J Hum Genet (2002) 47:14–19

ORIGINAL ARTICLE

Aritoshi Iida · Susumu Saito · Akihiro Sekine
Chihiro Mishima · Yuri Kitamura · Kimie Kondo
Satoko Harigae · Saori Osawa · Yusuke Nakamura

Catalog of 77 single-nucleotide polymorphisms (SNPs) in the carbohydrate sulfotransferase 1 (*CHST1*) and carbohydrate sulfotransferase 3 (*CHST3*) genes

Received: September 18, 2001 / Accepted: October 28, 2001

Abstract Individual phenotypes with respect to drug response or toxicity often result from genetic variations that alter drug metabolism. We have been focusing on genomic loci that encode various enzymes and transporters involved in the metabolism of drugs, and have described more than 1200 single-nucleotide polymorphisms (SNPs) and other variations. Regarding the carbohydrate sulfotransferase (*CHST*) gene family, we have already constructed high-density SNP maps of three genomic segments that included *CHST2*, *CHST4*, and *CHST5*, providing a total of 28 SNPs for those loci. In the present study, we screened DNA from 48 healthy Japanese volunteers for SNPs at the *CHST1* and *CHST3* gene loci, by means of direct sequencing combined with a polymerase chain reaction method for amplifying genomic DNA, and characterized 77 SNPs and four insertion–deletion polymorphisms. The collection of human variations presented here adds to the archive of tools now available for investigating complex genetic diseases, population migration patterns, and a variety of pharmacogenetic possibilities.

Key words Single-nucleotide polymorphisms (SNPs) · Insertion–deletion polymorphisms · High-density SNP map · Carbohydrate sulfotransferase 1 (*CHST1*) · Carbohydrate sulfotransferase 3 (*CHST3*) · Japanese population

A. Iida · S. Saito · A. Sekine · C. Mishima · Y. Kitamura · K. Kondo · S. Harigae · S. Osawa · Y. Nakamura
Laboratory for Genotyping, RIKEN SNP Research Center, Tokyo, Japan

Y. Nakamura (✉)
Laboratory of Molecular Medicine, Human Genome Center, Institute of Medical Science, The University of Tokyo, 4-6-1 Shirokanedai, Minato-ku, Tokyo 108-8639, Japan
Tel. +81-3-5449-5372; Fax +81-3-5449-5433
e-mail: yusuke@ims.u-tokyo.ac.jp

Introduction

Sulfotransferases constitute a large family of enzymes that catalyze transfer of the sulfonate group from 3′ phosphoadenosine 5′ phosphosulfate to a suitable hydroxyl group or other moiety on an acceptor substrate to form either a sulfate ester or a sulfamate (Falany and Wilborn 1994). In general, two classes of sulfotransferases are recognized, cytosolic sulfotransferases and carbohydrate sulfotransferases (CHSTs), which are distinct both structurally and with respect to biological function. Sulfotransferases in the cytosol play important roles in the biotransformation, detoxification, or elimination of many compounds, including hormones, neurotransmitters, bile acids, and catabolic end products of drugs. In contrast, CHSTs are resident transmembrane enzymes of the Golgi network that recognize glycans attached to lipids and proteins passing through the secretory pathway (see reviews by Bowman and Bertozzi 1999; Habuchi 2000; Hemmerich and Rosen 2000). CHSTs are known to play a fundamental role in extracellular signaling and adhesion by generating unique ligands from a carbohydrate scaffold. In this sense, these enzymes might be similar in function to the tyrosine sulfotransferase, also resident in Golgi compartments, that modulates the activity of both secreted and membrane-associated proteins by sulfating tyrosine.

Human CHST1 was originally isolated as the homolog of chick chondroitin 6-sulfotransferase (C6ST) that catalyzes sulfation of chondroitin, keratan sulfate, and sialyl N-acetyllactosamine oligosaccharides (Fukuta et al. 1997; Mazany et al. 1998; Li and Tedder 1999). CHST1 shows 34% homology with chick C6ST on the amino acid level and may function to reconstitute the high endothelial cell ligand for L-selectin (Bistrup et al. 1999). Human CHST3 encodes a type II transmembrane protein that shares 74% and 36% identity, respectively, with chick C6ST and human CHST1 (Fukuta et al. 1998; Tsutsumi et al. 1998). Recombinant human CHST3 displays C6ST activity with a marked specificity for a GlcA–GalNAc sequence. In addition, CHST3 catalyzes sulfation of position 6 of the Gal

residue of keratan sulfate, as do chick C6ST and human CHST1.

