Review

Vertebrate scavenger receptor class B member 2 (SCARB2): comparative studies of a major lysosomal membrane glycoprotein

Roger S Holmes

School of Biomolecular and Physical Sciences, Griffith University, Nathan, QLD, Australia

Received on April 5, 2012; Accepted on May 24, 2012; Published on June 16, 2012

Correspondence should be addressed to Roger S Holmes; Phone: +61 7 37355077, Fax: +61 7 37357773, E-mail: r.holmes@griffith.edu.au

Abstract

Scavenger receptor class B member 2 (SCARB2) (also called LIMP-2, CD36L2 or LGP85) is a major lysosomal membrane glycoprotein involved in endosomal and lysosomal biogenesis and maintenance. SCARB2 acts as a receptor for the lysosomal mannose -6-phosphate independent targeting of β-glucuronidase and enterovirus 71 and influences Parkinson's disease and epilepsy. Genetic deficiency of this protein causes deafness and peripheral neuropathy in mice as well as myoclonic epilepsy and nephrotic syndrome in humans. Comparative SCARB2 amino acid sequences and structures and *SCARB2* gene locations were examined using data from several vertebrate genome projects. Vertebrate SCARB2 sequences shared 43-100% identity as compared with 30-36% sequence identities with other CD36-like superfamily members, SCARB1 and CD36. At least 10 N-glycosylation sites were conserved among most vertebrate SCARB2 proteins examined. Sequence alignments, key amino acid residues and conserved predicted secondary structures were examined, including cytoplasmic, transmembrane and external lysosomal membrane sequences: cysteine disulfide residues, thrombospondin (THP1) binding sites and 16 proline and 20 glycine conserved residues, which may contribute to short loop formation within the exomembrane SCARB2 sequences. Vertebrate *SCARB2* genes contained 12 coding exons. The human *SCARB2* gene contained a CpG island (CpG100), ten microRNA-binding sites and several transcription factor binding sites (including PPARA) which may contribute to a higher level (2.4 times average) of gene expression. Phylogenetic analyses examined the relationships and potential evolutionary origins of the vertebrate *SCARB2* gene with vertebrate *SCARB1* and *CD36* genes. These suggested that *SCARB2* originated from duplications of the *CD36* gene in an ancestral genome forming three vertebrate *CD36* gene family members: *SCARB1*, *SCARB2* and *CD36*.

Introduction

Scavenger receptor class B member 2 (SCARB2) [also called SRB2, lysosomal membrane glycoprotein 2 (LIMP-2), cluster of differentiation 36 like-2 (CD36L2) or lysosomal glycoprotein 85 (LGP85)] is one of at least three members of the collagen type 1 receptor (thrombospondin) CD36-like family that is an integrated lysosomal and endosomal membrane protein of many tissues and cells of the body (Calvo *et al*. 1995, Fujita *et al*. 1992, Ogata *et al*. 1994, Tabuchi *et al*. 1997). SCARB2 plays a major role in lysosomal and endosomal membrane organization (Gamp *et al*. 2003, Kuronita *et al*. 2002), cytosolic protein turnover (Roszek & Gniot-Szulzycka 2005), phagosomal trafficking and macrophage activation (Carrasco-Marin *et al*. 2011). SCARB2 also serves as a lysosomal membrane surface receptor of thrombospondins and other lipids (Tserentsoodol *et al.* 2006), mannose-6-phosphate independent targeting of β-glucuronidase (Blanz *et al*. 2010, Reczek *et al*. 2007, Sleat *et al*. 2006, Velayti *et al*. 2011) and enterovirus 71 (involved in the pathogenesis of hand, foot and mouth disease) (HFMD) (Lin *et al*. 2012, Yamayoshi *et al*. 2009, Yamayoshi & Koike 2011). In addition to HFMD, SCARB2 has been implicated in diseases such as cardiac myocyte hypertrophy (Schroen *et al*. 2007) and progressive myoclonus epilepsy and renal failure syndrome (Berkovic *et al*. 2008, Chaves *et al*. 2011, Hopfner *et al*. 2011, Rubboli *et al*. 2011) and has been

recognized as a candidate gene for autism (Ilu *et al*. 2010) and Parkinson's disease (Do *et al*. 2011, Michelakakis *et al*. 2012).

SCARB1 (also called CLA1, SRB1 and CD36L1) is a second member of the CD36 family that serves as a homo-oligomeric plasma membrane cell surface glycoprotein receptor for high density lipoprotein cholesterol (HDL), other phospholipid ligands and chylomicron remnants (Acton *et al*. 1996, Bultel-Brienne *et al*. 2002, Connelly *et al*. 2004, Holmes and Cox 2012, Kent & Stylianou 2011, Marsche *et al*. 2003). A third member of the CD36 family, CD36 (also called SCARB3, fatty acyltranslocase [FAT] and glycoprotein 88 [GP88]) is an integral membrane protein of many tissues of the body which plays a role in fatty acyl translocation and as a multiple ligand cell surface receptor of oxidized low density lipoprotein cholesterol (LDL) (Martin *et al*. 2007, Tandon *et al*. 1989), and has been implicated in several diseases including insulin resistance, diabetes, atherosclerosis and malaria (Adachi & Tsujimoto 2006, Collot-Teixeira *et al*. 2007, Gautum & Banerjee 2011, Martin *et al*. 2007, Ren 2012, Simantov & Silverstein 2003).

The gene encoding human SCARB2 (*SCARB2*) is on chromosome 4, encoded by 12 coding exons (Calvo *et al*. 1995) and localized between the genes encoding nucleoporin (*NUP54*) and protein family 47 (*FAM47E*) (Kent *et al*. 2003). In addition, *SCARB2* has been assigned to chromosome 8 in pigs (Kim *et al*. 2006) and to chromosome 5 in mice, but designated as *Scarb2* in the latter genome (Tabuchi *et al*. 1997). *SCARB2* is ubiquitously expressed in various cells and tissues of the body, including kidney glomerular tubules (Berkovic *et al*. 2008, Desmond *et al.* 2011), liver (Tabuchi *et al*. 1997, Zhang *et al*. 2007), retinal ganglia and photoreceptor outer segments (Tserentsoodol *et al*. 2006), ureter epithelial cells (Gamp *et al.* 2003), metastatic pancreas islet cells (Fujita *et al*. 1992), the cardiac intercalated disc (Schroen *et al*. 2007) and phagocytes (Carrasco-Marin *et al*. 2011). Studies of *Scarb2*$^{-}$/*Scarb2*$^{-}$ knock out mice have shown that SCARB2-deficiency causes ureteric pelvic junction obstruction, deafness and peripheral neuropathy (Gamp *et al*. 2003), renal tubular proteinuria (Desmond *et al*. 2011), an inability to mount a hypertrophic response to increased blood pressure, due to the absence of SCARB2 in the cardiac intercalated disc (Schroen *et al*. 2007) and macrophage -related defects in immunity to infection (Carrasco-Marin *et al*. 2011). Human clinical studies have also examined *SCARB2* polymorphisms associated with diseases causing mutations of β-glucocerebrosidase binding, which is defective in Gaucher disease (Blanz *et al*. 2010, Velayati *et al*. 2011); and collapsing focal and segmental glomerular sclerosis (FSGS) and myoclonic epilepsy (Balreira *et al*. 2008, Berkovic *et al*. 2008, Chaves *et al*. 2011, Desmond *et al*. 2011, Dibbens *et al*. 2011, Hopfner *et al*. 2011, Rubboli *et al*. 2011).

This paper reports the predicted gene structures and amino acid sequences for several vertebrate *SCARB2* genes and proteins, the secondary structures for vertebrate SCARB2 proteins, several potential sites for regulating human *SCARB2* gene expression and the structural, phylogenetic and evolutionary relationships for these genes and enzymes with those for vertebrate *SCARB2*, *SCARB1* and *CD36* gene families.

Materials and Methods

Vertebrate *SCARB2* gene and protein identification

BLAST (Basic Local Alignment Search Tool) studies were undertaken using web tools from the National Center for Biotechnology Information (NCBI) (http://blast.ncbi.nlm.nih.gov/Blast.cgi) (Altschul *et al*. 1997). Protein BLAST analyses used human and mouse SCARB2 amino acid sequences previously described (Fujita *et al*. 1992, Ogata *et al*. 1994, Calvo *et al*. 1995, Tabuchi *et al*. 1997) (Table 1). Non-redundant protein sequence databases for several vertebrate genomes were examined using the blastp algorithm from sources previously described (Holmes 2012). This procedure produced multiple BLAST 'hits' for each of the protein databases which were individually examined and retained in FASTA format, and a record kept of the sequences for predicted mRNAs and encoded SCARB2-like proteins. Predicted SCARB2-like protein sequences were obtained in each case and subjected to analyses of predicted protein and gene structures.

BLAT (Blast-like Alignment Tool) analyses were subsequently undertaken for each of the predicted SCARB2 amino acid sequences using the UC Santa Cruz Genome Browser (Kent *et al*. 2003) with the default settings to obtain the predicted locations for each of the vertebrate *SCARB2* genes, including predicted exon boundary locations and gene sizes. BLAT analyses were similarly undertaken for vertebrate *SCARB1* and *CD36* genes using previously reported sequences in each case (see Table 1; Holmes & Cox 2012). Structures for human *SCARB2*, mouse *Scarb2* and rat *Scarb2* transcripts were obtained using the AceView website to examine predicted gene and protein structures (Thierry-Mieg & Thierry-Mieg 2006). Predictions for human, mouse and rat *SCARB2* CpG islands, miRNA binding sites and transcription factor binding sites were obtained using the UC Santa Cruz Genome Browser (Kent *et al*. 2003).

Table 1 (continued on page 4). Vertebrate SCARB2, SCARB1 and CD36 Genes and Proteins. RefSeq: the reference amino acid sequence; [1]predicted Ensembl amino acid sequence; na-not available; GenBank IDs are derived NCBI http://www.ncbi.nlm.nih.gov/genbank/; Ensembl ID was derived from Ensembl genome database http://www.ensembl.org ; UNIPROT refers to UniprotKB/Swiss-Prot IDs for individual CD36-like proteins (see http://kr.expasy.org); Un-refers to unknown chromosome; bps refers to base pairs of nucleotide sequences; the number of coding exons are listed; gene expression levels are in bold; % identities are shown in bold.

