Genetic variability in proteoglycan biosynthetic genes reveals new facets of heparan sulfates diversity. A systematic review and analysis

Mohand Ouidir Ouidja, Denis S.F. Biard[a,b], Sandrine Chantepie[a], Xavier Laffray[a], Gael Le Douaron[a], Minh-Bao Huynh[a], Nicolas Rebergue[a], Auriane Maïza[a], Karla Rubio[a,c,1], Oscar González-Velasco[d], Guillermo Barreto[a,c,1], Javier De Las Rivas[d], and Dulce Papy-Garcia[a,*]

[a] Univ Paris Est Creteil, Glycobiology, Cell Growth and Tissue Repair Research Unit (Gly-CRRET), Creteil, France. [b] - CEA, Institut de Biologie François Jacob (IBFJ), SEPIA, Université Paris-Saclay, Fontenay-aux-Roses, France. [c] Lung Cancer Epigenetic, Max-Planck-Institute for Heart and Lung Research, Bad Nauheim, Germany. [d] Bioinformatics and Functional Genomics Group, Cancer Research Center (CiC-IMBCC, CSIC/USAL/IBSAL), University of Salamanca (USAL), Salamanca, Spain.
[1] Present address: Université de Lorraine, CNRS, Laboratoire IMoPA, UMR 7365, F-54000 Nancy, France.

*** Correspondence to Dulce Papy-Garcia:** Gly-CRRET, University Paris Est Creteil, 61 Av General de Gaulle, 94010 Créteil, France. papy@u-pec.fr

Keywords: Systematic review and analysis, heparan sulfate proteoglycans, core protein, linker tetrasaccharide, heparan sulfate, biosynthetic genes, clinical variants, expression, sulfotransferases, stimuli.

Abstract

Proteoglycans are complex macromolecules formed of glycosaminoglycan chains covalently linked to core proteins through a linker tetrasaccharide common to heparan sulfate proteoglycans (HSPG) and chondroitin sulfate proteoglycans (CSPG). Biosynthesis of a single proteoglycan requires the expression of dozens of genes, which together create the large structural and functional diversity reflected by the numerous diseases or syndromes associated to their genetic variability. Among proteoglycans, HSPG are the most structurally and functionally complex. To decrease this complexity, we retrieved and linked information on pathogenic variants, polymorphism, expression, and literature databases for 50 genes involved in the biosynthesis of HSPG core proteins, heparan sulfate (HS) chains, and their linker tetrasaccharide. This resulted in a new gene organization and biosynthetic pathway representation in which the phenotypic continuum of disorders as linkeropathies and other pathologies could be predictable. Moreover, ubiquitous *NDST1*, *GLCE*, *HS2ST1*, and *HS6ST1* appeared to generate ubiquitous heparan sulfate (HS) sequences essential for normal development and homeostasis, whereas the tissue restricted *NDST2-4*, *HS6ST2-3*, and *HS3ST1-6* appeared to generate specialized HS sequences mainly involved in responsiveness to stimuli. Supported by data on genetic polymorphism and clinical variants, we afford a new vision of HSPG involvement in homeostasis, disease, vulnerability to disease, and behavioral disorders.

1. Introduction

Proteoglycans are a family of macromolecules involved in virtually all aspects of cell biology and physiology[1]. These complex molecules consist in a core protein carrying one or several glycosaminoglycan (GAG) chains, either chondroitin sulfates (CS), heparan sulfates (HS), keratan sulfates (KS), or combinations of them. In HS proteoglycans (HSPG) and CS proteoglycans (CSPG), the glycanic chain is connected to their core proteins trough a common GAG-core protein linker tetrasaccharide (GAG-CP) (Fig. 1)[1-3]. Both the identity of the core protein and the structure of the glycanic chain confer to proteoglycans a great functional diversity, as reflected by the multiplicity of diseases or syndromes resulting from defects on genes coding for their biosynthetic machinery. Among proteoglycans, HS proteoglycans (HSPG) display the greater structural and functional diversity. This is principally due to the diversity and differential expression of the numerous genes required for their synthesis (Fig. 1). Interestingly, whereas various of these genes are essential for normal development and homeostasis, several others are not, suggesting that some, but not all, HSPG play essential biological roles depending on the core protein and associated HS chains generated in each tissue and cell by its HSPG biosynthetic machinery.

he biosynthesis of HSPG starts in the endoplasmic reticulum with the production of the core protein coded by a single gene (for instance *HSPG2*, *AGR*, *COLXVIIA*, *GPCs*, or *SDCs* that code for full time HSPG core proteins). At the Golgi, the core protein is glycosylated to first form the GAG-CP linker tetrasaccharide. This requires the expression of several genes coding for different enzymes including glycosyltransferases (*XYLT1*, *XYLT2*, *B4GALT7*, *B3GALT6*, *B3GAT3*)[4,5], kinases (*FAM20B*)[6], phosphatases (*PXYLP1*)[7], and nucleosidases (*CANT1*)[8], as well as nucleoside transporters (*SCL26A2, SLC35A2*, and *SLC35A3*)[9] and ion transporters (*SLC10A7*)[10] (Fig. 1).

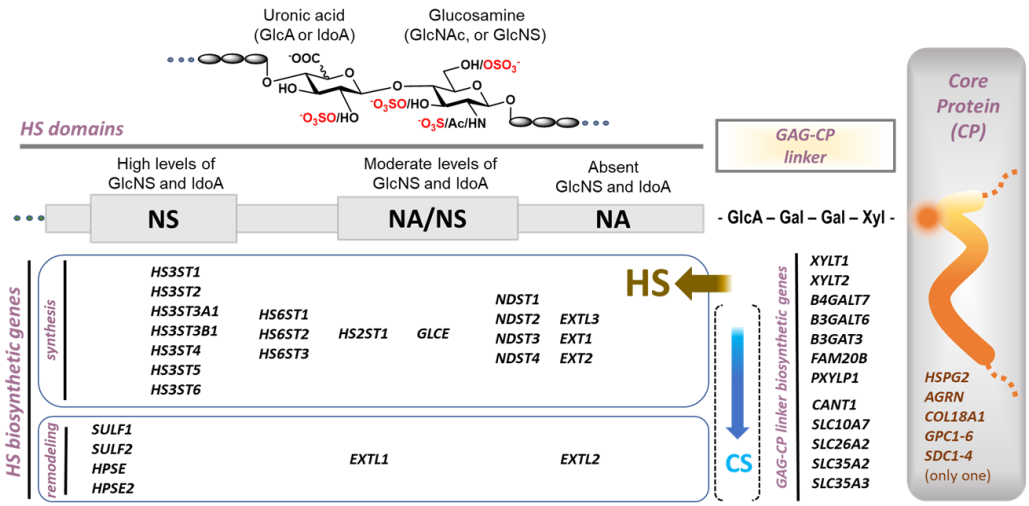


Fig. 1. Schematic representation of a HSPG and their biosynthetic genes. HSPG are made of a core protein (CP) carrying a glycosaminoglycan-CP linker tetrasaccharide (GAG-CP linker) in which either heparan sulfate (HS) or chondroitin sulfate (CS) chains are elongated. The HS chain is formed by constitutive disaccharides composed of a uronic acid (GlcA or IdoA) and a glucosamine (GlcNAc, or GlcNS) carrying different sulfation and IdoA contents. NS domains carry high *N*-sulfation and IdoA levels; NA/NS domains carry intermediary *N*-sulfation, *N*-acetylation, and IdoA levels; NA domains are not sulfated and carry not IdoA. The GAG-CP linker biosynthetic genes are common to HSPG and CSPG biosynthetic machineries.

After formation of the GAG-CP linker, HS chain initiation starts through action of EXTL3 (coded by *EXTL3*), which adds a *N*-acetyl glucosamine (GlcNAc) to the non-reducing end of the GAG-CP linker. HS chain elongation continue by sequential addition of GlcA and GlcNAc catalyzed by EXT1/EXT2 (coded by *EXT1* and *EXT2*). During chain elongation, epimerization of glucuronic acid (GlcA) into iduronic acid (IdoA) is then carried by C5 epimerase (coded by *GLCE*)[11] and sulfation occurs at different position of the uronic acid and glucosamine residues. Sulfation of IdoA is assured by 2-*O*-sulfotransferase (*HS2ST1*), whereas sulfation of GlcNAc is assured by *N*-deacetyl-*N*-sulfotransferases (coded by *NDST1, NDST2, NDST3,* or *NDST4),* 6-*O*-sulfotransferases (coded by *HS6ST1, HS6ST2,* or *HS6ST3*), and 3-*O*-sulfotransferases (coded by *HS3ST1, HS3ST2, HS3ST3A1, HS3ST3B1, HS3ST4, HS3ST5,* or *HS3ST6*)[3,12]. During biosynthesis, the HS chain elongation can be stopped by the unproductive glycosyltransferases EXTL1 or EXTL2 (coded by *EXTL2* or *EXTL1*)[13], whereas after secretion of the HSPG at the extracellular space 6-*O*-sulfatases SULF1 and SULF2 (coded by *SULF1* and *SULF2*) and heparanase (coded by *HPSE*) can modify the HS chain structures[14-16]. Globally, this process leads to formation of different HS structural domains[3,17]. 'NA' domains carry unsulfated *N*-acetylated sequences, 'NA/NS' domains carry HS sequences with intermediary levels of sulfation and IdoA, and 'NS' domains carry high levels of sulfation and IdoA[3,17] (Fig. 1).

Regardless of their structural diversity, HS are frequently referred as a single ubiquitous molecule, possibly because their structural characterization is highly complex and yet beyond reach. Here, as an approach to reduce HSPG complexity, we propose a new strategy based in a PRISMA[18] systematic review and analysis of well stablished pathogenic variants and emerging genetic wide association studies (GWAS) linking HSPG biosynthetic machinery gene variability to disease, susceptibility to disease, altered response to instigating stimuli, or altered behaviors. To confirm the documented clinical variants and traits in human, we compared them to the phenotypes of corresponding gene null mouse, when available. Expression patterns in different organs or tissues were retrieved from gene expression databases and information on structural features, including substrate specificities for enzymes, or protein-protein interactions for core proteins and other biosynthetic proteins, were reviewed following the PRISMA guidelines[18]. Association of this body of information resulted in a new HSPG vision that gives a better understanding of the involvement of each biosynthetic gene in the generation of HS sequences carrying structures that can play either ubiquitous or specialized roles in the maintain of tissue homeostasis, disease, behavior disorders, proper tissue response to instigating stimuli, and tissue specific vulnerability to disease.

2. Methods

Guidelines for Systematic Reviews and Meta-Analyses (PRISMA 2020)[18] were followed to setup a step-wise method (Fig. 2) for selecting and filtering research items on 50 genes directly involved in the biosynthesis of HSPG (Table S1). For each HSPG biosynthetic gene, research items were

organized to screen clinical variants, single nucleotide polymorphism (SNP), expression levels in different organs of the human body, and literature records databases on substrate selectivities and protein-protein interactions.

2.1. Search in clinical variants and SNP databases

For each human gene (Table S1), clinical variants and SNP public databases accessible without restriction were consulted by using gene names previously confirmed by MeSH (Medical Subject Headings: https://www.ncbi.nlm.nih.gov/mesh/). Data on the impact of gene variability in health status were obtained from three main databases (Fig. 2A): 1) the ClinVar archive (database ClinVar: https://www.ncbi.nlm.nih.gov/clinvar/intro/), which gives information on correlations between genetic variations and overt phenotypes or health status with history of interpretations, 2) the Orphanet archive (Database Orphanet: https://www.orpha.net/), which considers clinical presentation based on published scientific articles and expert reviews, and 3) the dbSNP archive (database: https://www.ncbi.nlm.nih.gov/snp/), which is a free public archive for genetic variation within and across different species. For clinical variants, only monogenic variants (indel, deletions, duplications, insertions, and single nucleotide) were considered. Clinical significance was further confirmed in web platforms including Online Mendelian Inheritance in Man (OMIM) (https://www.omim.org/) and the Database of Genomic Variants Archive (DGVa database: https://www.ebi.ac.uk/dgva/). Information on clinical variants overt phenotypes in human were confirmed by literature research on case reports and on reports on the corresponding gene null mice (Methods section 2.3 and Fig. 2C-2D).

2.2. Search in expression databases

Information on the basal expression of the HSPG biosynthetic genes (Table S1) in different tissues or organs of the healthy adult human body was retrieved from open access transcriptomic data bases including: 1) the RNA-seq 32Uhlen project (database 32Uhlen: http://www.proteinatlas.org/humanproteome), which analyzed 32 different tissues from 122 human individuals, 2) the RNA-Seq CAGE (Cap Analysis of Gene Expression) in the RIKEN FANTOM5 project (database FANTOM5: http://fantom.gsc.riken.jp/data/), which analyzed several healthy adult human tissues, and 3) the ENCODE strand-specific RNA-seq of 13 human tissues from Michael Snyder's lab (Database ENCODE: https://www.encodeproject.org/). All these RNA-seq databases are considered in the expression atlas (http://www.ebi.ac.uk/gxa)[19], as they use standardized methods that allow the determination of baseline gene expression across controlled biological replicates checked for high quality (good quality raw files and good quality reference genome build are available). The original raw and processed data files can be found in the ArrayExpress platform (https://www.ebi.ac.uk/arrayexpress/), in which the raw single-channel microarray intensities were normalized either using an oligo package from Bioconductor (Affymetrix data) or by quantile normalization by the limma package (Agilent data). The data quality was assessed using the arrayQualityMetrics package in R. Quality-trimmed reads were mapped to the latest version of the

reference genome from Ensembl using TopHat2. For each HSPG biosynthetic gene, the number of mapped reads per gene were summed (raw counts) using htseq-count and represented as average transcripts per million (TPM). When transcripts were detected in a tissue, regardless of their level of expression, a gene was considered to be basally expressed in that tissue. To illustrate gene expression, the best TPM value from the different databases was included in the analysis.

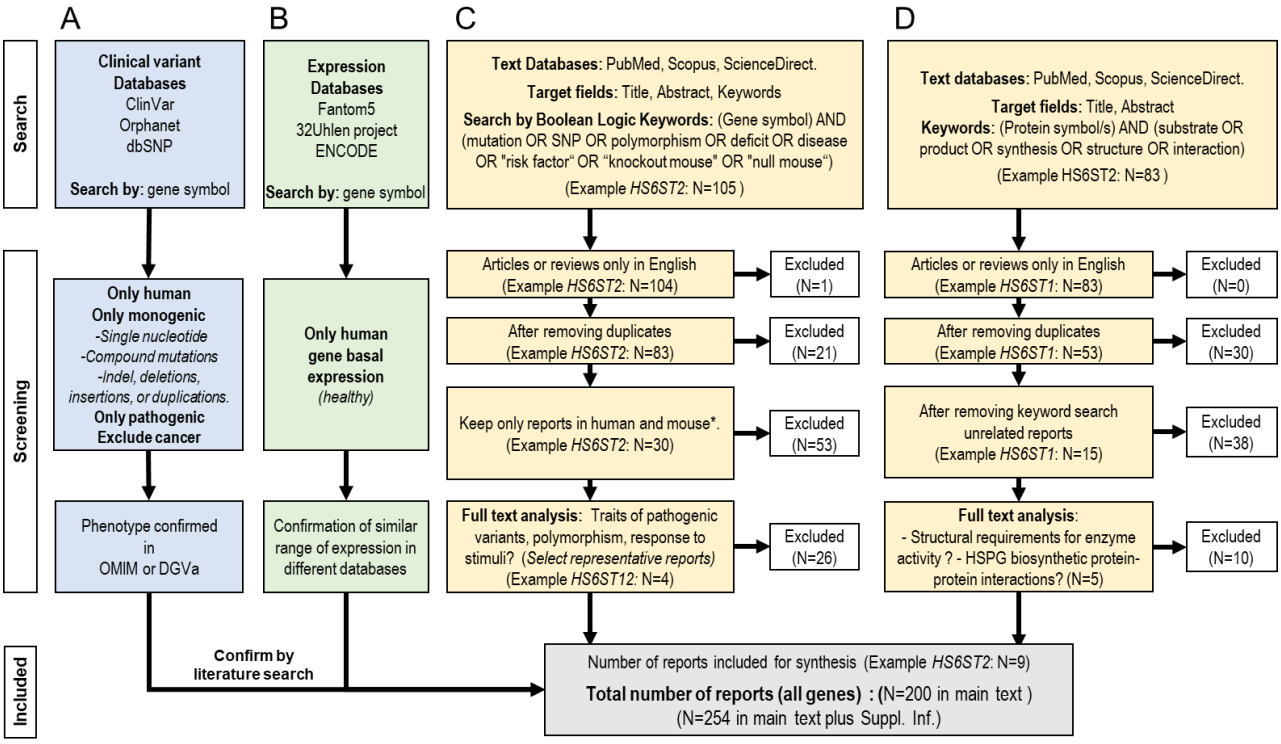


Fig. 2. Schematic representation for collecting articles process following PRISMA[18] recommendations. **A)** Clinical variants databases were used to document the effect of HSPG biosynthetic genes variabilities in health status. **B)** Expression databases were used to document organs or tissues in which transcript for each of the genes can be detected under basal homeostasis (healthy status). **C)** Bibliographic databases were used to document reports on the effect of genetic variability in health status and to expression in organs and tissues. **D)** Bibliographic databases were also used to document proteins interactions between biosynthetic machinery proteins and substrates selectivities for biosynthetic enzymes.

2.3. Search in literature databases

PRISMA guidelines[18] were used to carry two campaigns of literature research for each gene. The first campaign was performed to confirm information obtained from clinical variants databases (section 2.1 and Fig. 2A) by search of clinical variants case reports and reports on the corresponding gene null mouse (Fig. 2C). The second campaign was performed to retrieve information on each gene product functionally, including structural features for substrate specificity and selectivity for enzymes, and on protein-protein interactions for core proteins and other biosynthetic proteins (Fig. 2D). Three literature search database thesauruses (PubMed, Scopus, and Science Direct) were employed for a systematic search facilitating an extensive inclusion (Fig. 2C and 2D). For each search engine, 2 subsets of pool keywords and Boolean logic were conducted for each of the 45 genes. The first keywords subset ((‘gene symbol’) AND (mutation OR SNP OR polymorphism OR

deficit OR disease OR "risk factor" OR “knockout mouse” OR "null mouse")) was used to review the effect of genetic variability in health status (Fig. 2C). The second keywords and Boolean logic subset was used to review the substrates specificities for enzymes or protein-protein interactions ((‘protein name’ OR ‘protein symbol’) AND (substrate OR “substrate structure” OR “substrate specificity” OR interaction OR “interaction specificity”)) (Fig. 2D). References were screened in accordance with title and abstract. Only research articles and literature reviews were considered. Only reports in English language were included. We did not specify a time frame in our literature search. An email alarm system, on each database, was created for notifications on new publications related to the search terms. All reports sorted in the three databases were exported to a common file for each gene in a bibliographic management system (EndNote) and an automatic deduplication tool was applied to discard duplicated reports. Then, reports in species other than human and mouse and reports in cancer cell lines were systematically excluded, unless any example of SNP could be found in human. A pilot user study was then conducted for examining full text reports and representative references were selected and pooled in a single bibliographic management file using EndNote. The reference list of the retrieved reports was also explored, articles non available were requested to authors via email, otherwise references could not be included. Finally, our systematic review included a synthesis of 254 peer-reviewed reports that were checked and review as representative of the full retrieved data (200 reports cited in main text and additional 54 only cited in Suppl. Inf.).

