Update of the human secretoglobin (*SCGB*) gene superfamily and an example of 'evolutionary bloom' of androgen-binding protein genes within the mouse *Scgb* gene superfamily

Brian C. Jackson,[1] David C. Thompson,[2] Mathew W. Wright,[3] Monica McAndrews,[4] Alfred Bernard,[5] Daniel W. Nebert[6] and Vasilis Vasiliou[1]**

[1]Molecular Toxicology and Environmental Health Sciences Program, Department of Pharmaceutical Sciences, University of Colorado Anschutz Medical Center, Aurora, CO 80045, USA
[2]Department of Clinical Pharmacy, University of Colorado Anschutz Medical Center, Aurora, CO 80045, USA
[3]HUGO Gene Nomenclature Committee (HGNC), EMBL-EBI, Wellcome Trust Genome Campus, Cambridge, CB10 1SD, UK
[4]The Jackson Laboratory, Mouse Genome Informatics, 600 Main Street, Bar Harbor, ME 04609–1500, USA
[5]Unit of Toxicology and Applied Pharmacology, Catholic University of Louvain, Brussels, Belgium
[6]Department of Environmental Health and Center for Environmental Genetics (CEG), University of Cincinnati Medical Center, Cincinnati, OH 45267–0056, USA
**Correspondence to:* Tel: +1 303 724 3520(V.V); Fax: +1 303 724 7266(V.V.); E-mail: vasilis.vasiliou@ucdenver.edu; dan.nebert@uc.edu

Date received (in revised form): 25th October 2011

Abstract

The secretoglobins (SCGBs) comprise a family of small, secreted proteins found in animals exclusively of mammalian lineage. There are 11 human *SCGB* genes and five pseudogenes. Interestingly, mice have 68 *Scgb* genes, four of which are highly orthologous to human *SCGB* genes; the remainder represent an 'evolutionary bloom' and make up a large gene family represented by only six counterparts in humans. SCGBs are found in high concentrations in many mammalian secretions, including fluids of the lung, lacrimal gland, salivary gland, prostate and uterus. Whereas the biological activities of most individual SCGBs have not been fully characterised, what already has been discovered suggests that this family has an important role in the modulation of inflammation, tissue repair and tumorigenesis. In mice, the large *Scgb1b* and *Scgb2b* gene families encode the androgen-binding proteins, which have been shown to play a role in mate selection. Although much has been learned about SCGBs in recent years, clearly more research remains to be done to allow a better understanding of the roles of these proteins in human health and disease. Such information is predicted to reveal valuable novel drug targets for the treatment of inflammation, as well as designing biomarkers that might identify tissue damage or cancer.

Keywords: *SCGB, secretoglobin, evolutionary bloom, androgen-binding, protein, nomenclature, gene family*

Introduction

The secretoglobins (SCGBs) comprise a family of secreted proteins found in mammals and marsupials. The first discovered SCGB was found in rabbits and was first called blastokinin,[1] then later uteroglobin[2] and is now designated SCGB1A1 (in some early literature, the SCGB family is referred to as the 'uteroglobin' family). Eventually, the term

'secretoglobin' was coined to refer to the characteristics that all family members have in common. The 'secreto-' portion of the name indicates that these proteins are secreted. A second reason was proposed for the suffix 'globin'; their functions had largely remained a secret (Lehrer, R., personal communication). This suffix was given because secretoglobins form dimers consisting of two four-α helix-bundle monomers, creating a hydrophobic binding pocket, reminiscent of the globin-fold, which is an eight-α-helix bundle with a pocket for a molecule such as a heme group.[3]

Secretoglobins are found at high levels in many secretions, including uterine, prostatic, pulmonary, lacrimal and salivary glands, with any specific secretoglobin often being expressed in more than one tissue. For example, mRNA expression of every SCGB family member (except SCGB1D2) has been demonstrated in human airways.[4] In general, the physiological and pathophysiological functions of most individual SCGBs remain to be defined. Nevertheless, roles currently ascribed to SCGBs include lung maintenance and repair, immune modulation and, at least in rodents, mate selection. Some SCGB family members, such as mammaglobin, have been successfully used as epithelial cancer biomarkers.

SCGBs are small (~10 kDa in humans) proteins that dimerise before secretion. Dimers are resistant to proteases, heat and pH.[5,6] The crystal structures of several SCGBs have been resolved, including those of rabbit and rat uteroglobin (Protein Data Bank identifiers [PDB ID]:1UTG, 2UTG, 1UTR), rat Clara-cell specific protein (CCSP) (PDB ID:1CCD) and feline CH2 (Feld-1) (PDB ID:1PUO, 1ZKR, 2EJN).[7] These proteins contain four α-helical structures and assemble into homo- or hetero-dimers orientated in anti-parallel fashion, held together by covalent disulphide bonds (via one to three conserved Cys residues) and non-covalent interactions.[8]

The uteroglobin (UGB) dimer forms an internal hydrophobic cavity, located at the interface between the two subunits; this is the location of binding of hydrophobic ligands, including steroid hormones, some polychlorinated biphenyl metabolites, retinoids and various eicosanoid mediators of inflammation.[9,10] UGB's subunits consist of four α-helices which do not form a canonical four-helix bundle motif but, rather, a boomerang-shaped structure. The subunits are connected in an anti-parallel fashion to form a dimer in which helices 3 and 4 are involved in the dimer interface. In the structure of SCGB1A1, six residues (Phe6, Leu13, Tyr21, Phe28, Met41 and Ile63) in each subunit have been identified as being particularly important to this aspect of UGB structure.[11] All of these, except Phe28, are accessible to the ligand, which probably functions in maintaining the dimer interface. The other five are involved in ligand binding. The aromatic residues Phe6 and Tyr21 are critical to this binding and cannot be replaced by aliphatic amino acids. Conversely, Leu13 is accessible to solvent in the hydrophobic pocket and is commonly substituted by aromatic amino acids. This suggests that Leu13 may be involved in determining ligand specificity.