SCARB2 Gene	Species	RefSeq ID [1]Ensembl/NCBI	GenBank ID	UNIPROT ID	Amino acids	Chromosome location	Coding Exons (strand)	Gene Size bps	Subunit MW	Gene Expression Level	% Identity with human SCARB1	% Identity with human SCARB2	% Identity with human CD36
Human	*Homo sapiens*	NM_005506	BT006939	Q53Y63	478	4:77,084,378-77,134,696	12 (-ve)	50,316	54,290	**3.2**	34	**100**	35
Chimpanzee	*Pantroglodytes*	[1]XP_517214.2	na	na	478	4:77,134,567-77,135,377	12 (+ve)	50,633	54,309	na	34	**100**	35
Orangutan	*Pongo abelii*	[1]XP_002814942.1	na	na	478	4:79,503,577-79,555,051	12 (-ve)	51,475	54,331	na	33	**99**	34
Rhesus	*Macaca mulatta*	[1]XP_001096458.1	na	na	478	5:53,378,629-54,329,383	12 (+ve)	50,755	54,281	na	33	**99**	35
Marmoset	*Callithrix jacchus*	[1]XP_002745738.1	na	na	478	3:118,471,528-118,533,432	12 (+ve)	61,905	54,215	na	32	**97**	34
Mouse	*Mus musculus*	NM_007644	BC029073	O35114	478	5:92,875,330-92,934,334	12 (-ve)	59,005	54,044	**3.6**	32	**85**	35
Rat	*Rattus norvegicus*	NM_054001	BC061853	P27615	478	14:17,119,648-17,168,942	12 (+ve)	51,485	54,091	**1.0**	32	**86**	36
Panda	*Ailuropoda melanoleuca*	[1]XP_002924780.1	na	na	478	*GL193185.1:272,540-348,835	12 (-ve)	76,296	54,059	na	32	**87**	34
Cow	*Bos taurus*	NM_001102153.1	BC149935	A6QQP4	478	6:92,797,818-92,879,565	12 (-ve)	81,748	53,985	na	33	**88**	34
Dog	*Canis familaris*	[1]XP_535612.4	na	F1PGJ5	478	32:3,756,732-3,808,324	12 (-ve)	51,593	54,015	na	31	**86**	34
Pig	*Sus scrofa*	NM_001244155.1	na	F1RYT3	478	8:61,268,560-61,335,361	12 (-ve)	66,802	54,015	na	34	**88**	35
Rabbit	*Oryctolagus cuniculus*	[1]XP_002717150.1	na	G1SJL6	478	15:74169938-74233068	12 (+ve)	63,131	53,588	na	34	**90**	33
Elephant	*Loxodonta africana*	[1]XP_003414189.1	na	G3T2P7	478	30:21,030,405-21,106,415	12 (+ve)	76,011	53,981	na	33	**87**	35
Chicken	*Gallus gallus*	[1]XP_42093.1	BX931548	na	481	4:51,411,268-51,429,620	12 (+ve)	18,353	53,907	na	30	**59**	33
Lizard	*Anolis carolinensis*	[1]XP_003221897.1	na	G1KMI7	482	5:146,878,187-146,901,887	12 (+ve)	23,701	53,969	na	33	**53**	33
Frog	*Xenopus tropicalis*	NM_001016577.2	BC171025	B7ZTX2	483	GL173943:30,875-63,699	12 (-ve)	32,825	54,586	na	31	**54**	35
Zebrafish	*Danio rerio*	NM_173259.1	BC162407	Q8JQR8	474	5: 63,942,096-63,955,449	13 (+ve)	13,354	60,234	na	31	**43**	33

Table 1 (continued from previous page).

SCARB1 Gene	Species	RefSeq ID [1]Ensembl/NCBI	GenBank ID	UNIPROT ID	Amino acids	Chromosome location	Coding Exons (strand)	Gene Size bps	Subunit MW	Gene Expression Level	% Identity with human SCARB1	% Identity with human SCARB2	% Identity with human CD36
Human	*Homo sapiens*	NM_00505	BC022087	Q8WVT0	509	12:125,267,232-125,348,266	12 (-ve)	81,035	56,973	**13.7**	**100**	29	31
Mouse	*Mus musculus*	NM_001205082.1	BC004656	Q61009	509	5:125,761,478-125,821,252	12 (-ve)	63,985	56,754	**5.1**	**79**	29	29
Chicken	*Gallus gallus*	[1]XP_415106	na	na	503	15:4,543,054-4,558,954	12 (+ve)	15,901	55,918	na	**57**	28	31
Zebrafish	*Danio rerio*	NM_198121	BC044516	E7FB50	496	11:21,526,513-21,572,478	12 (-ve)	45,684	55,742	na	**51**	28	30
CD36 Gene													
Human	*Homo sapiens*	NM_001001547	BC008406	P16671	472	7:80,275,645-80,303,732	12 (+ve)	72,231	53,053	**6.6**	31	30	**100**
Mouse	*Mus musculus*	NM_001159555.1	BC010262	Q08857	472	5:17,291,543-17,334,712	12 (-ve)	43,170	52,698	**4.2**	30	31	**83**
Chicken	*Gallus gallus*	[1]ENSGALG8439	AJ719746	F1NER9	471	1:12,077,308-12,107,415	12 (-ve)	30,108	52,624	na	30	32	**61**
Zebrafish	*Danio rerio*	NP_001002363.1	BC076048	Q6DHC7	465	4:21,594,449-21,606,961	12 (-ve)	12,513	51,590	na	31	31	**53**
Lancelet	*Branchiostoma floridae*	[1]XP_002609178.1	na	na	480	Un:534,334,234-534,343,082	12 (+ve)	8,849	54,141	na	34	35	35
Sea squirt	*Ciona intestinalis*	[1]XP_002127015.1	na	na	523	09p:2,872,362-2,873,903	1 (-ve)	1,542	58,009	na	26	33	31
Nematode	*Caenorhabditis elegans*	NM_067224	na	Q9XTT3	534	III:12,453,609-12,456,726	8 (+ve)	3,118	60,182	**4.6**	21	26	24
Fruit fly	*Drosophila melanogaster*	NP_523859	na	na	520	2R:20,864,606-20,867,116	6 (-ve)	#2,511	58,663	na	20	23	26

Predicted Structures and Properties of Vertebrate SCARB2 and other CD36-like Proteins

Predicted secondary structures for vertebrate SCARB2 proteins, human SCARB1 and CD36, lancelet (*Branchiostoma floridae*) CD36, sea squirt (*Ciona intestinalis*) CD36 and a fruit fly (*Drosophila melanogaster*) epithelial membrane protein (FBpp0072309) were obtained using the PSIPRED v2.5 web site tools provided by Brunel University (McGuffin *et al*. 2000). Molecular weights, N-glycosylation sites (Gupta & Brunak 2002) and predicted transmembrane, cytosolic and exocellular sequences for vertebrate SCARB2 proteins were obtained using Expasy web tools (http://au.expasy.org/tools/pi_tool.html).

Comparative Human *SCARB2* and Mouse *Scarb2* Gene Expression

The genome browser (http://genome.ucsc.edu) (Kent *et al*. 2003) was used to examine GNF Expression Atlas 2 data using various expression chips for human *SCARB2* and mouse *Scarb2* genes, respectively (Su *et al*. 2004) (http://biogps.gnf.org). Gene array expression 'heat maps' were examined for comparative gene expression levels among human and mouse tissues showing high (red), intermediate (black), and low (green) expression levels.

Phylogeny Studies and Sequence Divergence

Alignments of vertebrate SCARB2, SCARB1 and CD36 sequences were assembled using BioEdit v.5.0.1 using the default settings (Hall 1999). Alignment of ambiguous regions, including the amino and carboxyl termini, were excluded prior to phylogenetic analysis, yielding alignments of 431 residues for comparisons of vertebrate SCARB2 sequences with human, mouse, chicken and zebra-fish SCARB1 and CD36 sequences with the lancelet (*Branchiostoma floridae*) CD36 sequence (Table 1). Evolutionary distances and phylogenetic trees were calculated as previously described (Holmes 2012). Tree topology was reexamined by the boot-strap method (100 bootstraps were applied) of resampling and only values that were highly significant (≥95) are shown (Felsenstein 1985).

Results and Discussion

Alignments of Vertebrate SCARB2 Amino Acid Sequences

The deduced amino acid sequences for cow (Co) (*Bos taurus*), chicken (Ch) (*Gallus gallus*), lizard (Li) (*Anolis carolensis*), frog (Fr) (*Xenopus tropicalis*) and

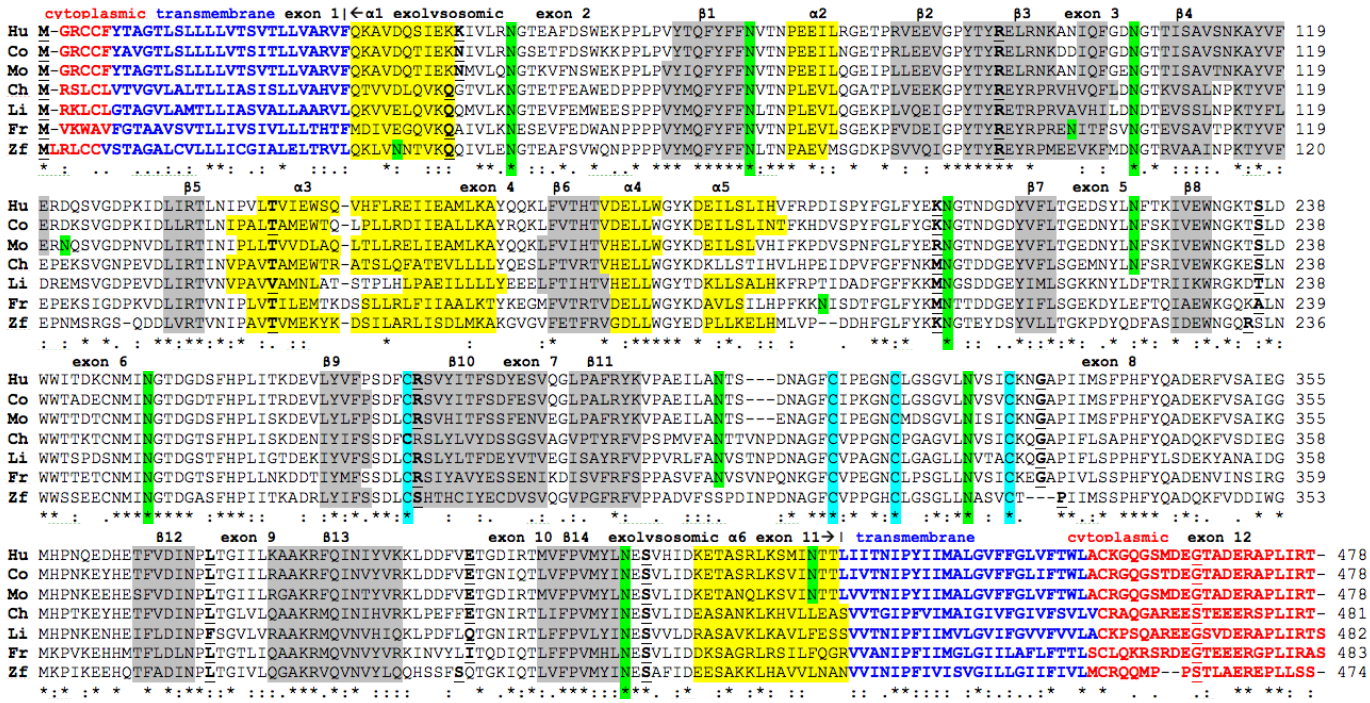

Figure 1. Amino Acid Sequence Alignments for Vertebrate SCARB2 Sequences. See Table 1 for sources of SCARB2 sequences; * shows identical residues for SCARB2 subunits; : similar alternate residues; . dissimilar alternate residues; predicted cytoplasmic residues are shown in red; predicted transmembrane residues are shown in blue; N-glycosylated and potential N-glycosylated Asn sites are highlighted in green; predicted disulfide bond Cys residues are highlighted in cyan blue; predicted α-helices for vertebrate SCARB2 are highlighted in yellow and numbered in sequence from the start of the predicted exoplasmic domain; predicted β-sheets are highlighted in grey and also numbered in sequence; bold underlined font shows residues corresponding to known or predicted exon start sites; exon numbers refer to human *SCARB2* gene exons; exolysosomic refers to the predicted SCARB2 sequence external to the lysosomal membrane.

zebrafish (Zf) (*Danio rerio*) SCARB2 are shown in Figure 1 together with the previously reported sequences for human (Hu) and mouse (Mo) SCARB2 (Table 1) (Fujita et al. 1992, Tabuchi et al. 1997). Alignments of human with other vertebrate SCARB2 sequences examined were between 43-100% identical, suggesting that these are products of the same family of genes, whereas comparisons of sequence identities of vertebrate SCARB2 proteins with human SCARB1 and CD36 proteins exhibited lower levels of sequence identities (30-36%), indicating that these are members of distinct *CD36*-like gene families (Table 1). The amino acid sequences for mammalian SCARB2 contained 478 residues while chicken (*Gallus gallus*), lizard (*Anolis carolensis*), frog (*Xenopus tropicalis*) and zebrafish (*Danio rerio*) SCARB2 contained 481, 482, 483 and 474 amino acids, respectively (Table 1; Figure 1).