3. Results

3.1. Essential and non-essential HSPG biosynthetic genes

A first clustering of the HSPG biosynthetic genes was performed on the bases of their requirement for normal embryonic development and homeostasis, as reported by clinical variants databases. Information on gene essentialness were confirmed by literature records in cases reports and/or from reports on the corresponding gene null mouse (see extended review on genes in Suppl. Inf.). Genes considered essential for normal embryonic development and homeostasis were those for which genetic variability in human is causative of diseases or syndromes, and which corresponding null mouse dies or shows overt phenotypes, consistent with that observed in clinical variants. Genes for which genetic variability has not been reported as causative of human diseases or syndromes, and/or for which the corresponding null mouse is viable, fertile, lives normally, and shows non-overt phenotypes in the absence of stimuli, were considered non-essential for normal embryonic development and homeostasis (Suppl. Inf.). However, most of the ‘Non-essential’ genes show some genetic variability associations with altered responsiveness to stimuli, altered behaviors, or are involved in vulnerability to altered conditions health conditions or behaviors (Suppl. Inf.). Based in these criteria, 2 clusters of genes were formed, the first cluster is referred as ‘Essential’, and the second as ‘Non-essential’ (Fig. 3). To avoid loss of clustering, HSPG biosynthetic genes involved in tumor development and growth was not considered, as most HSPG are altered in cancer[16,20,21].

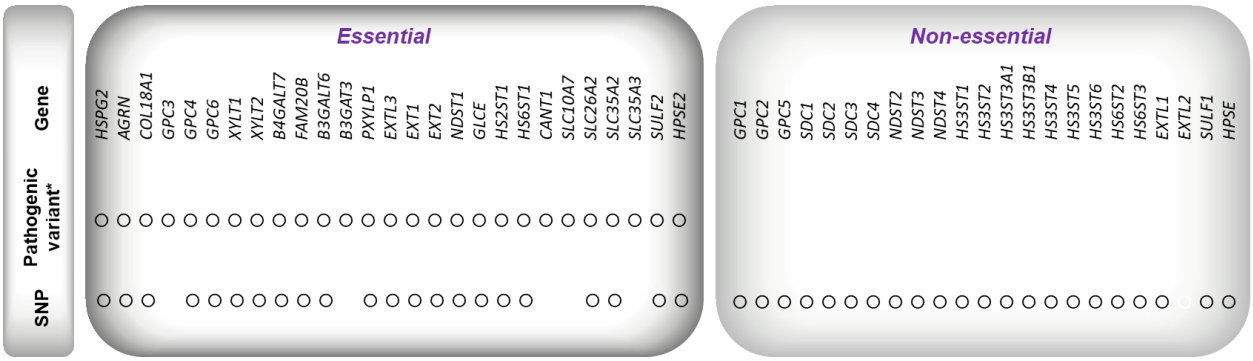


Fig. 3. First clustering of HSPG genes depending on their requirement for normal development and homeostasis. 'Essential' genes are those for which genetic variability in human is causative of diseases or syndromes and/or the corresponding null mouse dies or shows overt phenotypes consistent with that observed in clinical variants. *They can show SNP. 'Non-essential' genes are those for which genetic variability has not been reported as causative of human diseases or syndromes and/or for which the corresponding null mouse is viable, fertile, lives normally, and shows non-overt phenotypes in the absence of stimuli, although many of them show genetic variability associations with altered responsiveness to stimuli, altered behaviors, or are involved in vulnerability to altered conditions health conditions or behaviors.

3.2. Ubiquitous and tissue restricted HSPG biosynthetic machinery genes

To get additional information on Essential and Non-essential HSPG biosynthetic genes in relation with their basal expression in human, genes were re-clustered depending on their expression in healthy organs or tissues, as documented in RNAseq databases (see Methods). Genes which transcripts were detectable in all analyzed organs were considered as 'widely expressed', whereas genes which expression was only detected in some tissues/organs were considered as 'restrictedly expressed' (Fig. 4 and Supp. Inf.). This transcription-based re-clustering resulted in the HSPG biosynthetic genes organization in four groups of genes (Fig. 4): Group I, Group III, Group III, and Group IV, each of which is described below (simplified clustering is presented in Fig. 5). A review on each gene is detailed in Suppl. Inf.

3.2.1. Group I - Essential HSPG biosynthetic genes that are ubiquitously expressed

The Group I of HSPG biosynthetic genes clustered widely expressed genes that are essential for normal development and homeostasis (Fig. 4, simplified Fig. 5, and Suppl. Inf). This included genes coding for pericellular core proteins *HSPG2*[22-25], *AGRN*[26-29], and *COL18A1*)[30-32], for some membrane-associated core proteins (*GPC3*[33-35], *GPC4*[36,37], *GPC6*[38-40]), and for all the genes coding for the GAG-CP linker biosynthetic proteins, including glycosyl transferases (*XYLT1*[41,42], *XYLT2*[43], *B4GALT7*[44,45], *B3GALT6*[46,47], and *B3GAT3*[48]), kinase FAM20B (*FAM20B*)[49], phosphatase (*PXYLP1*)[7], ion or nucleotide transporters (*SLC10A7*[10], *SLC26A2*[50], *SCL35A2*[51], and *SCL35A3*[9,52]), or other Golgian enzymes (*CANT1*)[8]. Importantly, Group I also clustered genes coding for the glycosyltransferases that assure HS chain initiation and elongation (*EXTL3*[53,54], *EXT1*[55], *EXT2*[56]), for epimerase (*GLCE*)[57], and for three sulfotransferases (*NDST1, HS2ST1,* and *HS6ST1*)[58-61]. Moreover, *SULF2,* which codes for one of the extracellular 6-*O*-sulfatase that generates biologically

active HS sequences resulting from the cleavage of 6-*O*-sulfated groups[14,62], also clustered in Group I. Altogether, the enzymes coded by the ubiquitous HS chain biosynthetic genes can then produce ubiquitous HS sequences (uHS), which fine structures might depend on the expression levels of each biosynthetic gene in each particular cell or tissue[63-65], as well as in their transcriptional regulation and translational control[66]. Together, this body of data suggests that the Group I HSPG biosynthetic machinery can, by itself, produce HSPG carrying biologically relevant HS sequences or chains that are essential for normal development and homeostasis.

3.2.2. Group II - Essential tissue restricted HSPG biosynthetic genes

Group II clustered HSPG biosynthetic genes that, although essential, show tissue restricted expression. Based in current available data and literature records, only *HPSE2* clustered in this group (Fig. 4, Fig. 5, and Suppl. Inf). *HPSE2* codes for the enzymatically inactive heparanase 2. Although it does not affect HS structures, heparanase 2 can bind to these polysaccharides with higher affinity than the enzymatically active heparanase (Group III), avoiding its enzymatic action and maintaining long HS chains in their biological environment[67,68]. *HPSE2* is here reported as the only gene in Group II, but the evolving literature and clinical reports might further enrich the number of genes in this group.

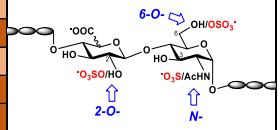
3.2.3. Group III - Non-essential HSPG biosynthetic genes that are ubiquitously expressed

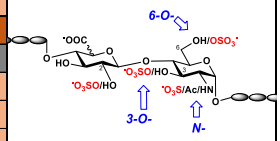
Group III clustered HSPG biosynthetic genes that although widely expressed are non-essential for normal development and homeostasis (Fig. 4, Fig. 5, and Suppl. Inf.). However, genes in Group III appeared to be required for appropriate tissue responsiveness to stimuli and, for several of them, SNP is associated with increased or decreased vulnerability to disease, or to altered conditions or behaviors (Supp. Inf). Interestingly, all syndecans core proteins coding genes (*SDC1*[69-71], *SDC2*[72], *SDC3*[73-76], *and SDC4*[77-79]) and one glypican (*GPC1*[80,81]) clustered in this group (Fig. 4). Concerning the HS chain biosynthetic genes, none of these clustered in this group. Instead, Group III clustered genes involved in the control of HS levels and quality, including *EXTL2*[82,83], *HPSE*[84-86] and *SULF1*[70,87-89] (Fig. 5). Globally, data from GWAS, knockout mice, and literature research indicate that genes in Group III are not essential for normal development and homeostasis but are required for proper response to stimuli and or involved in vulnerability or resistance to disease and altered behaviors.

3.2.4. Group IV - Non-essential HSPG biosynthetic genes that are restrictedly expressed

Group IV of HSPG biosynthetic machinery genes clustered all genes that are expressed in a tissue restricted manner (Fig. 4) and which genetic variability is not causative of lethality or disease, although SNP might be associated to resistance or vulnerability to pathologic conditions, to altered response to stimuli, and to altered behaviors (Fig. 4, Fig. 5, and Suppl. Inf). Genes in Group IV included those coding for two core proteins (*GPC2*[90-92] and *GPC5*[93-95]), three NDSTs (*NDST2*[96-98],

GROUP			CHROMOSOME	GENE \ TISSUE	SMOOTH MUSCLE	SKIN	ADIPOSE TISSUE	URINARY BLADER	OVARY	PROSTATE	TESTIS	ENDOMETRIUM	PLACENTA	LUNG	KIDNEY	PANCREAS	LIVER	HEART	STOMACH	SMALL INTESTINE	DUODENUM	COLON	LYMPH NODE	SPLEEN	ADRENAL GLANT	TYROID GLAND	SALIVARY GLAND	SKELETAL MUSCLE	BRAIN	COMMENT
ESSENTIAL	I	uCP	1	*HSPG2*	116	31	140	62	20	51	16	72	63	55	19	6	5	31	27	18	17	29	19	43	28	59	18	14	6	Essential ubiquitous core proteins (uCP)
			1	*AGRN*	7	15	14	18	4	23	21	15	16	57	48	8	4	4	17	16	12	18	5	15	3	30	19	2	10	
			21	*COL18A1*	25	36	35	38	99	56	26	49	37	26	80	23	128	14	15	29	34	24	18	66	41	22	12	4	10	
			X	*GPC3*	12	9	150	23	32	16	18	28	1114	220	79	10	2	12	9	7	4	7	6	7	49	15	7	0.6	18	
			X	*GPC4*	64	3	16	20	45	23	29	47	100	66	43	11	3	5	27	22	26	33	10	19	26	23	17	14	6	
			13	*GPC6*	22	1	20	40	27	23	5	9	2	13	18	1	28	5	13	10	6	20	1	7	8	6	3	0.5	16	
		uTetr	16	*XYLT1*	19	13	10	11	26	8	8	14	20	6	2	1	0.5	3	7	12	9	11	16	15	34	9	2	0.6	9	Essential ubiquitous GAG-CP linker (uTetr)
			17	*XYLT2*	21	22	9	15	24	26	32	21	23	16	11	13	3	13	33	10	9	17	19	21	12	20	10	3	13	
			5	*B4GALT7*	10	22	10	11	16	16	17	13	15	14	14	28	4	13	9	9	10	13	19	18	12	17	13	3	19	
			1	*FAM20B*	69	38	48	53	69	51	28	82	53	68	14	32	19	29	35	30	24	36	36	40	65	74	17	11	43	
			1	*B3GALT6*	16	17	16	15	19	17	10	23	21	22	11	4	6	8	13	9	8	10	17	24	12	24	6	3	9	
			11	*B3GAT3*	20	38	17	27	38	44	37	42	48	34	14	21	8	9	27	27	26	16	65	52	30	39	18	2	20	
			3	*PXYLP1*	9	6	4	10	13	13	19	14	17	6	7	2	3	5	4	6	6	5	11	8	5	9	4	0.8	12	
		uHS	8	*EXTL3*	29	17	11	17	16	13	23	22	24	14	7	3	3	21	7	6	6	7	9	16	24	18	6	6	24	Essential ubiquitous HS (uHS) elongation & sulfation
			8	*EXT1*	32	31	16	35	23	22	12	28	1	34	10	4	15	10	15	41	45	28	13	24	17	28	5	1	3	
			11	*EXT2*	100	41	43	58	71	56	50	97	122	52	29	15	17	19	33	48	34	41	31	44	36	73	24	4	6	
			5	*NDST1*	58	80	55	42	59	36	56	36	36	119	27	16	32	28	16	31	22	27	19	124	53	86	41	9	21	
			15	*GLCE*	19	24	17	21	38	16	13	23	32	24	22	4	17	6	16	48	29	27	10	12	29	15	12	2	9	
			1	*HS2ST1*	22	21	15	28	35	25	25	32	18	25	17	5	14	7	14	20	16	19	31	31	53	35	7	2	18	
			2	*HS6ST1*	19	35	25	62	46	35	58	25	36	33	49	13	11	7	13	25	16	22	20	69	54	25	13	11	31	
		CQ	20	*SULF2*	199	24	34	75	214	33	37	202	60	26	33	7	28	15	20	21	30	46	50	30	13	30	7	4	8	*Likely NA/NS domains*
	II	CQ	10	*HPSE2*	4	0.6	1	28	nd	19	5	15	3	5	2	nd	nd	nd	0.7	9	7	6	nd	nd	nd	nd	0.7	nd	3	Essential tissue restricted QC
NON-ESSENTIAL	III	uCP	2	*GPC1*	22	97	17	21	23	69	64	21	50	18	8	5	8	33	8	5	6	8	2	10	35	9	11	34	15	Non-essentail ubiquitous core proteins (uCP)
			2	*SDC1*	5	529	12	220	4	67	10	5	75	186	114	21	249	1	143	144	129	109	7	17	43	73	25	nd	11	
			8	*SDC2*	116	24	106	97	268	97	37	152	169	91	67	13	227	40	36	36	23	44	24	38	58	666	12	6	42	
			1	*SDC3*	186	22	49	71	33	61	15	72	40	57	10	7	15	19	19	28	20	24	46	263	190	31	13	3	87	
			20	*SDC4*	196	378	48	178	77	145	182	200	108	469	284	172	636	344	331	90	83	200	30	62	140	135	126	21	137	
		QC	1	*EXTL2*	26	11	12	22	38	29	22	38	47	17	11	8	17	6	8	7	7	10	8	17	43	22	6	3	19	Non essentail ubiquitous QC (uCQ)
			8	*SULF1*	119	2	22	89	35	79	25	32	8	18	12	1	2	9	4	20	13	21	16	3	5	30	2	1	2	
			4	*HPSE*	4	15	4	8	9	2	6	12	5	11	2	0.8	2	0.5	4	6	3	11	11	16	8	3	4	nd	7	
	IV	rCP	7	*GPC2*	1	9	0.6	0.7	0.9	1	19	2	0.5	0.6	nd	nd	nd	0.9	nd	0.9	0.7	4	3	2	0.5	nd	0.9	nd	13	Non-essential tissue restricted core proteins (rCP)
			13	*GPC5*	nd	nd	nd	nd	0.9	nd	16	nd	0.6	3	11	nd	0.9	0.5	nd	nd	nd	0.6	nd	nd	4	1	nd	nd	5	
		rHS	10	*NDST2*	5	6	3	3	4	4	4	5	nd	6	1	1	0.9	0.8	2	2	2	2	7	8	3	4	2	nd	nd	Non essential tissue restricted sulfation (rHS)
			X	*HS6ST2*	6	0.5	0.9	6	45	6	8	14	35	1	21	2	nd	nd	nd	nd	nd	2	nd	nd	6	0.6	nd	3	1	
			13	*HS6ST3*	7	nd	0.9	18	nd	1	1	9	nd	nd	2	1	nd	nd	2	0.5	0.9	nd	nd	nd	nd	6	nd	nd	40	
			4	*HS3ST1*	3	nd	1	12	61	3	3	5	2	3	2	nd	nd	nd	4	2	2	3	3	4	3	1	1	nd	2	
			16	*HS3ST2*	2	7	4	2	9	2	3	0.8	2	16	nd	nd	1	1	1	nd	nd	2	5	0.8	4	0.8	nd	nd	20	
			17	*HS3ST3A1*	0.5	nd	nd	0.8	7	6	0.7	0.9	6	0.7	1	nd	0.9	nd	0.8	0.9	0.7	0.6	4	13	3	nd	nd	nd	nd	
			17	*HS3ST3B1*	1	3	2	6	12	6	2	3	2	4	4	0.9	36	1	3	3	3	1	21	11	3	2	7	nd	1	
			4	*NDST3*	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	0,5	nd	1	nd	0.5	nd	nd	nd	nd	2	1	nd	nd	nd	2	
			4	*NDST4*	0.9	nd	nd	1	5	0.6	0.9	0.6	nd	nd	nd	nd	nd	nd	0.7	nd	nd	nd	nd	3	nd	nd	nd	nd	1	
			16	*HS3ST4*	nd	nd	nd	0.6	nd	3	1	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	5	nd	nd	nd	nd	11	*Likely NS domains*
			6	*HS3ST5*	0.5	0.7	nd	0.7	nd	nd	0.5	nd	nd	0.8	nd	nd	nd	nd	nd	nd	nd	nd	nd	nd	0.9	nd	nd	2	4	
			16	*HS3ST6*	nd	108	nd	5	nd	0.5	0.6	nd	3	0.7	nd	nd	0.8	nd	nd	nd	nd	0.6	nd	nd	nd	nd	nd	nd	nd	
		QC	1	*EXTL1*	4	2	nd	0.6	3	0.7	1	4	2	nd	nd	nd	nd	3	nd	nd	nd	nd	0.7	4	3	1	nd	6	7	Non-essential tissue restricted QC
REFERENCE GENES			12	*GAPDH*	1854	1009	558	1741	686	974	768	739	1763	577	831	177	600	1989	599	1205	1201	1174	1248	775	2140	1105	532	6797	894	
			7	*ACTB*	7777	1375	1836	4533	2306	2972	1113	5617	3463	4197	799	188	1257	426	2088	2521	1763	2267	7360	5164	1371	2631	846	176	1545	





>100 TPM
10-100 TPM
1-10 TPM
<1 TPM
Below cutoff

Fig. 4. The human HSPG biosynthetic machinery organized by gene essentiality and expression levels. HSPG biosynthetic genes were first separated in 'Essential' or 'Non-essential' for normal development and homeostasis depending on their requirement for normal human development and homeostasis and/or the phenotype of the corresponding gene null mouse (Suppl. Inf). Then, genes were re-clustered depending on their expression (expressed in transcript par million TPM) in a variety of human organs or tissues, as documented from RNAseq databases (FANTOM5: http://fantom.gsc.riken.jp/data/; ENCODE: https://www.encodeproject.org/; and 32Uhlen: http://www.proteinatlas.org/humanproteome/tissue+specific). Genes were organized in 4 mains groups depending on their basal expression levels in the adult human. CQ stands for 'Quality Control' that can arrest HS elongation during synthesis or modify HS structure after secretion to the extracellular space.