Sources of secretoglobin genes and proteins

Protein sequences for human SCGBs were accessed from Uniprot[12] through the HUGO Gene Nomenclature Committee website (http://www.genenames.org). Sequences for mouse SCGBs were retrieved from the National Center for Biotechnology Information (NCBI) gene database (http://www.ncbi.nlm.nih.gov/gene), and from the 'supplementary data' of Laukaitis *et al.*[13] Sequences were aligned with T-COFFEE using the most accurate mode, which combines multiple sources of sequence homology and structural information, where available.[14]

Human gene family members

As is commonly the case for a newly discovered family of proteins, SCGBs were originally named based on the location in which they were most highly expressed; this led to the same SCGB often being 'rediscovered' and named multiple times. In 2000, a standard nomenclature was established, when all proteins in the family were named SCGBs

and assigned family and subfamily names.[3] The nomenclature system was based on that used for the cytochrome P450[15,16] and nuclear hormone receptor[17] superfamilies, and was guided by the phylogenetic relationships of known SCGB family members, assembled by Ni and colleagues.[18] This provided a convenient and systematic naming system for an entire superfamily. In this report, the most common names used for each protein are listed, along with their standardised names. The human genome contains 11 *SCGB* genes and five pseudogenes (Figure 1).

SCGB1A1 subfamily

UGB, also known as blastokinin and CCSP (SCGB1A1), was initially discovered in the rabbit uterus and is the founding family member.[2] For this reason, more information about its biology is available than for many of the other SCGBs. These proteins differ from other SCGBs in that they are homodimers — that is, they are composed of two identical monomers and their subunits lack the middle Cys residue found in other SCGBs. In humans, high SCGB1A1 levels are found in peripheral airway surface fluid, where it is one of the most abundant proteins; it is also expressed in uterine endometrium and the prostate.[20] In the airways, SCGB1A1 is expressed in several cell types, especially Clara cells, and appears to play a role in immunomodulation through regulation of cell infiltration and in tissue repair after injury.[20]

SCGB1A1 may also exert anti-tumorigenic activity. For example, ablation of the mouse *Scgb1a1* gene in some strains is usually lethal and survivors develop tumours.[21] Conversely, recombinant SCGB1A1 inhibits proliferation and invasion of some cancer cell lines.[20] Studies of the *Scgb1a1*$^{(-/-)}$ knockout mouse suggest that SCGB1A1 may provide protection from oxidative stress and exert anti-inflammatory actions, in addition to providing resistance to pollutant-induced injury.[22] Interestingly, SCGB1A1 is initially downregulated to allow the body to respond to an infection.[23]

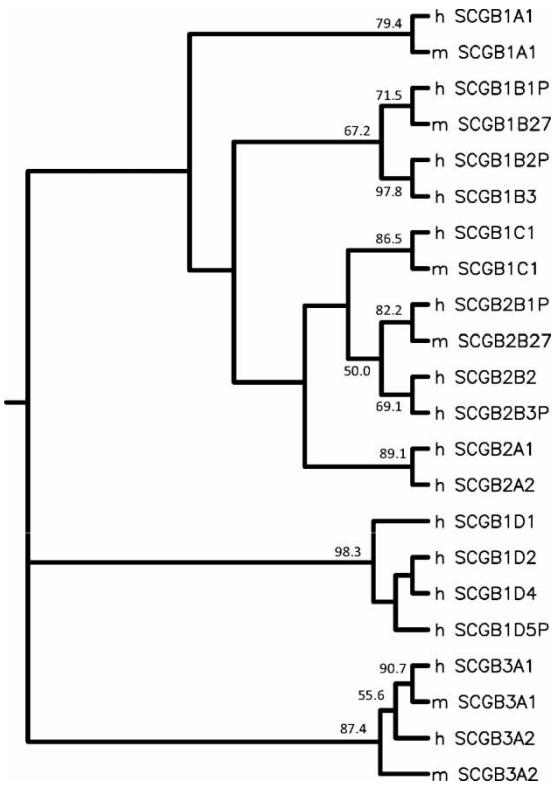

Figure 1. Phylogenetic tree of mouse (m) and human (h) SCGBs. For simplicity, and to avoid clutter, of the mouse androgen-binding protein (ABP) group, only SCGB1B27 (ABPA) and SCGB2B27 (ABPGB) are included. SCGB protein sequences were aligned using TCOFFEE[14] and analysed using nearest-neighbour-joining methods, as well as using 10,000 bootstrap replicates in the Phylip package.[19] Nodes with $\geq$50 per cent bootstrap confidence levels have been labelled.

SCGB1B subfamily

The human genome contains six genes that cluster phylogenetically with genes encoding mouse androgen-binding proteins (*Scgb1b* and *Scgb2b*). These genes were described based on genomic analysis, and have been given *SCGB4A* designations.[24] Based on phylogenetic clustering of their protein sequences, however, we propose that these genes be changed to *SCGB1B* and *SCGB2B* designations, to reflect their similarity to the mouse proteins. The *SCGB1B* subfamily includes *SCGB1B1P* (formerly *ABPA1P*), *SCGB1B2P* (formerly *SCGB4A1P*) and *SCGB1B3* (formerly *SCGB4A4*). *SCGB1B1P* and *SCGB1B2P* are predicted to have become pseudogenes, whereas *SCGB1B3* has no obvious inactivating mutations.

Interestingly, however, *SCGB1B2P* is the only *SCGB1B* member having evidence for expression in expressed sequence tag (EST) databases.

SCGB1C1

SCGB1C1 has been shown to be localised to Bowman's glands in the olfactory mucosa. Here, it is thought to act as an odorant-binding protein, with ligands appearing to be small, hydrophobic molecules.[25]

SCGB1D subfamily

SCGB1D1 and SCGB1D2 are also known as lipophilin A and lipophilin B, respectively. The lipophilins form heterodimers with SCGB2A proteins, which further associate to form tetramers. They have been identified in the prostatic fluid of rats and in the lacrimal gland fluid of humans and rabbits;[26] little is known about their function.

SCGB1D4 is widely distributed throughout the body; however, expression is particularly strong in the lymph node, tonsil, cultured lymphoblasts and ovary. It is inducible by interferon-γ in lymphoblast cells. SCGB1D4 appears to exhibit immunological functions, including regulation of chemotactic migration and invasion.[27] There is one pseudogene in this subfamily identified as *SCGB1D1P1*. We propose that it be renamed *SCGB1D5P*, keeping it in line with the other members of this subfamily.