Previous studies have reported several key regions and residues for human and mouse SCARB2 proteins (human SCARB2 amino acid residues were identified in each case). These included cytoplasmic N -terminal and C-terminal residues, residues 2-6 and 457-478, and N-terminal and C-terminal transmembrane helical regions: residues 7-30 and 432-456 (Fujita *et al*. 1992, Tabuchi *et al*. 1997). These motifs underwent significant changes in amino acid sequence but retained the predicted cytoplasmic and transmembrane properties in each case (Figure 1). These changes are in contrast to the N-terminal transmembrane sequences for vertebrate SCARB1 and CD36 sequences, for which several glycine residues were predominantly conserved, especially for CD36 Gly12, Gly16 and Gly24/Gly25 residues and for SCARB1 key N-terminal glycine residues (Gly15/Gly18/Gly25) (Holmes & Cox 2012), which form a dimerization motif in the N-terminal transmembrane domain and participate in forming SCARB1 oligomers (Gaidukov *et al.* 2011). There were no SCARB2 N-terminal transmembrane glycine residues conserved for the vertebrate sequences examined, although Gly10 was predominantly conserved with a Gly/Ala substitution observed for the frog SCARB2 sequence (Figure 1). A conserved glycine residue was observed for the vertebrate C-terminal transmembrane sequences (human SCARB2 Gly449) (Figure 1); however the role of this residue has not been investigated.

Comparative Sequences for Vertebrate SCARB2 N-Glycosylation Sites

Ten N-glycosylation sites for human SCARB2 have been previously identified for this protein (Figure 1; Table 2) (Lewandrowski *et al*. 2005), whereas eleven such sites have been reported for mouse SCARB2 (Tabuchi *et al*. 1997). All of these sites were predominantly retained among the 16 vertebrate SCARB2 sequences examined. Given the sequence conservation observed for these residues among the vertebrate SCARB2 sequences examined, it is apparent that they are essential for the function of vertebrate SCARB2 as a glycoprotein. The multiple N-glycosylation sites observed for vertebrate SCARB2 sequences were consistent with a major role for N-proteoglycan residues exposed on the external membrane surface of lysosomes in the performance of SCARB2 functions in binding various lipid molecules, and in their reported functions in maintaining the organization of lysosomal and endosomal membranes (Gamp *et al*. 2003; Kuronita *et al*. 2002). This is also supported by recent animal model studies, which demonstrated a key role for N-glycosylation in the recruitment of a related integrated membrane CD36-like family member (CD36) into cardiac membranes (Lauzier *et al*. 2011).

Vertebrate SCARB2 Cysteine Residues

Four conserved external lysosomal membrane vertebrate SCARB2 cysteine residues were observed: Cys274, Cys312, Cys318 and Cys329, which corresponded to four of six previously identified disulfide forming cysteine residues for bovine CD36 (Rasmussen *et al*. 1998). In contrast, ten cysteine residues of the vertebrate CD36 sequences were conserved, including two within each of the N- (Cys3 and Cys7) and C-terminal (Cys464 and Cys466) cytoplasmic sequences, and six within the vertebrate exoplasmic sequences (Cys243; Cys272; Cys311; Cys313; Cys322; and Cys333) (Holmes & Cox 2012). The CD36 N- and C-terminal conserved cytoplasmic cysteine residues have been shown to be palmitoylated (Tao *et al*. 1996), which may contribute to protein-protein interactions, protein trafficking and membrane localization (Salaun *et al*. 2010). These conserved cysteines are lacking in the vertebrate SCARB2 sequences (Figure 1), which suggests that S-palmitoyl cysteine residues do not play a role for this lysosomal membrane protein. Vertebrate SCARB1 exoplasmic sequences also contained only four conserved cysteine residues forming disulfide bridges (Cys281; Cys321; Cys323; and Cys334) (Holmes & Cox 2012) although another conserved SCARB1 cysteine (not observed in the CD36 and SCARB2 sequences) (human SCARB1 Cys384) serves a major role in lipid transfer activity (Papale *et al.* 2011, Yua *et al.* 2011) (Figure 2).

SCARB2 C-terminal Lysosomal Targeting Sequences

The targeting of SCARB2 to lysosomes has been previously shown to result from a Leu-Ile dipeptide motif

Table 2. Predicted N-glycosylation Sites for Vertebrate SCARB2 Sequences. Numbers refer to amino acids in the acid sequences, including N-asparagine; K-lysine; I-isoleucine; H-histidine; S-serine; T-threonine; Q-glutamine; D-aspartate; Y-tyrosine; and V-valine. Note that there are 12 potential sites identified, including 10 sites for human SCARB2. N-glycosylation sites were identified using the NetNGlyc 1.0 web server (http://www.cbs.dtu.dk/services/NetNGlyc/). Higher probability N-glycosylation sites are bold.

Vertebrate SCARB2	Species	Site 1	Site 2	Site 3	Site 4	Site 5	Site 6	Site 7	Site 8	Site 9	Site 10	Site 11	Site 12	Site 13	Site 14	No of Sites
Human	*Homo sapiens*		45 NGTE	68 NVTN		105 NGTT			206 NGTN	224 NFTK	249 NGTK	304 NTSD	325 NVSI	412 NESV	430 NTTL	10
Chimp	*Pan troglodytes*		45 NGTE	68 NVTN		105 NGTT			206 NGTN	224 NFTK	249 NGTK	304 NTSD	325 NVSI	412 NESV	430 NTTL	10
Orangutan	*Pongo abelii*		45 NGTE	68 NVTN		105 NGTT			206 NGTN	224 NFTK	249 NGTD	304 NTSD	325 NVSI	412 NESV	430 NTTL	10
Rhesus	*Macaca mulatta*		45 NGTE	68 NVTN		105 NGTT			206 NGTN	224 NFTK	249 NGTD	304 NTSD	325 NVSI	412 NESV	430 NTTL	10
Marmoset	*Callithrix jacchus*		45 NGTE	68 NVTN		105 NGTT			206 NGTN	224 NFTK	249 NGTD	304 NTSD	325 NVSI	412 NESV	430 NTTL	10
Mouse	*Mus musculus*		45 NGTK	68 NVTN		105 NGTT	122 NQSV		206 NGTN	224 NFSK	249 NGTD	304 NTSE	325 NISI	412 NESV	430 NTTL	11
Rat	*Rattus norvegicus*		45 NGTK	68 NVTN		105 NGTT	122 NQSV		206 NGTN	224 NFSK	249 NGTD	304 NTSE	325 NISI	412 NESV	430 NTTL	11
Cow	*Bos taurus*		45 NGSE	68 NVTN		105 NGTT			206 NGTN	224 NFSK	249 NGTD	304 NTSD	325 NVSV	412 NESV	430 NTTL	10
Dog	*Canis familaris*		45 NGSE	68 NVTN		105 NGTT	122 NQSV		206 NGTN	224 NFSK	249 NGTD	304 NTSD	325 NISI	412 NESV	430 NTTL	11
Pig	*Sus scrofa*		45 NGSE	68 NVTN		105 NGTT			206 NGTN	224 NFSK	249 NGTD	304 NTSD	325 NVSV	412 NESV	430 NTTM	10
Rabbit	*Oryctolagus cuniculus*		45 NGTE	68 NVTN		105 NGTT	122 NQSV		206 NGTN	224 NFTK	249 NGTD	304 NTSE	325 NASI	412 NESV	430 NTTM	11
Elephant	*Loxodonta africana*		45 NGTE	68 NVTN		105 NGTT	122 NLSI		206 NGTN	224 NFTK	249 NGTD	304 NTSE	325 NVSI	412 NESV	430 NTTL	11
Chicken	*Gallus gallus*		45 NGTE	68 NVTN		105 NGTK			206 NGTD	224 NFSR	249 NGTD	304 NTTV	328 NVSI	415 NESV		9
Lizard	*Anolis carolensis*		45 NGTE	68 NLTN		105 NDTE			206 NGSD		249 NGTD	304 NVST	328 NVTA	415 NESV		8
Frog	*Xenopus tropicalis*		45 NESE	68 NVTN	99NITF	105 NGTE		195 NISD	207 NTTD		250 NGTD	305 NVSV	329 NVSI	416 NESV		10
Zebrafish	*Danio rerio*	35 NNTV	46 NGTE	69 NLTN		106 NGTR	123 NMSR		204 NGTE		247 NGTD		326 NASV			8

in the Leu475-Ile476-Arg477-Thr478 C-terminal cytoplasmic sequence (Ogata & Fukuda 1994, Tabuchi *et al*. 1997). The deletion of the nine amino acids closer to the SCARB2 C-terminal transmembrane domain also abolished this lysosomal location suggesting that an extended cytoplasmic tail for this protein is required for this function. Figure 1 shows a comparison of vertebrate SCARB2 C-terminal sequences, with the 7 sequences examined showing a Leu-Ile or a Leu-Leu (zebrafish SCARB2) C-terminal sequence in each case, which is consistent with a proposal for a dileucine (or leucine-isoleucine) lysosomal sorting motifs reported for the COOH tails of SCARB2 (Sandoval *et al*. 2000) and GLUT4 (insulin-regulatable glucose transporter) (James *et al*. 1989). Tabuchi and coworkers (2000; 2002) have also reported that two acidic amino acids (Asp470-Glu471) in the COOH-terminal SCARB2 sequence play important roles in regulating the movement of this protein within the endocytic pathway. A comparison of vertebrate SCARB2 C-terminal sequences supports this hypothesis with acidic amino acids being predominantly conserved for these positions, with the exception of zebrafish SCARB2, which has an apparent human 470Asp→ zebrafish 466Ala substitution, but with an additional acidic amino acid (469Glu) further down the C-

