control alleles
				Probability of deleterious mutation	Confident *in-silico* hypotheses	
Ex.1	rs11248061	c.24C>A (A8A)	NA (0.42)	.		15,24,26,27,32,35,37,44,60,62,64, 68,69,70,72,73,75,76,79,80,81,82, 85,86,87,89,90, 92,93,98,99,102
Ex.1	rs10902762	c.60G>A (A20A)	NA (0.44)	.		15,24,26,27,32,35,37,44,60,62,64, 68,69,70,72,73,75,76,80,82,85,86, 87,89,90,92,93,98,99,102
Ex.1	rs10794537	c.99T>G (Q33H)	NA (0.19)	0.21	None	17,20,21,23,25,30,33,39,40,41,43, 44,45,46,52,53,54,55,57,60,61,63, 64,65,76,79,91,96,100,101
Intr. 2		c.299+6C>T✦	(0.04)	.		control alleles
Intr. 2		**c.299+7G>A**	(0.01)	.		control alleles
Intr. 2		c.299+9C>T✦✦	(0.00)	.		20
Intr. 2		c.300-44C>T✦	(0.07)	.		37,83,72
Ex.3	rs3755955	c.314G>A (R105Q)	0.15 (0.18)	0.12	None	24,28,29,35,45,51,53,59,60,67,71, 75
Ex.3	rs3755954	c.352C>T (L118L)	0.16 (0.14)	.		17,20,21,23,25,27,30,33,34,35,39,4 0,41,43,44,46,49,52,54,55,57,59,60 ,61,62,63,64,65,70,71,76,79,83,84, 91,96,100,101
Ex.4		**c.408C>T (A136A)**	(0.01)	.		control alleles
Ex.5	rs6815946	c.543T>C (N181N)	0.14 (0.18)	.		24,28,29,49,51, 59,60,67,71
Intr. 5		**c.589+20dupG**	(0.00)	.		22
Intr. 5	rs6829789	c.590-45G>C	NA (0.18)	.		24,28,29,51,53,59,60,67,71
Intr. 5	rs6848974	c.590-8C>T	NA (0.18)	.		19,24,28,29,45,49,51,53,59,60,67, 71
Intr. 6		**c.792+12C>T**	(0.01)	.		control alleles
Intr. 6		**c.792+14C>T**	(0.00)	.		17
Ex. 7	rs11480689	c.891C>T (N297N)	NA (0.04)	.		21,42,43,52
Ex. 7	rs6830825	c.942G>C (A314A)	NA (0.16)	.		19,24,28,49,51,59,60,67,71
Intr. 7	rs6811373	c.972+48A>G	0.14 (0.07)	.		28,29, 49,51,53,67,71
Intr. 7	rs6831021	c.973-45G>C	NA	.		28,29,49,51,53,67
Ex. 8	rs6831280	c.1081G>A (A361T)	0.21 (0.17)	0.13	None	28,49,53,59,60,67
Ex. 8	rs6836258	c.1164G>C (T388T)	NA (0.17)	.		28,49,53,59,60,67
Ex. 8		c.1174C>T (L392L)✦	(0.01)	.		control alleles
Intr. 8		c.1190-10insC✦	(0.15)	.		19,24,28,29,45,49,51,53,67,71
Ex. 9	rs115790973	c.1230C>G (T410T)	NA (0.16)	.		19,24,28,29,45,49,51,53, 59,60,67,71
Ex. 9		**c.1345C>A (H449N)**	(0.00)	0.40	None	46
Ex. 9	rs73066479	c.1360G>A (V454I)	NA (0.17)	0.10	None	19,24,28,29,45,49,51,53,59,60,67, 71
Intr. 9		c.1403+36T>C✦	(0.17)	.		19,24,45,51,59,60,67,71
Ex. 10	rs115929690	c.1467C>T (R489R)	NA (0.14)	.		19,24,45,49,51,53, 59,60,67,71
Ex. 10		**c.1473C>A (G491G)**	(0.01)	.		control alleles
Ex. 10		**c.1515C>G (R505R)**	(0.00)	.		19
Intr. 10		c.1524+41G>T✦	(0.11)	.		45,49,53,59,60,67,71
Intr. 10	rs113289555	c.1524+53G>T	NA (0.29)	.		18,19,22,24,26,29,31,32,36,37,39,4 7,48,50,51,55,56,58,65,66,68,69,70 ,74,75,76,77,81,82,85,86,88,90,92, 95,97,98,100,102
Intr. 10		c.1525-38T>C✦	(0.09)	.		28
Intr. 12	rs2305488	c.1727+72T>G	0.11	.		19,24,29,49,53,67
Intr. 12	rs2305489	c.1727+75G>T	0.04	.		21,42,43,52
Ex. 13		c.1771G>A (A591T)✦✦✦	(0.00)	0.15	None	17,19,22,23,30,33