In this study, we provide high-resolution maps of the human *CHST1* and *CHST3* gene loci, in which we detected a total of 77 single-nucleotide polymorphisms (SNPs) and four insertion–deletion polymorphisms among 96 chromosomes from a representative Japanese population sample.

Subjects and methods

Blood samples were obtained with written informed consent from 48 healthy Japanese volunteers for this study, which was approved by the ethical committee of the RIKEN SNP Research Center. On the basis of genomic sequences from the GenBank database (accession numbers *CHST1* NT-008982.3, *CHST3* AC073370.3, and AC022392.4), we designed primers to amplify each gene in its entirety, as well as 2 kb upstream of the first exon and downstream of the last exon. However, most regions corresponding to repetitive sequences were excluded. Protocols for polymerase chain reaction experiments and DNA sequencing are available from our worldwide website (http://snp.ims.u-tokyo.ac.jp/). The gene symbols *CHST1* and *CHST3* conform to official nomenclatures set out by HUGO/GDB Nomenclature Committees (http://www.gene.ucl.ac.uk/nomenclature/) and by Li and Tedder (1999).

Results and discussion

A total of 77 SNPs were identified within the *CHST1* and *CHST3* gene loci among 48 Japanese volunteers. The exonic organization of each gene and locations of identified SNPs are illustrated schematically in Fig. 1; detailed information is given in Table 1. After comparing our data with SNPs deposited in the dbSNP database in the National Center for