SCGB2A subfamily

SCGB2A1 is also known as lipophilin C. SCGB2A2 is also known as mammaglobin and is expressed in a highly tissue-specific manner in breast epithelium, where it forms heterodimers with SCGB1D2.[28]

SCGB2B subfamily

The *SCGB2B* gene subfamily includes *SCGB2B1P* (formerly *ABPBG1P*), *SCGB2B2* (formerly *SCGB4A2*, *SCGBL*) and *SCGB2B3P* (formerly *SCGB4A3P*). Of the *SCGB1B* and *SCGB2B* subfamilies, only *SCGBL* is listed in the HGNC database, but we propose a name change to include it in the SCGB superfamily-naming system. There is only a single reference to SCGB2B2 in the literature[24] but there is evidence for its expression in EST databases.

SCGB3A subfamily

SCGB3A1 and SCGB3A2, identified in 2002, have high structural homology[29] with SCGB1A1. Their expression appears to be localised principally to epithelial organs, such as the lung, mammary gland, trachea, prostate and salivary gland.[30] In the bronchial epithelium, expression is decreased after injury.[29] It has been proposed[29] that SCGB3A1 might have similar and overlapping expression and function with SCGB1A1.

SCGB3A1 is a candidate tumour-suppressor gene and a target gene for endothelial PAS domain protein 1 (EPAS1—formerly HIF2α).[31] SCGB3A1 expression is diminished in many human cancers (including lung, prostate, pancreatic and nasopharyngeal); hypermethylation of the *SCGB3A1* promoter has also been reported for many malignancies.[31] SCGB3A2 has been shown to be induced by T-helper cell 1 (Th1) cytokines but suppressed by proinflammatory and Th2 cytokines.[4] Any given cytokine can evoke different responses in different SCGBs.[4] Intranasal administration of recombinant SCGB3A2 suppresses allergen-induced lung inflammation, further highlighting similarities between SCGB3A2 and SCGB1A1.[32]

Mouse gene family members

Scgb1a1

This gene encodes mouse UGB and is orthologous to human *SCGB1A1*.

Scgb1c1

This gene encodes a protein that is the mouse equivalent of human SCGB1C1.

Scgb1b and *Scgb2b*: The androgen-binding protein (ABP) family

Sixty-four of the 68 mouse *Scgb* genes belong to a family that has been called the ABP family.[13] These proteins are heterodimers consisting of two distinct types of subunits, SCGB1B (previously called

ABPA-like), and SCGB2B (previously ABPBG-like). These were originally isolated from mouse saliva and described based on their ability to bind androgens.[33] ABPs have since been shown to be expressed in glands of the face and neck, as well as in the prostate and ovary.[34] The role of ABPs in communication is supported by the expression of many *Abpa (Scgb1b)* and *Abpbg (Scgb2b)* mRNAs in the brain (olfactory lobe), sensory organs (olfactory epithelium, vomeronasal organ), glands of the head and neck (parotid, sublingual, submaxillary and lacrimal) and sexual tissues (prostate and ovary and preputial and clitoral glands).[13]

Scgb3a genes

Scgb3a1 and *Scgb3a2* encode predicted proteins that align well with human SCGB3A1 and SCGB3A2 protein structures and are most likely orthologous to them.

Evolution

SCGB members have amino acid sequences that are highly divergent within the superfamily, complicating the identification of group members. To test whether all entries found were related to known SCGBs, a jackHMMER profile was created (an iterated sequence profile search, seeded with human SCGB1A1), which confirmed group membership for all human and mouse SCGBs[35] with an expected value of less than 0.001. *Homologene*,[36] a software program that analyses groups of homologous proteins across multiple species, currently recognises 21 SCGB clades (Figure 2). The *SCGB* genes encode proteins that all have a similar structure.[18] Despite high amino acid sequence divergence, many structural features (such as helical bundles and the ability to dimerise) are retained.[11,18] This is consistent with a highly flexible and rapidly evolving gene superfamily and is likely to have aided in the evolution of the diverse functions of the superfamily.

When SCGBs were named in 2000, six human SCGBs were described and divided into five groups, based on proposed evolutionary relationships.[18] Currently, there are 11 described human *SCGB* genes. Figure 1 shows mouse and human proteins on a phylogenetic tree for this family. In the case of the ABP proteins, we have used SCGB1B27 (ABPA27) and SCGB2B27 (ABPBG27) to represent the mouse SCGB1B and SCGB2B groups, respectively. Table 1 lists chromosomal locations of human *SCGB*s, and only those mouse genes that share orthology. Four human *SCGB*s have direct mouse orthologues; the ABP subfamily includes three human *SCGB1B*s versus 30 mouse *Scgb1b*s and three human *SCGB2B*s versus 34 mouse *Scgb2b*s. The human genome contains the *SCGB1D* and *SCGB2A* subfamilies, both of which are absent in the mouse.

The ABP (*Scgb1b/Scgb2b*) family contains genes for two different types of subunit, ABPA (SCGB1B) and ABPBG (SCGB2B),[37] located adjacent to each other on mouse chromosome 7 (Table 2). This 'recent, phylogenetically independent proliferation of close paralogs, or lineage specific gene family expansion' is an example of an 'evolutionary bloom'.[37] Another example of this has been most notably studied in the large and diverse cytochrome P450 family.[38] It has been suggested that these evolutionary blooms might represent simply a stochastic process.[37]

The genes that encode any *Scgb1b/Scgb2b* pair tend to be next to each other on the chromosome and orientated in a 'head-to-head' ($3'-5'|5'-3'$) fashion. These structures have been called 'modules'.[39] It appears that there was a single *Scbg1b–Scgb2b* module which has expanded dramatically in some species (64 genes in mouse, 43 in rabbit). In other species it has resulted largely in pseudogenes, such as those of the primate lineage, or been lost altogether in species such as the shrew and elephant.[13] Interestingly, in humans there are three such modules. Although at least two modules have become pseudogenes, it remains possible that the *SCGB1B2–SCGB2B2* module might be active, based on EST data. The mouse shows the most extensive expansion, which began in the ancestor of the genus *Mus*[13] after divergence from rat, less than 17 million years ago, and apparently has involved two different modes of duplication.[39]