dbSNP, the Single Nucleotide Polymorphism database, is available at http://www.ncbi.nlm.nih.gov/projects/SNP/; *GenBank-EMBL accession no. NM_000203.3; **GenBank-EMBL accession no. NP_000194; the minor allele frequencies were § obtained from dbSNP or ( )& calculated from 100 Caucasian control alleles (novel polymorphisms, in bold) ; NA: not available; #patient number according to Table 1. ✦Scott et al., (1995); ✦✦ Venturi et al., (2002); ✦✦✦neutral change shown in vitro by Beesley et al.(2001); ♦Li et al., (2009) and Mort et al., (2010).

Supp. Table S3. MutPred analysis for *IDUA* missense mutations

Mutation	Probability of deleterious mutation	Molecular mechanism altered	
		Confident Hypotheses	Very Confident Hypotheses
p.M1?	0.99		
p.G51D	0.97	Gain of relative solvent accessibility (P = 0.0215) Gain of solvent accessibility (P=0.0306)	
p.Y76C	0.73		
p.G84R	0.93	Gain of MoRF binding (P=0.0303) Gain of solvent accessibility (P=0.0456)	
p.R89W	0.96		
p.T103P	0.36	See details in Supp. Table S4	
p.E178K	0.97	Gain of MoRF binding (P=0.0143) Gain of ubiquitination at E178 (P = 0.0258) Loss of catalytic residue at W180 (P = 0.0315)	Gain of methylation at E178 (P=0.0036)
p.F188L*	0.95		
p.S423R*	0.92	Gain of methylation at S423 (P=0.0279)	
p.G219E	0.89	Gain of solvent accessibility (P=0.0145) Gain of relative solvent accessibility (P=0.0215)	
p.G265R	0.95	Gain of methylation at G265 (P=0.0219) Loss of helix (P=0.028) Gain of loop (P=0.0312) Gain of MoRF binding (P=0.033)	
p.E276K	0.76		
p.W306L	0.82	Loss of catalytic residue at P309 (P=0.0327)	
p.A327P	0.84	Gain of relative solvent accessibility (P=0.0215) Gain of methylation at K324 (P=0.0487)	
p.N348K	0.85	Loss of sheet (P=0.0181) Gain of ubiquitination at N348 (P=0.0192) Gain of methylation at N348 (P=0.0295) Loss of stability (P=0.0381)	
p.P385R	0.82	Gain of relative solvent accessibility (P=0.0215) Loss of sheet (P=0.0315) Loss of loop (P=0.0374) Gain of helix (P=0.0425)	Gain of solvent accessibility (P=0.0055) Gain of MoRF binding (P=0.0081)
p.L396P	0.83	Loss of helix (P=0.0167) Gain of loop (P=0.0195)	Loss of stability (P=0.0047)
p.A436P	0.45	See details in Supp. Table S4	
p.L490P	0.86	Gain of disorder (P=0.0154) Loss of helix (P=0.028) Loss of stability (P=0.0441)	
p.R492P	0.83		Loss of MoRF binding (P = 0.0011)
p.P496R	0.95		Gain of MoRF binding (P = 0.0078)
p.P533R	0.94		Gain of MoRF binding (P = 0.0068)
p.L535F	0.59		

*mutations in *cis* on the same allele; MoRFs=Molecular Recognition Features

Supp. Table S4. Bioinformatic analysis of novel missense variants of uncertain pathogenicity

Nucleotide substitution	Amino acid change	Conservation base level (phyloP)	MutPred probability of deleterious mutation	SIFT Prediction (score)	PolyPhen2 (score)	Splicing disruption
c.307A>C	p.T103P	Conserved (0.4)	0.36	Tolerated (0.5)	Benign (0.0)	FALSE
c.1306G>C	p.A436P	Conserved (2.8)	0.45	Tolerated (0.2)	Possibly damaging (0.26)	FALSE

Supp. Table S5. Prediction of consequences of novel missense mutations using the 3D homology model of the IDUA protein constructed by Rempel et al. [2005]