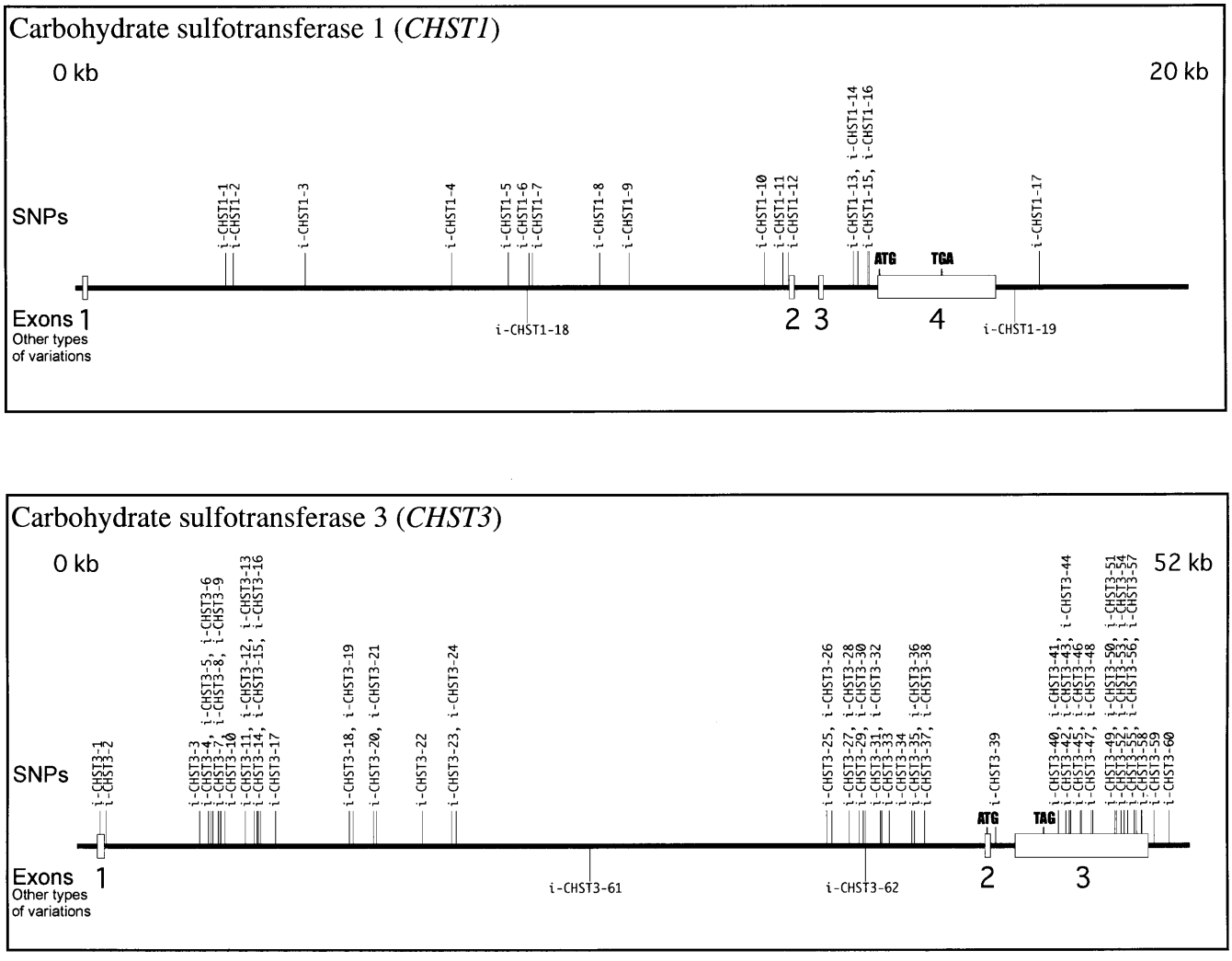

Fig. 1. Genomic organization and localization of single-nucleotide polymorphisms (*SNPs*) in the carbohydrate sulfotransferase 1 (*CHST1*) and carbohydrate sulfotransferase 3 (*CHST3*) genes. Exons and introns are represented by *vertical rectangles* and *horizontal lines*, respectively. The SNPs are indicated *above* each gene (designations correspond to the ones in the left-most column on Table 1). Locations of other types of variations are indicated *below* each gene

Table 1. Characterization of variations in *CHST1* and *CHST3* gene loci

ID	Region	Position[a]	Flanking sequence	Variation (5' to 3')[c]	Flanking sequence[b]	Repetitive sequence[d]	Identity to dbSNP
Carbohydrate sulfotransferase 1 (*CHST1*)							
i-CHST1-1	Intron 1	2475	taaatggagaaaataacacc	G/A	acctgatagcattgttgtga	+	dbSNP ID:rs895734
i-CHST1-2	Intron 1	2612	aaactccccaagcatgctca	C/A	ctagatccttaccctaggtc		dbSNP ID:rs895733
i-CHST1-3	Intron 1	3900	gccctgcccccactcccaga	C/G	ttgcggccctccagcccctt		
i-CHST1-4	Intron 1	6520	cctcccccagaggagctggg	C/T	acactggggccttgtgttgt		
i-CHST1-5	Intron 1	7534	attgtgtgttggcatactgc	T/C	cacatggaaggatgctctag		dbSNP ID:rs895729
i-CHST1-6	Intron 1	7911	ttttccttaggaagaaaaac	G/A	ccttgctgttttatgcattt		dbSNP ID:rs2012019
i-CHST1-7	Intron 1	7963	aaaacattcatgggggatta	G/C	tgctggctacgtcagagtca		
i-CHST1-8	Intron 1	9173	gcgctgccacagatcaggcc	G/A	aggtgggggacagaaatgcc		
i-CHST1-9	Intron 1	9701	cccagaattctgaatacagc	A/G	gcgatgacgggactacgagg		
i-CHST1-10	Intron 1	12132	aacagatccacaggaccaga	C/A	agcaaaggggaggaacatgc		
i-CHST1-11	Intron 1	12465	atgcagggaaggggcttggc	G/A	caaaactgtcaactgagata		
i-CHST1-12	intron 1	12561	atgctccctggtccactttc	G/A	ctttgagtttcaggtagctg		
i-CHST1-13	Intron 3	529	ccatggtctgcaggggtcct	T/G	catgctcaggggattggggt		
i-CHST1-14	Intron 3	617	agaggacagaggaaagagga	C/A	cacctggagaactgggcgcc		
i-CHST1-15	Intron 3	796	aagaggcttccgcagctgtc	C/T	gcaggttaaatcctggggtg		
i-CHST1-16	Intron 3	818	caggttaaatcctggggtgc	A/G	aggaatgtttgttcagctcc		
i-CHST1-17	3' Flanking region	762	ataactggtacaggtttact	G/C	gtgtctacactggcagagaa		
i-CHST1-18	Intron 1	7874	gttttccccttgccttgcct	T/del	cattttcatcacctcatttt		
i-CHST1-19	3' Flanking region	335–349	cacactgccacacctggcta	$(T)_{12-15}$	ggattttagtagagacgggg	+	
Carbohydrate sulfotransferase 3 (*CHST3*)							
i-CHST3-1	5' Untranslated region	–294	tccagcgtgccgaccggccc	C/G	gcagcgcctccatccctccg		
i-CHST3-2	Intron 1	96	gcgtccaggcgcgcgcgcca	G/A	actttggagggagaaggggg		
i-CHST3-3	Intron 1	4467	agagaagaatggggcagagc	C/G	ggagcagccaggggaggtga		
i-CHST3-4	Intron 1	4853	ggatgagcactgcccagctg	A/G	tccctgcccaccttccacag		
i-CHST3-5	Intron 1	4965	tccactgcagaggggacaca	G/C	tgaccaggacggaagttggg		
i-CHST3-6	Intron 1	5046	gggctgtccatctttgtacc	C/T	ctggttccatcccagtgcct	+	
i-CHST3-7	Intron 1	5300	ccttttcttctctaaggcct	A/G	aagagatgacagaatgctgc		
i-CHST3-8	Intron 1	5354	agcgcgtggactccacagcg	G/A	ggtgtggggtggcccctggc		
i-CHST3-9	Intron 1	5428	gacacgcttcagccctctgt	C/G	tctattgccccaaatctggc		