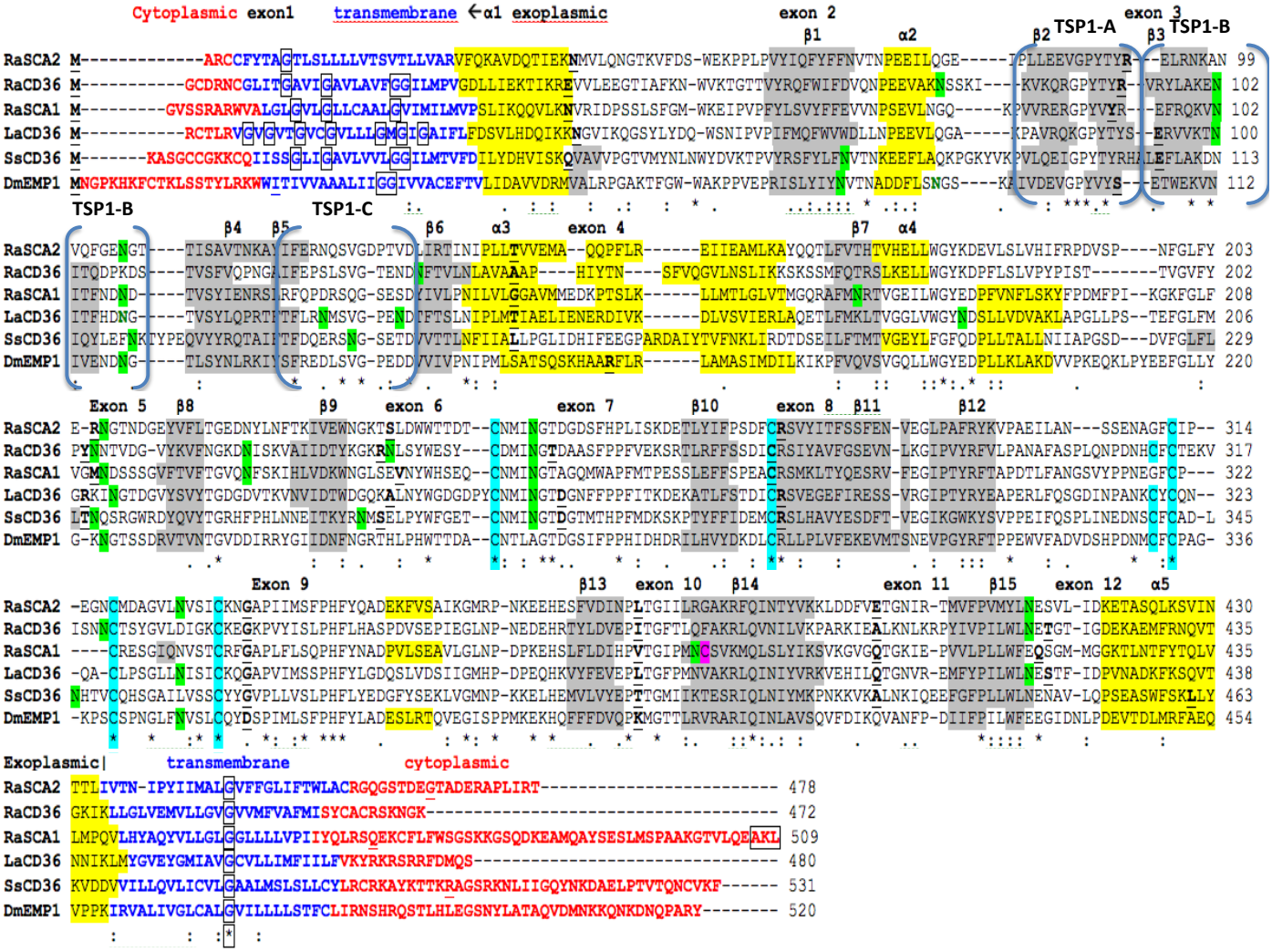

Figure 2. Amino Acid Sequence Alignments for Rat (Ra) SCARB2 (SCA2), SCARB1 (SCA1) and CD36 Sequences with Lancelet (La), Sea Squirt (Ss) and Fruit Fly (DmEMP1) CD-36 Like Sequences. See Table 1 for sources of rat SCARB2, SCARB1 and CD36 sequences as well as other CD36-like sequences; * shows identical residues for CD36-like subunits; : similar alternate residues; . dissimilar alternate residues; predicted cytoplasmic residues are shown in red; predicted transmembrane residues are shown in blue; N-glycosylated and potential N-glycosylated Asn sites are highlighted in green; free-SH Cys involved in lipid transfer for rat SCARB1 is highlighted in pink; predicted disulfide bond Cys residues are highlighted in cyan blue; predicted α-helices for vertebrate CD36-like sequences are highlighted in yellow and numbered in sequence from the start of the predicted exoplasmic domain; predicted β-sheets are highlighted in grey and also numbered in sequence; bold underlined font shows residues corresponding to known or predicted exon start sites; exon numbers are shown; AKL refers to final three C -terminal residues for rat SCARB1 which bind a PDZ domain-containing protein (PDZK1); transmembrane conserved glycines are shown as G; TSP1-A, TSP1-B and TSP1-C represent motifs identified by Crombie & Silverstein 1998.

terminal cytoplasmic tail (Figure 1). The intracellular trafficking of lysosomal membrane proteins has been extensively investigated in recent years, and these studies have shown that sorting from the Golgi or the plasma membrane into the endosomes and lysosomes represents one pathway which is mediated by short COOH-terminal cytoplasmic sequences (Hunziker & Geuze 1996). The dileucine cytoplasmic signal is one such pathway, however others also serve to target lysosomal proteins, such as cholesterol ester lipase (LIPA) which contains a C-terminal Arg-Lys dipeptide sequence (Sleat *et al.* 2006) and the C-terminal tyrosine based lysosomal targeting signal reported for lysosome-associated membrane glycoprotein-1 (LAMP-1) (Akasaki *et al*. 2010, Höning *et al*. 1996).

Predicted Secondary Structures for Vertebrate SCARB2

Predicted secondary structures for vertebrate SCARB2 sequences were examined (Figure 1), particularly for the sequences external to the lysosomal membrane (residues 28-433 for rat SCARB2) (Figure 2). α-Helix and β-sheet structures were similar in each case, with a α-helix extending beyond the N-terminal and C-terminal transmembrane regions in each case: α1 and α6. A consistent sequence of predicted secondary structure was also observed for each of the vertebrate SCARB2 sequences: N-terminal cytoplasmic sequence--N-terminal transmembrane sequence--α1--β1--α2--β2--β3--β4--β5--α3--β6--α4--α5--β7--β8--β9--β10--β11--β12--β13--β14--α6--C-terminal trans-membrane sequence--C-terminal cytoplasmic sequence. Further description of the secondary and tertiary structures for SCARB2 must await the three dimensional structure of this protein, particularly for the external lysosomal membrane region which directly binds lipids and contributes towards the organization and maintenance of the lysosomal membranes (Gamp *et al*. 2003; Kuronita

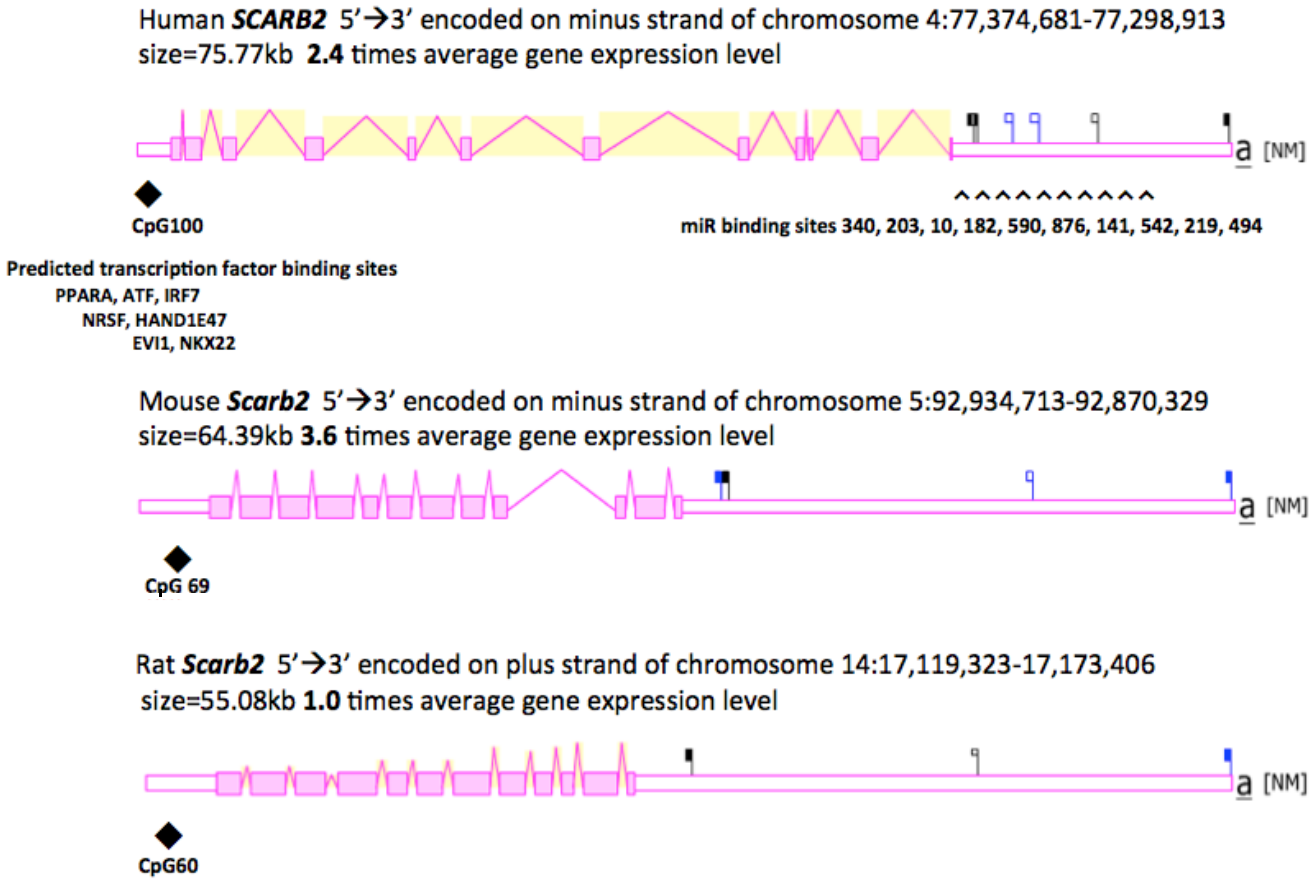

Figure 3. Gene Structures and Major Transcripts for the Human, Mouse and Rat *SCARB2* Genes. Derived from the AceView website http://www.ncbi.nlm.nih.gov/IEB/Research/Acembly/; [31] mature isoforms are shown with capped 5'- and 3'- ends for the predicted mRNA sequences; NM refers to the NCBI reference sequence; exons are in pink; the directions for transcription are shown as 5' → 3'; black squares show predicted CpG island sites at or near the 5'untranslated regions of the genes; the symbol ^ shows 10 predicted microRNA binding sites observed at or near the *SCARB2* 3'untranslated regions; sizes of mRNA sequences are shown in kilobases (kb); predicted transcription factor binding sites (TFBS) for human *SCARB2* are shown: PPARA- peroxisome proliferator-activated receptor-α ; ATF-activation transcription factor; IRF7-interferon regulatory factor; HAND1-heart- and neural crest derivatives-expressed protein system; EVI1-ecotropic viral integration site; NKX 2.2-

et al. 2002).