NDST3[99-101], *and NDST4*[102,103]), two HS6STs (*HS6ST2*[104,105], *HS6ST3*[106,107]), all HS3STs (*HS3ST1*[108-110], *HS3ST2*[111,112], *HS3ST3A1*[113-115], *HS3ST3B1*[115], *HS3ST4*[116], *HS3ST5*[117], and *HS3ST6*[118,119]), and one HS levels regulating genes (EXTL1[120]). Based in current available data and literature records, it appeared that, at least in the absence of instigating stimuli, these genes are non-causative of disease and are not essential for normal development and homeostasis (Supp. Inf). Accordingly, their corresponding gene null mice, when available, develop normally, are fertile, and show non deleterious phenotypes (Supp. Inf). Depending on the gene, mice can show inappropriate response to instigating stimuli during processes including wound healing, angiogenesis, inflammation, immunity, infection, and response to neurological stimuli leading to altered adaptative behaviors (as alertness and satiety during feeding, and response to stress) (Suppl. Inf.). Moreover, SNP in Group IV genes, along or in combination with SNP in other loci, have been associated to vulnerability or resistance to diseases, to infection, and to impaired angiogenesis, inflammation, immunological response, altered behaviors, etc. (Fig. 3 and Suppl. Inf.). It is to note that all the HS sulfotransferases coding genes that did not cluster in Group I clustered in this group, including *NDST2*, *HS6ST2*, *HS6ST3*, and all *HS3STs*. Thus, based in information from knockout mice, GWAS data, and literature research, it appeared that the tissue restricted rHS sequences are those required for appropriate adaptive behaviors and responsiveness to stimuli in tissues in which each gene is specifically expressed.

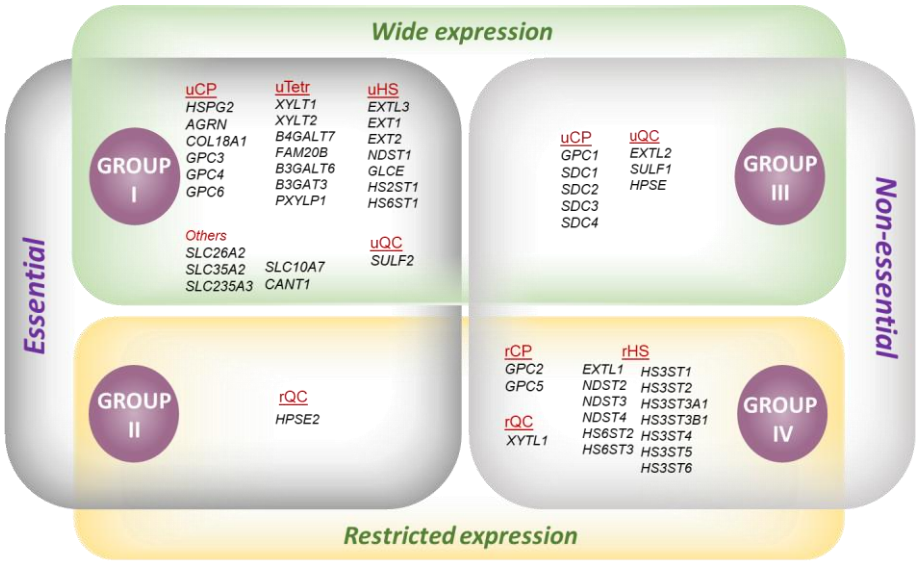


Fig. 5. Simplified organization of the HSPG biosynthetic machinery genes. After a first clustering in 'Essential' or 'Non-Essential', GSPG biosynthetic genes were clustered depending on whether they show restricted expression or wide expression in human tissues under basal physiologic conditions. This dual organization allowed formation of four groups of genes: **Group I** integrates Essential and widely expressed genes coding for ubiquitous HSPG core proteins (uCP), ubiquitous GAG-CP linker tetrasaccharide (uTetr), and ubiquitous HS sulfate (uHS) biosynthetic enzymes and transporters. **Group II** integrates an Essential and tissue-restricted gene involved in the control of post-synthetic HS modifications (indicated as rCQ, for tissue restricted quality control). **Group III** integrates 'Non-essential widely expressed genes' coding for ubiquitous HSPG core proteins (uCP) and widely expressed genes involved in HS post-synthetic modifications (uQC). **Groupe IV** integrates 'Non-essential restrictedly expressed genes' coding for tissue restricted core proteins (rCP), tissue restricted HS (rHS) biosynthetic enzymes, and tissue restricted genes involved in post-synthetic modifications (indicated as rCQ, for tissue restricted quality control).

3.3. Functional aspects of the HSPG core proteins

3.3.1. Essential pericellular core proteins

Pericellular HSPG are essential constituents of basement membranes, the highly specialized extracellular matrices that tie the epithelium to connective, endothelial, muscular, neuronal, and other tissues[121]. The main pericellular HSPG core proteins are perlecan, agrin, and collagen 18, respectively coded by *HSPG2*, *AGRN*, and *COL18A1*, which clustered in Group I (Table 1, Fig. 5, and Suppl. Inf.). Perlecan is mainly expressed in skeletal, musculoskeletal, and cardiovascular basement membranes[23,122]. Hence, *HSPG2* variants affect these tissues with severities that depend on the variant, as observed in Stuve-Wiedemann, Schwartz–Jampel and lethal Kniest-Like syndromes[23,25], which traits are recapitulated in the *Hspg2*$^{-/-}$ mouse[22]. Similarly, clinical variants in *AGRN*, which codes for the agrin core protein characteristically expressed in neuromuscular junctions[122], result in a myasthenic syndrome, recapitulated in the *Agrn*$^{-/-}$ mouse[27,29]. Collagen XVIII additionally supports basement membranes in some tissues[32], as suggested by the traits observed in *COL18A1* variants that lead to eye defects, occipital encephalocele (Knobloch syndrome), and other overt phenotypes recapitulated in the *Col18a1*$^{-/-}$ mouse[30,31]. In addition to the importance of pathogenic variants, SNP has been associated with risk to altered health condition or disease (Table 1, Fig. 3). For instance, *HSPG2* SNP has been associated with tardive dyskinesia[123] whereas *COL18A1* SNP has been associated with pelvic organ prolapse[124]. Globally, these data confirm the essentialness of ubiquitous pericellular HSPG core proteins, and the importance of their likely-ubiquitous HS chains, in the pericellular matrix organization[1].

3.3.2. Essential cell membrane-associated core proteins

Among HSPG membrane-associated core proteins coding genes, only *GPC3, GPC4,* and *GPC6* clustered in Group I (Table 1, Fig. 5, and Supp. Inf.). By keeping the HS chains very close to cell membranes, glypicans are known to bring HS binding proteins near to their high affinity receptors[125]. Proteins requiring this proximity includes fibroblast growth factors, bone morphogenetic proteins, wints, hedgehogs, and several other proteins involved in cell signaling[125]. Among *GPCs* pathogenic variants, those in *GPC3* are the most severe, as shown by pre- and post-natal overgrowth with skeletal, visceral, and neurological abnormalities observed in the X-linked Simpson-Golabi-Behmel syndrome resulting from *GPC3* variability[34] and recapitulated in the *Gpc3*$^{-/-}$ mice[33]. Similar traits are observed in *GPC4* variants responsible of Keipert syndrome[36], characterized by craniofacial and digital abnormalities, deafness, and synaptic impairment, and in *GPC6* variants responsible of omodysplasia-1[37] with facial dysmorphism, short stature, and shortened limbs, recapitulated in the knockout mice[39]. In addition to pathogenic genetic variability, SNP in *GPC4* and *GPC6* has been associated to risk to altered conditions (Table 1 and Suppl. Inf). These data confirm the essentialness of the ubiquitous *GPC3*, *GPC4*, *GPC6* in normal development and homeostasis and suggest that

Table 1. Genes coding for ubiquitous and tissue restricted HSPG core proteins. Genes are organized based on their expression in human tissues and traits of their clinical variants and/or phenotype of null mice.

	Gene	Impact of genetic variants in human	Phenotype of the null mouse	Most affected organs or tissues	Ref.
GROUP I Essential Widely expressed	*HSPG2*	From mild to lethal diseases: Lethal Kniest-like syndrome, Schwartz Jampel syndrome type 1, Stuve-Wiedemann syndrome, dyssegmental dysplasia of Silverman-Handmaker type (various degrees of bone dysplasia, cardiovascular and pulmonary defects, myotonia, hyperthermia...). SNP associations with tardive dyskinesia and intracranial aneurysm.	Embryonic lethality (E10-12) with severe chondrodysplasia, defective skeletal and cephalic development.	Cartilage, bone, bone marrow, cardiac muscle, brain, vasculature, etc.	22-25, 123
	AGRN	Congenital myasthenic syndrome with distal muscle weakness and atrophy with or without synaptic dysfunction. SNP associations with myasthenia gravis.	Embryonic lethality or die at birth due to respiratory failure, altered growth plate, central nervous system defects, and other defects.	Neuromuscular junctions, muscle, brain, growth plate.	26,27, 29,126
	COL18A1	Knobloch syndrome (severe eye and encephalocele defects). SNP association with pelvic organ prolapse.	Viable but with defects in the vasculature, ocular tissue, kidney proximal tubules, heart valves, epidermis, choroid plexus.	Eyes, kidney, heart, vasculature, brain, and epidermis.	30-32,124, 127
	GPC3	Simpson-Golabi-Behmel syndrome (pre/post-natal overgrowth, skeletal, facial, visceral, neurological anomalies). Wilms tumor (nephroblastoma).	Perinatally die with kidney and lung dysplasia, cystic kidneys, etc. Recapitulates Simpson-Golabi-Behmel syndrome.	Bone, brain, kidney, lung, muscle, etc.	33,34
	GPC4	Keipert Syndrome (craniofacial and digital abnormalities, cognitive impairment, deafness, etc.). SNP association with autism spectrum disorders.	Craniofacial and digital abnormalities, behavioral alterations.	Brain, cranial bone, limb, hearing.	36,37
	GPC6	Omodysplasia (facial dysmorphism, short stature, shortened limbs…). SNP association with osteoporosis and multiple sclerosis.	Die at birth with severe facial dysmorphism and short bones.	Skeleton, facial bones.	38-40
GROUP III Non-essential Restricted expression	*GPC2*	No pathogenic variants reported. SNP as part of polygenic risk suggestive of Alzheimer's disease.	Mouse has not been generated. However, transcript and protein levels increase when neurogenesis is stimulated in the adult mice brain and decrease when neurogenesis is ablated.	Brain.	91,128
	GPC5	No pathogenic variants reported. SNP association with risk for multiple sclerosis, autoimmune thyroid disease, and acquired nephrotic syndrome.	Mouse has not been generated.	Kidney and testis.	93-95
GROUP IV Non-essential Widely expressed	*GPC1*	No pathogenic variants reported. SNP association with risk of schizophrenia and biliary atresia.	Viable, fertile, healthy. However, slightly reduced brain size.	Brain.	80,81, 129
	SDC1	No pathogenic variants reported. SNP association with risk of coronary plaque in patients with coronary artery disease and risk of biliary atresia.	Viable, fertile, and healthy. However, altered response to stimuli during wound healing, inflammation, fibrosis, re-vascularization, infection…	Several tissues, altered response to instigating agents and angiogenesis.	69,70, 130
	SDC2	No pathogenic variants reported. SNP association with risk of posttraumatic stress disorder.	Mouse has not been generated. However, altered adaptative responses during tissue recovery after injury (fibrosis, inflammation).	Several tissues, altered response to instigating agents and angiogenesis.	72,130-132
	SDC3	No pathogenic variants reported. SNP association with resistance to obesity and risk to metabolic syndrome and female hyperandrogenism.	Viable, fertile, and healthy. However, altered resistance to obesity and altered feeding behavior.	Brain, immune and adrenal systems.	73-76, 130,133
	SDC4	No pathogenic variants reported. SNP association with miss-regulation of whole-body energy metabolism and with longevity and lipid profile in healthy elderly.	Viable, fertile, and healthy life. However, altered reactivity to stimuli related to wound healing.	Several tissues: altered response to instigating agents and angiogenesis.	77,79, 134

glypican, at least glypicans 3, 4 and 6, are the main HSPG that carry the HS involved in essential cell signaling. Whether these HS are made by the also ubiquitous HS biosynthetic genes is unknown.

3.3.3. Non-essential cell membrane-associated HSPG core proteins

Excepting the essential *GPC3*, *GPC4*, *GPC6,* all other cell membrane-associated core proteins coding genes (*GPC1*, *GPC2*, *GPC5*, *SDC1*, *SDC2*, *SDC3*, and *SDC4*) clustered in Groups III and IV, both characterized for being non-essential for normal development but required for appropriated response to stimuli, as during wound healing, angiogenesis, inflammation, immune response, normal behaviors, etc. (Table 1, Fig. 5, and Supp. Inf.). For instance, SNP in *GPC1* loci was associated to biliary atresia[80] and schizophrenia[81], whereas *GPC2*, expressed in brain (Fig. 4) was involved in proper response to neurogenic stimuli[91]. Similarly, the kidney, testis, and adrenal and thyroid glands predominant *GPC5* (Fig. 4) was involved in podocyte response to injury and nephropathy progression in type 2 diabetes patients[135] and SNP was associated with acquired nephrotic syndrome, autoimmune thyroid disease, and risk for multiple sclerosis[93-95]. Similarly, syndecans are required for appropriate tissue response to inflammation, infection, wound healing, angiogenesis, neurological stimuli, etc.[136-138]. In agreement with the role of *SCD* genes in appropriate response to stimuli, the *Sdc1*$^{-/-}$ mouse showed altered capacity to recover from microbial infection or ischemia[139] and the *Sdc4*$^{-/-}$ mouse showed delayed wound healing and impaired angiogenesis[134]. Although the *Sdc2*$^{-/-}$ mouse has not yet been generated, syndecan 2 has been involved in fibrosis and inflammation[131,132]. Moreover, SNPs in some of these genes have been associated with risk to develop diseases involving altered adaptative behaviors. For instance, *SDC1* polymorphism has been associated with coronary plaque formation in coronary artery disease patients[69], with multiple sclerosis[70], and with renal malfunction in adolescents with excess weight[71]. *SDC2* polymorphism has been associated to posttraumatic stress disorders in post war veterans[72]. *SDC3* polymorphism has been associate to satiety dysregulation and obesity[74,76], metabolic syndromes[73], and schizophrenia[75]. *SDC4* polymorphism was associated with risk of metabolic syndromes and obesity[77] and with longevity in healthy elderly Italian[79]. Together, these data indicate that Group III and IV core proteins, and possibly their HS chains, are required for appropriate responsiveness to stimuli and adaptative behaviors rather than for the regulation of essential physiological processes.

3.4. The GAG-CP-linker tetrasaccharide pathway first assures HSPG over CSPG biosynthesis

After synthesis of the core protein, the GAG-CP linker tetrasaccharide (Fig. 1) is synthetized by an ubiquitous biosynthetic machinery involving several genes which, with exception of XYLT1 and XYLT2[3], have not paralogue genes able to assure compensation. Because this machinery is common to HSPG and CSPG, it can then be expected that defects in any of the ubiquitous GAG-CP

linker biosynthetic genes will affect production of both proteoglycans in all tissues. However, a skeletal phenotype predominantly affecting CS is mainly observed in linkeropathies, the family of diseases caused by variability in GAG-CP linker biosynthetic genes[140]. Here, the integrative analysis of clinical variants, gene expression, enzymes selectivities, and protein clusters data (Table 2, Fig. 4, and Supp. Inf.) indicates that the GAG-CP linker biosynthetic pathway is organized to preferentially assure production of HSPG over that of CSPG (Fig. 6A), as detailed below.

3.4.1. The GAG-CP linker biosynthetic machinery

Biosynthesis of GAG-CP linker is initiated by addition of a xylose (Xyl) from UDP-Xyl to specific serine residues in core proteins (Ser-CP)[2,3]. This is assured by XYLT1 during development and by both XYLT1 and XYLT2 after bird[42] (Fig. 6A and Supp. Inf.). During development, *XYLT1* is highly expressed by chondroblasts[42], the cells that secrete a CSPG aggrecan rich matrix that forms the cartilage template upon which bone is formed[141]. Indeed, as endochondral ossification of the cartilage elements ensues, the expression of aggrecan by differentiating chondrocytes predominates over expression of other proteoglycan core proteins[142]. Thus, to assure all proteoglycans glycosylation in this high aggrecan content, XYLT1 binds to aggrecan core protein with lower affinity than to other core proteins[41,143], including to the small leucine rich (SLRP) CSPG decorin that stabilizes collagen fibrils in soft tissues, as skin and cornea[144]. The higher affinity of XYLT1 for decorin over aggrecan splits the GAG-CP linker biosynthetic pathway in a CSPG-decorin path and a HSPG/CSPG-aggrecan path (Fig. 6A), in which rich aggrecan synthesis is assured by an abundant substrate pool. Thus, during development, a defective *XYLT1* should primarily assure the decorin path, principally affecting the CSPG-aggrecan path. Accordingly, *XYLT1* variability results in Desbuquois dysplasia type 2 (known as Baratela–Scott syndrome)[41], characterized by severe skeletal growth retardation with multiple dislocations, joint laxity, advanced carpal ossification, with no skin overt phenotypes, indicating that the CSPG-decorin path is unaffected (Table 2 and Supp. Inf.). After birth, under physiologic conditions, *XYLT1* expression decreases and *XYLT2* is expressed. XYLT2 shows similar affinity for both decorin and aggrecan core proteins, possibly because after birth the abundant aggrecan synthesis is not any more required and HSPG and CSPG syntheses need to be balanced. Although at different extents depending on the tissue, both XYLT1 and XYLT2 are expressed after birth (Fig. 4), Thus, defects in *XYLT2* should not affect development and after birth they might be compensated by *XYLT1*, at least in some extent and in the tissues in which the enzyme is expressed. Accordingly, during development *XYLT2* variants show no clinical significance, but after birth they cause the post-developmental spondyloocular disorder characterized by osteoporosis, cataract, hearing, and learning impairment[43,135].