Table 1. Continued

ID	Region	Position[a]	Flanking sequence	Variation (5' to 3')[c]	Flanking sequence[b]	Repetitive sequence[d]	Identity to dbSNP
i-CHST3-10	Intron 1	5621	ctgtggcttccctgggccct	A/G	ggaaatttatcactgaggtt		dbSNP ID:rs874692
i-CHST3-11	Intron 1	6555	gagtggggcactgctggaag	G/C	ttctggttcctgctttgttc		
i-CHST3-12	Intron 1	6990	aaacacactgggccaccccc	G/A	tccccgcactgtgactacac		
i-CHST3-13	Intron 1	7133	ctgagggcctgtcctgcagg	T/G	ttgatgtgtctgaagaggcc		
i-CHST3-14	Intron 1	7161	gtctgaagaggccccgagaa	T/C	agaaatctagaacctgccag		
i-CHST3-15	Intron 1	7199	cagtcacgaagcagtgtcac	C/T	caccagaggatgaagaactg		
i-CHST3-16	Intron 1	7316	cttgcatctggtgtaggtgc	C/T	tgggggtagcgtgcccagga		
i-CHST3-17	Intron 1	7967	gacaggaaccccaccccgag	T/G	gatgtctggccctgtgacct		
i-CHST3-18	Intron 1	11412	gcttgcacttctgattcatt	C/T	tgcagtcactggctctttgt		
i-CHST3-19	Intron 1	11591	ccctggaagggcctcactgc	G/A	gtgactcattacccagcatg		
i-CHST3-20	Intron 1	12541	acccacacagcatgaatggg	G/C	ccagccccagcctgcccgct		
i-CHST3-21	Intron 1	12672	gtagccacagctggggctgt	G/C	gggtcagggcatggcaaggg		
i-CHST3-22	Intron 1	14809	ggatgtgtagggtttgggct	C/T	ggccttaagggatgggtgga		
i-CHST3-23	Intron 1	16161	gatgctggtcaggcattgtc	G/A	ttgggatctttaacaccacc		
i-CHST3-24	Intron 1	16385	tatttagcatgtgggtttca	A/C	ctttctgtttttcaaaggg		
i-CHST3-25	Intron 1	33638	gacttgggccacgtccttgg	G/C	catgaatcttggtctatgtc		
i-CHST3-26	Intron 1	33878	agcaagaaagtgtgctcccc	C/T	acagccccactcaggcataa		dbSNP ID:rs730334
i-CHST3-27	Intron 1	34690	agcacacatggagctttccc	G/A	cagtgggtttcagcgctccc		dbSNP ID:rs1880685
i-CHST3-28	Intron 1	35145	agggaagccgaagcctcact	T/C	gctggggcttgcctggcctc		
i-CHST3-29	Intron 1	35340	tgtgaagttttgcccacagt	T/C	ggtggccatggttcgcaccg		
i-CHST3-30	Intron 1	35436	gccactcatgtatggagcaa	T/C	tgccttttttcttcctctt		
i-CHST3-31	Intron 1	36150	ccatagaagaggctgggcct	G/T	aggaagccagggaagcagga		
i-CHST3-32	Intron 1	36194	ggtgtggggaggccagcagg	G/A	gtgtgggcctcagcggggag		
i-CHST3-33	Intron 1	36561	ctctggtgtttgctgtcaat	A/G	tgcagagtgctggacaaaac		dbSNP ID:rs751450
i-CHST3-34	Intron 1	37602	ctggaacagcaacttaaaaa	A/T	agaaatagtccctggaaggg		
i-CHST3-35	Intron 1	37725	gggtagccagggcagctccc	C/T	gacccgcaC/Gctgcctttt		
i-CHST3-36	Intron 1	37734	gcagctcccC/Tgacccgca	C/G	ctgccttttcacccctctcc		
i-CHST3-37	Intron 1	38208	gccattctagatgcgagtcc	C/T	gactttggggT/Cgcttgca		
i-CHST3-38	Intron 1	38219	cgagtccC/Tgactttgggg	T/C	gcttgcattctgggaaggga		dbSNP ID:rs2219837
i-CHST3-39	Intron 2	255	ctacagctgtgaaaggttag	A/G	caagatacttaacatttctg	+	
i-CHST3-40	3' Untranslated region	2202	acacctcagaggagcctgtg	C/A	ttaacatttgtaggattatt		
i-CHST3-41	3' Untranslated region	2569	aggcctcatctggggtaggg	C/G	caagaggaaagtacagagtg		
i-CHST3-42	3' Untranslated region	2717	ctggaattcctccttagggc	C/T	ctgggaagagtattgcttaa		