Conserved Proline and Glycine Residues within the SCARB2 External Lysosomal Membrane Domain

Figure 1S (see supplementary data) shows the alignment of 7 vertebrate SCARB2 amino acid sequences for the external lysosomal membrane domain with colors depicting the properties of individual amino acids and conservation observed for some of these protein sequences. In addition to the key vertebrate SCARB2 amino acids detailed previously, others were also conserved, including 16 proline residues. Prolines play a major role in protein folding and protein-protein interactions, involving the cyclic pyrrolidine amino acid side chain, which may introduce turns (or kinks) in the polypeptide chain as well as having destabilizing effects on α-helix and β-strand conformations (MacArthur & Thornton 1991). In addition, the presence of sequential prolines within a protein sequence may confer further restriction in folding conformation and create a distinctive structure, such as that reported for the mammalian Na^+/H^+ exchanger, which plays a major role in cation transport (Kreiger *et al.* 2005). Sequential prolines (P1 and P2: Pro57-Pro58) were conserved for 6 of 7 vertebrate SCARB2 sequences examined, which may confer a distinctive conformation in this region supporting the lipid receptor functions for this protein. For three of the vertebrate SCARB2 sequences examined, four sequential proline residues were observed (chicken, frog and zebrafish: Pro57-Pro58-Pro59-Pro60) which may contribute further to the distinctive folding conformation in this region. Regions of water exposed proteins with high levels of proline residues are often sites for protein-protein interactions (Kay *et al.* 2000) and these residues may significantly contribute to the binding of lipids by the external lysosomal membrane region of SCARB2. Similar results have been recently reported for vertebrate SCARB1 and CD36 exoplasmic regions, however 30 and 17 conserved proline residues were observed, respectively in these proteins (Holmes & Cox 2012).

Figure 1S (see supplementary data) also shows conservation of 20 glycine residues for these vertebrate SCARB2 external domains of lysosomal membranes, which due to their small size, may be essential for static turns, bends or close packing in the domain, or required for conformational dynamics during lipid receptor on-off switching, as in the case of the aspartate receptor protein (Coleman *et al.* 2005). Both proline and glycine residues are frequently found in turn and loop structures of proteins, and usually influence short loop formation within proteins containing between 2 and 10 amino acids (Kreiger *et al.* 2005). Evidence for these short loop structures within vertebrate SCARB2 external lysosomal membrane sequences was evident from the predicted secondary structures for vertebrate SCARB2 (Figure 1), with proline and/or glycine residues found at the start or end of the following structures: β1 (Pro57; Pro58; Pro60), α2 (Pro72), β2 (Gly78 -Pro81), β3 (Gly87/Pro88), β4 (Gly106), α3 (Pro139), α5 (Gly179), β7 (Gly211; Gly218), β8 (Gly233), β11 (Gly211) and β12 (Pro371) (Figure 1).

Alignments of Rat SCARB2, SCARB1 and CD36 with other CD36 Sequences

The amino acid sequences for rat SCARB2, SCARB1 and CD36 (see Table 1) are aligned in Figure 2. The sequences were 30-33% identical and showed similarities in several key features and residues, including cytoplasmic N-terminal and C-terminal residues; N-terminal and C-terminal trans-membrane helical regions; disulfide bond forming residues, previously identified for bovine CD36: Cys243-Cys311, Cys272-Cys333 and Cys313-Cys322 (Rasmussen *et al.* 1998); several predicted N-glycosylation sites for rat SCARB2 (11 sites) (Table 2), rat SCARB1 (10 sites) and CD36 (9 sites) (Holmes & Cox 2012), of which only one was shared between these sequences (N-glycosylation site 10) (Table 2); similar thrombospondin-1 binding sites (designated as TSP1-A, TSP1-B and TSP1-C) previously reported for SCARB2, SCARB2 and CD36 (Crombie & Silverstein 1998) and similar predicted secondary structures previously identified for SCARB1 and CD36 (Holmes & Cox 2012) (Figure 1). Sequence comparisons for the TSP1 binding sites confirmed the presence of a protein kinase C consensus phosphorylation site (Crombie & Silverstein 1998) within the TSP1-A site (GPYTYR) for rat SCARB2, CD36 and SCARB1 (but with a substitution at one site: GPYVYR). The current studies demonstrated other common features of these sites; namely the predicted β-sheet secondary structures for TSP1-A (β2) and TSP1-B (β3), the absence of predicted secondary structure for TSP1-C (located between β5 and β6) and the predicted N-glycosylation sites observed for TSP1-B. The Cys384 residue, for which the free-SH group plays a major role in SCARB1-mediated lipid transport (Yua *et al.* 2011), was unique to SCARB1, being replaced by other residues for the corresponding SCARB2 and CD36 proteins (Gly379 and Phe383, respectively). N-terminal transmembrane glycine residues, which play a role in the formation of SCARB1 oligomers (Gaidukov *et al.* 2011), were also observed for the rat CD36 sequence, although twin-glycines (Gly23-Gly24) were observed for the vertebrate CD36 sequences (Holmes & Cox 2012). In contrast, only one of these glycines (Gly10)

was observed for the rat SCARB2 N-terminal transmembrane sequence. These results suggest that rat SCARB2, SCARB1 and CD36 proteins share several important properties, features and conserved residues, including being membrane-bound with cytoplasmic and transmembrane regions, N-glycosylated at specific sites and have similar secondary structures but each is sufficiently different to serve distinct functions external to the respective membrane surfaces.

Alignments were also prepared for the predicted lancelet (*Branchiostoma floridae*) and sea squirt (*Ciona intestinalis*) CD36-like sequences (Table 1) and a major epithelial membrane protein (EMP) from the fruit fly (*Drosophila melanogaster*) (FBpp0072309) (Nichols & Vogt 2008) with the rat SCARB2, SCARB1 and CD36 sequences (Figure 2). The lancelet, sea squirt and fruit fly sequences examined shared many features with the CD36-like rat sequences, including the N- and C-terminal cytoplasmic and transmembrane sequences, similarities in predicted secondary structures, positional identities for five conserved cysteine residues (indicating conservation of at least 2 disulfide bridges for these proteins), predicted N-glycosylation sites (including one which is shared across all 6 CD-like sequences; site 10 in Table 2) and transmembrane glycine residues, which were observed in both the N- and C-terminal sequences, although with only a single glycine residue for the human SCARB2 sequence.

Gene Locations and Exonic Structures for Vertebrate *SCARB2* Genes

Table 1 summarizes the predicted locations for vertebrate *SCARB2* genes based upon BLAT interrogations of several vertebrate genomes using the reported human and mouse SCARB2 sequences (Fujita *et al*. 1992, Tabuchi *et al*. 1997) and the predicted sequences for other vertebrate and fruit fly *CD36*-like genes derived from the UC Santa Cruz genome browser (Kent *et al.* 2003). The predicted vertebrate *SCARB2* genes were transcribed on either the positive strand (e.g. chimpanzee, rhesus monkey, marmoset, rat, chicken, lizard and zebrafish genomes) or the negative strand (e.g. human, orangutan, mouse, cow, pig, opossum, chicken and frog genomes). Figure 1 summarizes the predicted exonic start sites for human, mouse, cow, chicken, lizard, frog and zebrafish *SCARB2* genes with each having 12 coding exons, in identical or similar positions to those reported for the human *SCARB2* gene (Kent *et al*. 2003).

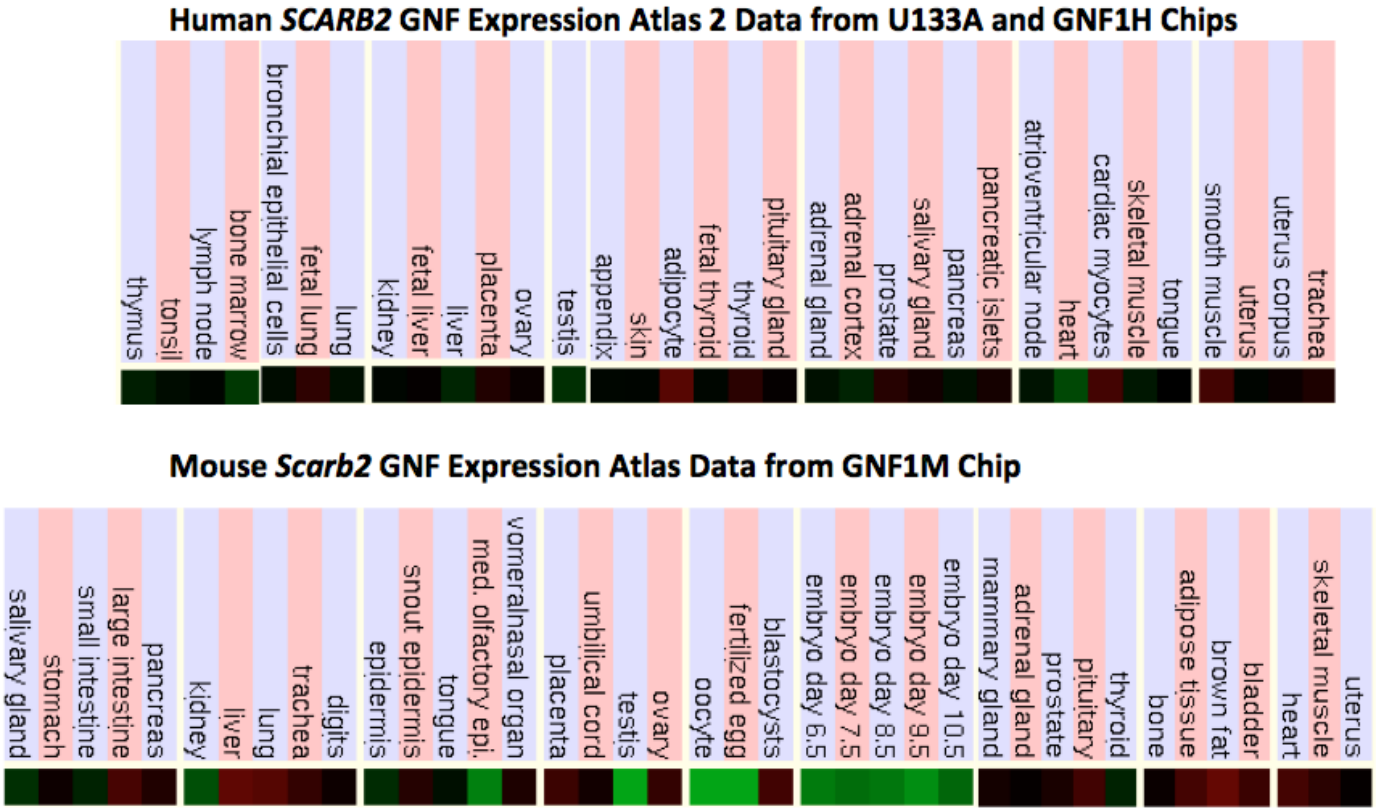

Figure 4. Comparative Tissue Expression for Human and Mouse *SCARB2* Genes. Expression 'heat maps' (GNF Expression Atlas 2 data) (http://biogps.gnf.org) (Su et al, 2004) were examined for comparative gene expression levels among human and mouse tissues for *SCARB2* genes showing high (red); intermediate (black); and low (green) expression levels. Derived from human and mouse genome browsers (http://genome.ucsc.edu) (Kent *et al*. 2003).