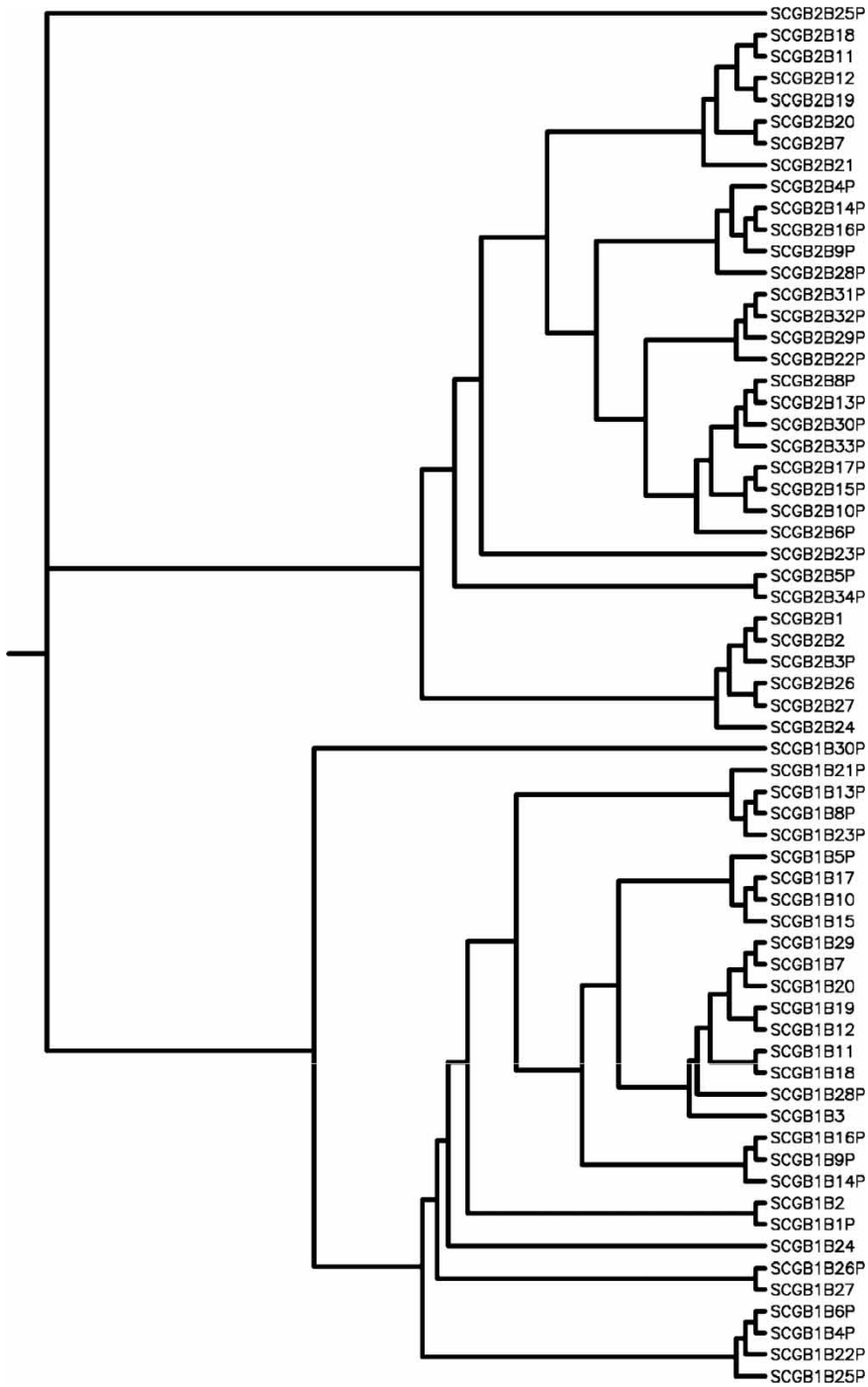

Figure 2. Phylogenetic tree of mouse androgen-binding proteins. Protein sequences were aligned using TCOFFEE[14] and analysed using nearest-neighbour-joining methods, as well as using 10,000 bootstrap replicates in the Phylip package.[19]

Association of SCGBs with disease

SCGBs have been linked to multiple disease states, either as participants or as biomarkers. SCGB1A1 may serve as an early biomarker for lung injury, owing to the regenerative role of cells that secrete SCGB1A1.[20,40] In addition, SCGB1A1 may act as a tumour suppressor[41] and has been shown to be up- or downregulated in various human lung cancers.[42]

Table 1. Comparison of human secretoglobin genes (*SCGBs*) with only those mouse *Scgbs* that share orthology. All data include genes, common names and chromosomal locations. Genes that are likely to share homology are placed on the same line

Gene symbol	Human chr location	Synonyms	Mouse homologue	Mouse chr location
SCGB1A1	11q11-qter	Uteroglobin (UGB), blastokinin, Clara-cell specific 10-kDa protein (CC10), Clara-cell specific 16-kD protein (CC16), Clara-cell specific protein (CCSP), Clara-cell phospholipid-binding protein (CCPBP), and urinary protein 1 (UP1)	*Scgblal*	19
*SCGB1B1P**	19	Abpa1P	*Scgblb* subfamily	7
*SCGB1B2P**	19	SCGB4A1P		
*SCGB1B3**	19	SCGB4A4		
SCGB1C1	11p15.5	RYD5	*Scgblcl*	7
SCGB1D1	11q13	Lipophilin A (LIPA)		
*SCGB1D5P**	4q32.3	SCGB1D1P1		
SCGB1D2	11q13	Lipophilin B (LIPB)		
SCGB1D4	11q12.3	IFN-γ inducible SCGB (IIS)		
SCGB2A1	11q13	Lipophilin C, mammaglobin 2 (MGB2), mammaglobin B, UGB3 and lacryglobin		
SCGB2A2	11q13	Mammaglobin 1 (MGB1), mammaglobin A, UGB2		
*SCGB2B1P**	19	Abpbg1P	*Scgb2b* subfamily	7
*SCGB2B2**	19	SCGB4A2, SCGBL		
*SCGB2B3P**	19	SCGB4A3P cytokine high in normal-1 (HIN1), pneumo		
SCGB3A1	5q35-qter	Secretory protein 2 (PnSP-2), and uteroglobin-related protein 2 (UGRP2)	*Scgb3a1*	11
SCGB3A2	5q32	Pneumo secretory protein 1 (PnSP-1), and uteroglobin-related protein 1 (UGRP1)	*Scgb3a2*	18

Abbreviations: IFN, interferon
* Indicates new *Scgb* designation.