On the continuation of the GAG-CP linker biosynthesis, B4GALT7 adds galactose (Gal) to Xyl-(Ser-CP)[145] (Fig. 6A). B4GALT7 shows similar affinity for aggrecan and decorin core proteins, indicating that defective B4GALT7 should affect both CSPG-decorin and HSPG/CSPG-aggrecan paths. Accordingly, *B4GALT7* variants lead to B4GALT7-related Ehlers-Danlos syndrome

(spondylodysplastic EDS type 1)[45], which traits are like those caused by *XYLT1* mutations (CSPG-aggrecan path affected) with additional soft tissues overt phenotypes including hyperextensible, soft, thin, translucent, and doughy skin (CSPG-decorin path is affected) (Table 2 and Supp. Inf.)[45].

Table 2. Ubiquitous GAG-CP linker tetrasaccharide biosynthetic genes. Genes are organized based on their expression in human tissues and traits of their clinical variants and/or phenotype of null mice.

	Gene	Impact of genetic variants in human	Phenotype of the null mouse	Most affected organs or tissues	Ref.
GROUP I Essential Widely expressed	XYLT1	Desbuquois dysplasia (DBQD) type 2 (also known as Baratela–Scott syndrome), characterized by with short stature, joint laxity, advanced carpal ossification, mental retardation.	Non-available. However, a mouse carrying a *Xylt1* mutation shows chondrocyte premature maturation, early ossification, and dwarfism.	Connective and skeletal tissue, brain.	41,42
	XYLT2	Spondyloocular syndrome (post-developmental osteoporosis, cataract, hearing and learning impairment).	Develops normally but show post-developmental defects in liver, biliary epithelial and renal cysts, low body weight, adipose tissue loss and lipodystrophy.	Skeletal, adipose, and connective tissues, liver, eye, bone, brain, kidney.	43,135, 147
	B4GALT7	Spondylodysplastic Ehlers-Danlos Syndrome (spEDS)-progeroid type (connective and skeletal tissues with radioulnar synostosis) and Reunion Island Larsen-like syndrome (large spectrum).	Not reported.	Bone, cartilage, skin, brain, eye.	44,45, 148
	FAM20B	Short limb dysplasia resembling XYLT1-related DBQD with mid-face and thoracic hypoplasia (leading to respiratory failure), very short stature, mesomelic limbs shortening and multiple joint dislocations.	Dies as embryo and shows severely stunted growth, multisystem organ hypoplasia, and delayed development of the skeletal system, eyes, lung, gastrointestinal tract, and liver.	Multiorgan hypoplasia, delayed growth and development.	49,149, 150
	B3GALT6	spEDS progeroid type 2 (affecting connective and skeletal systems with sever skeletal affections), spondyloepimetaphyseal dysplasia with joint laxity, learning defects.	Not reported.	Bone, cartilage, skin tendons, ligaments, brain.	46,47
	B3GAT3	Larsen-like syndrome with variable phenotypes similar to Antley-Bixler, Shprintzen-Goldberg, and Geroderma osteodysplastica syndromes.	Embryonic lethality before the 8-cell stage because failed cytokinesis.	Bone, cartilage, skin tendons, ligaments, heart, brain.	48
	CANT1	Various skeletal phenotypes related to DBQD and Larsen syndrome with growth retardation, short extremities, progressive scoliosis, joint laxity, severe prenatal and postnatal growth retardation, etc.	Shows the continuum of skeletal dysplasia phenotypes including Desbuquois dysplasia and multiple epiphyseal dysplasia.	Skeletal and connective tissues, brain.	8,151
	SLC10A7	Shortened long bones, growth plate and tooth enamel anomalies.	Short stature, amelogenesis imperfecta, and skeletal dysplasia with scoliosis.	Skeletal and connective tissues, brain.	10
	PXYLP1	No pathogenic variants reported possibly because lethal. SNP association with good prognosis in cancer immunotherapies.	Not reported.	Unknown	152

In the HSPG/CSPG-aggrecan path, the resultant Gal-Xyl-(Ser-CP) is then phosphorylated by FAM20B to form Gal-Xyl(P)-(Ser-CP) (Fig. 6A), which is the preferred B3GALT6 substrate that additionally promotes the HSPG/CSPG-aggrecan path. Indeed, in further steps of the GAG-CP linker synthesis, phosphorylated products will become the preferred substrates that nourish the HSPG path[49,146]. This is consistent with the high lethality observed in the rare *FAM20B* clinical variants, which show a phenotype like that observed in *XYLT1*-Desbuquois dysplasia (CSPG-aggrecan path is affected) but additionally affecting many other organs[49] (HSPG path is affected) (Fig. 6A). In the next step of the GAG-CP synthesis, B3GALT6 transfers a second Gal to Gal-Xyl(P)-(Ser-CP) in the HSPG/CSPG-aggrecan path and to Gal-Xyl-(Ser-CP) in the CSPG-decorin path[145], nourishing both pathways (Fig. 6). Again, the phosphorylated product is the preferred B3GAT3 substrate that will then favor the HSPG path over the CSPG-aggrecan and CSPG-decorin paths (see below). Accordingly, *B3GALT6* mutations lead to B3GALT6-related spondylodysplastic Ehlers-Danlos syndrome (spEDS), characterized by bone dysplasia, joint laxity (CSPG-aggrecan path is affected), and mild skin hyper elasticity[47] (CSPG-decorin path is affected) (Table 2). Subsequently, depending on substrate pool levels, B3GAT3 will act alone or in association with PXYLP1, for which it shows a high affinity. Indeed, when complexed to PXYLP1, for which B3GAT3 shows the high affinity, GlcA addition occurs followed by immediate Xyl dephosphorylation to form GlcA-Gal-Gal-Xyl-(Ser-CP), which is the preferred substrate for EXTL3 that starts HS chain initiation[7,146,153] (Fig. 6A and 6B). Abundance of GlcA-Gal-Gal-Xyl-(Ser-CP) substrate pool allows CSGALNACT1 action on substrates that did not integrate the HS path. Moreover, in associating with B3GAT3 un-complexed PXYLP1, CSGALNACT1 can also act in Gal-Gal-Xyl(P)-(Ser-CP), which is the preferred substrate for CSGALNACT1, starting CS chain initiation[7,146,153]. Thus, higher expression levels of B3GAT3 compared to PXYLP1 (Fig. 4) and high affinity between PXYLP1 and B3GAT3 can favor equilibrium through the HS synthesis. PXYLP1 un-complexed B3GAT3 (Fig. 6A). Therefore, in defective *EXTL3*, its substrate excess allows the initiation of CS chain synthesis by CSGALNACT1 along or in association with PXYLP1[146,153], favoring aggrecan synthesis (Fig. 6A and 6B)[7,146,153]. Thus, the phosphorylated and non-phosphorylated substrate pools and the levels of expression of B3GAT3, PXYLP1, XYLT3, and CSGALNACT1, at least, will participate to the selection of the proteoglycan that will be preferentially elongated[146]. Globally, based in literature data, clinical variants and expression databases, our analysis indicate that the CSPG and HSPG overlapping paths can concomitantly occur with crossed pathways, which preferences are driven by abundancy of substrates pools, substrates selectivities, levels of expression, and protein-protein interactions. In case of defective genes , the preferred maintain of the HSPG path is supported by the mixed skeletal and soft tissue phenotypes observed in *B3GAT3* clinical variants responsible of Larsen, Antley-Bixler, Shprintzen-Goldberg, Geroderma osteodysplastica, and spEDS[48] and other linkeropathies (Tables 2 and 3).

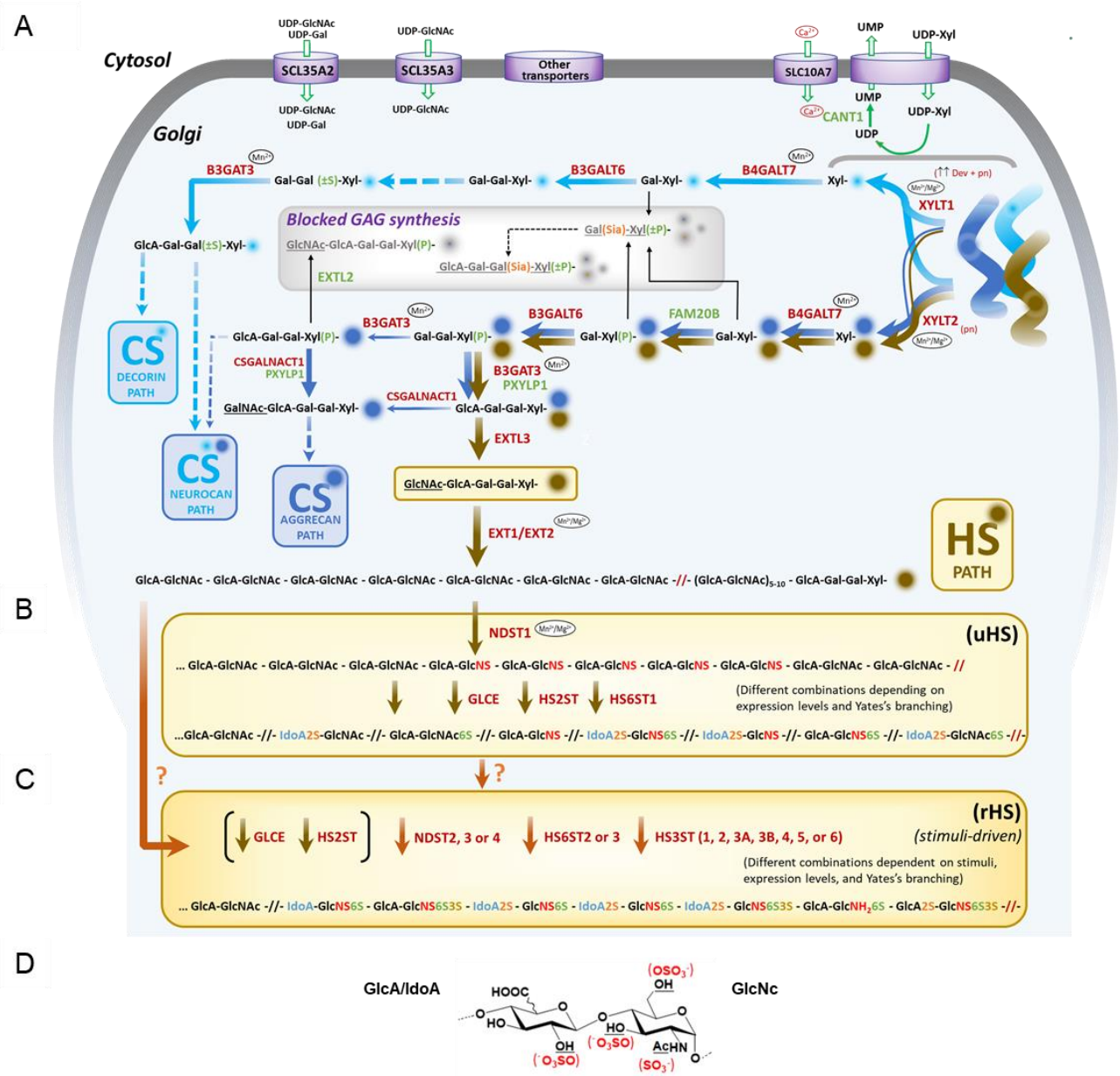


Fig. 4. Schematic representation of the HSPG biosynthetic pathway. The HSPG path (brown arrows) overlaps with the CSPG-aggrecan path (marine blue arrows). The CSPG-decorin path (light blue arrows) is spliced from HSPG/CSPG path at early steps. Black arrows designate paths that stop GAG elongation. Arrow width designs preferred paths based in substrate preferences (see Gene review), wider arrows indicating preferred paths. Substrate pools guarantees maintenance of minor paths (narrow arrows). Dotted arrows indicate non-detailed steps. Green empty arrows indicate UDP-sugars or ion transport. Some nucleotide transporters or ion transports are represented. Ions as Ca^{2+}, Mn^{2+} and/or Mg^{2+} are indicated when required for enzymatic activity. **A)** Biosynthesis of the GAG-CP linker tetrasaccharide common to HSPG and CSPG.), Excepting XYLT1 and XYLT2, all enzymes and transports are ubiquitously expressed during development (dev) and after birth (aft. birth). XYLT1 is highly expressed during development (↑↑dev) and its expression decreases aft. birth and preferentially binds to low molecular weight CSPG core proteins, as decorin (leading to favored CSPG-decorin path). XYLT2 is expressed aft. birth and shows similar binding to HSPG and CSPG core proteins (leading to both CSPG-decorin and HSPG/CSPG-aggrecan paths). Xyl is not phosphorylated in the CSPG-decorin path. **B)** Ubiquitously expressed enzymes drive the synthesis of ubiquitous HS sequences (uHS). **C)** Tissue restricted enzymes form HS sequences (rHS) likely during response to stimuli. rHS might contain high NS domains formed either from uHS or directly from unsulfated chains). **D)** Representative HS disaccharide indicating GlcA, IdoA and GlcN units and sulfation positions (in red).

3.4.2. Other enzymes and transporters

In addition to variability in GAG-CP linker biosynthetic genes, genetic variability in ion and nucleotide transporters, or in other Golgian proteins that allow availability of sugar nucleotides and ions, can dramatically alter the biosynthesis of the GAG-CP linker[154,155]. For instance, the calcium activated nucleotidase-1 (CANT1) hydrolyses the uridine diphosphate nucleoside (UDP), which inhibits glycosyltransferases (as XYLTs)[155] (Fig. 6A). Accordingly, *CANT1* clinical variants lead to Desbuquois dysplasia traits like those observed in *XYLT1* and *B4GAT7* variants[8,151] (Table 2). Because CANT1 is calcium dependent, variants in the Ca^{2+} transporter SLC10A7 result in skeletal dysplasia with *CANT1* and *XYLT1* overlapping traits[10] (Table 2). Similarly, variants in genes coding for sulfate ion transporters (*SLC26A2*)[156], nucleotide (UDP-GlcNAc and UDP-Gal) transporters (*SCL35A2* and *SCL35A3*), or ions (Mg^{2+} and Mn^{2+}) transporters required for proper enzymatic activities at the Golgi (Fig. 6A), can result in linkeropathies-like phenotypes[9,155,157].

3.5. Essential and non-essential HS sequences

3.5.1. Ubiquitous HS are essential for normal development and homeostasis

After biosynthesis of the GAG-CP linker, formation of the HS chain starts by the action of the glycosyltransferase EXTL3[2], which adds a GlcNAc residue to the dephosphorylated GlcA-Gal-Gal-Xyl-(Ser-CP) product of B3GAT3/PXYLP1 (Fig. 6B). EXTL3 deficiency results in reduced HSPG and increased CSPG levels, confirming that abundant GlcA-Gal-Gal-Xyl-(Ser-CP) and GlcA-Gal-Gal-Xyl(P)-(Ser-CP) substrate pools respectively favor CSGALNACT1 and CSGALNACT1/PXYLP1 activities, thus driving CS initiation (Fig. 6A). Accordingly, lack of *Extl3* in mice causes embryonic lethality possibly due to the lack of HS chain synthesis. In human, biallelic mutations in the *EXTL3* gene result in a neuro-immuno-skeletal dysplasia syndrome[54], suggesting an imbalanced HSPG/CSPG-aggrecan paths with decreased HS synthesis and increased syntheses of CSPG aggrecan and neurocan (Fig. 6A and 6B, Table 3, and Supp. Inf.). Following continuation of the HS chain synthesis, the EXT1/EXT2 hetero-oligomeric complex elongates the HS chain by alternated addition GlcNAc and GlcA to the GlcNAc-GlcA-Gal-Gal-Xyl-(Ser-CP) substrate, which initiates HS chain elongation[2] (Fig. 6B). In *Ext1/Ext2* depleted mice, lack of HS leads to gastrulation failure and embryonic death[55,56], agreeing the essentialness of HS chains carried by proteoglycans. In human, *EXT1* or *EXT2* variants lead to hereditary multiple exostoses (Table 3), characterized by the formation of cartilage-capped bony outgrowths[158], suggestive of the accumulation of EXT1/EXT2 GlcNAc-GlcA-Gal-Gal-Xyl-(Ser-CP) substrate, which is the EXTL3 product. Thus, this in turn results in increased GlcA-Gal-Gal-Xyl-(Ser-CP) pool, which nourishes the CSPG-aggrecan path (Fig. 6A). Thus, in the context of GlcA-Gal-Gal-Xyl-(Ser-CP) substrate pool excess, CSPG-aggrecan can be accumulated in tissues in which its synthesis is rich, explaining the epiphyseal growth plates resulting in cartilage-capped bony outgrowths. On the continuation of HS elongation, HS chain maturation is

assured by the ubiquitous NDST1, HS2ST, HS6ST1, and C5-epimerase (*GLCE*), which respectively assure basal *N*-deacetyl-*N*-sulfation, 2-*O*-sulfation, 6-*O*-sulfation, and GlcA/IdoA epimerization[2]. Interestingly, NDST1 and HS6ST1 substrate specificities suggest that, in association with HS2ST and GLCE, these enzymes can form sequences carrying NA/NS domains[59,159-163] (Supp. Inf.). Indeed, NDST1 can generate HS domains with lower level of *N*-sulfation than NDST2[164] while HS6ST1 shows higher selectivity for non-2-*O*-sulfated residues (hexA-GlcNS), as confirmed in the *Hs6st1* null mouse[59]. According with their essentialness, uHS made of NA/NS domains might then be those that interact and regulate HBP required for normal development and homeostasis[17,165]. Accordingly, *Ndst1*$^{-/-}$, *Hs2st1*$^{-/-}$, and *Glce*$^{-/-}$ mice perinatally die with severe developmental abnormalities[57,60,61], and the *Hs6st1*$^{-/-}$ mouse cannot thrive[59] (Table 3).

Table 3. Ubiquitous heparan sulfate (uHS) biosynthetic genes. Genes are organized based on their expression in human tissues and on clinical variants traits and/or phenotype of the corresponding null mice.