Figure 3 shows the predicted structures for the major human, mouse and rat *SCARB2/Scarb2* transcripts (Thierry-Mieg & Thierry-Mieg 2006). The transcripts were ~2kbs in length with 12 exons present for the mRNA transcripts and in each case, a non-coding 5'-untranslated sequence and an extended 3'-untranslated region (UTR) were observed. The human *SCARB2* genome sequence contained several predicted transcription factor binding sites (TFBS), including PPARA (peroxisome proliferator-activated receptor-α), which plays a major role in kidney proximal tubule development and maintenance (Kamijo *et al*. 2002) where SCARB2 is highly expressed (Berkovic *et al*. 2008, Desmond *et al.* 2011); NRSF (neuron-restrictive silencing factor), which functions as a neuronal cell repressor (Kim et al. 2004); ATF (activating transcription factor 1) which mediates heme oxygenase induction by heme and drives macrophage adaptation to intraplaque hemorrhage during atherosclerosis (Boyle *et al*. 2012); IRF7 (interferon regulatory factor-7) which is critical for the regulation of inflammatory responses in the central nervous system (Salem *et al.* 2011); HAND1 (heart- and neural crest derivatives-expressed protein system) which drives ongoing expression of cardiac-specific genes (Riley *et al*. 1998); EVI1 (ecotropic viral integration site), which is a complex transcription factor with multiple functions (Buonamici *et al*. 2003) and NKX22 (homeobox protein Nkx-2.2) which contributes to the expression of genes that play a role in axonal guidance (Holz *et al*. 2010).

Figure 3 also reports the presence of a CpG island within the promoter region of the human, mouse and rat *SCARB2* genes (CpG100, CpG69 and CpG60, respectively). CpG islands were also observed within the promoter regions of other vertebrate *SCARB2* genes, including marmoset (CpG106), pig (CpG71) and lizard (CpG420) (data not shown). CpG islands are typically found within the gene promoter for housekeeping genes (Saxanov *et al*. 2006). Elango & Yi (2011) have also proposed that larger CpG islands are associated with gene promoters showing a broad range of gene expressions and contain more RNA polymerase II binding sites than other promoters. Conse-

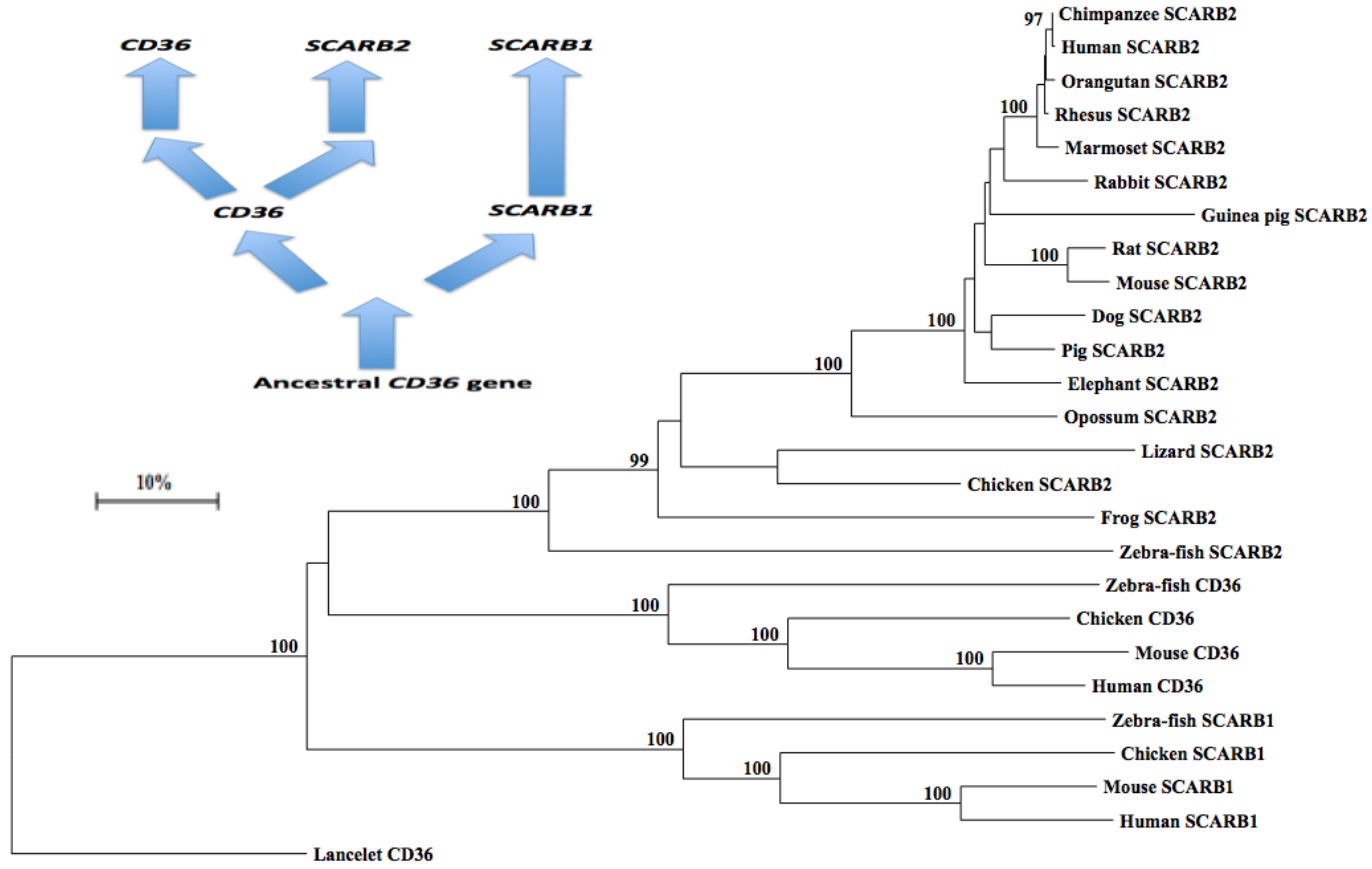

Figure 5: Phylogenetic Tree of Vertebrate SCARB2 Amino Acid Sequences with Human, Mouse, Chicken and Zebrafish SCARB1 and CD36 Sequences. The tree is labeled with the SCARB-like name and the name of the animal and is 'rooted' with the lancelet CD36 sequence. Note the 3 major clusters corresponding to the *SCARB2, SCARB1* and *CD36* gene families. A genetic distance scale is shown. The number of times a clade (sequences common to a node or branch) occurred in the bootstrap replicates are shown. Only replicate values of 95 or more, which are highly significant are shown with 100 bootstrap replicates performed in each case.

quently, the presence of CpG100 and the transcription factor binding sites observed within the *SCARB2* gene may contribute significantly to the broad tissue expression observed for *SCARB2* transcripts. The human *SCARB2* transcript also contained an extended 3'-noncoding segment with ten predicted miRNA binding sites (miR-340, 203, 10, 182, 590, 876, 141, 542, 219 and 494), which is well in excess of the usual number of such sites. miRNAs have been reported to function as post-transcriptional regulators that bind to complementary sequences on target messenger RNA transcripts (mRNAs), which result in translational repression or target degradation and gene silencing (Bartel 2009).

Comparative Human and Mouse *SCARB2* Tissue Expression

Figure 4 presents 'heat maps' showing comparative gene expression for various human and mouse tissues obtained from GNF Expression Atlas Data using the U133A and GNF1H (human) and GNF1M (mouse) chips (http://genome.ucsc.edu; http://biogps.gnf.org) (Su *et al.* 2004). These data support a broad and high level of tissue expression for human and mouse *SCARB2,* particularly for adipose tissue, cardiac myocytes, skeletal muscle and liver which is consistent with previous reports for these genes (Calvo *et al*. 1995, Fujita *et al*. 1992, Ogata *et al*. 1994, Tabuchi *et al*. 1997). Much lower levels of mouse *Scarb2* gene expression were observed in oocytes and during early embryonic development. Overall, however, human and mouse *SCARB2* tissue expression levels were 2-4 times higher than the average level of gene expression, which supports the key role played by this protein in lysosomal and endosomal membranes of the body.

Phylogeny of vertebrate SCARB2 and related CD36-like sequences

A phylogenetic tree (Figure 5) was calculated by the progressive alignment of 17 vertebrate SCARB2 amino acid sequences with human, mouse, chicken and zebrafish SCARB1 and CD36 sequences 'rooted' with the lancelet (*Branchiostoma floridae*) CD36 sequence (see Table 1). The phylogram showed clustering of the SCARB2 sequences into groups which were consistent with their evolutionary relatedness as well as groups for human, mouse, chicken and zebrafish SCARB1 and CD36 sequences, which were distinct from the lancelet CD36 sequence. These groups were significantly different from each other (with bootstrap values of ~ 100/100). It is apparent from this study of vertebrate CD-like genes and proteins that this is an ancient protein for which a proposed common ancestor for the CD36, *SCARB1* and *SCARB2* genes may have predated the appearance of fish > 500 million years ago (Donohue & Benton 2007). In parallel with the evolution of SCARB2 and other CD36-like proteins (SCARB1 and CD36), thrombospondins (TSPs) are also undergoing evolutionary changes in their structures and functions (Bentley & Adams 2010), with gene duplication events proposed at the origin of deuterostomes.