Table 2. The mouse androgen-binding protein (ABP) family, complete with the newly proposed *Scgb* nomenclature. Indicated are the recommended gene symbol, previously published symbol, chromosomal location (Chr), strand, and start and end locations. These data were adapted from Laukaitis *et al.*[13] Genomic locations were updated from Build 36 (mm8) of the mouse to Build 37 (mm9) using the University of California, Santa Cruz (UCSC) BLAT tool. Records from the National Center for Biotechnology Information (NCBI) gene database that correspond to mouse ABPs were aligned to these records using TCOFFEE[14] and analysed using nearest-neighbour-joining methods, as well as using 10,000 bootstrap replicates in the Phylip package;[19] analogous records are placed on the same line. In some cases, the start and end locations are different from those reported by Laukaitis *et al.*, presumably because of differences in gene prediction algorithms

Gene symbol	Adapted from Laukaitis (2008)[13]			Location		NCBI record		Location	
	Previous symbol	Chr	Strand	Start	End	Gene symbol	GI number	Start	End
Scgblblp	B6_a1ψ	7	−	32,040,110	32,041,296	*Gm4360*	100043323	32,040,902	32,041,296
Scgb2b1	B6_bg1	7	+	32,048,775	32,050,776				
Scgb1b2	B6_a2	7	−	32,075,661	32,076,819	*Abph*	57426	32,075,538	32,076,835
Scgb2b2	B6_bg2	7	+	32,087,800	32,089,833	*Abpe*	381970	32,087,788	32,089,996
Scgb2b3p	B6_bg3ψ	7	−	32,144,059	32,147,091	*Gm4362*	100043326	32,144,057	32,147,091
Scgb1b3	B6_a3	7	+	32,160,626	32,161,811	*Gm5325*	384585	32,160,617	32,161,935
Scgb2b4p	B6_bg4ψ	7	−	32,186,052	32,188,197				
Scgb1b4p	B6_a4ψ	7	+	32,194,693	32,195,875				
Scgb1b28p	B6_a28ψ	7	+	32,228,841	32,230,016				
Scgb2b5p	B6_bg5ψ	7	−	32,285,202	32,287,735				
Scgb1b5p	B6_a5ψ	7	+	32,292,459	32,292,831				
Scgb2b28p	B6_b28ψ	7	−	32,330,062	32,331,828				
Scgb2b6p	B6_bg6ψ	7	−	32,402,759	32,404,545	*Gm19607*	100503242	32,402,757	32,404,545
Scgb1b6p	B6_a6ψ	7	+	32,410,618	32,411,808				
Scgb2b7	B6_bg7	7	−	32,488,962	32,490,757	*Gm4684*	100043836	32,488,798	32,490,772
Scgb1b7	B6_a7	7	+	32,497,700	32,498,874	*Gm6662*	626305	32,497,690	32,498,998
Scgb2b8p	B6_bg8ψ	7	−	32,563,799	32,565,577				
Scgb1b8p	B6_a8ψ	7	+	32,573,297	32,574,476				
Scgb2b9p	B6_bg9ψ	7	−	32,700,256	32,702,404				
Scgb1b9p	B6_a9ψ	7	+	32,709,361	32,710,545				
Scgb2b29p	B6_bg29ψ	7	−	32,798,052	32,799,818				
Scgb2b10p	B6_bg10ψ	7	−	32,877,166	32,878,958	*Gm6513*	624584	32,877,164	32,878,958
Scgb1b10	B6_a10	7	+	32,885,887	32,887,066	*Gm4384*	100043355	32,885,888	32,887,069
Scgb2bg30p	BG_bg30ψ	7	−	32,931,261	32,933,051				
Scgb2b11	B6_bg11	7	−	32,994,356	32,996,142	*Gm4687*	100043842	32,994,354	32,996,142

Continued

Table 2. Continued

Gene symbol	Adapted from Laukaitis (2008)[13]			Location		NCBI record		Location	
	Previous symbol	Chr	Strand	Start	End	Gene symbol	GI number	Start	End
Scgb1b11	B6_a11	7	+	33,007,875	33,009,053				
Scgb2b12	B6_bg12	7	−	33,110,463	33,112,248	*Gm9138*	668379	33,110,305	33,112,248
Scgb1b12	B6_a12	7	+	33,119,216	33,120,398	*Gm9140*	668381	33,119,217	33,120,401
Scgb1b29	B6_a29	7	+	33,226,559	33,227,738	*Gm5326*	384589	33,226,560	33,227,863
Scgb2b13p	B6_bg13ψ	7	−	33,292,522	33,294,299				
Scgb1b13p	B6_a13ψ	7	+	33,302,029	33,303,208				
Scgb2b14p	B6_bg14ψ	7	−	33,432,671	33,434,820				
Scgb1b14p	B6_a14ψ	7	+	33,441,783	33,442,968				
Scgb2bg31p	BG_bg31ψ	7	−	33,530,315	33,532,081				
Scgb2b15p	B6_bg15ψ	7	−	33,612,867	33,614,659	*Gm6504*	624439	33,612,702	33,614,672
Scgb1b15	B6_a15	7	+	33,621,588	33,622,767	*Gm4399*	100043376	33,621,589	33,622,770
Scgb2b16p	B6_bg16ψ	7	−	33,659,561	33,661,710				
Scgb1b16p	B6_a16ψ	7	+	33,668,673	33,669,858	*Gm4404*	100043383	33,668,674	33,669,068
Scgb2bg32p	BG_bg32ψ	7	−	33,757,214	33,758,980				
Scgb2b17p	B6_bg17ψ	7	−	33,839,766	33,841,558	*Gm9223*	668526	33,839,601	33,841,571
Scgb1b17	B6_a17	7	+	33,848,483	33,849,662	*Gm4406*	100043388	33,848,484	33,849,665
Scgb2bg33p	BG_bg33ψ	7	−	33,896,690	33,898,477				
Scgb2b18	B6_bg18	7	−	33,957,075	33,958,861	*Gm4692*	100043856	33,957,073	33,958,861
Scgb1b18	B6_a18	7	+	33,970,586	33,971,764				
Scgb2b19	B6_bg19	7	−	34,063,553	34,065,338	*Gm5894*	545947	34,063,551	34,065,338
Scgb1b19	B6_a19	7	+	34,072,325	34,073,507	*Gm5632*	434676	34,072,326	34,073,510
Scgb2b20	B6_bg20	7	−	34,149,526	34,151,319	*Abpd*	494519	34,149,362	34,151,338
Scgb1b20	B6_a20	7	+	34,158,268	34,159,452	*Gm5895*	545948	34,158,259	34,159,577
Scgb2bg34p	BG_bg34ψ	7	−	34,228,045	34,229,836				
Scgb2b21	B6_bg21	7	−	34,303,665	34,305,459	*Gm6725*	626970	34,303,663	34,305,459
Scgb1b21p	B6_a21ψ	7	+	34,312,394	34,313,577				
Scgb2b22p	B6_bg22ψ	7	−	34,389,300	34,391,071				
Scgb1b22p	B6_a22ψ	7	+	34,399,640	34,400,823				
Scgb2b23p	B6_bg23ψ	7	−	34,410,332	34,412,120	*Scgb2b1-ps*	353109	34,410,165	34,412,120