Table 1. Continued

ID	Region	Position[a]	Flanking sequence	Variation (5' to 3')[c]	Flanking sequence[b]	Repetitive sequence[d]	Identity to dbSNP
i-CHST3-43	3' Untranslated region	2753	cttaacgcaggatgtgctgg	G/A	tgttttgtttcgggctttta		
i-CHST3-44	3' Untranslated region	2800	gcttggtgtctttcttgttt	C/T	atggctgtgtttttgctttt		
i-CHST3-45	3' Untranslated region	3283	ccgagggctgcccagctctg	C/T	ttctggtttcctggacaatt		
i-CHST3-46	3' Untranslated region	3327	ctgtcagatacggcccattg	T/C	aaacccagagggctgcattt		
i-CHST3-47	3' Untranslated region	3787	gttccccatgtggaggtcgg	A/G	ggggctgggactggggaggg		
i-CHST3-48	3' Untranslated region	3860	ggccctgctaatgtggacag	T/C	agactttatccctccttctt		
i-CHST3-49	3' Untranslated region	4915	ccagatgtgcatagaagcca	G/A	tctctgtcacatacaccgca		
i-CHST3-50	3' Untranslated region	4993	taaagcaaatttaggctttt	G/A	tccttctgcaatacatgcac		
i-CHST3-51	3' Untranslated region	5223	ggaaggagcttcagcaggag	G/A	tccttcccagaaggttgatt		dbSNP ID:rs1871450
i-CHST3-52	3' Untranslated region	5370	tcatacctgtaatcccagca	G/T	ttggggaggccaaggtggga	+	dbSNP ID:rs1871451
i-CHST3-53	3' Untranslated region	5545	ccattcccaaagtcagaaag	T/C	gaagccagatctcaagggct		dbSNP ID:rs731027
i-CHST3-54	3' Untranslated region	5859	caaaagcacaaagcagaatt	G/C	gcaacttcacT/Atgtctca		dbSNP ID:rs730722
i-CHST3-55	3' Untranslated region	5870	cagaattG/Cgcaacttcac	T/A	tgtctcaagagctccaagat		dbSNP ID:rs1871452
i-CHST3-56	3' Untranslated region	5971	ttccaaggctacagacatgg	C/T	gccatcctcacaggcctagc		dbSNP ID:rs730720
i-CHST3-57	3' Untranslated region	6208	atttcatgtctgcatggtac	G/A	agacacccctcacG/Agca		
i-CHST3-58	3' Untranslated region	6223	tacG/Aagacacccctcac	G/A	gcatacactgccatggtatg		dbSNP ID:rs12418
i-CHST3-59	3' Flanking region	281	agacaggagtgttgggccag	C/T	ggtcagggggcctggggatg	+	
i-CHST3-60	3' Flanking region	997	acctcttaaagtatttgagc	C/T	ggtgcctgtcatcccaacct		
i-CHST3-61	Intron 1	22595	cgggagcaggaaaaaaaaaa	A/del	gaataagaagaaaagaggct		
i-CHST3-62	Intron 1	35423–35424	gctcatgctcacagccactc	AT/del	gtatggagcaaT/Ctgccttt		

del, deletion polymorphism; SNP, single-nucleotide polymorphism

[a]Nucleotide numbering is according to the mutation nomenclature (den Dunnen and Antonarakis 2000)