Conclusions

The results of the present study indicate that vertebrate *SCARB2* genes and encoded proteins represent a distinct gene and protein family of *CD36*-like proteins which share key conserved sequences that have been reported for other CD36-like proteins (SCARB1 and CD36) previously studied (Acton *et al.* 1996, Bultel-Brienne *et al.* 2002, Connelly *et al.* 2004, Fujita *et al*. 1992, Holmes & Cox 2012, Kent *et al.* 2011, Kuronita *et al.* 2002, Lin *et al.* 2012, Marsche *et al.* 2003, Ogata & Fukuda 1994, Tabuchi *et al.* 1997). SCARB2 has a unique property among these proteins in serving major roles within endosomal and lysosomal membranes of various cells and tissues of the body. *SCARB2* is encoded by a single gene among the vertebrate genomes studied and is highly expressed in human and mouse tissues, particularly in adipose tissue, cardiac myocytes, skeletal muscle and liver, and usually contained 12 coding exons. Predicted secondary structures for vertebrate CD36 proteins showed strong similarities with other CD36-like proteins, SCARB1 and CD36. Three major structural domains were apparent for vertebrate SCARB2, including the N-terminal and C-terminal cytoplasmic domains, the N-terminal and C-terminal transmembrane domains, the external lysosomal membrane domain, two disulfide bridges and several N-glycosylation sites for glycan binding, which are apparently essential for membrane recruitment. Phylogenetic studies using 17 vertebrate SCARB2 sequences with human, mouse, chicken and zebrafish SCARB1 and CD36 sequences indicated that the *CD36* gene has appeared early in evolution, prior to the appearance of bony fish more that 500 million years ago, and has undergone at least two gene duplication events.

Conflicts of Interest and Responsibility for Contents

The author declares no conflicts of interest and is fully responsible for the writing and completion of the studies undertaken.

Acknowledgements

The expert assistance of Dr Bharet Patel of Griffith University with the phylogeny studies is acknowledged.

References

Acton S, Osgood D, Donoghue M, Corella D, Pocovi M, Cenarro A, Mozas P, Keilty J, Squazzo S, Woolf EA & Ordovas JM 1999 Association of polymorphisms at the SR-BI gene locus with plasma lipid levels and body mass index in a white population**.** *Arterioscler Thromb Vasc Biol* **19** 1734-1743.

Adachi H & Tsujimoto M 2006 Endothelial scavenger receptors. *Prog Lipid Res* **45** 379-404.

Akasaki K, Suenobu M, Mukaida M, Michihara A & Wada I 2010 COOH-terminal isoleucine of lysosome-associated membrane protein-1 is optimal for its efficient targeting to dense secondary lysosomes. *J Biochem* **148** 669-679.

Altschul F, Vyas V, Cornfield A, Goodin S, Ravikumar TS, Rubin EH & Gupta E 1997 Basic local alignment search tool. *J Mol Biol* **215** 403-410.

Balreira A, Gaspar P, Caiola D, Chaves J, Beirão I, Lima JL, Azevedo JE & Miranda MC 2008 A nonsense mutation in the LIMP-2 gene associated with progressive myoclonic epilepsy and nephrotic syndrome. *Hum Mol Genet* **17** 2238-2243.

Bartel DP 2009 MicroRNAs: target recognition and regulatory functions. *Cell* **136** 215-233.

Bentley AA & Adams JC 2010 The evolution of thrombospondins and their ligand-binding activities. *Mol Biol Evol* **27** 2187-2197.

Berkovic SF, Dibbens LM, Oshlack A, Silver JD, Katerelos M, Vears DF, Lüllmann-Rauch R, Blanz J, Zhang KW, Stankovich J, Kalnins RM, Dowling JP, Andermann E, Andermann F, Faldini E, D'Hooge R, Vadlamudi L, Macdonell RA, Hodgson BL, Bayly MA, Savige J, Mulley JC, Smyth GK, Power DA, Saftig P & Bahlo M 2008 Array-based gene discovery with three unrelated subjects shows SCARB2/LIMP-2 deficiency causes myoclonus epilepsy and glomerulosclerosis. *Am J Hum Genet* **82** 673-684.

Blanz J, Groth J, Zachos C, Wehling C, Saftig P & Schwake M 2010 Disease-causing mutations within the lysosomal integral membrane protein type 2 (LIMP-2) reveal the nature of binding to its ligand beta-glucocerebrosidase. *Hum Mol Genet* **19** 563-572.

Boyle JJ, Johns M, Kampfer T, Nguyen AT, Game L, Schaer DJ, Mason JC & Haskard DO 2012 Activating Transcription Factor 1 Directs Mhem Atheroprotective Macrophages Through Coordinated Iron Handling and Foam Cell Protection *Circ Research* **110** *20-33.*

Bultel-Brienne S, Lestavel S, Pilon A, Laffont I, Tailleux A, Fruchart JC, Siest G & Clavey V 2002 Lipid free apolipoprotein E binds to the class B Type I scavenger receptor I (SR-BI) and enhances cholesteryl ester uptake from lipoproteins. *J Biol Chem* **277** 36092-36099.

Buonamici S, Chakraborty S, Senyuk V & Nucifora G 2003 The role of EVI1 in normal and leukemic cells. *Blood Cells Mol Dis* **31** 206-212.

Calvo D, Dopazo J, & Vega MA 1995 The CD36, CLA-1 (CD36L1), and LIMPII (CD36L2) gene family: cellular distribution, chromosomal location, and genetic evolution. Genomics **25** 100-106.

Carrasco-Marín E, Fernández-Prieto L, Rodriguez-Del Rio E, Madrazo-Toca F, Reinheckel T, Saftig P & Alvarez-Dominguez C 2011 LIMP-2 links late phagosomal trafficking with the onset of the innate immune response to Listeria monocytogenes: a role in macrophage activation. *J Biol Chem* **286** 3332-3341.

Chaves J, Beirão I, Balreira A, Gaspar P, Caiola D, Sá-Miranda MC & Lima JL 2011 Progressive myoclonus epilepsy with nephropathy C1q due to SCARB2/LIMP-2 deficiency: clinical report of two siblings. *Seizure* **20** 738-740.

Coleman MD, Bass RB, Mehan RS & Falke JJ 2005 Conserved glycine residues in the cytoplasmic domain of the aspartate receptor play essential roles in kinase coupling and on-off switching. *Biochem* **44** 7687-7695.

Collot-Teixeira S, Martin J, McDermott-Roe C, Poston R & McGregor JL 2007 CD36 and macrophages in atherosclerosis. *Cardiovasc Res* **75** 468-477.

Connelly MA & Williams DL 2004 Scavenger receptor BI: A scavenger receptor with a mission to transport high density lipoprotein lipids**.** *Curr Opinion Lipid* **5** 287-295.

Crombie R & Silverstein R 1998 Lysosomal integral membrane protein II binds thrombospondin-1. Structure-function homology with the cell adhesion molecule CD36 defines a conserved recognition motif. *J Biol Chem* **273** 4855-4863.

Desmond MJ, Lee D, Fraser SA, Katerelos M, Gleich K, Martinello P, Li YQ, Thomas MC, Michelucci R, Cole AJ, Saftig P, Schwake M, Stapleton D, Berkovic SF & Power DA 2011 Tubular proteinuria in mice and humans lacking the intrinsic lysosomal protein SCARB2/Limp-2. *Am J Physiol Renal Physiol* **300** F1437-1447.

Dibbens LM, Karakis I, Bayly MA, Costello DJ, Cole AJ & Berkovic SF 2011 Mutation of SCARB2 in a patient with progressive myoclonus epilepsy and demyelinating peripheral neuropathy. *Arch Neurol* **68** 812-813.

Do CB, Tung JY, Dorfman E, Kiefer AK, Drabant EM, Francke U, Mountain JL, Goldman SM, Tanner CM, Langston JW, Wojcicki A & Eriksson N 2011 Web-based genome-wide association study identifies two novel loci and a substantial genetic component for Parkinson's disease. *PLoS Genet* **7** e1002141.

Donoghue PCJ & Benton MJ 2007 Rocks and clocks: calibrating the tree of life using fossils and molecules. *Trends Genet* **22** 424-431.

Elango N & Yi SV 2011 Functional relevance of CpG island length for regulation of gene expression. *Genetics* **187** 1077-1083.

Felsenstein J 1986 Confidence limits on phylogenies: an approach using the bootstrap. *Evolution* **39** 783-789.

Fujita H, Takata Y, Kono A, Tanaka Y, Takahashi T, Himeno M & Kato K 1992 Isolation and sequencing of a cDNA clone encoding the 85 kDa human lysosomal sialoglycoprotein (hLGP85) in human metastatic pancreas islet tumor cells. *Biochem Biophys Res Commun* **184** 604-611.

Gaidukov L, Nager AR, Xu S, Penman M & Krieger M 2011 Glycine dimerization motif in the N-terminal transmembrane domain of the high density lipoprotein receptor SR-BI required for normal receptor oligomerization and lipid transport. *J Biol Chem* **286** 18452-18464.

Gamp AC, Tanaka Y, Lüllmann-Rauch R, Wittke D, D'Hooge R, De Deyn PP, Moser T, Maier H, Hartmann D, Reiss K, Illert AL, von Figura K & Saftig P 2003 LIMP-2/LGP85 deficiency causes ureteric pelvic junction obstruction, deafness and peripheral neuropathy in mice. *Hum Mol Genet* **12** 631-646.

Gautam S & Banerjee M 2011 The macrophage Ox-LDL receptor, CD36 and its association with type II diabetes mellitus. *Mol Genet Metab* **102** 389-398.

Gupta R & Brunak S 2002 Prediction of glycosylation across the human proteome and the correlation to protein function. *Pac Symp Biocomput* **7** 310-322.

Hall TA 1999 BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp. Ser* **41** 95-99.

Holmes RS 2012 Vertebrate patatin-like phospholipase domain-containing protein 4 (*PNPLA4*) genes and proteins: a gene with a role in retinol metabolism. *3 Biotech* doi:10.1007/s13205-012-0063-7.

Holmes RS & Cox LA 2012 Comparative studies of vertebrate scavenger receptor class B type 1 (SCARB1): a high density lipoprotein (HDL) binding protein. *Res. Reports Biochem* **2** 1016.

Holz A, Kollmus H, Ryge J, Niederkofler V, Dias J, Ericson J, Stoeckli ET, Kiehn O & Arnold HH 2010 The transcription factors Nkx2.2 and Nkx2.9 play a novel role in floor plate development and commissural axon guidance. *Development* **137** 4249-4260.

Höning S, Griffith J, Geuze HJ & Hunziker W 1996 The tyrosine-based lysosomal targeting signal in lamp-1 mediates sorting into Golgi-derived clathrin-coated vesicles. *EMBO J* **15** 5230-5239.

Hopfner F, Schormair B, Knauf F, Berthele A, Tölle TR, Baron R, Maier C, Treede RD, Binder A, Sommer C, Maihöfner C, Kunz W, Zimprich F, Heemann U, Pfeufer A, Näbauer M, Kääb S, Nowak B, Gieger C, Lichtner P, Trenkwalder C, Oexle K & Winkelmann J 2011 Novel SCARB2 mutation in action myoclonus-renal failure syndrome and evaluation of SCARB2 mutations in isolated AMRF features. *BMC Neurol* **11** 134.

Hunziker W & Geuze HJ 1996 Intracellular trafficking of lysosomal membrane proteins. *Bioessays* **18** 379-389.

IIu I, Vorsanova SG, Saprina EA & IuB I 2010 Identification of candidate genes of autism on the basis of molecular cytogenetic and in silico studies of the genome organization of chromosomal regions involved in unbalanced rearrangements. *Genetika* **46** 1348-1351.

James DE, Strube M & Mueckler M 1989 Molecular cloning and characterization of an insulin-regulatable glucose transporter. *Nature* **338** 83-87.