Continued

Table 2. Continued

Gene symbol	Adapted from Laukaitis (2008)[13]			Location		NCBI record		Location	
	Previous symbol	Chr	Strand	Start	End	Gene symbol	GI number	Start	End
Scgb1b23p	B6_a23ψ	7	+	34,420,180	34,421,351				
Scgb2b24	B6_bg24	7	−	34,522,368	34,524,304	*Abpz*	233090	34,522,212	34,524,314
Scgb1b24	B6_a24	7	+	34,528,817	34,529,998	*Gm12769*	100043860	34,528,818	34,530,122
Scgb2b25p	B6_bg25ψ	7	−	34,638,053	34,639,837	*Gm12777*	627079	34,639,065	34,639,837
Scgb1b25p	B6_a25ψ	7	+	34,652,959	34,654,103				
Scgb2b26	B6_bg26	7	−	34,728,187	34,729,983	*Abpg*	110187	34,728,016	34,730,004
Scgb1b26p	B6_a26ψ	7	+	34,743,823	34,745,002	*Gm12759*	100043864	34,743,824	34,745,005
Scgb2b27	B6_bg27	7	−	34,797,105	34,798,885	*Abpb*	233099	34,796,951	34,798,961
Scgb1b27	B6_a27	7	+	34,806,586	34,807,768	*Abpa*	11354	34,806,586	34,807,900
Scgb1b30p	B6_a30ψ	7	+	34,880,532	34,885,724	*Gm12776*	100043868	34,880,450	34,885,856

Proteomic analysis of lacrimal gland fluid has revealed that patients with dry eyes have a decrease in SCGB1D1, SCGB1D2 and SCGB2A1 expression; the condition 'dry eyes' may be caused by post-translational modifications.[43] In addition, SCGB1D2 has been reported to be upregulated in breast cancer, making it a potential marker for this type of malignancy.[44] In this context, panels of autoantibodies to tumour-associated antigens in breast cancer include SCGB1D2, which, when combined with others, may have diagnostic potential. SCGB1D2 is also downregulated in pituitary adenomas.[45]

SCGB2A1 has been shown to be a prognostic marker in epithelial ovarian cancer[46,47] and endometrial cancer.[48] Because SCGB2A2 expression is highly specific to breast epithelial tissue, it has been proposed as a marker for detecting breast cancer metastases to sentinel lymph nodes and distant tissues.[49–51] SCGB2A1 overexpression has also been evaluated as a marker for breast cancer, with mixed conclusions.[28,52–55]

SCGB3A1 has been shown to be differentially expressed in smokers with lung cancer.[56] Its decreased expression has been correlated with increased tumour burden in non-small-cell lung cancer.[31] A *SCGB3A2* polymorphism has been associated with increased asthma risk in a Japanese population.[57,58] In chronic rhinosinusitis, SCGB3A2 levels in sino-nasal tissue are inversely correlated with the total number of infiltrating inflammatory cells, as well as scores of symptom severity.[4]

Conclusions

The SCGBs represent an intriguing family of biologically active proteins. The relatively recent revelations of anti-inflammatory and immunomodulatory functions, together with their potential as cancer biomarkers, underscore their physiological and pathophysiological importance. However, a great deal more needs to be elucidated regarding the actions of individual SCGBs. Further studies directed at characterising the individual SCGBs are necessary, the results of which are likely to yield valuable targets for therapeutic intervention.

One of the most intriguing characteristics of the mammalian 'Abp' genes, the *Scgb1b/Scgb2b* subset of the *SCGB* gene family, is their evolutionarily independent expansions (so-called 'evolutionary blooms') in a number of mammalian lineages. Discovery of the reason for these blooms may lead

to a better understanding of how these SCGBs function in different mammals.

Acknowledgments

We thank our colleagues, especially Drs Robert Karn, Christine Laukaitis, Robert Lehrer, Beverly Chilton, Anil Mukherjee, Lucio Miele, Shioko Kimura, Joerg Klug and Richard Emes, for valuable discussions during the course of writing this review and for careful readings of this manuscript. This work was supported by grants NIH F31 AA20728 (B.C.J), NIH NHGRI P41 HG03345 and Wellcome Trust grant O81979/Z/07/Z (M.W.R.), NIH NHGR1 HG 00330 (M.M.), NIH R01 EY017963 (V.K.V) and NIH P30 ES006096 (D.W.N.).