MUTATION	PROTEIN POSITION	MODELLING ANALYSIS PREDICTION
p.Y76C	Solvent accessible	Large aromatic and polar side chain substituted by a smaller, highly reactive one.
p.G84R	Close to active site	Loss of main chain flexibility and of solvent accessible surface. Gain of charge.
p.T103P	Solvent accessible	Introduction of a new proline within a long loop (G101-N110) protruding into the solvent. Acquisition of loop rigidity. Decrease in residue polarity.
p.F188L*	Active site	F188 belongs to an amino acid cluster comprising aromatic residues (F143, W180, W306, Y343) and residues bearing indole rings (H56, H91, H185). The newly acquired 188L residue would maintain the hydrophobicity but would give rise to a loss of aromaticity. F188 is considered to be an active site residue
p.S423R*	Core	Increase in size, alteration in charge. Disruption of local protein fold due to steric hindrance
p.G219E	Core	Increase in size, alteration in charge. Destabilisation of the local protein fold due to steric hindrance
p.E276K	Solvent accessible	Alteration in charge (negative to positive). Change of the protein surface electrostatic potential
p.W306L	Core	W306 belongs to an amino acid cluster comprising aromatic residues (F143, W180, W306, Y343) and residues bearing indole rings (H56, H91, H185). The newly acquired 306L residue would maintain the hydrophobicity but would give rise to a loss of aromaticity.
p.N348K	Solvent accessible	Alteration in surface local charge.
p.P385R	Core	Loss of local rigidity. Destabilisation of the local protein fold.
p.L396P	Solvent accessible	Loss of long and branched hydrophobic side chain (leucine). Loss of local flexibility in the edge of a turn. Loss of local protein fold.
p.A436P	Solvent accessible	Gain of main chain rigidity (436P) which would disrupt the super-secondary structure of the protein owing to its inability to form the main chain H-bonds that are required for β-sheet formation.

*mutations occur in *cis* on the same allele

Supp Table S6. Genotype-phenotype correlation: comparison with existing data from the literature

GENOTYPES*		NO. OF PATIENTS ANALYZED AND (PHENOTYPE)		REFERENCES
		THIS STUDY	OTHER STUDIES	
INFORMATIVE	[p.W402X]+[p.W402X]	12 (H)	46 (H)	Bunge et al. (1994); Gort et al. (1998); Hein et al. (2003); Li et al, (2002); Matte et al. (2003); Scott et al. (1992); Vazna et al. (2009)
	[p.Q70X]+[p.Q70X]	9 (H)	26 (H)	Bunge et al. (1994); Gort et al. (1998); Hein et al. (2003); Scott et al. (1992); Vazna et al. (2009); Venturi et al. (2002); Voskoboeva et al. (1998)
	[p.W402X]+[p.Q70X]	4 (H)	19 (H)	Bunge et al. (1994); Li et al. (2002); Scott et al. (1992)
	[p.W402X]+[c.1650+5G>A]	1 (H)	2 (H)	Vazna et al. (2009); Venturi et al. (2002);
	[p.W402X]+[p.A327P]	1 (H)	2 (H)	Bunge et al. (1994); Vazna et al. (2009)
	[p.A327P]+[p.A327P]	1 (H)	1 (H-H/S)	Gatti et al. (1997)
	[p.G51D]+[p.G51D]	2 (H)	1 (H)	Gatti et al. (1997)
	[p.Q70X]+[c.1333_1335del3]	1 (S)	1 (S)	Bunge et al. (1994)
	[p.L490P)+[p.L490P]	1 (S)	3 (H/S)	Beesley et al. (2001); Tieu et al. (1995)
PARTIALLY INFORMATIVE	[p.P533R]+[p.P533R]	2 (H/S)	8 (§)	Alif et al. (1999); Gatti et al. (1997); Laradi et al. (2005); Matte et al. (2003); Scott et al. (1992)
	[p.P533R]+[p.P533R]	1 (H)	11 (§)	Alif et al. (1999); Hein et al. (2003)
	[p.G51D]+[p.P533R]	1 (H/S)	2 (§)	Gatti et al. (1997); Venturi et al. (2002)
	[c.46_57del12]+[c.46_57del12]	2 (H/S)	2 (H)	Bunge et al. (1994); Venturi et al. (2002)

H=Hurler; S=Scheie; *Note that only the comparable genotypes with existing data are reported in this table. For the other genotypes, refer to Table 1; §various phenotypes