	Gene	Impact of genetic variants in human	Phenotype of the null mouse	Most affected organs or tissues	Ref.
GROUP I Essential Widely expressed	*EXTL3*	Severe autosomal recessive skeletal dysplasia, including epispondylo-metaphyseal dysplasia, with developmental delay, immunodeficiency, and neuromotor and brain development delay.	Embryonically lethal at around 9 days post-fecundation with HS non-detected and with over production of CS in chondrogenic tissues.	All tissues.	54,167
	EXT1	Hereditary multiple exostoses (multiple cartilaginous tumors), scoliosis, seizures, macrocephaly, defects in pancreas, lung, heart, cornea, etc.	*Ext1*$^{-/-}$*Ext2*$^{-/-}$ mice die at embryo due to gastrulation failure. *Ext1*$^{-/-}$ form several exostoses.	Cartilage and bone.	55,158
	EXT2	Hereditary multiple exostoses (less severe that that caused by EXT1).	*Ext1*$^{-/-}$*Ext2*$^{-/-}$ mouse dies at embryo due to gastrulation failure. *Ext1*$^{-/-}$ mouse forms several exostoses.	Cartilage and bone.	55,56, 158
	NDST1	Compound heterozygous mutations cause developmental delays, muscular hypotonia, ataxia, history of seizures, intellectual disability, epilepsy, gastroesophageal reflux, minor malformation, etc. Partially compensated by NDST2 except in brain (expressing NDST3 and 4), and skeletal muscle.	*Ndst1*$^{-/-}$ perinatally dies by respiratory failure and shows skeletal, brain, and heart defects. Heterozygous mouse shows severe developmental defects of the forebrain and forebrain-derived structures.	Brain, lung, skeletal and hearth muscles, stomach, etc.	58,61
	GLCE	No pathogenic variants reported possibly because lethal. SNP associations with hypertension, and cerebrovascular events.	Neonatally lethal with severe developmental abnormalities on kidney, lung, skeleton, spleen, thymus, lymph node, etc.	Several organs: kidney, skeleton, thymus, lung, lymph node, spleen, etc.	11,57, 168
	HS2ST1	No pathogenic variants reported possibly because lethal. SNP has been associated to n birds and pig has been associated with low metabolic rate and longevity in pigs and birds.	Neonatally lethal with onset of abnormalities after mid-gestation leading to traits including complete failure of kidney development.	Several tissues, particularly kidney.	60,169
	HS6ST1	No pathogenic variants reported possibly because lethal. SNP association with delayed puberty and idiopathic hypogonadotrophic hypogonadism.	Lethal at late embryonic stages, with abnormalities in lung morphology, angiogenesis, and retinal axon guidance. If viable, mice showed growth retardation and lung defects.	Several organs: lung, kidney, cartilage, vascular system, reproduction organs, etc.	59

Moreover, no pathogenic variants are known in human for *HS2ST1*, *HS6ST1*, or *GLCE*, possibly because lethal. However, individuals with compound heterozygous *NDST1* mutations can survive, although with brain and muscle overt phenotypes[58]. This is possibly due to NDST2 redundance, as *NDST2* is not expressed in brain and skeletal muscle (Fig. 4)[58,166]. Together, these data support that uHS sequences synthetized by the Group I HSPG biosynthetic machinery could be those which, depending on their level of expression in each tissue, form HS sequences able to interact and regulate the activities of heparin binding proteins that assure tissue basal functions and homeostasis, as growth factors, morphogens, and other pericellular and cell membrane-associated proteins centrally involved in normal development and homeostasis.

3.5.2. Non-essential HS in the appropriate response to stimuli

NDST2, NDST3, NDST4, HS6ST2, HS6ST3, and all HS3STs, are tissue restricted sulfotransferases coded by genes clustered in Group IV (Fig. 4, Fig. 5, and Suppl. Inf). No pathogenic variants are known for any of these genes and the corresponding knockout mice, when available, develops normally, are fertile, and live healthy (Table 4, and Supp. Inf.). Although it cannot be excluded that absence of deleterious phenotypes can be due to compensation phenomena, polymorphisms in most of these genes have been associated with resistance or risk to altered conditions affecting the organs in which they are expressed (Tables 4,see Suppl. Inf.). For instance, in brain, SNP in *NDST3* has been associated with risk of schizophrenia and bipolar disorders[99,101], SNP in *NDST4* has been associated with reading difficulty and language impairment[102], SNP in *HS6ST2* has been associated with intellectual disability[105], SNP in *HS6ST3* has been associated with dysregulation of satiety[101,106], and SNP in *HS3ST1* has been associated with to Alzheimer's disease[108,110]. Moreover, the involvement of some of these genes in the appropriate response to neurological stimuli is starting to be documented. For instance, the *Ndst3*$^{-/-}$ mouse exhibited subtle impaired anxiety with not compensatory effects from other *Ndst*[100] while HS3ST2 expression was increased in the rat pineal gland after adrenergic stimulation[170], indicating a role of these enzymes in adaptative neurological functions.

In peripheral tissues, several polymorphisms have been associated, along or in combination with other genes, with altered conditions or susceptibility to disease (Table 4). For instance in coronary artery diseases (*HS3ST1*), ulcerative and Crohn's colitis (*HS3ST2*)[111], chronic obstructive pulmonary disease (*HS3ST2*)[112], and hereditary angioedema (*HS3ST6*)[119]. The importance of these genes in the regulation of response to stimuli is also supported by available knockout mice and biomedical research. For instance, the *Hs3st1* null mouse showed a strong proinflammatory phenotype[109], *HS3ST3B1* was shown to be involved in T lymphocytes activation[171], *HS3ST3A1* expression was found increased during normotensive and pre-eclamptic pregnancies[113], and *NDST2* was essential for mast cells responsiveness during inflammation and during innate and adaptive immunity toward pathogens[172]. Moreover, 3-*O*-sulfated HS (3S-HS) are well known to increase susceptibility to

infection by pathogens[12]. Accordingly, polymorphism in both *HS3ST3A1 and HS3STB1* has been associated with risk to *P. falciparum* infection[173] and some 3S-HS structures trigger HSV infection[12].

Table 4. **Tissue restricted heparan sulfate (rHS) biosynthetic genes.** Genes are organized based on their expression in human tissues, traits of clinical variants, and/or phenotype of the corresponding null mice.

	Gene	Impact of genetic variants in human	Phenotype of the null mouse	Most affected organs or tissues	Ref.
GROUP IV Non-essential Partial restricted expression	*NDST2*	No pathogenic variants reported. SNP association with risk of coronary artery disease and chronic kidney disease.	Viable, healthy, and fertile. Mast cell shows decreased sulfation in HS.	Mast cells	176,177
	HS6ST2	No pathogenic variants reported. SNP association with risk of X-linked intellectual disability and myopia.	Viable, healthy, and fertile.	Brain and cartilage.	105
	HS6ST3	No pathogenic variants reported. SNP association with decreased risk to obesity and to increased diabetic retinopathy.	Not reported.	Brain, eye.	107,178
	HS3ST1	No pathogenic variants reported. SNP association with risk of arteriosclerosis, coronary artery diseases, and Alzheimer's disease.	Viable, healthy, and fertile but show proinflammatory phenotype when submitted to pathologic stimuli.	Vascular system and brain.	179,180
	HS3ST2	No pathogenic variants reported. Expression is increased in Alzheimer's disease. SNP association with risk of schizophrenia.	Viable, healthy, and fertile. Responds to adrenergic response to light stimuli.	Brain.	12,170, 181
	HS3ST3A1	No pathogenic variants reported. SNP association with risk of HSV1 and *P. falciparum* infection.	Not reported.	Vascular system.	173,182
	HS3ST3B1	No pathogenic variants reported. SNP association with risk of *P. falciparum* infection.	Not reported. But overexpression affects immunity (peripheral T lymphocytes and Jurkat T cells adaptive response to stimuli).	Immune system.	173,182, 183
GROUP IV Non-essential High restricted expression	*NDST3*	No pathogenic variants reported. SNP association with risk of mental disorders as schizophrenia and bipolar disorders.	Develop normally, are fertile, and show only subtle hematological and behavioral abnormalities, with reduced anxiety.	Brain.	99-101
	NDST4	No pathogenic variants reported. SNP association with risk of reading disability and language impairment.	No pathological outcomes, but some anomalies can be observed in histology images of the proximal colon, where the gene is not expressed .	In cartilage, brain, immune system.	102,103
	HS3ST4	No pathogenic variants reported. SNP association with risk of altered verbal declarative memory, age-related macular degeneration, schizophrenia, and Alzheimer's disease.	Not reported.	Brain.	116,117, 178,184
	HS3ST5	No pathogenic variants reported. SNP association with risk of intellectual disability and altered gray matter volume in schizophrenia.	Not reported.	Brain.	117,185
	HS3ST6	No pathogenic variants reported. SNP association with risk of altered response to stimuli that triggers angioedema.	Not reported.	Skin, vascular system.	118,119

3.5.3. HS chain remodeling and biosynthesis arrest

After biosynthesis, HSPG are secreted to the cell surface or to the pericellular space, where their HS chains can be remodeled by SULF1, SULF2[14], and/or by heparanase[15], which activity can be negatively regulated by the enzymatically inactive heparanase 2[67] (Table 5). Although no clinical variants are known for *SULF2*, the *Sulf2*$^{-/-}$ mouse shows lethality[14] (included in Group I), whereas in human, *HPSE2* variants result in pos-developmental syndromes (included in Group II), and other altered conditions (Suppl. Inf.)[68,174,175]. This indicates that both SULF2 and HPSE2 are required for normal homeostasis and basal response to stimuli. The involvement of HPSE2 in response to stimuli was confirmed in the homozygous *Hpse2* gene trap mouse, which developed normally but showed deficient responses to chemical and electrical field stimulation and showed low weight gain after weaning and died early in life[174,175]. In the other side, both *Sulf1*$^{-/-}$ and *hpse*$^{-/-}$ mice develop and live normally, are anatomically normal fertile, and show non overt phenotypes[14,186] (included in Group III). Accordingly, non-pathogenic variants are known in human(Table 5). However, SNP has been associated with chronic graft-versus-host disease[84], supporting its clustering in Group III. Accordingly, as for other genes in Group III, nonpathogenic *SULF1* variants are known, but polymorphism has been associated to multiple sclerosis[70], fetus failure during *in vitro* fertilisation[88], and preeclampsia[89] (Table 5).

Table 5. **HS remodeling enzymes coding genes.** Genes are organized based on their expression in human tissues, clinical variants traits, and/or phenotype of the corresponding null mice.

	Gene	Impact of genetic variants in human	Phenotype of the null mouse	Most affected organs or tissues	Ref
GROUP I Essential Widely expressed	*SULF2*	No pathogenic variants reported. SNP associations with risk to altered regulation of lipoprotein metabolism, fasting, and postprandial triglycerides clearance.	Significant lethality with reduction in brain mass during neuronal development with some kidney, lung, and skeletal defects.	Various, including brain, kidney, lung, skeletal tissue, etc.	168,187-189
GROUP II Essential restricted expression	*HPSE2*	No pathogenic variants reported. SNP high associating with urofacial syndrome and that separates migraine with and without depression.	Viable, anatomically normal, and fertile.	Urinary system, facial tissue, brain.	68,174, 175
GROUP III Non-essential Widely expressed	*EXTL2*	No pathogenic variants reported.	Normal embryonic development. Healthy during adult life but show altered recovery after tissue injury.		82,83
	SULF1	No pathogenic variants reported. SNP association with multiple sclerosis, fetus failure in IVF technique, and preeclampsia.	Non-overt phenotype, viable with no phenotypic or histological defects.		70,88, 89
	HPSE	No pathogenic variants reported. SNP association with chronic graft-versus-host disease.	Viable, anatomically normal, and fertile.		84,86, 186
GROUP IV Non-essential restricted expression	*EXTL1*	No pathogenic variants reported. SNP association with muscle lipid composition, although without pathologic impact.	Primary cultures of B-cells from mouse overexpressing Extl1 showed altered maturation capacities.		120,190

Moreover, although widely distributed, *HPSE* and SULF1 basal expressions are low (Fig. 4), possibly because it is mainly required during tissue remodeling, as during wound healing, inflammation, and tumor growth and metastasis[15]. Interestingly, this is also the case for HPSE2 (Fig. 4), which main action could also be regulatory during endogenous stimuli. Together, these reports suggest that arrest of HS biosynthesis and HS chain remodeling predominantly occurs when cells and tissues are submitted to instigating or adaptative conditions, and that their SNP might predispose disease.

3.5.4. Stimuli driven arrest of HS chain initiation or elongation

During biosynthesis, EXTL1 (Group IV) and EXTL2 (Group III) can control HS levels by respectively blocking chain elongation or chain initiation when the biosynthetic machinery is overactivated, for instance during wound healing, during adaptative immunological response of cells, or following other internal or external stimuli[13]. No pathogenic variants have been reported for neither of these genes and no genetic polymorphism has been associated with disease (Table 5). Indeed, to date, only polymorphism in *EXTL1* has been associated with muscle lipid composition, although without pathologic impact[120]. However, *in vitro*, primary cultures of B-cells from mouse overexpressing *Extl1* showed altered maturation capacities[190] (Table 5), suggesting that but the enzyme might be involved in generation of environments or processed assuring an appropriate responsiveness to stimuli, including immunological stimuli[190]. Similarly, the *Extl2*$^{-/-}$ mouse showed non overt phenotype, although presented impaired recovery following liver and kidney injury[82,83]. These data suggest that arrest of HS chain initiation or elongation is not essential during development and homeostasis but is required for appropriate response to stimuli and that SNP can predispose to altered conditions with or without clinical significance.

4. Discussion

HSPGs are among the more complex molecules in biology due to their HS structural and functional diversity. This diversity results from the complexity of their biosynthetic machinery coded by dozens of genes, which unclear relation to each other has hidden advances in the understanding of proteoglycans biology. Here, by following the PRISMA guidelines for systematic reviews and meta-analysis[18], we retrieved information for the 50 main human HSPG biosynthetic genes, including those shared with the CSPG biosynthesis through the GAG-CP-linker tetrasaccharide, and for other genes involved in Golgian homeostasis. Analysis of the documented relationship between human pathogenic variants, SNPs associations with altered conditions or susceptibility to disease, null mice phenotypes, gene expression profiles, and substrates specificities of enzymes coded by the biosynthetic genes, resulted in the clustering in 4 Groups of genes organized depending on their essentialness and expression (Fig. 5). Globally, Group I integrates ubiquitous genes which by

themselves can produce complete HSPG, as the group included genes coding for cores protein, all the GAG-CP linker biosynthetic enzymes, and glycosyltransferases and three sulfotransferases involved in the synthesis of a complete HS chain. Concerning the core proteins, Group I included genes coding for all the full time pericellular HSPG core proteins (perlecan, agrin, and collagen 18)[1] and for 3 membrane-associated glypicans (glypicans 3, 4, and 6)[125], indicating that this group can afford both essential pericellular HSPG involved structure and function of the pericellular matrix and and cell membrane-associated HSPG involved in cell signaling.

Concerning the GAG-CP linker tetrasaccharide coding genes, analysis of their clinical variants, null mice phenotypes, and substrate selectivities of their coded enzymes, resulted in a new organization of the GSG-CP linker biosynthetic pathway in which the HSPG synthesis is primarily assured over that of CSPG and in which the phenotypic continuum of disorders as linkeropathies and other pathologies is understandable[52,140,191]. Moreover, this new representation of the GAG-CP linker biosynthetic genes supports the idea that HSPG biosynthetic enzymes can assemble into 'GAGosmes'[64,192], or 'Proteoglycanomes'[193], which consider the integration of proteoglycan core proteins in the protein complexes that assure proteoglycans synthesis in the Golgi. Indeed, common protein modules in the ubiquitous/essential pericellular perlecan and agrin (Group I) are different from those shared by the ubiquitous/non-essential syndecans (Group III)[1], but common to those in CSPGs aggrecan and neurocan, suggesting that agrin, perlecan, aggrecan, and neurocan can circulate through a common HSPG/CSPG path during GAG-CP linker synthesis, as proposed in Fig. 6A. Moreover, aggrecan and neurocan modules are different from those carried by decorin[1], supporting the splitting of the CSPG-decorin path from the HSPG/CSPG-aggrecan path under interaction with XYLT1 at early steps of the GAG-CP linker biosynthesis (Fig. 6A). Likewise, syndecans modules are different from those carried by agrin, perlecan, aggrecan, neurocan, and decorin[1], suggesting that the tissue restricted syndecans might integrate protein complexes different to those in which the ubiquitous core proteins are integrated. Thus, it will be important to investigate whether the tissue restricted HS biosynthetic enzymes in Groups III and IV contain protein modules that could associate them to syndecans or to other tissue restricted core proteins. Answering to this question will give information on whether rHS can be carried by uCP (Fig. 6C), as syndecans, but not only.

Concerning the HS chain, the EXTL3, EXT1, and EXT2 glycosyltransferases coding genes comprehensibly clustered in Group I, which contains genes that are essential for normal development and homeostasis (EXTL3, EXT1, and EXT2 are required for constructing the HSPG polysaccharide). Moreover, Group I clustered the ubiquitous *NDST1*, *GLCE*, *HS2ST1*, and *HS6ST1*, which appeared to be involved in the generation of uHS sequences that can then be those essential for normal development and homeostasis. In the other side, the tissue restricted *NDST2-4*, *HS6ST2-3*, and *HS3ST1-6* appeared to generate specialized rHS sequences mainly involved in responsiveness to stimuli. Based in their substrate specificities[59,159-163], NDST1 and HS6ST1, in

association with HS2ST and GLCE, likely form uHS NA/NS domains (Supp. Inf.), which sulfation level in each position might vary depending on these ubiquitous gene expression levels in the different tissues or cells[63-65]. Indeed, NDST1 generates HS domains with lower level of *N*-sulfation than NDST2[164] while HS6ST1 shows higher selectivity for non-2-*O*-sulfated residues (hexA-GlcNS), as confirmed in the *Hs6st1* null mouse[59]. In the other side, mostly based in the presence of 3-*O*-sulfation and higher levels of *N*- and 6-*O*-sulfations that result from NDST2 *and* HS6ST2 substrate specificities[12,97,100,159,161-163,177,194-199], it appeared that specialized rHS are formed by NS domains[12,162,177]. The existence of different synthetic paths leading to NA/NS or NS domains agrees the HS biosynthetic tree proposed by Yates *el al.* in which different enzymes specificities and substrate abundances can determine the HS synthesis direction either through a major branch assuring commonly occurring HS disaccharides or a minor branch assuring least common disaccharides[200]. Separation of the two branches seems to be associated to the relative abundance of two sets of structures resulting from the C5-epimerase preference to convert GlcA-GlcNS to IdoA-GlcNS compared to converting GlcA-GlcNAc to IdoA-GlcNAc[65,200]. Thus, because they are made by essential biosynthetic machinery, uHS NA/NS sequences might be those that are essential for normal development and homeostasis, whereas the specialized rHS sequences generated by enzymes coded by Group IV genes (*NDST2*, *NDST3*, *NDST4*, *HS6ST2*, *HS6ST3*, and all *HS3STs*) might be involved in specific biological functions. Accordingly, they are required for assuring adaptative behaviors and/or for proper responsiveness to instigating stimuli in the tissues or cells in which they are expressed.