References

1. Daniel, J.C. (2000), 'Discovery and perspectives from the blastokinin era', *Ann. NY Acad. Sci.* Vol. 923, pp. 1–8.
2. Beier, H.M. (2000), 'The discovery of uteroglobin and its significance for reproductive biology and endocrinology', *Ann. NY Acad. Sci.* Vol. 923, pp. 9–24.
3. Klug, J., Beier, H.M., Bernard, A., Chilton, B.S. *et al.* (2000), 'Uteroglobin/Clara cell 10-kDa family of proteins: Nomenclature committee report', *Ann. NY Acad. Sci.* Vol. 923, pp. 348–354.
4. Lu, X., Wang, N., Long, X.B., You, X.J. *et al.* (2011), 'The cytokine-driven regulation of secretoglobins in normal human upper airway and their expression, particularly that of uteroglobin-related protein 1, in chronic rhinosinusitis', *Respir. Res.* Vol. 12, p. 28.
5. Pilon, A.L. (2000), 'Rationale for the development of recombinant human CC10 as a therapeutic for inflammatory and fibrotic disease', *Ann. NY Acad. Sci.* Vol. 923, pp. 280–299.
6. Karn, R.C. and Laukaitis, C.M. (2003), 'Characterization of two forms of mouse salivary androgen-binding protein (ABP): Implications for evolutionary relationships and ligand-binding function', *Biochemistry* Vol. 42, pp. 7162–7170.
7. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G. *et al.* (2000), 'The Protein Data Bank', *Nucleic Acids Res.* Vol. 28, pp. 235–242.
8. Bally, R. and Delettre, J. (1989), 'Structure and refinement of the oxidized P21 form of uteroglobin at 1.64 A resolution', *J. Mol. Biol.* Vol. 206, pp. 153–170.
9. Stripp, B.R., Lund, J., Mango, G.W., Doyen, K.C. *et al.* (1996), 'Clara cell secretory protein: A determinant of PCB bioaccumulation in mammals', *Am. J. Physiol.* Vol. 271, pp. L656–L664.
10. Mukherjee, A.B., Zhang, Z. and Chilton, B.S. (2007), 'Uteroglobin: A steroid-inducible immunomodulatory protein that founded the secretoglobin superfamily', *Endocr. Rev.* Vol. 28, pp. 707–725.
11. Callebaut, I., Poupon, A., Bally, R., Demaret, J.P. *et al.* (2000), 'The uteroglobin fold', *Ann. NY Acad. Sci.* Vol. 923, pp. 90–112.
12. UniProt Consortium (2011), 'Ongoing and future developments at the Universal Protein Resource', *Nucleic Acids Res.* Vol. 39, pp. D214–D219.
13. Laukaitis, C.M., Heger, A., Blakley, T.D., Munclinger, P. *et al.* (2008), 'Rapid bursts of androgen-binding protein (Abp) gene duplication occurred independently in diverse mammals', *BMC Evol. Biol.* Vol. 8, p. 46.
14. Notredame, C., Higgins, D.G. and Heringa, J. (2000), 'T-Coffee: A novel method for fast and accurate multiple sequence alignment', *J. Mol. Biol.* Vol. 302, pp. 205–217.
15. Nebert, D.W. and Nelson, D.R. (1991), 'P450 gene nomenclature based on evolution', *Methods Enzymol.* Vol. 206, pp. 3–11.
16. Nebert, D.W., Adesnik, M., Coon, M.J., Estabrook, R.W. *et al.* (1987), 'The P450 gene superfamily: Recommended nomenclature', *DNA*, Vol. 6, pp. 1–11.
17. Nuclear Receptors Nomenclature Committee (1999), 'A unified nomenclature system for the nuclear receptor superfamily', *Cell* Vol. 97, pp. 161–163.
18. Ni, J., Kalff-Suske, M., Gentz, R., Schageman, J. *et al.* (2000), 'All human genes of the uteroglobin family are localized on chromosome 11q12.2 and form a dense cluster', *Ann. NY Acad. Sci.* Vol. 923, pp. 25–42.
19. Felsenstein, J. (1989), 'PHYLIP – Phylogeny Inference Package (Version 3.2)', *Cladistics*, Vol. 5, pp. 164–166.
20. Wong, A.P., Keating, A. and Waddell, T.K. (2009), 'Airway regeneration: The role of the Clara cell secretory protein and the cells that express it', *Cytotherapy* Vol. 11, pp. 676–687.
21. Zhang, Z., Kundu, G.C., Yuan, C.J., Ward, J.M. *et al.* (1997), 'Severe fibronectin-deposit renal glomerular disease in mice lacking uteroglobin', *Science* Vol. 276, pp. 1408–1412.
22. Stripp, B.R., Reynolds, S.D., Plopper, C.G., Boe, I.M. *et al.* (2000), 'Pulmonary phenotype of CCSP/UG deficient mice: A consequence of CCSP deficiency or altered Clara cell function?', *Ann. NY Acad. Sci.* Vol. 923, pp. 202–209.
23. Hayashida, S., Harrod, K.S. and Whitsett, J.A. (2000), 'Regulation and function of CCSP during pulmonary Pseudomonas aeruginosa infection in vivo', *Am. J. Physiol. Lung Cell. Mol. Physiol.* Vol. 279, pp. L452–L459.
24. Emes, R.D., Riley, M.C., Laukaitis, C.M., Goodstadt, L. *et al.* (2004), 'Comparative evolutionary genomics of androgen-binding protein genes', *Genome Res.* Vol. 14, pp. 1516–1529.
25. Dear, T.N., Boehm, T., Keverne, E.B. and Rabbitts, T.H. (1991), 'Novel genes for potential ligand-binding proteins in subregions of the olfactory mucosa', *EMBO J.* Vol. 10, pp. 2813–2819.
26. Lehrer, R.I., Nguyen, T., Zhao, C., Ha, C.X. *et al.* (2000), 'Secretory lipophilins: A tale of two species', *Ann. NY Acad. Sci.* Vol. 923, pp. 59–67.
27. Choi, M.S., Ray, R., Zhang, Z. and Mukherjee, A.B. (2004), 'IFN-gamma stimulates the expression of a novel secretoglobin that regulates chemotactic cell migration and invasion', *J. Immunol.* Vol. 172, pp. 4245–4252.
28. Goedegebuure, P.S., Watson, M.A., Viehl, C.T. and Fleming, T.P. (2004), 'Mammaglobin-based strategies for treatment of breast cancer', *Curr. Cancer Drug Targets* Vol. 4, pp. 531–542.
29. Reynolds, S.D., Reynolds, P.R., Pryhuber, G.S., Finder, J.D. *et al.* (2002), 'Secretoglobins SCGB3A1 and SCGB3A2 define secretory cell subsets in mouse and human airways', *Am. J. Respir. Crit. Care Med.* Vol. 166, pp. 1498–1509.
30. Porter, D., Lahti-Domenici, J., Torres-Arzayus, M., Chin, L. *et al.* (2002), 'Expression of high in normal-1 (HIN-1) and uteroglobin related protein-1 (UGRP-1) in adult and developing tissues', *Mech. Dev.* Vol. 114, pp. 201–204.
31. Mazumdar, J., Hickey, M.M., Pant, D.K., Durham, A.C. *et al.* (2010), 'HIF-2alpha deletion promotes Kras-driven lung tumor development', *Proc. Natl. Acad. Sci. USA* Vol. 107, pp. 14182–14187.
32. Chiba, Y., Kurotani, R., Kusakabe, T., Miura, T. *et al.* (2006), 'Uteroglobin-related protein 1 expression suppresses allergic airway inflammation in mice', *Am. J. Respir. Crit. Care Med.* Vol. 173, pp. 958–964.
33. Dlouhy, S.R. and Karn, R.C. (1983), 'The tissue source and cellular control of the apparent size of androgen binding protein (Abp), a mouse salivary protein whose electrophoretic mobility is under the control of sex-limited saliva pattern (Ssp)', *Biochem. Genet.* Vol. 21, pp. 1057–1070.
34. Laukaitis, C.M., Dlouhy, S.R., Emes, R.D., Ponting, C.P. *et al.* (2005), 'Diverse spatial, temporal, and sexual expression of recently duplicated androgen-binding protein genes in Mus musculus', *BMC Evol. Biol.* Vol. 5, p. 40.
35. Johnson, L.S., Eddy, S.R. and Portugaly, E. (2010), 'Hidden Markov model speed heuristic and iterative HMM search procedure', *BMC Bioinformatics* Vol. 11, p. 431.