Kamijo Y, Hora K, Tanaka N, Usuda N, Kiyosawa K, Nakajima T, Gonzalez FJ & Aoyama T 2002 Identification of functions of peroxisome proliferator-activated receptor alpha in proximal tubules. *J Am Soc Nephrol* **13** 1691-1702.

Kay BK, Williamson MP & Sudol M 2000 The importance of being proline: the interaction of proline-rich motifs in signaling proteins with their cognate domains. *FASEB J* **14** 231-241.

Kent AP & Stylianou IM 2011 Scavenger receptor class B member 1 protein: hepatic regulation and its effects on lipids, reverse cholesterol transport, and atherosclerosis. *Hep Med: Evid Res* **3** 29-44.

Kent WJ, Sugnet CW & Furey TS 2003 The human genome browser at UCSC. *Genome Res* **12** 994-1006.

Kim CS, Hwang CK, Choi HS, Song KY, Law PY, Wei LN & Loh HH 2004 Neuron-restrictive silencer factor (NRSF) functions as a repressor in neuronal cells to regulate the mu opioid receptor gene. *J Biol Chem* **279** 46464-46473.

Kim JW, Zhao SH, Uthe JJ, Bearson SM & Tuggle CK 2006 Assignment of the scavenger receptor class B, member 2 gene (SCARB2) to porcine chromosome 8q11-->q12 by somatic cell and radiation hybrid panel mapping. *Cytogenet Genome Res* **112** 342H.

Krieger A, Möglich A & Kiefhaber T 2005 Effect of proline and glycine residues on dynamics and barriers of loop formation in polypeptide F. *J Am Chem Soc* **127** 3346-3352.

Kuronita T, Eskelinen EL, Fujita H, Saftig P, Himeno M & Tanaka Y 2002 A role for the lysosomal membrane protein LGP85 in the biogenesis and maintenance of endosomal and lysosomal morphology. *J Cell Sci* **115** 4117-4131.

Lauzier B, Merlen C, Vaillant F, McDuff J, Bouchard B, Beguin PC, Dolinsky VW, Foisy S, Villeneuve LR, Labarthe F, Dyck JR, Allen BG, Charron G & Des Rosiers C 2011 Post-translational modifications, a key process in CD36 function: lessons from the spontaneously hypertensive rat heart. *J Mol Cell Cardiol* **51** 99-108.

Lewandrowski U, Moebius J, Walter U & Sickmann A 2006 Elucidation of N-glycosylation sites on human platelet proteins: a glycoproteomic approach. *Mol Cell Proteomics* **5** 226-233.

Lin YW, Lin HY, Tsou YL, Chitra E, Hsiao KN, Shao HY, Liu CC, Sia C, Chong P & Chow YH 2012 Human SCARB2-Mediated Entry and Endocytosis of EV71. *PLoS One* **7** e30507.

MacArthur MW & Thornton JM 1991 Influence of proline residues on protein conformation. *J Mol Biol* **218** 397-412.

Marsche G, Zimmermann R, Horiuchi S, Tandon NN, Sattler W & Malle E 2003 Class B scavenger receptors CD36 and SR-BI are receptors for hypochlorite-modified low density lipoprotein**.** *J Biol Chem* **278** 47562-47570.

Martin CA, Longman E, Wooding C, Hoosdally SJ, Ali S, Aitman TJ, Gutmann DA, Freemont PS, Byrne B & Linton KJ 2007 Cd36, a class B scavenger receptor, functions as a monomer to bind acetylated and oxidized low-density lipoproteins. *Protein Sci* **16** 2531-2541.

McGuffin LJ, Bryson K & Jones DT 2000 The PSIPRED protein structure prediction server. *Bioinform* **16** 404-405.

Michelakakis H, Xiromerisiou G, Dardiotis E, Bozi M, Vassilatis D, Kountra PM, Patramani G, Moraitou M, Papadimitriou D, Stamboulis E, Stefanis L, Zintzaras E & Hadjigeorgiou GM 2012 Evidence of an association between the scavenger receptor class B member 2 gene and Parkinson's disease. *Mov Disord* doi:10.1002/mds.24886.

Nichols Z & Vogt RG 2008 The SNMP/CD36 gene family in Diptera, Hymenoptera and Coleoptera: Drosophila melanogaster, D. pseudoobscura, Anopheles gambiae, Aedes aegypti, Apis mellifera, and Tribolium castaneum. *Insect Biochem Mol Biol* **38** 398-415.

Ogata S & Fukuda M 1994 Lysosomal targeting of Limp II membrane glycoprotein requires a novel Leu-Ile motif at a particular position in its cytoplasmic tail. *J Biol Chem* **269** 5210-5217.

Papale GA, Hanson PJ & Sahoo D 2011 Extracellular disulfide bonds support scavenger receptor class B type I-mediated cholesterol transport. *Biochem* **50** 6245-6254.

Rasmussen JT, Berglund L, Rasmussen MS & Petersen TE 1998 Assignment of disulfide bridges in bovine CD36. *Eur J Biochem* **257** 488-494.

Reczek D, Schwake M, Schröder J, Hughes H, Blanz J, Jin X, Brondyk W, Van Patten S, Edmunds T & Saftig P 2007 LIMP-2 is a receptor for lysosomal mannose-6-phosphate-independent targeting of beta-glucocerebrosidase. *Cell* **131** 770-783.

Ren Y 2012 Peroxisome proliferator-activator receptor γ: a link between macrophage CD36 and inflammation in malaria infection. *PPAR Res* **2012** 640769.

Riley P, Anson-Cartwright L & Cross JC 1998 The Hand1 bHLH transcription factor is essential for placentation and cardiac morphogenesis. *Nat Genet* **18** 271-275.

Roszek K & Gniot-Szulzycka J 2005 Lysosomal membrane glycoprotein lamp2a--receptor for chaperone-mediated degradation of cytosolic proteins. Postepy Biochem **51** 88-94.

Rubboli G, Franceschetti S, Berkovic SF, Canafoglia L, Gambardella A, Dibbens LM, Riguzzi P, Campieri C, Magaudda A, Tassinari CA & Michelucci R 2011 Clinical and neurophysiologic features of progressive myoclonus epilepsy without renal failure caused by SCARB2 mutations. *Epilepsia* **52** 2356-2363.

Salaun C, Greaves J & Chamberlain LH 2010 The intracellular dynamic of protein palmitoylation. *J Cell Biol* **191** 1229-1238.

Salem M, Mony JT, Løbner M, Khorooshi R & Owens T 2011 Interferon regulatory factor-7 modulates experimental autoimmune encephalomyelitis in mice. *J Neuroinflammation* **8** 181.

Sandoval IV, Martinez-Arca S, Valdueza J, Palacios S & Holman GD 2000 Distinct reading of different structural determinants modulates the dileucine-mediated transport steps of the lysosomal membrane protein LIMPII and the insulin-sensitive glucose transporter GLUT4. *J Biol Chem* **275** 39874-39885.

Saxonov S, Berg P & Brutlag DL 2006 A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc Natl Acad Sci USA* **103** 1412-1417.

Schroen B, Leenders JJ, van Erk A, Bertrand AT, van Loon M, van Leeuwen RE, Kubben N, Duisters RF, Schellings MW, Janssen BJ, Debets JJ, Schwake M, Høydal MA, Heymans S, Saftig P & Pinto YM 2007 Lysosomal integral membrane protein 2 is a novel component of the cardiac intercalated disc and vital for load-induced cardiac myocyte hypertrophy. *J Exp Med* **204** 1227-1235.

Simantov R & Silverstein RL 2003 CD36: a critical anti-angiogenic receptor. *Front Biosci* **8** s874-882.

Sleat DE, Zheng H, Qian M & Lobel P 2006 Identification of sites of mannose 6-phosphorylation on lysosomal proteins. *Mol Cell Prot* **5** 686-701.

Su AI, Wiltshire T, Batalov S, Lapp H, Ching KR, Block D, Zhang J, Soden R, Hayakawa M, Kreiman G, Cooke MP, Walker JR & Hogenesch JB 2004 A gene atlas of the human and mouse protein encoding transcriptomes. *Proc Natl Acad Sci USA* **101** 6062-6067.

Tandon NN, Lipsky RH, Burgess WH & Jamieson GA 1989 Isolation and characterization of platelet glycoprotein IV (CD36). *J Biol Chem* **264** 7570-7575.

Tabuchi N, Akasaki K, Sasaki T, Kanda N &Tsuji H **1997** Identification and characterization of a major lysosomal membrane glycoprotein, LGP85/LIMP II in mouse liver. *J Biochem* **122** 756-763.

Tabuchi N, Akasaki K & Tsuji H 2000 Two acidic amino acid residues, Asp(470) and Glu(471), contained in the carboxyl cytoplasmic tail of a major lysosomal membrane protein, LGP85/LIMP II, are important for its accumulation in secondary lysosomes. *Biochem Biophys Res Commun* **270** 557-563.

Tabuchi N, Akasaki K & Tsuji H 2002 Ile (476), a constituent of di-leucine-based motif of a major lysosomal membrane protein, LGP85/LIMP II, is important for its proper distribution in late endosomes and lysosomes. *Biochem Biophys Res Commun* **295** 149-156.

Tao N, Wagner SJ & Lublin DM 1996 CD36 is palmitoylated on both N- and C-terminal cytoplasmic tails. *J Biol Chem* **271** 22315-22320.

Thierry-Mieg D & Thierry-Mieg J 2006 AceView: A comprehensive cDNA-supported gene and transcripts annotation. *Genome Biol* **7** S12.

Tserentsoodol N, Gordiyenko NV, Pascual I, Lee JW, Fliesler SJ & Rodriguez IR 2006 Intraretinal lipid transport is dependent on high density lipoprotein-like particles and class B scavenger receptors. *Mol Vis* **12** 1319-1333.

Velayati A, DePaolo J, Gupta N, Choi JH, Moaven N, Westbroek W, Goker-Alpan O, Goldin E, Stubblefield BK, Kolodny E, Tayebi N & Sidransky E 2011 A mutation in SCARB2 is a modifier in Gaucher disease. Hum Mutat 32 1232-1238.

Yamayoshi S, Yamashita Y, Li J, Hanagata N, Minowa T, Takemura T & Koike S 2009 Scavenger receptor B2 is a cellular receptor for enterovirus 71. *Nat Med* **15** 798-801.

Yamayoshi S & Koike S 2011 Identification of a human SCARB2 region that is important for enterovirus 71 binding and infection. *J Virol* **85** 4937-4946.

Yua M, Romera KA, Nielanda TJF, Xua S, Saenz-Vashb V, Penmana M, Yesilaltaya A, Carrb SA & Kriegera M 2011 Exoplasmic cysteine Cys384 of the HDL receptor SR-BI is critical for its sensitivity to a small-molecule inhibitor and normal lipid transport activity. *Proc Natl Acad Sci USA* **108** 12243–12248.

Zhang X, Moor AN, Merkler KA, Liu Q & McLean MP 2007 Regulation of alternative splicing of liver scavenger receptor class B gene by estrogen and the involved regulatory splicing factors. *Endocrinology* **148** 5295-5304.