Ideas and Speculation

This work highlighted the possible existence of uHS likely tailored by NDST1, HS6ST1 in associating with HS2ST and GLCE, which together can form NA/NS domains. However, experimental evidence directly demonstrating that uHS sequences in ubiquitous HSPG are formed of NA/NS require to be established. Nevertheless, because of their essentialness, one can speculate that these uHS sequences are those that interact and regulate HBP required for normal development and homeostasis, including matrix proteins (collagen, laminin, etc.), growth factors (FGFs, VEGF, PDGF, BDNF, GDNF, etc.), morphogens (Wint, hedgehog, etc.), lipid binding proteins (apolipoproteins, lipoprotein lipase, etc.), cell adhesion proteins (P and L-selectin, integrins, etc.), among others[17,165]. Interestingly, several of these factors have been involved in cancer development and growth. Indeed, if NA/NS sequences are those that under physiological conditions interact with these factors and potentiated their activities, their excess might have over-signaling related oncogenic effects. Interestingly, SULF2 is a 6-*O*-sulfatase that shows strong pro-oncogenic activity and that also clustered in Group I. Thus, by de-sulfating NS sequences, SULF2 can generate biologically active NA/NS-like sequences that potentiate growth factors activities. If this is the case SULF2 will be pro-oncogenic, as is the case[14]. Although compensation phenomena among SULF1 and SULF2 can be

argued, it noteworthy that while SULF2 is a pro-oncogene, SULF1 shows anti-oncogenic activity[14], suggesting differential functions. Accordingly, SULF1 might inactivate excess of NA/NS sequences by de-sulfating uHS and thus generating under-sulfated 6-*O*-desulfated NA/NS domains unable to promote HBP trophic effects[14,62]. Thus, it is possible that by generating biologically active NA/NS-like sequences, SULF2 favors growth factors activities acting as pro-oncogene, whereas SULF1 participate to inactivation of active NA/NS sequences by generating under sulfated and inactive NA/NS domains, acting as anti-oncogene. This hypothesis requires large research to be confirmed or discarded. Similarly, based in information from the knockout mice and GWAS data, it appeared that specialized rHS sequences are those required for appropriate responsiveness to stimuli and adaptive behaviors in tissues in which each gene is specifically expressed. An additional question of interest is whether these specialized rHS are those that afford to HS chains their fine structural specificity, which will then not be directly related to fundamental processes in life, but mostly to specific tissue responsiveness to stimuli, resistance or vulnerability to disease or altered behaviors, as those observed in complex neurological disorders. Moreover, it is possible that rHS sequences (their biosynthetic genes knockout mice lives healthy) are carried by Group II and III core proteins (their biosynthetic genes knockout mice lives healthy), whereas uHS sequences (their biosynthetic genes knockout mice die) are carried by the ubiquitous core proteins (their biosynthetic genes knockout mice die). Answering to these questions will facilitate research aiming to understand the differential role of HSPG core proteins and the HS chains that they carry in a large variety of biological and pathological context. Thus, although to date it is not yet possible to sequence single HS chains isolated from their physiological contexts, the selective purification of core proteins from tissues or cells will make available their specific HS chains, in which the research of NA/NS or NS sequences could be performed. Finally, as HS chains are long, one should consider that the specialized rHS sequences might be those that gives to cells the capacity to generate specific HS interactomes with cell membrane proteins, or matrix proteins. The involvement of Group IV rHS biosynthetic genes in complex pathologies, as neurologic or behavioral disorders, might represent a solution for the identification of pathologies associating specific rHS interactomes. We are currently working in this direction.

In conclusion, by analyzing clinical variants and transcriptomic databases and literature records for biochemical and functional information on HSPG biosynthetic machinery genes, we decrypted essential features of proteoglycans and HS structural and functional diversity. Our analysis considers new concepts non developed in previous works, for instance those in which HS structures are defined by the HSPG biosynthetic genes expression levels[63-65]. From this analysis, we propose a clinical-variant and substrate selectivity-based proteoglycan biosynthetic pathway in which HSPG are primarily assured over CSPG, and in which the HS chains or sequences are characterized for carrying either an ubiquitous ‘sulfation level specificity’ (uHS) or a specialized ‘sulfation pattern

specificity' (rHS), depending on the tissue in where they are made. This work opens new routes of research for a better understanding of the roles of HSPG in physiology and pathology.

5. Acknowledgments

This work has received funding from the ANR SkelGAG and from the European Union's Horizon 2020 Research and Innovation Program (grant agreement No 737390). G. Barreto was funded by the "Centre National de la Recherche Scientifique" (CNRS, France), "Délégation Centre-Est" (CNRS-DR6), the "Lorraine Université" (LU, France) through the initiative "Lorraine Université d'Excellence" and the dispositive "Future Leader" and the "Deutsche Forschungsgemeinschaft" (DFG, Bonn, Germany) (BA 4036/4-1). We thank ArrestAD partners, particularly Prof. T. van Kuppevelt (Radboud University, the Netherlands) and Prof. D. Fernig (University of Liverpool, UK), to Prof. P. Albanese, Dr. A. Fifre (University Paris Est Créteil, France), Prof. J-P. Li, Prof. U. Lindahl, and Prof. L. Kjellen (Uppsala University, Sweden), for interesting discussions.

6. Competing interests

The authors declare that they have no competing interests.

7. Author contributions

DPG designed the study and wrote the manuscript. DPG, MOO, BDSF, and JDR analyzed the data. DPG, GB, BDSF, and MMO participated to data organization. DPG, MOO, GB, NR, and AM participated to figures design. SC, MBH, XL, GLD, and OGV participated to data analysis and literature review. All authors discussed results and commented the manuscript for improvement.

References

1 Iozzo, R. V. & Schaefer, L. Proteoglycan form and function: A comprehensive nomenclature of proteoglycans. *Matrix Biol* **42**, 11-55, doi:10.1016/j.matbio.2015.02.003 (2015).

2 Li, J. P. & Kusche-Gullberg, M. Heparan Sulfate: Biosynthesis, Structure, and Function. *International review of cell and molecular biology* **325**, 215-273, doi:10.1016/bs.ircmb.2016.02.009 (2016).

3 Sarrazin, S., Lamanna, W. C. & Esko, J. D. Heparan sulfate proteoglycans. *Cold Spring Harbor perspectives in biology* **3**, doi:10.1101/cshperspect.a004952 (2011).

4 Gao, J. & Huang, X. Recent advances on glycosyltransferases involved in the biosynthesis of the proteoglycan linkage region. *Advances in carbohydrate chemistry and biochemistry* **80**, 95-119, doi:10.1016/bs.accb.2021.10.003 (2021).

5 Sasarman, F. *et al.* Biosynthesis of glycosaminoglycans: associated disorders and biochemical tests. *Journal of inherited metabolic disease* **39**, 173-188, doi:10.1007/s10545-015-9903-z (2016).

6 Wen, J. *et al.* Xylose phosphorylation functions as a molecular switch to regulate proteoglycan biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **111**, 15723-15728, doi:10.1073/pnas.1417993111 (2014).

7 Koike, T., Izumikawa, T., Sato, B. & Kitagawa, H. Identification of phosphatase that dephosphorylates xylose in the glycosaminoglycan-protein linkage region of proteoglycans. *The Journal of biological chemistry* **289**, 6695-6708, doi:10.1074/jbc.M113.520536 (2014).

8 Paganini, C. *et al.* Calcium activated nucleotidase 1 (CANT1) is critical for glycosaminoglycan biosynthesis in cartilage and endochondral ossification. *Matrix Biol* **81**, 70-90, doi:10.1016/j.matbio.2018.11.002 (2019).

9 Paganini, C., Gramegna Tota, C., Superti-Furga, A. & Rossi, A. Skeletal Dysplasias Caused by Sulfation Defects. *International journal of molecular sciences* **21**, doi:10.3390/ijms21082710 (2020).

10 Dubail, J. *et al.* SLC10A7 mutations cause a skeletal dysplasia with amelogenesis imperfecta mediated by GAG biosynthesis defects. *Nature communications* **9**, 3087, doi:10.1038/s41467-018-05191-8 (2018).

11 Li, J. P. Glucuronyl C5-epimerase an enzyme converting glucuronic acid to iduronic acid in heparan sulfate/heparin biosynthesis. *Progress in molecular biology and translational science* **93**, 59-78, doi:10.1016/s1877-1173(10)93004-4 (2010).

12 Thacker, B. E., Xu, D., Lawrence, R. & Esko, J. D. Heparan sulfate 3-O-sulfation: a rare modification in search of a function. *Matrix Biol* **35**, 60-72, doi:10.1016/j.matbio.2013.12.001 (2014).

13 Busse-Wicher, M., Wicher, K. B. & Kusche-Gullberg, M. The exostosin family: proteins with many functions. *Matrix Biol* **35**, 25-33, doi:10.1016/j.matbio.2013.10.001 (2014).

14 El Masri, R., Seffouh, A., Lortat-Jacob, H. & Vives, R. R. The "in and out" of glucosamine 6-O-sulfation: the 6th sense of heparan sulfate. *Glycoconjugate journal* **34**, 285-298, doi:10.1007/s10719-016-9736-5 (2017).

15 Masola, V., Bellin, G., Gambaro, G. & Onisto, M. Heparanase: A Multitasking Protein Involved in Extracellular Matrix (ECM) Remodeling and Intracellular Events. *Cells* **7**, doi:10.3390/cells7120236 (2018).

16 Mayfosh, A. J., Nguyen, T. K. & Hulett, M. D. The Heparanase Regulatory Network in Health and Disease. *International journal of molecular sciences* **22**, doi:10.3390/ijms222011096 (2021).

17 Xie, M. & Li, J. P. Heparan sulfate proteoglycan - A common receptor for diverse cytokines. *Cell Signal* **54**, 115-121, doi:10.1016/j.cellsig.2018.11.022 (2019).

18 Page, M. J. *et al.* The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Rev Esp Cardiol (Engl Ed)* **74**, 790-799, doi:10.1016/j.rec.2021.07.010 (2021).

19 Papatheodorou, I. *et al.* Expression Atlas: gene and protein expression across multiple studies and organisms. *Nucleic Acids Res* **46**, D246-D251, doi:10.1093/nar/gkx1158 (2018).

20 Filippi, L. & Braat, A. J. Theragnostics in primary and secondary liver tumors: the need for a personalized approach. *The quarterly journal of nuclear medicine and molecular imaging : official publication of the Italian Association of Nuclear Medicine (AIMN) [and] the International Association of Radiopharmacology (IAR), [and] Section of the So* **65**, 353-370, doi:10.23736/s1824-4785.21.03407-5 (2021).

21 Marques, C., Reis, C. A., Vivès, R. R. & Magalhães, A. Heparan Sulfate Biosynthesis and Sulfation Profiles as Modulators of Cancer Signalling and Progression. *Frontiers in oncology* **11**, 778752, doi:10.3389/fonc.2021.778752 (2021).

22 Arikawa-Hirasawa, E., Watanabe, H., Takami, H., Hassell, J. R. & Yamada, Y. Perlecan is essential for cartilage and cephalic development. *Nature genetics* **23**, 354-358, doi:10.1038/15537 (1999).

23 Gubbiotti, M. A., Neill, T. & Iozzo, R. V. A current view of perlecan in physiology and pathology: A mosaic of functions. *Matrix Biol* **57-58**, 285-298, doi:10.1016/j.matbio.2016.09.003 (2017).

24 Liu, H. & Huang, W. The Association between Genes Polymorphisms of Heparan Sulfate Proteoglycan 2 (HSPG2) and Chondroitin Sulfate Proteoglycan 2 (CSPG2) and Intracranial Aneurysm Susceptibility: A Meta-Analysis. *Iranian journal of public health* **48**, 1945-1951 (2019).

25 Martinez, J. R., Dhawan, A. & Farach-Carson, M. C. Modular Proteoglycan Perlecan/HSPG2: Mutations, Phenotypes, and Functions. *Genes* **9**, doi:10.3390/genes9110556 (2018).

26 Daniels, M. P. The role of agrin in synaptic development, plasticity and signaling in the central nervous system. *Neurochemistry international* **61**, 848-853, doi:10.1016/j.neuint.2012.02.028 (2012).

27 Gautam, M. *et al.* Defective neuromuscular synaptogenesis in agrin-deficient mutant mice. *Cell* **85**, 525-535, doi:10.1016/s0092-8674(00)81253-2 (1996).

28 Hausser, H. J., Ruegg, M. A., Brenner, R. E. & Ksiazek, I. Agrin is highly expressed by chondrocytes and is required for normal growth. *Histochemistry and cell biology* **127**, 363-374, doi:10.1007/s00418-006-0258-2 (2007).

29 Nicole, S. *et al.* Agrin mutations lead to a congenital myasthenic syndrome with distal muscle weakness and atrophy. *Brain : a journal of neurology* **137**, 2429-2443, doi:10.1093/brain/awu160 (2014).

30 Fukai, N. *et al.* Lack of collagen XVIII/endostatin results in eye abnormalities. *The EMBO journal* **21**, 1535-1544, doi:10.1093/emboj/21.7.1535 (2002).

31 Heljasvaara, R., Aikio, M., Ruotsalainen, H. & Pihlajaniemi, T. Collagen XVIII in tissue homeostasis and dysregulation - Lessons learned from model organisms and human patients. *Matrix Biol* **57-58**, 55-75, doi:10.1016/j.matbio.2016.10.002 (2017).

32 Seppinen, L. & Pihlajaniemi, T. The multiple functions of collagen XVIII in development and disease. *Matrix Biol* **30**, 83-92, doi:10.1016/j.matbio.2010.11.001 (2011).

33 Cano-Gauci, D. F. *et al.* Glypican-3-deficient mice exhibit developmental overgrowth and some of the abnormalities typical of Simpson-Golabi-Behmel syndrome. *The Journal of cell biology* **146**, 255-264, doi:10.1083/jcb.146.1.255 (1999).

34 Cottereau, E. *et al.* Phenotypic spectrum of Simpson-Golabi-Behmel syndrome in a series of 42 cases with a mutation in GPC3 and review of the literature. *American journal of medical genetics. Part C, Seminars in medical genetics* **163c**, 92-105, doi:10.1002/ajmg.c.31360 (2013).

35 Iglesias, B. V. *et al.* Expression pattern of glypican-3 (GPC3) during human embryonic and fetal development. *Histology and histopathology* **23**, 1333-1340, doi:10.14670/hh-23.1333 (2008).

36 Amor, D. J. *et al.* Pathogenic Variants in GPC4 Cause Keipert Syndrome. *American journal of human genetics* **104**, 914-924, doi:10.1016/j.ajhg.2019.02.026 (2019).

37 Dowling, C. & Allen, N. J. Mice Lacking Glypican 4 Display Juvenile Hyperactivity and Adult Social Interaction Deficits. *Brain plasticity (Amsterdam, Netherlands)* **4**, 197-209, doi:10.3233/bpl-180079 (2018).

38 Campos-Xavier, A. B. *et al.* Mutations in the heparan-sulfate proteoglycan glypican 6 (GPC6) impair endochondral ossification and cause recessive omodysplasia. *American journal of human genetics* **84**, 760-770, doi:10.1016/j.ajhg.2009.05.002 (2009).

39 Capurro, M. *et al.* Glypican-6 promotes the growth of developing long bones by stimulating Hedgehog signaling. *The Journal of cell biology* **216**, 2911-2926, doi:10.1083/jcb.201605119 (2017).

40 Kemp, J. P. *et al.* Identification of 153 new loci associated with heel bone mineral density and functional involvement of GPC6 in osteoporosis. *Nature genetics* **49**, 1468-1475, doi:10.1038/ng.3949 (2017).

41 Bui, C. *et al.* XYLT1 mutations in Desbuquois dysplasia type 2. *American journal of human genetics* **94**, 405-414, doi:10.1016/j.ajhg.2014.01.020 (2014).

42 Mis, E. K. *et al.* Forward genetics defines Xylt1 as a key, conserved regulator of early chondrocyte maturation and skeletal length. *Developmental biology* **385**, 67-82, doi:10.1016/j.ydbio.2013.10.014 (2014).

43 Taylan, F. *et al.* Spondyloocular Syndrome: Novel Mutations in XYLT2 Gene and Expansion of the Phenotypic Spectrum. *Journal of bone and mineral research : the official journal of the American Society for Bone and Mineral Research* **31**, 1577-1585, doi:10.1002/jbmr.2834 (2016).

44 Caraffi, S. G. *et al.* Severe Peripheral Joint Laxity is a Distinctive Clinical Feature of Spondylodysplastic-Ehlers-Danlos Syndrome (EDS)-B4GALT7 and Spondylodysplastic-EDS-B3GALT6. *Genes* **10**, doi:10.3390/genes10100799 (2019).

45 Mihalic Mosher, T. *et al.* Expansion of B4GALT7 linkeropathy phenotype to include perinatal lethal skeletal dysplasia. *European journal of human genetics : EJHG*, doi:10.1038/s41431-019-0464-8 (2019).

46 Malfait, F. *et al.* Defective initiation of glycosaminoglycan synthesis due to B3GALT6 mutations causes a pleiotropic Ehlers-Danlos-syndrome-like connective tissue disorder. *American journal of human genetics* **92**, 935-945, doi:10.1016/j.ajhg.2013.04.016 (2013).

47 Van Damme, T. *et al.* Biallelic B3GALT6 mutations cause spondylodysplastic Ehlers-Danlos syndrome. *Human molecular genetics* **27**, 3475-3487, doi:10.1093/hmg/ddy234 (2018).

48 Ritelli, M. *et al.* Further Defining the Phenotypic Spectrum of B3GAT3 Mutations and Literature Review on Linkeropathy Syndromes. *Genes* **10**, doi:10.3390/genes10090631 (2019).

49 Kuroda, Y. *et al.* A novel gene (FAM20B encoding glycosaminoglycan xylosylkinase) for neonatal short limb dysplasia resembling Desbuquois dysplasia. *Clinical genetics* **95**, 713-717, doi:10.1111/cge.13530 (2019).

50 Spyroglou, A. *et al.* Diastrophic dysplasia sulfate transporter (SLC26A2) is expressed in the adrenal cortex and regulates aldosterone secretion. *Hypertension (Dallas, Tex. : 1979)* **63**, 1102-1109, doi:10.1161/hypertensionaha.113.02504 (2014).

51 Asselta, R. *et al.* X Chromosome Contribution to the Genetic Architecture of Primary Biliary Cholangitis. *Gastroenterology* **160**, 2483-2495.e2426, doi:10.1053/j.gastro.2021.02.061 (2021).

52 Taylan, F. & Mäkitie, O. Abnormal Proteoglycan Synthesis Due to Gene Defects Causes Skeletal Diseases with Overlapping Phenotypes. *Hormone and metabolic research = Hormon- und Stoffwechselforschung = Hormones et metabolisme* **48**, 745-754, doi:10.1055/s-0042-118706 (2016).

53 Osman, N. M., Kagohashi, Y., Udagawa, J. & Otani, H. Alpha1,4-N-acetylglucosaminyltransferase encoding gene EXTL3 expression pattern in mouse adult and developing tissues with special attention to the pancreas. *Anatomy and embryology* **207**, 333-341, doi:10.1007/s00429-003-0348-z (2003).

54 Yamada, S. Specific functions of Exostosin-like 3 (EXTL3) gene products. *Cellular & molecular biology letters* **25**, 39, doi:10.1186/s11658-020-00231-y (2020).

55 Lin, X. *et al.* Disruption of gastrulation and heparan sulfate biosynthesis in EXT1-deficient mice. *Developmental biology* **224**, 299-311, doi:10.1006/dbio.2000.9798 (2000).