[b]5′ to 3′ flanking sequence to each variation are denoted by small letters

[c]Variation is shown by capital letter

[d]+ indicates the variation is located with a repetitive sequence

Biotechnology Information, we considered 61 of the 77 SNPs (79%) to be novel as of the middle of July 2001. The extent of each genomic sequence we screened for SNPs in this study was 15.3 kb for the *CHST1* gene locus and 26.6 kb for *CHST3* gene locus. Distributions of SNPs at the *CHST1* and *CHST3* gene loci averaged 1 per 900 bp and 443 bp, respectively, but distribution was uneven along both transcription units, as Fig. 1 makes clear. We have found no SNPs within coding elements of *CHST1*, *CHST3*, or other *CHSTs*, suggesting that members of this gene family are under strong selective pressure to conserve sequences in those regions. Other genes that encode drug-metabolizing enzyme genes, e.g., microsomal glutathione S-transferases, are similarly conserved (Iida et al. 2001a–c). In addition, we also identified four insertion–deletion polymorphisms within the *CHST1* and *CHST3* gene loci (Fig. 1 and Table 1).

Altogether, we have collected a total of 105 SNPs among five *CHST* genes in the Japanese population. We hope that the virtual experiments made possible by our SNP catalog will accelerate certain aspects of genomic and pharmacogenetic research.

References

Bistrup A, Bhakta S, Lee JK, Belov YY, Gunn MD, Zuo FR, Huang CC, Kannagi R, Rosen SD, Hemmerich S (1999) Sulfotransferases of two specificities function in the reconstitution of high endothelial cell ligands for L-selectin. J Cell Biol 145:899–910

Bowman KG, Bertozzi CR (1999) Carbohydrate sulfotransferases: mediators of extracellular communication. Chem Biol 6:R9–R22

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

Falany CN, Wilborn (1994) Biochemistry of cytosolic sulfotransferases involved in bioactivation. In: Anders MW, Dekant W (eds) Advances in pharmacology, conjugation-dependent carcinogenicity and toxicity of foreign compounds, vol. 27. Academic, San Diego, pp 301–329

Fukuta M, Inazawa J, Torii T, Tsuzuki K, Shimada E, Habuchi O (1997) Molecular cloning and characterization of human keratan sulfate Gal-6-sulfotransferase. J Biol Chem 272:32321–32328

Fukuta M, Kobayashi Y, Uchimura K, Kimata K, Habuchi O (1998) Molecular cloning and expression of human chondroitin 6-sulfotransferase. Biochim Biophys Acta 1399:57–61

Habuchi O (2000) Diversity and functions of glycosaminoglycan sulfotransferases. Biochim Biophys Acta 1474:115–127

Hemmerich S, Rosen SD (2000) Carbohydrate sulfotransferases in lymphocyte homing. Glycobiology 10:849–856

Iida A, Sekine A, Saito S, Kitamura Y, Kitamoto T, Osawa S, Mishima C, Nakamura Y (2001a) Catalog of 320 single nucleotide polymorphisms (SNPs) in 20 oxidoreductase and sulfotransferae genes. J Hum Genet 46:225–240

Iida A, Saito S, Sekine A, Kitamoto T, Kitamura Y, Mishima C, Osawa S, Kondo K, Harigae S, Nakamura Y (2001b) Catalog of 434 single nucleotide polymophisms (SNPs) in the alcohol dehydrogenase genes, glutathione S-transferase genes, and NADH ubiquinone oxidoreductase genes. J Hum Genet 46:385–407

Iida A, Saito S, Sekine A, Harigae S, Osawa S, Mishima C, Kondo K, Kitamura Y, Nakamura Y (2001c) Catalog of 46 single-nucleotide polymorphisms (SNPs) in the microsomal glutathione S-transferase 1 (*MGST1*) gene. J Hum Genet (in press)

Li X, Tedder TF (1999) CHST1 and CHST2 sulfotransferases expressed by human vascular endothelial cells: cDNA cloning, expression, and chromosomal localization. Genomics 55:345–347

Mazany KD, Peng T, Watson CE, Tabas I, Williams KJ (1998) Human chondroitin 6-sulfotransferase: cloning, gene structure, and chromosomal localization. Biochim Biophys Acta 1407:92–97

Tsutsumi K, Shimakawa H, Kitagawa H, Sugahara K (1998) Functional expression and genomic structure of human chondroitin 6-sulfotransferase. FEBS Lett 441:235–241