36. Sayers, E.W., Barrett, T., Benson, D.A., Bolton, E. *et al.* (2010), 'Database resources of the National Center for Biotechnology Information', *Nucleic Acids Res.* Vol. 38, pp. D5–D16.
37. Waterston, R.H., Lindblad-Toh, K., Birney, E., Rogers, J. *et al.* (2002), 'Initial sequencing and comparative analysis of the mouse genome', *Nature* Vol. 420, pp. 520–562.
38. Feyereisen, R. (2011), 'Arthropod CYPomes illustrate the tempo and mode in P450 evolution', *Biochim. Biophys. Acta* Vol. 1814, pp. 19–28.
39. Karn, R.C. and Laukaitis, C.M. (2009), 'The mechanism of expansion and the volatility it created in three pheromone gene clusters in the mouse (Mus musculus) genome', *Genome Biol. Evol.* Vol. 1, pp. 494–503.
40. Shijubo, N., Kawabata, I., Sato, N. and Itoh, Y. (2003), 'Clinical aspects of Clara cell 10-kDa protein/uteroglobin (secretoglobin 1A1)', *Curr. Pharm. Des.* Vol. 9, pp. 1139–1149.
41. Patierno, S.R., Manyak, M.J., Fernandez, P.M., Baker, A. *et al.* (2002), 'Uteroglobin: A potential novel tumor suppressor and molecular therapeutic for prostate cancer', *Clin. Prostate Cancer* Vol. 1, pp. 118–124.
42. Sjodin, A., Guo, D., Sorhaug, S., Bjermer, L. *et al.* (2003), 'Dysregulated secretoglobin expression in human lung cancers', *Lung Cancer* Vol. 41, pp. 49–56.
43. Versura, P., Nanni, P., Bavelloni, A., Blalock, W.L. *et al.* (2010), 'Tear proteomics in evaporative dry eye disease', *Eye (Lond.)* Vol. 24, pp. 1396–1402.
44. Culleton, J., O'Brien, N., Ryan, B.M., Hill, A.D. *et al.* (2007), 'Lipophilin B: A gene preferentially expressed in breast tissue and upregulated in breast cancer', *Int. J. Cancer* Vol. 120, pp. 1087–1092.
45. Sjodin, A., Guo, D., Lund-Johansen, M., Krossnes, B.K. *et al.* (2005), 'Secretoglobins in the human pituitary: High expression of lipophilin B and its down-regulation in pituitary adenomas', *Acta Neuropathol.* Vol. 109, pp. 381–386.
46. Tassi, R.A., Calza, S., Ravaggi, A., Bignotti, E. *et al.* (2009), 'Mammaglobin B is an independent prognostic marker in epithelial ovarian cancer and its expression is associated with reduced risk of disease recurrence', *BMC Cancer* Vol. 9, p. 253.
47. Tassi, R.A., Bignotti, E., Rossi, E., Falchetti, M. *et al.* (2007), 'Overexpression of mammaglobin B in epithelial ovarian carcinomas', *Gynecol. Oncol.* Vol. 105, pp. 578–585.
48. Tassi, R.A., Bignotti, E., Falchetti, M., Calza, S. *et al.* (2008), 'Mammaglobin B expression in human endometrial cancer', *Int. J. Gynecol. Cancer* Vol. 18, pp. 1090–1096.
49. Janku, F., Kleibl, Z., Novotny, J., Tesarova, P. *et al.* (2004), 'Mammaglobin A, a novel marker of minimal residual disease in early stages breast cancer', *Neoplasma* Vol. 51, pp. 204–208.
50. Aihara, T., Fujiwara, Y., Miyake, Y., Okami, J. *et al.* (2000), 'Mammaglobin B gene as a novel marker for lymph node micrometastasis in patients with abdominal cancers', *Cancer Lett.* Vol. 150, pp. 79–84.
51. Douglas-Jones, A.G. and Woods, V. (2009), 'Molecular assessment of sentinel lymph node in breast cancer management', *Histopathology* Vol. 55, pp. 107–113.
52. Zehentner, B.K. and Carter, D. (2004), 'Mammaglobin: A candidate diagnostic marker for breast cancer', *Clin. Biochem.* Vol. 37, pp. 249–257.
53. Sjodin, A., Ljuslinder, I., Henriksson, R. and Hedman, H. (2008), 'Mammaglobin and lipophilin B expression in breast tumors and their lack of effect on breast cancer cell proliferation', *Anticancer Res.* Vol. 28, pp. 1493–1498.
54. Zafrakas, M., Petschke, B., Donner, A., Fritzsche, F. *et al.* (2006), 'Expression analysis of mammaglobin A (SCGB2A2) and lipophilin B (SCGB1D2) in more than 300 human tumors and matching normal tissues reveals their co-expression in gynecologic malignancies', *BMC Cancer*, Vol. 6, p. 88.
55. Brown, N.M., Stenzel, T.T., Friedman, P.N., Henslee, J. *et al.* (2006), 'Evaluation of expression based markers for the detection of breast cancer cells', *Breast Cancer Res. Treat.* Vol. 97, pp. 41–47.
56. Beane, J., Vick, J., Schembri, F., Anderlind, C. *et al.* (2011), 'Characterizing the impact of smoking and lung cancer on the airway transcriptome using RNA-Seq', *Cancer Prev. Res. (Phila)*, Vol. 4, pp. 803–817.
57. Inoue, K., Wang, X., Saito, J., Tanino, Y. *et al.* (2008), 'Plasma UGRP1 levels associate with promoter G-112A polymorphism and the severity of asthma', *Allergol. Int.* Vol. 57, pp. 57–64.
58. Niimi, T., Munakata, M., Keck-Waggoner, C.L., Popescu, N.C. *et al.* (2002), 'A polymorphism in the human UGRP1 gene promoter that regulates transcription is associated with an increased risk of asthma', *Am. J. Hum. Genet.* Vol. 70, pp. 718–725.