56 Stickens, D., Zak, B. M., Rougier, N., Esko, J. D. & Werb, Z. Mice deficient in Ext2 lack heparan sulfate and develop exostoses. *Development (Cambridge, England)* **132**, 5055-5068, doi:10.1242/dev.02088 (2005).

57 Li, J. P. *et al.* Targeted disruption of a murine glucuronyl C5-epimerase gene results in heparan sulfate lacking L-iduronic acid and in neonatal lethality. *The Journal of biological chemistry* **278**, 28363-28366, doi:10.1074/jbc.C300219200 (2003).

58 Armstrong, L. *et al.* A girl with developmental delay, ataxia, cranial nerve palsies, severe respiratory problems in infancy-Expanding NDST1 syndrome. *American journal of medical genetics. Part A* **173**, 712-715, doi:10.1002/ajmg.a.37621 (2017).

59 Habuchi, H. *et al.* Mice deficient in heparan sulfate 6-O-sulfotransferase-1 exhibit defective heparan sulfate biosynthesis, abnormal placentation, and late embryonic lethality. *The Journal of biological chemistry* **282**, 15578-15588, doi:10.1074/jbc.M607434200 (2007).

60 Merry, C. L. *et al.* The molecular phenotype of heparan sulfate in the Hs2st-/- mutant mouse. *The Journal of biological chemistry* **276**, 35429-35434, doi:10.1074/jbc.M100379200 (2001).

61 Ringvall, M. & Kjellen, L. Mice deficient in heparan sulfate N-deacetylase/N-sulfotransferase 1. *Progress in molecular biology and translational science* **93**, 35-58, doi:10.1016/s1877-1173(10)93003-2 (2010).

62 Chiu, L. T. *et al.* Trisaccharide Sulfate and Its Sulfonamide as an Effective Substrate and Inhibitor of Human Endo-O-sulfatase-1. *Journal of the American Chemical Society* **142**, 5282-5292, doi:10.1021/jacs.0c00005 (2020).

63 Lindahl, U. & Kjellen, L. Pathophysiology of heparan sulphate: many diseases, few drugs. *J Intern Med* **273**, 555-571, doi:10.1111/joim.12061 (2013).

64 Kreuger, J. & Kjellén, L. Heparan sulfate biosynthesis: regulation and variability. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **60**, 898-907, doi:10.1369/0022155412464972 (2012).

65 Huang, Y. F., Mizumoto, S. & Fujita, M. Novel Insight Into Glycosaminoglycan Biosynthesis Based on Gene Expression Profiles. *Frontiers in cell and developmental biology* **9**, 709018, doi:10.3389/fcell.2021.709018 (2021).

66 Grobe, K. & Esko, J. D. Regulated translation of heparan sulfate N-acetylglucosamine N-deacetylase/n-sulfotransferase isozymes by structured 5'-untranslated regions and internal ribosome entry sites. *The Journal of biological chemistry* **277**, 30699-30706 (2002).

67 Pinhal, M. A. S., Melo, C. M. & Nader, H. B. The Good and Bad Sides of Heparanase-1 and Heparanase-2. *Advances in experimental medicine and biology* **1221**, 821-845, doi:10.1007/978-3-030-34521-1_36 (2020).

68 Petschner, P. *et al.* A replication study separates polymorphisms behind migraine with and without depression. *PloS one* **16**, e0261477, doi:10.1371/journal.pone.0261477 (2021).

69 Nemoto, T. *et al.* Endothelial glycocalyx and severity and vulnerability of coronary plaque in patients with coronary artery disease. *Atherosclerosis* **302**, 1-7, doi:10.1016/j.atherosclerosis.2020.04.014 (2020).

70 Okolicsanyi, R. K., Bluhm, J., Miller, C., Griffiths, L. R. & Haupt, L. M. An investigation of genetic polymorphisms in heparan sulfate proteoglycan core proteins and key modification enzymes in an Australian Caucasian multiple sclerosis population. *Human genomics* **14**, 18, doi:10.1186/s40246-020-00264-6 (2020).

71 Saboia, Z., Meneses, G. C., Martins, A. M. C., Daher, E. F. & Silva Junior, G. B. Association between syndecan-1 and renal function in adolescents with excess weight: evidence of subclinical kidney disease and endothelial dysfunction. *Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas* **51**, e7174, doi:10.1590/1414-431x20177174 (2018).

72 Ashley-Koch, A. E. *et al.* Genome-wide association study of posttraumatic stress disorder in a cohort of Iraq-Afghanistan era veterans. *Journal of affective disorders* **184**, 225-234, doi:10.1016/j.jad.2015.03.049 (2015).

73 Chang, B. C., Hwang, L. C. & Huang, W. H. Positive Association of Metabolic Syndrome with a Single Nucleotide Polymorphism of Syndecan-3 (rs2282440) in the Taiwanese Population. *International journal of endocrinology* **2018**, 9282598, doi:10.1155/2018/9282598 (2018).

74 Ha, E. *et al.* Positive association of obesity with single nucleotide polymorphisms of syndecan 3 in the Korean population. *The Journal of clinical endocrinology and metabolism* **91**, 5095-5099, doi:10.1210/jc.2005-2086 (2006).

75 Moons, T. *et al.* Clock genes and body composition in patients with schizophrenia under treatment with antipsychotic drugs. *Schizophrenia research* **125**, 187-193, doi:10.1016/j.schres.2010.10.008 (2011).

76 Schüring, A. N. *et al.* Role of syndecan-3 polymorphisms in obesity and female hyperandrogenism. *Journal of molecular medicine (Berlin, Germany)* **87**, 1241-1250, doi:10.1007/s00109-009-0529-1 (2009).

77 De Luca, M. *et al.* A conserved role for syndecan family members in the regulation of whole-body energy metabolism. *PloS one* **5**, e11286, doi:10.1371/journal.pone.0011286 (2010).

78 Nagy, N. *et al.* The altered expression of syndecan 4 in the uninvolved skin of venous leg ulcer patients may predispose to venous leg ulcer. *Wound repair and regeneration : official publication of the Wound Healing Society [and] the European Tissue Repair Society* **16**, 495-502, doi:10.1111/j.1524-475X.2008.00394.x (2008).

79 Rose, G. *et al.* Metabolism and successful aging: Polymorphic variation of syndecan-4 (SDC4) gene associate with longevity and lipid profile in healthy elderly Italian subjects. *Mechanisms of ageing and development* **150**, 27-33, doi:10.1016/j.mad.2015.08.003 (2015).

80 Bai, M. R. *et al.* Association of common variation in ADD3 and GPC1 with biliary atresia susceptibility. *Aging* **12**, 7163-7182, doi:10.18632/aging.103067 (2020).

81 Potkin, S. G. *et al.* A genome-wide association study of schizophrenia using brain activation as a quantitative phenotype. *Schizophrenia bulletin* **35**, 96-108, doi:10.1093/schbul/sbn155 (2009).

82 Nadanaka, S., Kagiyama, S. & Kitagawa, H. Roles of EXTL2, a member of the EXT family of tumour suppressors, in liver injury and regeneration processes. *The Biochemical journal* **454**, 133-145, doi:10.1042/bj20130323 (2013).

83 Purnomo, E. *et al.* Glycosaminoglycan overproduction in the aorta increases aortic calcification in murine chronic kidney disease. *Journal of the American Heart Association* **2**, e000405, doi:10.1161/jaha.113.000405 (2013).

84 Martin, P. J. *et al.* Replication of associations between genetic polymorphisms and chronic graft-versus-host disease. *Blood* **128**, 2450-2456, doi:10.1182/blood-2016-07-728063 (2016).

85 Ostrovsky, O. *et al.* The HPSE Gene Insulator-A Novel Regulatory Element That Affects Heparanase Expression, Stem Cell Mobilization, and the Risk of Acute Graft versus Host Disease. *Cells* **10**, doi:10.3390/cells10102523 (2021).

86 Ostrovsky, O., Shimoni, A., Rand, A., Vlodavsky, I. & Nagler, A. Genetic variations in the heparanase gene (HPSE) associate with increased risk of GVHD following allogeneic stem cell transplantation: effect of discrepancy between recipients and donors. *Blood* **115**, 2319-2328, doi:10.1182/blood-2009-08-236455 (2010).

87 Oshima, K. *et al.* Loss of endothelial sulfatase-1 after experimental sepsis attenuates subsequent pulmonary inflammatory responses. *American journal of physiology. Lung cellular and molecular physiology* **317**, L667-l677, doi:10.1152/ajplung.00175.2019 (2019).

88 Taghizadeh, E. *et al.* SULF 1 gene polymorphism, rs6990375 is in significant association with fetus failure in IVF technique. *Iranian journal of reproductive medicine* **13**, 215-220 (2015).

89 Zhang, P. *et al.* Polymorphisms in Sex Hormone Metabolism Genes and Risk of Preeclampsia in Taiyuan, China. *Gynecologic and obstetric investigation* **83**, 179-186, doi:10.1159/000478931 (2018).

90 Kamimura, K. & Maeda, N. Glypicans and Heparan Sulfate in Synaptic Development, Neural Plasticity, and Neurological Disorders. *Front Neural Circuits* **15**, 595596, doi:10.3389/fncir.2021.595596 (2021).

91 Lugert, S. *et al.* Glypican-2 levels in cerebrospinal fluid predict the status of adult hippocampal neurogenesis. *Scientific reports* **7**, 46543, doi:10.1038/srep46543 (2017).

92 Luxardi, G. *et al.* Glypicans are differentially expressed during patterning and neurogenesis of early mouse brain. *Biochem Biophys Res Commun* **352**, 55-60, doi:10.1016/j.bbrc.2006.10.185 (2007).

93 Chorąży, M. *et al.* Analysis of chosen SNVs in GPC5, CD58 and IRF8 genes in multiple sclerosis patients. *Advances in medical sciences* **64**, 230-234, doi:10.1016/j.advms.2018.12.004 (2019).

94 Guan, Y. *et al.* The Role of Cell Growth-Related Gene Copy Number Variation in Autoimmune Thyroid Disease. *Biological trace element research* **195**, 409-416, doi:10.1007/s12011-019-01880-7 (2020).

95 Okamoto, K. *et al.* Common variation in GPC5 is associated with acquired nephrotic syndrome. *Nature genetics* **43**, 459-463, doi:10.1038/ng.792 (2011).

96 Bush, K. T. *et al.* N-sulfation of heparan sulfate regulates early branching events in the developing mammary gland. *The Journal of biological chemistry* **287**, 42064-42070, doi:10.1074/jbc.M112.423327 (2012).

97 Ledin, J. *et al.* Enzymatically active N-deacetylase/N-sulfotransferase-2 is present in liver but does not contribute to heparan sulfate N-sulfation. *The Journal of biological chemistry* **281**, 35727-35734, doi:10.1074/jbc.M604113200 (2006).

98 Ledin, J. *et al.* Heparan sulfate structure in mice with genetically modified heparan sulfate production. *The Journal of biological chemistry* **279**, 42732-42741, doi:10.1074/jbc.M405382200 (2004).

99 Lencz, T. *et al.* Genome-wide association study implicates NDST3 in schizophrenia and bipolar disorder. *Nature communications* **4**, 2739, doi:10.1038/ncomms3739 (2013).

100 Pallerla, S. R. *et al.* Altered heparan sulfate structure in mice with deleted NDST3 gene function. *The Journal of biological chemistry* **283**, 16885-16894, doi:10.1074/jbc.M709774200 (2008).

101 Wang, L. *et al.* Association study of NDST3 gene for schizophrenia, bipolar disorder, major depressive disorder in the Han Chinese population. *American journal of medical genetics. Part B, Neuropsychiatric genetics : the official publication of the International Society of Psychiatric Genetics* **177**, 3-9, doi:10.1002/ajmg.b.32573 (2018).

102 Eicher, J. D. *et al.* Genome-wide association study of shared components of reading disability and language impairment. *Genes, brain, and behavior* **12**, 792-801, doi:10.1111/gbb.12085 (2013).

103 Jao, T. M. *et al.* Alteration of colonic epithelial cell differentiation in mice deficient for glucosaminyl N-deacetylase/N-sulfotransferase 4. *Oncotarget* **7**, 84938-84950, doi:10.18632/oncotarget.12915 (2016).

104 Guo, Y. *et al.* Downregulation of HS6ST2 by miR-23b-3p enhances matrix degradation through p38 MAPK pathway in osteoarthritis. *Cell death & disease* **9**, 699, doi:10.1038/s41419-018-0729-0 (2018).

105 Paganini, L. *et al.* A HS6ST2 gene variant associated with X-linked intellectual disability and severe myopia in two male twins. *Clinical genetics* **95**, 368-374, doi:10.1111/cge.13485 (2019).

106 Iwase, M. *et al.* Differential Effect of Polymorphisms on Body Mass Index Across the Life Course of Japanese: The Japan Multi-Institutional Collaborative Cohort Study. *Journal of epidemiology* **31**, 172-179, doi:10.2188/jea.JE20190296 (2021).

107 Wang, K. S., Wang, L., Liu, X. & Zeng, M. Association of HS6ST3 gene polymorphisms with obesity and triglycerides: gene x gender interaction. *Journal of genetics* **92**, 395-402 (2013).

108 Desikan, R. S. *et al.* Polygenic Overlap Between C-Reactive Protein, Plasma Lipids, and Alzheimer Disease. *Circulation* **131**, 2061-2069, doi:10.1161/circulationaha.115.015489 (2015).

109 Shworak, N. W., HajMohammadi, S., de Agostini, A. I. & Rosenberg, R. D. Mice deficient in heparan sulfate 3-O-sulfotransferase-1: normal hemostasis with unexpected perinatal phenotypes. *Glycoconjugate journal* **19**, 355-361, doi:10.1023/a:1025377206600 (2002).

110 Witoelar, A. *et al.* Meta-analysis of Alzheimer's disease on 9,751 samples from Norway and IGAP study identifies four risk loci. *Scientific reports* **8**, 18088, doi:10.1038/s41598-018-36429-6 (2018).

111 Azuara, D. *et al.* Novel methylation panel for the early detection of neoplasia in high-risk ulcerative colitis and Crohn's colitis patients. *Inflammatory bowel diseases* **19**, 165-173, doi:10.1002/ibd.22994 (2013).

112 Yu, H., Guo, W., Liu, Y. & Wang, Y. Immune Characteristics Analysis and Transcriptional Regulation Prediction Based on Gene Signatures of Chronic Obstructive Pulmonary Disease. *International journal of chronic obstructive pulmonary disease* **16**, 3027-3039, doi:10.2147/copd.s325328 (2021).

113 Amraoui, F. *et al.* Placental expression of heparan sulfate 3-O-sulfotransferase-3A1 in normotensive and pre-eclamptic pregnancies. *Placenta* **36**, 1218-1224, doi:10.1016/j.placenta.2015.09.008 (2015).

114 Huang, J. *et al.* Identification of gene and microRNA changes in response to smoking in human airway epithelium by bioinformatics analyses. *Medicine* **98**, e17267, doi:10.1097/md.0000000000017267 (2019).

115 Joubert, B. R., Franceschini, N., Mwapasa, V., North, K. E. & Meshnick, S. R. Regulation of CCR5 expression in human placenta: insights from a study of mother-to-child transmission of HIV in Malawi. *PloS one* **5**, e9212, doi:10.1371/journal.pone.0009212 (2010).

116 Spencer, K. L. *et al.* Dissection of chromosome 16p12 linkage peak suggests a possible role for CACNG3 variants in age-related macular degeneration susceptibility. *Investigative ophthalmology & visual science* **52**, 1748-1754, doi:10.1167/iovs.09-5112 (2011).

117 Wang, Q. *et al.* Genome-wide association analysis with gray matter volume as a quantitative phenotype in first-episode treatment-naive patients with schizophrenia. *PloS one* **8**, e75083, doi:10.1371/journal.pone.0075083 (2013).

118 Bork, K. *et al.* Novel hereditary angioedema linked with a heparan sulfate 3-O-sulfotransferase 6 gene mutation. *J Allergy Clin Immunol*, doi:10.1016/j.jaci.2021.01.011 (2021).

119 Santacroce, R., D'Andrea, G., Maffione, A. B., Margaglione, M. & d'Apolito, M. The Genetics of Hereditary Angioedema: A Review. *Journal of clinical medicine* **10**, doi:10.3390/jcm10092023 (2021).

120 Jiang, Z., Michal, J. J., Wu, X. L., Pan, Z. & MacNeil, M. D. The heparan and heparin metabolism pathway is involved in regulation of fatty acid composition. *International journal of biological sciences* **7**, 659-663, doi:10.7150/ijbs.7.659 (2011).

121 Pastor-Pareja, J. C. Atypical basement membranes and basement membrane diversity - what is normal anyway? *Journal of cell science* **133**, doi:10.1242/jcs.241794 (2020).

122 McCarthy, K. J. The Basement Membrane Proteoglycans Perlecan and Agrin: Something Old, Something New. *Current topics in membranes* **76**, 255-303, doi:10.1016/bs.ctm.2015.09.001 (2015).

123 Greenbaum, L., Alkelai, A., Zozulinsky, P., Kohn, Y. & Lerer, B. Support for association of HSPG2 with tardive dyskinesia in Caucasian populations. *The pharmacogenomics journal* **12**, 513-520, doi:10.1038/tpj.2011.32 (2012).

124 Li, L. *et al.* Genetic polymorphisms in collagen-related genes are associated with pelvic organ prolapse. *Menopause (New York, N.Y.)* **27**, 223-229, doi:10.1097/gme.0000000000001448 (2020).

125 Fico, A., Maina, F. & Dono, R. Fine-tuning of cell signaling by glypicans. *Cell Mol Life Sci* **68**, 923-929, doi:10.1007/s00018-007-7471-6 (2011).

126 Topaloudi, A. *et al.* Myasthenia gravis genome-wide association study implicates AGRN as a risk locus. *Journal of medical genetics*, doi:10.1136/jmedgenet-2021-107953 (2021).

127 Rygh, C. B. *et al.* Image-based assessment of microvascular function and structure in collagen XV- and XVIII-deficient mice. *The Journal of physiology* **592**, 325-336, doi:10.1113/jphysiol.2013.263574 (2014).

128 Sierksma, A. *et al.* Novel Alzheimer risk genes determine the microglia response to amyloid-β but not to TAU pathology. *EMBO molecular medicine* **12**, e10606, doi:10.15252/emmm.201910606 (2020).

129 Jen, Y. H., Musacchio, M. & Lander, A. D. Glypican-1 controls brain size through regulation of fibroblast growth factor signaling in early neurogenesis. *Neural development* **4**, 33, doi:10.1186/1749-8104-4-33 (2009).

130 Chung, H., Multhaupt, H. A., Oh, E. S. & Couchman, J. R. Minireview: Syndecans and their crucial roles during tissue regeneration. *FEBS letters* **590**, 2408-2417, doi:10.1002/1873-3468.12280 (2016).

131 Hong, H. *et al.* Up-regulation of syndecan-2 in proximal colon correlates with acute inflammation. *FASEB journal : official publication of the Federation of American Societies for Experimental Biology*, fj201900561R, doi:10.1096/fj.201900561R (2019).

132 Tsoyi, K. *et al.* Syndecan-2 Attenuates Radiation-induced Pulmonary Fibrosis and Inhibits Fibroblast Activation by Regulating PI3K/Akt/ROCK Pathway via CD148. *American journal of respiratory cell and molecular biology* **58**, 208-215, doi:10.1165/rcmb.2017-0088OC (2018).

133 Stacey, D., Ciobanu, L. G. & Baune, B. T. A systematic review on the association between inflammatory genes and cognitive decline in non-demented elderly individuals. *European*

neuropsychopharmacology : the journal of the European College of Neuropsychopharmacology **27**, 568-588, doi:10.1016/j.euroneuro.2015.12.017 (2017).

134 Echtermeyer, F. *et al.* Delayed wound repair and impaired angiogenesis in mice lacking syndecan-4. *The Journal of clinical investigation* **107**, R9-r14, doi:10.1172/jci10559 (2001).

135 Sivasami, P. *et al.* Adipose tissue loss and lipodystrophy in xylosyltransferase II deficient mice. *International journal of obesity (2005)* **43**, 1783-1794, doi:10.1038/s41366-019-0324-1 (2019).

136 Afratis, N. A. *et al.* Syndecans - key regulators of cell signaling and biological functions. *The FEBS journal* **284**, 27-41, doi:10.1111/febs.13940 (2017).

137 Agere, S. A., Kim, E. Y., Akhtar, N. & Ahmed, S. Syndecans in chronic inflammatory and autoimmune diseases: Pathological insights and therapeutic opportunities. *Journal of cellular physiology* **233**, 6346-6358, doi:10.1002/jcp.26388 (2018).

138 Gopal, S. Syndecans in Inflammation at a Glance. *Frontiers in immunology* **11**, 227, doi:10.3389/fimmu.2020.00227 (2020).

139 Stepp, M. A., Pal-Ghosh, S., Tadvalkar, G. & Pajoohesh-Ganji, A. Syndecan-1 and Its Expanding List of Contacts. *Advances in wound care* **4**, 235-249, doi:10.1089/wound.2014.0555 (2015).

140 Mizumoto, S., Yamada, S. & Sugahara, K. Human genetic disorders and knockout mice deficient in glycosaminoglycan. *BioMed research international* **2014**, 495764, doi:10.1155/2014/495764 (2014).

141 Hayes, A. J. & Melrose, J. Aggrecan, the Primary Weight-Bearing Cartilage Proteoglycan, Has Context-Dependent, Cell-Directive Properties in Embryonic Development and Neurogenesis: Aggrecan Glycan Side Chain Modifications Convey Interactive Biodiversity. *Biomolecules* **10**, 1244, doi:10.3390/biom10091244 (2020).

142 Lauing, K. L. *et al.* Aggrecan is required for growth plate cytoarchitecture and differentiation. *Developmental biology* **396**, 224-236, doi:10.1016/j.ydbio.2014.10.005 (2014).

143 Ponighaus, C. *et al.* Human xylosyltransferase II is involved in the biosynthesis of the uniform tetrasaccharide linkage region in chondroitin sulfate and heparan sulfate proteoglycans. *The Journal of biological chemistry* **282**, 5201-5206, doi:10.1074/jbc.M611665200 (2007).

144 Halper, J. Proteoglycans and diseases of soft tissues. *Advances in experimental medicine and biology* **802**, 49-58, doi:10.1007/978-94-007-7893-1_4 (2014).

145 Gulberti, S. *et al.* Phosphorylation and sulfation of oligosaccharide substrates critically influence the activity of human beta1,4-galactosyltransferase 7 (GalT-I) and beta1,3-glucuronosyltransferase I (GlcAT-I) involved in the biosynthesis of the glycosaminoglycan-protein linkage region of proteoglycans. *The Journal of biological chemistry* **280**, 1417-1425, doi:10.1074/jbc.M411552200 (2005).

146 Klein, J. A., Meng, L. & Zaia, J. Deep Sequencing of Complex Proteoglycans: A Novel Strategy for High Coverage and Site-specific Identification of Glycosaminoglycan-linked Peptides. *Mol Cell Proteomics* **17**, 1578-1590, doi:10.1074/mcp.RA118.000766 (2018).

147 Condac, E. *et al.* Polycystic disease caused by deficiency in xylosyltransferase 2, an initiating enzyme of glycosaminoglycan biosynthesis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 9416-9421, doi:10.1073/pnas.0700908104 (2007).

148 Tsutsui, Y., Ramakrishnan, B. & Qasba, P. K. Crystal structures of beta-1,4-galactosyltransferase 7 enzyme reveal conformational changes and substrate binding. *The Journal of biological chemistry* **288**, 31963-31970, doi:10.1074/jbc.M113.509984 (2013).

149 Liu, X. *et al.* Inactivation of Fam20b in the neural crest-derived mesenchyme of mouse causes multiple craniofacial defects. *European journal of oral sciences* **126**, 433-436, doi:10.1111/eos.12563 (2018).

150 Vogel, P. *et al.* Amelogenesis imperfecta and other biomineralization defects in Fam20a and Fam20c null mice. *Veterinary pathology* **49**, 998-1017, doi:10.1177/0300985812453177 (2012).

151 Huber, C. *et al.* Identification of CANT1 mutations in Desbuquois dysplasia. *American journal of human genetics* **85**, 706-710, doi:10.1016/j.ajhg.2009.10.001 (2009).

152 Liu, J., Xu, W., Li, S., Sun, R. & Cheng, W. Multi-omics analysis of tumor mutational burden combined with prognostic assessment in epithelial ovarian cancer based on TCGA database. *International journal of medical sciences* **17**, 3200-3213, doi:10.7150/ijms.50491 (2020).

153 Izumikawa, T. *et al.* GlcUAbeta1-3Galbeta1-3Galbeta1-4Xyl(2-O-phosphate) is the preferred substrate for chondroitin N-acetylgalactosaminyltransferase-1. *The Journal of biological chemistry* **290**, 5438-5448, doi:10.1074/jbc.M114.603266 (2015).

154 Orellana, A., Moraga, C., Araya, M. & Moreno, A. Overview of Nucleotide Sugar Transporter Gene Family Functions Across Multiple Species. *Journal of molecular biology* **428**, 3150-3165, doi:10.1016/j.jmb.2016.05.021 (2016).

155 Parker, J. L. & Newstead, S. Gateway to the Golgi: molecular mechanisms of nucleotide sugar transporters. *Current opinion in structural biology* **57**, 127-134, doi:10.1016/j.sbi.2019.03.019 (2019).

156 Anthony, S., Munk, R., Skakun, W. & Masini, M. Multiple epiphyseal dysplasia. *The Journal of the American Academy of Orthopaedic Surgeons* **23**, 164-172, doi:10.5435/jaaos-d-13-00173 (2015).

157 Winawer, M. R. *et al.* Somatic SLC35A2 variants in the brain are associated with intractable neocortical epilepsy. *Annals of neurology* **83**, 1133-1146, doi:10.1002/ana.25243 (2018).

158 Pacifici, M. The pathogenic roles of heparan sulfate deficiency in hereditary multiple exostoses. *Matrix Biol* **71-72**, 28-39, doi:10.1016/j.matbio.2017.12.011 (2018).

159 Debarnot, C. *et al.* Substrate binding mode and catalytic mechanism of human heparan sulfate d-glucuronyl C5 epimerase. *Proceedings of the National Academy of Sciences of the United States of America* **116**, 6760-6765, doi:10.1073/pnas.1818333116 (2019).

160 Dou, W., Xu, Y., Pagadala, V., Pedersen, L. C. & Liu, J. Role of Deacetylase Activity of N-Deacetylase/N-Sulfotransferase 1 in Forming N-Sulfated Domain in Heparan Sulfate. *The Journal of biological chemistry* **290**, 20427-20437, doi:10.1074/jbc.M115.664409 (2015).

161 Habuchi, H. *et al.* The occurrence of three isoforms of heparan sulfate 6-O-sulfotransferase having different specificities for hexuronic acid adjacent to the targeted N-sulfoglucosamine. *The Journal of biological chemistry* **275**, 2859-2868, doi:10.1074/jbc.275.4.2859 (2000).

162 Jemth, P. *et al.* Oligosaccharide library-based assessment of heparan sulfate 6-O-sulfotransferase substrate specificity. *The Journal of biological chemistry* **278**, 24371-24376, doi:10.1074/jbc.M212155200 (2003).

163 Sheng, J., Liu, R., Xu, Y. & Liu, J. The dominating role of N-deacetylase/N-sulfotransferase 1 in forming domain structures in heparan sulfate. *The Journal of biological chemistry* **286**, 19768-19776, doi:10.1074/jbc.M111.224311 (2011).

164 Kjellen, L. Glucosaminyl N-deacetylase/N-sulphotransferases in heparan sulphate biosynthesis and biology. *Biochemical Society transactions* **31**, 340-342, doi:10.1042/bst0310340 (2003).

165 Billings, P. C. & Pacifici, M. Interactions of signaling proteins, growth factors and other proteins with heparan sulfate: mechanisms and mysteries. *Connective tissue research* **56**, 272-280, doi:10.3109/03008207.2015.1045066 (2015).

166 Reuter, M. S. *et al.* NDST1 missense mutations in autosomal recessive intellectual disability. *American journal of medical genetics. Part A* **164a**, 2753-2763, doi:10.1002/ajmg.a.36723 (2014).

167 Takahashi, I. *et al.* Important role of heparan sulfate in postnatal islet growth and insulin secretion. *Biochem Biophys Res Commun* **383**, 113-118, doi:10.1016/j.bbrc.2009.03.140 (2009).

168 Kunnas, T., Solakivi, T., Määttä, K. & Nikkari, S. T. Glucuronic Acid Epimerase (GLCE) Variant rs3865014 (A>G) Is Associated with BMI, Blood Hemoglobin, Hypertension, and Cerebrovascular Events, the TAMRISK Study. *Annals of human genetics* **80**, 332-335, doi:10.1111/ahg.12166 (2016).

169 Lee, H. *et al.* Whole Genome Analysis of the Red-Crowned Crane Provides Insight into Avian Longevity. *Molecules and cells* **43**, 86-95, doi:10.14348/molcells.2019.0190 (2020).

170 Borjigin, J. *et al.* Diurnal pineal 3-O-sulphotransferase 2 expression controlled by beta-adrenergic repression. *The Journal of biological chemistry* **278**, 16315-16319, doi:10.1074/jbc.M300828200 (2003).

171 Vanpouille, C. *et al.* The heparin/heparan sulfate sequence that interacts with cyclophilin B contains a 3-O-sulfated N-unsubstituted glucosamine residue. *The Journal of biological chemistry* **282**, 24416-24429, doi:M701835200 [pii]

10.1074/jbc.M701835200 (2007).

172 Rönnberg, E., Melo, F. R. & Pejler, G. Mast cell proteoglycans. *The journal of histochemistry and cytochemistry : official journal of the Histochemistry Society* **60**, 950-962, doi:10.1369/0022155412458927 (2012).

173 Nguyen, N. T. *et al.* Genetic and enzymatic characterization of 3-O-sulfotransferase SNPs associated with Plasmodium falciparum parasitaemia. *Glycobiology* **28**, 534-541, doi:10.1093/glycob/cwy038 (2018).

174 Osorio, S., Rivillas, N. D. & Martinez, J. A. Urofacial (ochoa) syndrome: A literature review. *Journal of pediatric urology* **17**, 246-254, doi:10.1016/j.jpurol.2021.01.017 (2021).

175 Woolf, A. S., Stuart, H. M. & Newman, W. G. Genetics of human congenital urinary bladder disease. *Pediatric nephrology (Berlin, Germany)* **29**, 353-360, doi:10.1007/s00467-013-2472-1 (2014).

176 Chen, H., Wang, T., Yang, J., Huang, S. & Zeng, P. Improved Detection of Potentially Pleiotropic Genes in Coronary Artery Disease and Chronic Kidney Disease Using GWAS Summary Statistics. *Frontiers in genetics* **11**, 592461, doi:10.3389/fgene.2020.592461 (2020).

177 Deligny, A. *et al.* NDST2 (N-Deacetylase/N-Sulfotransferase-2) Enzyme Regulates Heparan Sulfate Chain Length. *The Journal of biological chemistry* **291**, 18600-18607, doi:10.1074/jbc.M116.744433 (2016).

178 Huang, Y. C. *et al.* Genome-wide association study of diabetic retinopathy in a Taiwanese population. *Ophthalmology* **118**, 642-648, doi:10.1016/j.ophtha.2010.07.020 (2011).

179 Espinosa, A. *et al.* Exploring Genetic Associations of Alzheimer's Disease Loci With Mild Cognitive Impairment Neurocognitive Endophenotypes. *Frontiers in aging neuroscience* **10**, 340, doi:10.3389/fnagi.2018.00340 (2018).

180 Shworak, N. W., Kobayashi, T., de Agostini, A. & Smits, N. C. Anticoagulant heparan sulfate to not clot--or not? *Progress in molecular biology and translational science* **93**, 153-178, doi:10.1016/s1877-1173(10)93008-1 (2010).

181 Sepulveda-Diaz, J. E. *et al.* HS3ST2 expression is critical for the abnormal phosphorylation of tau in Alzheimer's disease-related tau pathology. *Brain : a journal of neurology* **138**, 1339-1354, doi:10.1093/brain/awv056 (2015).

182 Atkinson, A., Garnier, S., Afridi, S., Fumoux, F. & Rihet, P. Genetic variations in genes involved in heparan sulphate biosynthesis are associated with Plasmodium falciparum parasitaemia: a familial study in Burkina Faso. *Malaria journal* **11**, 108, doi:10.1186/1475-2875-11-108 (2012).

183 Mackinnon, M. J. *et al.* Environmental Correlation Analysis for Genes Associated with Protection against Malaria. *Molecular biology and evolution* **33**, 1188-1204, doi:10.1093/molbev/msw004 (2016).

184 Howard, S. R. *et al.* HS6ST1 Insufficiency Causes Self-Limited Delayed Puberty in Contrast With Other GnRH Deficiency Genes. *The Journal of clinical endocrinology and metabolism* **103**, 3420-3429, doi:10.1210/jc.2018-00646 (2018).

185 Hayashi, S. *et al.* Comprehensive investigation of CASK mutations and other genetic etiologies in 41 patients with intellectual disability and microcephaly with pontine and cerebellar hypoplasia (MICPCH). *PloS one* **12**, e0181791, doi:10.1371/journal.pone.0181791 (2017).

186 Poon, I. K. *et al.* Mice deficient in heparanase exhibit impaired dendritic cell migration and reduced airway inflammation. *European journal of immunology* **44**, 1016-1030, doi:10.1002/eji.201343645 (2014).

187 Holst, C. R. *et al.* Secreted sulfatases Sulf1 and Sulf2 have overlapping yet essential roles in mouse neonatal survival. *PloS one* **2**, e575, doi:10.1371/journal.pone.0000575 (2007).

188 Hassing, H. C. *et al.* SULF2 strongly prediposes to fasting and postprandial triglycerides in patients with obesity and type 2 diabetes mellitus. *Obesity (Silver Spring, Md.)* **22**, 1309-1316, doi:10.1002/oby.20682 (2014).

189 Matikainen, N. *et al.* Genetic variation in SULF2 is associated with postprandial clearance of triglyceride-rich remnant particles and triglyceride levels in healthy subjects. *PloS one* **8**, e79473, doi:10.1371/journal.pone.0079473 (2013).

190 Duchez, S. *et al.* Glycotranscriptome study reveals an enzymatic switch modulating glycosaminoglycan synthesis during B-cell development and activation. *European journal of immunology* **41**, 3632-3644, doi:10.1002/eji.201140865 (2011).

191 Mizumoto, S., Yamada, S. & Sugahara, K. Mutations in Biosynthetic Enzymes for the Protein Linker Region of Chondroitin/Dermatan/Heparan Sulfate Cause Skeletal and Skin Dysplasias. *BioMed research international* **2015**, 861752, doi:10.1155/2015/861752 (2015).

192 Esko, J. D. & Selleck, S. B. Order out of chaos: assembly of ligand binding sites in heparan sulfate. *Annual review of biochemistry* **71**, 435-471, doi:10.1146/annurev.biochem.71.110601.135458 (2002).

193 Gesslbauer, B., Rek, A., Falsone, F., Rajkovic, E. & Kungl, A. J. Proteoglycanomics: tools to unravel the biological function of glycosaminoglycans. *Proteomics* **7**, 2870-2880, doi:10.1002/pmic.200700176 (2007).

194 Aikawa, J., Grobe, K., Tsujimoto, M. & Esko, J. D. Multiple isozymes of heparan sulfate/heparin GlcNAc N-deacetylase/GlcN N-sulfotransferase. Structure and activity of the fourth member, NDST4. *The Journal of biological chemistry* **276**, 5876-5882, doi:10.1074/jbc.M009606200 (2001).

195 Habuchi, H. *et al.* Biosynthesis of heparan sulphate with diverse structures and functions: two alternatively spliced forms of human heparan sulphate 6-O-sulphotransferase-2 having different

expression patterns and properties. *The Biochemical journal* **371**, 131-142, doi:10.1042/bj20021259 (2003).

196 Lawrence, R. *et al.* The principal neuronal gD-type 3-O-sulfotransferases and their products in central and peripheral nervous system tissues. *Matrix Biol* **26**, 442-455, doi:10.1016/j.matbio.2007.03.002 (2007).

197 Li, Y. J. *et al.* Characterization of heparan sulfate N-deacetylase/N-sulfotransferase isoform 4 using synthetic oligosaccharide substrates. *Biochimica et biophysica acta. General subjects* **1862**, 547-556, doi:10.1016/j.bbagen.2017.11.016 (2018).

198 Mochizuki, H. *et al.* Characterization of a heparan sulfate 3-O-sulfotransferase-5, an enzyme synthesizing a tetrasulfated disaccharide. *The Journal of biological chemistry* **278**, 26780-26787, doi:10.1074/jbc.M301861200 (2003).

199 Sheng, J., Xu, Y., Dulaney, S. B., Huang, X. & Liu, J. Uncovering biphasic catalytic mode of C5-epimerase in heparan sulfate biosynthesis. *The Journal of biological chemistry* **287**, 20996-21002, doi:10.1074/jbc.M112.359885 (2012).

200 Rudd, T. R. & Yates, E. A. A highly efficient tree structure for the biosynthesis of heparan sulfate accounts for the commonly observed disaccharides and suggests a mechanism for domain synthesis. *Molecular bioSystems* **8**, 1499-1506, doi:10.1039/c2mb25019e (2012).